

**31ST EUROPEAN SYMPOSIUM
ON COMPUTER
AIDED PROCESS ENGINEERING**

ESCAPE-31

Edited by
METIN TÜRKAY AND RAFIQU L GANI



COMPUTER-AIDED CHEMICAL ENGINEERING, 50

31ST EUROPEAN SYMPOSIUM ON
COMPUTER AIDED PROCESS
ENGINEERING

PART A

31ST EUROPEAN SYMPOSIUM ON
COMPUTER AIDED PROCESS
ENGINEERING

PART A

Edited by

Metin Türkay
Koç University

Rafiqul Gani
PSE for SPEED



ELSEVIER

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

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

Copyright © 2021 Elsevier B.V. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

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

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

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

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN (Part A): 978-0-323-98325-9

ISBN (Set) : 978-0-323-88506-5

ISSN: 1570-7946

For information on all Elsevier publications visit our website at <https://www.elsevier.com/>



Publisher: Susan Dennis

Acquisition Editor: Anita Koch

Editorial Project Manager: Lena Sparks

Production Project Manager: Paul Prasad Chandramohan

Designer: Alan Studholme

Typeset by STRAIVE

Contents

Preface	xv
PROCESS-PRODUCT DESIGN AND INTEGRATION	
1. Explanatory and Predictive Analysis of Naphtha Splitter Products <i>Ahmet Can Serfidan, Metin Türkay</i>	1
2. Multi-objective optimization of the supply chain for the production of biofuels and high value-added products in Mexico: importance of the water footprint <i>Yulissa M. Espinoza-Vázquez, Fernando Israel Gómez-Castro, José María Ponce-Ortega</i>	7
3. Synthesis of Combined Heat and Mass Exchange Networks Via a Trust Region Filter Optimisation Algorithm Including Detailed Unit Designs <i>Saif R. Kazi, Michael Short, Lorenz T. Biegler</i>	13
4. Computer-Aided Refrigerant Design: New Developments <i>Nichakorn Kuprasertwong, Orakotch Padungwatanaroj, Arisa Robin, Kornkanok Udomwong, Anjan Kumar Tula, Lingyu Zhu, Liyang Zhou, Binhui Wang, Shuhua Wang, Rafiqul Gani</i>	19
5. Comparative assessment of black liquor upgraded gasification in integrated kraft pulp, methanol and dimethyl ether production plants <i>Meire Ellen Gorete Ribeiro Domingos, Daniel Flórez-Orrego, Moisés Teles dos Santos, Silvio de Oliveira Junior</i>	25
6. Sustainable Process Design and Synthesis for HDPE Recycling <i>Xiang Zhao, Fengqi You</i>	31
7. A Novel Process Design for Waste Respirator Processing <i>Xiang Zhao, Fengqi You</i>	37
8. Design and optimization of a shell-and-tube heat exchanger using the univariate marginal distribution algorithm <i>Oscar D. Lara Montaña, Fernando I. Gómez-Castro, Claudia Gutierrez- Antonio</i>	43
9. Synthesis of Microgels with Minimal Final Concentration of Initiator <i>Falco Jung, Johannes M. M. Faust, Adel Mhamdi, Andrij Pich, Alexander Mitsos</i>	51
10. Novel Dual Reactive Distillation Process for Eco-efficient Lactic Acid Recovery and Purification from Fermentation Broth <i>Isabel Pazmiño Mayorga, A.A. Kiss, M. Jobson</i>	57

11. Exergy Analysis of Vacuum Distillation Unit <i>Sena Kurban, Gizem Kuşoğlu Kaya, Serdar Yaman</i>	63
12. Computer-Aided Solvent Design Integrated with a Machine Learning-based Atom Contribution Method <i>Qilei Liu, Lei Zhang, Jian Du, Rafiqul Gani</i>	69
13. A versatile modelling system for integrated chemical product design problems <i>Shiyang Chai, Lei Zhang, Jian Du, Anjan K Tula, Rafiqul Gani, Mario R Eden</i>	75
14. An integrated system of multi effect distillation and wind power system - Evaluation of total energy saving <i>O.M.A. Al-hotmani, Mudhar. A. Al-Obaidi, Y. M. John, Raj. Patel, Iqbal M. Mujtaba</i>	81
15. Modelling and Performance evaluation of the air heated and water heated Humidification-Dehumidification desalination process using solar thermal energy <i>Damson Kaunga, Raj Patel, Iqbal M Mujtaba</i>	87
16. Enhancement of energy saving of reverse osmosis system of Arab Potash Company via a wind energy system <i>Alanood A. Alsarayreh, Mudhar. A. Al-Obaidi, Raj. Patel, Iqbal M. Mujtaba</i>	95
17. Multi-objective optimization for the incorporation of safety and reliability considerations in process design <i>Andrea P. Ortiz-Espinoza, Yixin Ye, Ignacio E. Grossmann, Arturo Jiménez-Gutiérrez</i>	101
18. Integrating Reliability and Uncertainty in Process Synthesis <i>Ying Chen, Yixin Ye, Ignacio E. Grossmann, Bingzhen Chen</i>	107
19. Optimization of Graded Bed Reactors for Syngas to Olefin (STO) Processes <i>Can Ekici, Lorenz T. Biegler, Christopher Ho, Joseph F. DeWilde, Dylan Kipp, Paul M. Witt</i>	115
20. Gas Lift Optimization for Optimum Oil Production from a Well Platform <i>Rahul Sudhanshu, Nitin Dutt Chaturvedi</i>	123
21. Techno-economic evaluation of sorption enhanced steam gasification of PKS system for syngas using CaO for CO₂ capture <i>Muhammad Shahbaz, Ahmed AlNouss, Suzana Yusup, Gordon Mckay, Tareq-Al Ansari</i>	129
22. A Robust Optimization Approach for Hydrogen Allocation Network with Parametric Uncertainties <i>Gaurav Shukla, Nitin Dutt Chaturvedi</i>	135

23. Process Design and Techno-economic analysis of Hydrogen Production using Green Ammonia Imported from Australia- A Korea Case Study	
<i>Malik Sajawal Akhtar, Jay Liu</i>	141
24. Life Cycle Assessment of Hydrogen Production from Imported Green Ammonia: A Korea Case Study	
<i>Malik Sajawal Akhtar, Jay Liu</i>	147
25. Optimal Design of Distillation Columns Using a Combined Optimisation Approach	
<i>Dian Ning Chia, Fanyi Duanmu, Eva Sorensen</i>	153
26. Process and product design for the simultaneous synthesis of xylitol and sorbitol from biomass	
<i>Guillermo Galán, Mariano Martín, Ignacio E. Grossmann</i>	159
27. An approach for simultaneous computer-aided solvent design and process design for CO₂ chemical absorption process	
<i>Ye Seol Lee, Amparo Galindo, George Jackson, Claire S. Adjiman</i>	167
28. Simulation-based Derivative-free Optimization for Hybrid Separation Design	
<i>Shuang Xu, Selen Cremaschi, Mario R. Eden, Anjan K. Tula</i>	173
29. Material Screening for Thermochemical Energy Storage in Solar Power Systems	
<i>Ishan Baja, Xinyue Peng, Christos T. Maravelias</i>	179
30. Network design optimization of waste management systems: the case of plastics	
<i>Rafael Castro-Amoedo, Alina Dahmen, Ana Barbosa-Povoa, Francois Maréchal</i>	185
31. Rational Design of Ionic Liquid Phase-Change Material for Efficient Thermal Energy Storage	
<i>Teng Zhou, Huaiwei Shi, Kai Sundmacher</i>	191
32. Tailor-made solvents by integrated design of molecules and CO₂ absorption processes	
<i>Johannes Schilling, Madlen Hopp, Joachim Gross, André Bardow</i>	197
33. Small-scale GTL Applications with Heat Integration in Reforming and Fischer-Tropsch Stages	
<i>Lígia B. Dias, Frederico S. Coelho, Fernando G. Martins, José Carlos B. Lopes, Mariana G. Domingos</i>	203
34. Automated Process Synthesis Using Reinforcement Learning	
<i>Quirin Göttl, Dominik Grimm, Jakob Burger</i>	209

35. **Operational Analysis of a Proton-Conducting Solid Oxide Electrolysis Cell for Synthetic Fuel Production**
Jinjutha Bhichaiphab, Dang Saebea, Amornchai Arpornwichanop, Yaneeporn Patcharavorachot 215
36. **Octane Optimization with a Combined Machine Learning and Optimization Approach**
Ahmet Can Serfidan, Metin Türkay 221
37. **A Platform of Machine Learning-Based Next-Generation Property Estimation Methods for CAMD**
Abdulelah S. Alshehri, Anjan K. Tula, Lei Zhang, Rafiqul Gani, Fengqi You 227
38. **Machine Learning Analysis of the Feed Water Parameters Affecting Reverse Osmosis Membrane Operation**
Çağla Odabaşı, Pelin Döloğlu, Fatih Gülmez, Gizem Kuşoğlu, Ömer Çağlar 235
39. **Novel technologies for low-quality natural gas purification**
Giorgia De Guido, Matteo Gilardi, Laura A. Pellegrini 241
40. **The Influence of Cuttings Polydispersity on Wellbore Cleaning Operations using the Kinetic Theory of Granular Flow and the Dense Discrete Phase Model**
Emmanuel I. Epelle, Dimitrios I. Gerogiorgis 247
41. **A Digital Twin for Process Optimisation in Pharmaceutical Manufacturing**
Dimitrios I. Gerogiorgis, Daniel Castro-Rodriguez 253
42. **Development of Hybrid RTO approaches in the absence of a rigorous dynamic model by the use of Hammerstein model structures**
Pedro de A. Delou, Rodrigo Curvelo, Maurício B. de Souza Jr., Argimiro R. Secchi 259
43. **Study of alternative configurations for methanol purification**
Stefania Moioli, Laura A. Pellegrini 267
44. **Power-to-OME1 via Direct Oxidation of Methanol: Process Design and Global Flowsheet Optimization**
Dominik Bongartz, Jannik Burre, Anita L. Ziegler, Alexander Mitsos 273
45. **A Machine Learning Approach for Device Design from Materials and Operation Data**
Luis A. Briceno-Mena, Gokul Venugopalan, Christopher C. Arges, Jose A. Romagnoli 279
46. **A Computer-Aided Platform for Simultaneous Process Synthesis and Intensification**
Mohammed Sadaf Monjur, Salih Emre Demirel, Jianping Li, M. M. Faruque Hasan 287

47. **Studying the effect of solvent properties in treatment design within CO₂ Integration Networks**
Dhabia Al-Mohannadi, Sabla Y. Alnouri 295
48. **A Multi-generation system in Sabkha-Tah operating with nanofluids and a novel RO unit.**
Yasser M. Abdellatif, Eric C. Okonkwo Yusuf Bicer Tareq AlAnsari 301
49. **A POD-ROM Methodology for Optimal Structuring of a Non-isothermal Fixed-bed Reactor for Process Integration**
Katarzyna Bizon, Gaetano Continillo, Enrico A. Cuttillo, Alfonso D'Onofrio 307
50. **Effect of Pseudo-Homogeneous Concentration Based Kinetics on Hybrid Reactive Distillation Columns for Selectivity Engineering**
Deepshikha Singh, Rajdeep Mukherjee, Harinder Singh, Ankur Gaur, Shabih Ul Hasan 313
51. **An MILP model of post-combustion carbon capture based on detailed process simulation**
Lukas Weimann, Guus Dubbink, Louis van der Ham, Gert Jan Kramer, and Matteo Gazzani 319
52. **Numerical study of fertilizer granules dynamics within rotary drum granulator**
Safae Elmisaoui, Lhachmi Khamar, Saad Benjelloun, Mohamed Khamar, Jean-Michel Ghidaglia 327
53. **Application of PharmaPy in the digital design of the manufacturing process of an active pharmaceutical ingredient**
Daniel Casas-Orozco, Daniel Laky, Vivian Wang, Mesfin Abdi, Xin Feng, Erin Wood, Gintaras V. Reklaitis, Carl Laird, Zoltan K. Nagy 333
54. **Computer-Aided Screening of Deep Eutectic Solvent Systems for the Associative Extraction of α -Tocopherol from Deodorizer Distillate**
Zhen Song, Teng Zhou, Zhiwen Qi, Kai Sundmacher 341
55. **Investigating critical design criteria for photo-electrochemical reactors**
Mohammed Al-Breiki, Yusuf Bicer 347
56. **Advanced exergoeconomic analysis of a double effect evaporation process in an instant coffee plant**
Tinoco-Cacedo D.L., Feijoó-Villa E., Calle-Murillo J., Lozano-Medina A., Blanco-Marigorta A.M. 353
57. **Energy-efficient membrane cascades for industrial separations**
Jose Adrian Chavez Velasco, Zewei Chen, Radhakrishna Tumbalam Gooty, Mohit Tawarmalani and Rakesh Agrawal 359

58.	Techno-economic evaluation of Carbfix technology for carbon storage in the US <i>María Paz Etcheverry, Haoshui Yu, Emre Gençer</i>	365
59.	Computational Exploration of Adsorption Enhanced Haber-Bosch using MOFs and Ionic Liquid/MOFs <i>Amro M. O. Mohamed, Yusuf Bicer</i>	373
60.	Optimization of a Continuous Fluidized Bed Process for the Separation of Enantiomers by Preferential Crystallization <i>Nadiia Huskova, Michael Mangold, Andreas Seidel-Morgenstern</i>	381
61.	Methanol and power production from waste tyres: a process synthesis approach <i>Athi-enkosi Mavukwana, Ephraim Llane, James Fox, Baraka Celestin Sempuga</i>	387
62.	Analysing the robustness of multi-stage bioseparations to measurement errors <i>Robert Milton, Emma Chandler, Solomon F. Brown</i>	393
63.	Improvements in Methyl Ethyl Ketone Production Through Intensified Processes <i>Eduardo Sanchez-Ramirez, Heriberto Alcocer-Garcia, Ana Gabriela Romero-Garcia, Gabriel Contreras-Zarazua, Juan Gabriel Segovia-Hernandez</i>	399
64.	Flexibility Assessment of Spirits Distillation Processes: Focus on the Armagnac Distillation <i>Alessandro Di Pretoro, Michel Meyer, Marie-Claude Segur, Xavier Joulia</i>	407
DATA ANALYTICS, MODELING, OPTIMIZATION AND SIMULATION		
65.	OUTDOOR – An open-source superstructure construction and optimization tool <i>Philipp Kenkel, Timo Wassermann, Celina Rose, Edwin Zondervan</i>	413
66.	Probabilistic Modeling for Optimization of Bioreactors using Reinforcement Learning with Active Inference <i>Ernesto C. Martínez, Jong Woo Kim, Tilman Barz, Mariano N. Cruz Bournazou</i>	419
67.	SPH Simulation of the Rising of a Kerosene Bubble in a Water Column: Effect of the Injection Condition <i>M. Robles-Santacruz, C. E. Alvarado-Rodríguez, F.I. Gómez-Castro, A. R. Uribe-Ramírez</i>	425
68.	Novel method for properties prediction of pure organic compounds using machine learning <i>Nattasinee Chorbngam, Rathachai Chawuthai, Amata Anantpinijwatna</i>	431
69.	Self-learning surrogate models in superstructure optimization <i>Julia Granacher, Ivan Daniel Kantor, Michel Lopez, Francois Maréchal</i>	439

70.	Data-driven Modeling of an Industrial Ethylene Oxide Plant: Superstructure-based Optimal Design for Artificial Neural Networks	
	<i>Hasan Sildir, Sahin Sarrafi, Erdal Aydin</i>	445
71.	Novel Tool for Selecting Surrogate Modeling Techniques for Surface Approximation	
	<i>Bianca Williams, Selen Cremaschi</i>	451
72.	ProREFD: Tool for Automated Computer-Aided Refrigerant Design, Analysis, and Verification	
	<i>Kornkanok Udomwong, Arisa Robin, Nichakorn Kuprasertwong, Orakotch Padungwatanaroj, Anjan Kumar Tula, Lingyu Zhu, Liyang Zhou, Binhui Wange, Shuhua Wang, Rafiqul Gani</i>	457
73.	Event driven analysis to enhance model calibration of experiments with high offline sampling rates	
	<i>M. Adnan JOUNED, Julian KAGER, Christoph HERWIG, Tilman BARZ</i>	463
74.	Process System Engineering Tool Integration in the Context of Industry 4.0	
	<i>Manuel Rodríguez Hernández, Ismael Díaz Moreno</i>	469
75.	Benchmarking of Surrogate Models for the Conceptual Process Design of Biorefineries	
	<i>Nikolaus I. Vollmer, Resul Al, Gürkan Sin</i>	475
76.	Predicting the Estrogen Receptor Activity of Environmental Chemicals by Single-Cell Image Analysis and Data-driven Modeling	
	<i>Hari S. Ganesh, Burcu Beykal, Adam T. Szafran, Fabio Stossi, Lan Zhou, Michael A. Mancini, Efstratios N. Pistikopoulos</i>	481
77.	Data-driven prognostics for Lithium-ion battery health monitoring	
	<i>Sukanya G, Resmi Suresh, Raghunathan Rengaswamy</i>	487
78.	The Adaptive Radial Basis Method (ARBM): An application to the hydrodynamics of liquid-liquid dispersions	
	<i>Samer Alzyod</i>	493
79.	Process modeling and multi-criteria optimization of an industrial phosphoric acid wet-process	
	<i>Ilias Bouchkira, Abderrazak M. Latifi, Lhachmi Khamar, Saad Benjelloun</i>	499
80.	Towards Automated HAZOPs	
	<i>Calum Robinson, Solomon Brown, Joan Cordiner</i>	505
81.	Multi-objective optimization of co-processing of algae hydrothermal liquefaction oil and vacuum gas oil: a survey of algal oil co-processing ratio and gasoline selling price	
	<i>Shuai Zhang, Qingyu Lei, Le Wu, Yuqi Wang, Lan Zheng</i>	511

82. Modeling of a software for sizing a wastewater treatment plant <i>Anass Qoutbane, Btissam Khmajer, Essediya Cherkaoui, Mohamed Khamar, Abderrahman Nounah</i>	517
83. Joint Chance Constraint Approach based on Data-Driven Models for Optimization Under Uncertainty applied to the Williams-Otto Process <i>Byungjun You, Erik Esche, Joris Weigert, Jens-Uwe Repke</i>	523
84. CFD Simulations of High Efficiency Gas Cyclones: An Influence of Dustbin Geometry <i>Pitiwat Pechmanee, Apinan Namkanisorn, Santi Wattananusorn, Eakarach Bumrunghthaichaichan</i>	529
85. InvQMOM: A Simple Inversion Method that Works <i>Menwer Attarakih, Armin Fricke, Hans-Jörg Bart</i>	535
86. A Dual-Layer Attention-Based LSTM Network for Fed-batch Fermentation Process Modelling <i>Kai Liu, Jie Zhang</i>	541
87. Resilience and Survivability-aware Optimal Design and Operation of Interconnected Supply Chains <i>Spyridon D. Tsolas, M. M. Faruque Hasan</i>	549
88. A Hybrid Modeling Approach for Liquid Entrainment Fraction Predictions <i>Yushi Deng, Selen Cremaschi, Mario R. Eden, Carlos Avila, Haijing Gao, Ivan Mantilla</i>	555
89. Rethinking Balance Models Applied to Pfrs And Extracting New Insights <i>João Manzi</i>	561
90. Digital Twin Model Development for Chemical Plants Using Multiple Time-Steps Prediction Data-Driven Model and Rolling Training <i>Jia-Lin Kang, Somayeh Mirzaei, Yao-Chen Lee, Yao-Cheng Chuang, Marvin Frias, Cheng-Huang Chou, San-Jang Wang, David Shan Hill Wong, Shi-Shang Jang</i>	567
91. Parameter estimation and sensitivity analysis for a diesel hydro-processing model <i>Esin Iplik, Panagiotis Tsirikoglou, Ioanna Aslanidou, Konstantinos Kyprianidis</i>	573
92. Model Development of Amine Regeneration Process with Electrodialysis Reclamation Unit <i>Jonghun Lim, Jesung Lee, Hyungtae Cho, Junghwan Kim</i>	579
93. CO₂ methanation: Reactor modelling and parametric analysis <i>Erasmus Mancusi, Piero Bareschino, Annunziata Forgione, Claudio Tregambi, Francesco Pepe</i>	585

94. **An uncertainty-aware hybrid modelling approach using probabilistic machine learning**
Rasmus Fjordbak Nielsen, Nima Nazemzadeh, Martin Peter Andersson, Krist V. Gernaey, Seyed Soheil Mansouri 591
95. **Modelling and Dynamic Optimization of beer fermentation towards optimal flavor and operation**
Carina L. Gargalo, Alex Matesanz Salvadó, Ulrich Krühne, Krist V. Gernaey 599
96. **On the fitting of bathtub-shaped failure models to lifetime data for selective maintenance optimization**
Teemu J. Ikonen, Iiro Harjunkoski 605
97. **Sampling Space Reduction for Data-driven Modelling of Batch Distillation - Introducing Expert Process Knowledge through Operation Recipes**
Gerardo Brand-Rihm, Erik Esche, Jens-Uwe Repke 611
98. **On the acceleration of global optimization algorithms using a cutting plane decomposition and machine learning**
Asimina Marousi, Antonis Kokossis 617
99. **Towards a Digital Twin: A validated Digital Model of a Pilot Scale Bioreactor producing Green Fluorescent Protein (GFP)**
Ole Laustsen Lave, Carina L. Gargalo, Ryan Barton, Christoph Bayer, Isuru A. Udugama, Krist V. Gernaey 625
100. **Automated Box-Jenkins Methodology to Forecast the Prices of Crude Oil and Its Derivatives**
Ahmet Can Serfidan, Gurkan Ozkan, Metin Türkay 631
101. **Systematic Modelling of Transport Processes across Interfaces**
Robert Pujan, Heinz A. Preisig 637
102. **An identification approach to a reaction network for an ABE catalytic upgrade**
Elham Ketabchi, Tomas Ramirez Reina, Bogdan Dorneanu, Harvey Arellano-Garcia 643
103. **MOSKopt: A simulation-based data-driven digital twin optimizer with embedded uncertainty quantification**
Resul Al, Gürkan Sin 649
104. **Global Optimization of Mixed-Integer Polynomial Programs via Quadratic Reformulation**
Tanuj Karia, Claire S. Adjiman, Benoît Chachuat 655
105. **A data-based approach for the development of asynchronous and distributed optimization algorithms in complex engineering applications**
Ioannis Geroulis, Georgios Panayiotou, Antonis C. Kokossis 663

Preface

This book contains papers presented at the 31st European Symposium of Computer Aided Process Engineering (ESCAPE-31), held in Istanbul, Turkey, from June 6-9, 2021. The ESCAPE series started in 1992 at Elsinore, Denmark, on a strong foundation of 23 events of the European Federation of Chemical Engineers (EFCE) Working Party on Computer Aided Process Engineering (CAPE). The most recent symposia were organised in Eindhoven, The Netherlands 2019 (ESCAPE-29), and Milano, Italy 2020 (ESCAPE-30).

The ESCAPE series serves as a forum to bring together scientists, researchers, managers, engineers, and students from academia and industry, who are interested in CAPE and Process Systems Engineering (PSE). The scientific aim of the symposium is to present and review the latest developments in CAPE and/or PSE. This research area bridges fundamental chemical and biological sciences with various aspects of process and product engineering. The objective of ESCAPE-31 is to promote the use of a systems approach, highlighting the traditional CAPE-PSE topics as well as the emerging CAPE-PSE topics of current interest and future significance.

The main focus for ESCAPE-31 is on the methodical approaches in process systems engineering with emphasis on circular engineering economy, and sustainable product and/or process development. The papers at ESCAPE-31 are arranged in terms of the following themes: Process-product synthesis, design and integration; Methods, models and computational tools for PSE; Process control and operations; CAPE/PSE in energy/water/food nexus and sustainability; Process operations and supply chains; PSE in biological systems and processes; Education in CAPE/PSE & knowledge transfer.

A total of 320 contributed 6-pages papers are included in this book. All the papers have been reviewed and we thank the members of the international scientific committee for their evaluations, comments and recommendations. It was a very difficult task since we started with more than 452 submitted abstracts. The selection process involved review of abstracts, review of manuscripts and final selection of the revised manuscript.

We hope that this book will serve as a valuable reference document to the scientific and industrial community and that it will contribute to the progress in computer aided process and product engineering.

Metin Turkey
Rafiqul Gani

9 March 2021

Theme Coordinators

Theme	Name	Country
Process-product synthesis, design and integration	Edwin Zondervan	The Netherlands
	Dimitrios Gerogiorgis	UK
Methods, models and computational tools for PSE	Alexander Mitsos	Germany
	Erdal Aydin	Turkey
Process control and operations	Filip Logist	Germany
	Devrim Kaymak	Turkey
CAPE/PSE in energy/water/food nexus and sustainability	Michael Georgiadis	Greece
	Alper Uzun	Turkey
Process operations and supply chains	Ana Barbosa-Povoa	Portugal
	Jose Pinto	USA
PSE in biological systems and processes	Maria Papathanasiou	UK
	Halil Kavakli	Turkey
Education in CAPE/PSE & knowledge transfer	Antonio Espuña	Spain
	Seyed Mansouri	Denmark

National Organizing Committee

Metin Turkey (Turkey, Chair)
Erdal Aydin (Turkey, Co-chair)

Name	Affiliation	Name	Affiliation
Yaman Arkun	Koç University	Halil Kavakli	Koç University
Alper Uzun	Koç University	Burak Alakent	Bogazici University
Devrim Kaymak	Istanbul Tech. Univ.	Goktug Ahunbay	Istanbul Tech. Univ.
Halil Kalipcilar	METU	Enis Tolga Eroglu	KMO
İbrahim Akyurek	KMO		

International Scientific Committee

Metin Turkey (Turkey, Chair)
Rafiqul Gani (Denmark, Co-chair)

Name	Country	Name	Country
Elvis Ahmetović	Bosnia and Herzegovina	Burak Alakent	Turkey
Sabla Y. Alnouri	Lebanon	Yaman Arkun	Turkey
Olga Arsenyeva	Ukraine	Suttichai	Thailand
		Assabumrungrat	
Erdal Aydin	Turkey	Ana Barbosa-Povoa	Portugal
Andre Bardow	Switzerland	Mani Bhushan	India
Lorenz T. Biegler	USA	Giulia Bozzano	Italy
Jose Caballero	Spain	Pedro Castro	Portugal
Xi Chen	China	Shoukat Choudhury	Bangladesh

Selen Cremaschi	USA	Mario Eden	USA
Sebastian Engell	Germany	Antonio Espuña	Spain
Ferenc Friedler	Hungary	Emre Gencer	USA
Michael Georgiadis	Greece	Dimitrios Gerogiorgis	UK
Chrysanthos Gounaris	USA	Ignacio E. Grossmann	USA
Gonzalo Guillen-Gosalbez		Iiro Harjunoski	Germany
Faruque Hasan	USA	Vassily Hatzimanikatis	Switzerland
Gabriela Henning	Argentina	Rene Hoffmann	Austria
Elcin Icten	USA	Marianthi Ierapetritou	USA
Xavier Joulia	France	I A Karimi	Singapore
Halil Kavakli	Turkey	Devrim Kaymak	Turkey
Achim Kienle	Germany	Chris Kieslich	USA
Jiří Jaromír Klemesš	Czech Republic	Antonis Kokossis	Greece
Emilia Kondili	Greece	Zdravko Kravanja	Slovenia
Abderrazak Latifi	France	Jay Lee	South Korea
Patrick Linke	Qatar	Filip Logist	Germany
Montastruc Ludovic	France	Sandro Macchietto	UK
Emanuele Martelli	Italy	Henrique Anibal Santos de Matos	Portugal
Thokazani Majazi	South Africa	Flavio Manenti	Italy
Seyed Mansouri	Denmark	Christos Maravelias	USA
Francois Marechal	Switzerland	Mariano Martin	Spain
Ruth Misener	UK	Alexander Mitsos	Germany
Il Moon	South Korea	Iqbal Mujtaba	UK
Nuno Oliveira	Portugal	Lazaros Papageorgiou	UK
Stavros Papadokonstantakis		Maria Papathanasiou	UK
Jose Pinto	USA	Efstratios Pistikopoulos	N. USA
Valentin Plesu	Romania	Heinz A Preisig	Norway
Stefan Radl	Austria	Manojkumar Ramteke	India
Rex Reklaitis	USA	Mauricio Sales-Cruz	Mexico
Panos Seferlis	Greece	Gurkan Sin	Denmark
Sigurd Skogestad	Norway	Rajagopalan Srinivasan	India
Hirokazu Sugiyama	Japan	Anjan Tula	China
Alper Uzun	Turkey	Petar Varbanov	Czech Republic
Dimitrios Varvarezos	USA	Franz Winter	Austria
Yoshiyuki Yamashita	Japan	Yuan Yao	Taiwan
Fengqi You	USA	Shihong Yuan	China

Victor Zavala	USA	Lei Zhang	China
Jinsong Zhao	China	Teng Zhou	Germany
Edwin Zondervan	The Netherlands		

Conference Secretariat

Koç University, College of Engineering, Rumelifeneri Yolu, Sariyer 34450 Istanbul
Turkey

Explanatory and Predictive Analysis of Naphtha Splitter Products

Ahmet Can Serfidan,^{a,b} Metin Türkay,^b

^a*TUPRAS, Address, Petrol Cd. No:25 D:No:25, Turkey*

^b*Koc University, Rumelifeneri, Sariyer Rumeli Feneri Yolu*

Abstract

Refinery operations are always prone to optimization, and due to the increasingly adverse effects of COVID-19 on energy sectors, its importance increased significantly. This work aims to predict the naphtha column KPI parameters with high accuracy so that operators make corrective actions efficiently. Although linear regression provides acceptable results for prediction, this is not the case for top and bottom product C7 and C6 prediction in the central Naphtha Splitter column. First, we did gather all the available data to overcome this problem, which can affect the top and bottom products. Including upstream units that feed the column. Instead of one common technique (linear regression), we used five additional machine learning methods: Adaboost, support vectors, kNN, random forest, XGboosting. Since there are many measurements, however, very few samples need to reduce dimensions before modeling. We used BorutaSharp to select the essential features. We also use classification machine learning methods to categorize bottom products since there is no need to predict the value instead of whether the value is higher or lower than a constant. Overall, we achieved 30% higher accuracy than the traditional ways for the top product, and we reached to predict C6 content in the bottom with higher accuracy than 80%. Xgboost provides the best regression model, and stochastic gradient boosting yields the best classification model. After our implementation, the energy consumption is decreased significantly, and 100k\$/month is saved since we can monitor top and bottom products simultaneously.

Keywords: naphtha splitter column, quality estimator, soft sensor.

1. Introduction

In the era of COVID-19, some of the most negatively affected sectors are petrochemical, refining, energy, etc., due to the most significant demand destruction events the world has seen. The refining companies experienced a sharp drop in margins and experienced substantial financial losses. This time optimization, whether it is minimizing the product giveaway, maximizing the energy utilities, becomes the most critical topic.

Naphtha Splitter (NS) columns are used to separate light naphtha from heavy naphtha. The naphtha splitter column's key performance indicator is the qualities of the bottom and top products. Namely, the top product is light naphtha, and its quality indicator is hydrocarbon content, which is measured by ASTM D5134 method. This method provides with C1%, C2%, C3%, C4%, C5%, C6%, C7%, C7+%, where, for example, C4% means percent of C4 hydrocarbons, mainly 1,3-butadiene, 2-methylpropene, butanes and butanes. The most important measure is the C7% content in the top product since beyond a threshold value of C7%, profit decreases drastically. The bottom product is heavy

naphtha, and its quality indicator again hydrocarbon content. The most important quality parameter is the C6% content since a higher C6 fraction will strongly affect the downstream of the naphtha splitter column (catalytic cracking units), resulting in benzene more than regulation limits. The process operators control the unit by monitoring C6 and C7 compositions in the bottom and top products. However, the samples are only taken twice a week, so an operator's general tendency is to keep the column condition the same as before. And this yields to stay in the over-safe zone for the unit. In other words, the product gives away. This work aims to predict the KPI parameters with high accuracy to make corrective actions efficiently.

2. Methodology

Optimization of NS depends on the refinery configuration and operating characteristics of refineries. Despite the differences, the main aim is to find the best conditions to minimize product losses. There are many obstacles in optimizing primary NS columns. One of them is the uncertainty in-feed content. They generally have a very complex feed network. Some of the feed flows are usually unknown since the flow meter is placed before the pipeline junction. For example, the bottom product of the debutanizer in the hydrocracker unit can go either LSRN tank or unifier feed (cannot feed them simultaneously). Although feed flow rate is known, there is no direct data that indicates which unit is feed by that flow.

Also, there is no measurement for the tank, and the main NS operation cannot know which type of feed they are operating until they see the effect (temperature, liquid level changes, etc.)

Other than feed conditions, there are disturbances to columns like weather conditions. For example, in the summer, the operation conditions have to change to keep the pressure stable. Based on the disturbances above, even if there is no other change in the main NS column, its product quality will change significantly. You will see the last two years' C7 and C6 results of the main NS column below.

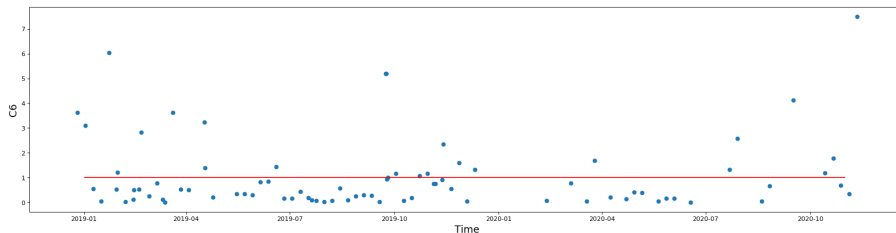


Figure 1 Bottom C6 Content

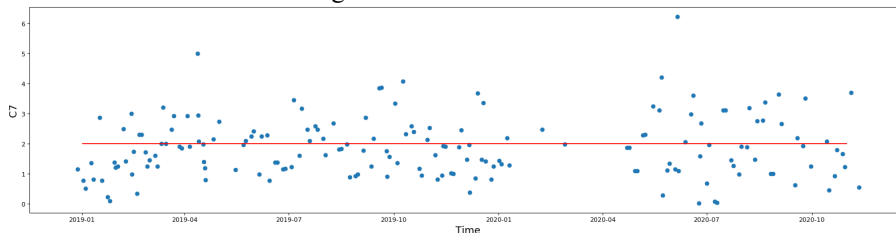


Figure 2 Top C7 Content

Here are our aims: (i) There is a hard process limit on C6, 1. As long as C6 content does not exceed this limit, there is no problem. Otherwise, the subsequent unit will have a benzene problem. This will yield to the environmental problem, which is unacceptable. (ii) There is no limit on C7. However, since it is impure to the isomerization reactor, the aim is to send it low as much as possible. Otherwise, the temperature needed to compensate for the impurity will increase, yielding enormous economic loss. As seen in the above graph, operations are carried out such that C7 content is 2.0 on average. As we decrease the C6 at the bottom, we will increase the C7 on top and vice versa. Since the acceptable limit is 1 for C6, the optimum condition for main NS: (i) Lower the energy given the column as long as C6 does not exceed the threshold, (ii) Keep C7 minimum as much as possible. Currently, there is no inferential for C6, and C7 inferential mean absolute error is around 1.5

Since there is no way to exceed the C6 limit, they try to give more energy columns more than necessary. As you see, there is almost no exceeding limit but also many zeros. This is mainly because they want to stay in a safe zone as much as they can.

In this project, I aim to predict C6 and C7 with great accuracy so that the main NS column will be operated optimally. And as previously discussed, it is needed to infer the tank's content. However, this requires the following preparations: (i) LSRN T95 inferential of each unit sent to the tank or main NS column (ii) Heuristic solution for which units sent their LSRN to the tank.

Five different units can feed the main NS column. Each has different operating conditions, different lsrn specifications, and different constraints. But all of them are binary distillation columns. To infer LSRN, I took those as possible candidates: input flow rates, input temperature, all column temperatures, all column pressures, tray temperature controller, steam feed flows, reflux flow rates, bottom flow rate, distillate flow rate. Also, I generated these features: reflux overfeed, temperature differences between bottom and top of the columns, distillate over the bottom. We were able to get 200-400 rows of data for each unit.

I used BarutoSharp to select the most important features. It is a wrapper feature selection method that uses both the Boruta feature selection algorithm with Shapley values. It successfully reduces the column number from 30 to 5-10[1]

After that, I used linear regression to predict LSRN T95. However, I want to keep minimum columns while minimizing the mean. Here is the algorithm for selecting the best features to minimize error with the minimum number of features.

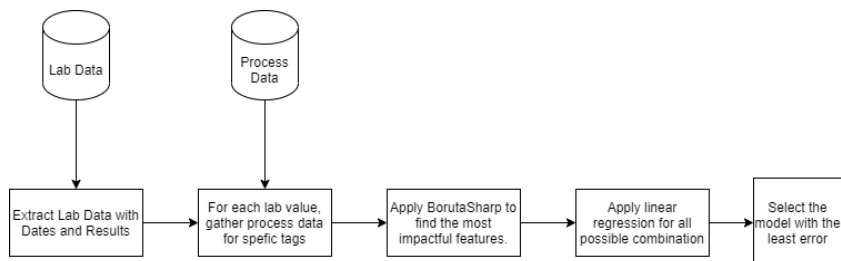


Figure 3 The algorithm for feed content prediction

The missing or misleading measurements are no surprise in refineries. In this optimization study, I have to also deal with this problem. Data by itself does not indicate whether the

flow is sent to the tank or some other unit. Here is the more detailed version of the feed network and all possible injections:

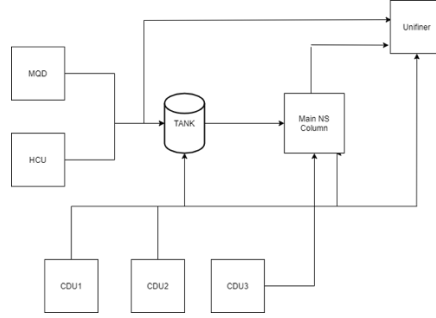


Figure 4 Feed Network

The heuristic optimization to find which feed goes to which unit is:

$$\begin{aligned} \text{minimize } F_{out} - (F_0 + \sum_{i=1}^n F_i * X_i) \\ X_i \in \{0,1\} \end{aligned}$$

Where F 's flows and X 's are binary variables indicating whether the flow is present or not.

After finding heuristic values for those unknowns and this data preparation process, we finally have the most important parameters that can affect top C7 content and bottom C6 content. Namely, the feed content and column operations. There might be other factors that can affect the target, like weather conditions. However, those disturbances will also change the column measurements. Thereby, even if we do not involve outlet temperature directly, we will know its change by one of the temperature/pressure measurements in the top part of the column.

To predict C7 and C6 content, I follow a very similar approach to the above methodology. But this time, in addition to linear regression [2], I make use of other powerful machine learning algorithms, specifically: Random Forest [3], AdaBoost, XGBoost [4], KNN, SVR[5]. Again, BorutaSharp is used to select the most important features, and then for each algorithm, I give all the important elements to the machine learning algorithm. Finally, this time, instead of trying all possible combinations of features, I tried different range of hyperparameters.

For the bottom content prediction, we followed a different approach. The main aim is not to exceed the process high limit. Since we will decrease the energy input to column bottom product quality impurity will get higher. We keep lowering the reboiler feed until we are close to the limit. So, instead of a regression problem, this becomes a classification one. To predict whether it is higher than the limit or lower, we used the following algorithms: Logistic Regression, kNN, Random Forest, SVM.

3. Results and Discussion

All of the algorithms are applied using python programming language between 2018 and 2020 of data gathered and implemented by our algorithms. The number of rows

As shown from the above table, all modeling algorithms yield almost the same result. However, explaining and updating Linear Regression is more accessible than the others in refinery conditions. Since lab results are taken daily, we can update the bias term if there is a significant process change.

And for the C6 prediction:

Table 3 C6 Classification Performance

Model	Accuracy
LogisticRegression	0.67
KNN	0.75
RF	0.78
SVM	0.80

4. Conclusions

Optimization in refinery has become one of the most important topics, and its importance is rapidly increasing due to the adverse effect of COVID-19 on the energy sector. We achieved this work to optimize column conditions by predicting the top and bottom product qualities with machine learning techniques. The winner algorithm in our case is Xgboost. And for C6 classification, it is SVM. Predicting the top and bottom content simultaneously gave us the leverage of optimizing the column operation without exceeding the process limits. In the previous case, since the column operations are at the safe-zone limits, we achieved 100k\$/month by lowering the energy cost.

References

- [1] Kursa, Miron B., and Witold R. Rudnicki. "Feature Selection with the Boruta package." *J Stat Softw* 36.11 (2010): 1-13.
- [2] Souza, F. A., Araújo, R., & Mendes, J. (2016). Review of soft sensor methods for regression applications. *Chemometrics and Intelligent Laboratory Systems*, 152, 69-79.
- [3] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- [4] Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015). Xgboost: an extreme gradient boosting. *R package version 0.4-2*, 1-4.
- [5] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222.

Multi-objective optimization of the supply chain for the production of biofuels and high value-added products in Mexico: importance of the water footprint

Yulissa M. Espinoza-Vázquez,^a Fernando Israel Gómez-Castro^{a,*}, José María Ponce-Ortega^b

^a *Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Campus Guanajuato, Universidad de Guanajuato, Noria Alta S/N, Col. Noria Alta, Guanajuato, Guanajuato, 36050, México.*

^b *División de Estudios de Posgrado, Facultad de Ingeniería Química, Universidad Michoacana de San Nicolás de Hidalgo, Francisco J. Múgica S/N, Morelia, Michoacán. 58060, México.*

fgomez@ugto.mx

Abstract

Mexico is a country with a wide agricultural production, which implies a high production of lignocellulosic residues. Those residues are potential raw materials to produce biofuels as bioethanol and biobutanol, similarly, such residues can be used to produce high value added bioproducts, as furfural, levulinic acid, among others. The use of biomass to produce only biofuels is not economically feasible; therefore, the use of a fraction of biomass to obtain bioproducts with high-added value may enhance the economy in the biofuels industry. To ensure obtaining as much profit as possible, a proper supply chain must be determined. Nevertheless, when considering biomass utilisation, the use of water must be taken into account, since these processes involve a large amount of that resource. Moreover, the availability of water is uncertain, thus this issue must be included in the decision-making process for such supply chain. For this reason, in this work a mathematical model representing the supply chain of the production of biofuels/bioproducts is proposed, considering economic and water footprint aspects. The model is relaxed to a MILP equivalent and solved using the software GAMS, aiming to the maximization of the profit and minimizing the water footprint, while satisfying as much as possible of the gasoline demand in the country. The water impact is measured through the ratio between the amount of water required by the process and the amount of water available at the plant location. Pareto-optimal solutions were obtained using the ϵ -constraint method. According to the results obtained, the maximum profit is 572,360.7 thousand dollars per year, with a water footprint of 0.584. For that case, there are two plants: Guanajuato and Queretaro; which produce levulinic acid and bioethanol. On the other hand, the minimum for water footprint is 0.021. For this case, a single facility is located in Veracruz, where levulinic acid and bioethanol are produced.

Keywords: optimization, water footprint, supply chain, biofuels.

1. Introduction

The use of fossil fuels as energy source has had an enormous cost for humanity and the environment, from air and water pollution to global warming. That is a reason to resort to the concept of biofuels, which are obtained from renewable biomass (Dermibas, 2017) and are expected to show lower greenhouse gas emissions than fossil fuels. In Mexico, there is a high potential for the production of biofuels, mainly from residues of corn, sugar cane, sorghum, wheat and barley. More than 50% of the waste is used for soil protection, 27% is fed to livestock and 20% is burned (Damián-Huato et al., 2013). Therefore, there is a major interest in evaluating the energy potential of waste that has low food quality, which is mostly burned in the field. Examples of biofuels are bioethanol, which can be mixed up to 10% by volume with gasoline (Costagliola et al., 2016), and biobutanol, which can be used in blends of up to 16% by volume (Zhang et al., 2016). Industrial biofuel production is not economically competitive. Therefore, in addition to the production of biofuels, the generation of high value-added products is necessary to improve the economic potential of the biomass-based industry. Among the many potential high value-added products, levulinic acid can be mentioned, which is a great platform for a variety of products such as levulinate esters, etcetera (Pileidis and Titirici, 2016). Another product with high added value is furfural, which is a platform for other products of industrial importance, such as furfuryl alcohol, tetrahydrofuran, etcetera (Bhogeswararao and Srinivas, 2015). Among the different biofuels and bioproducts that can be obtained from lignocellulose biomass, a given combination of products should have the highest profitability, using the greatest amount of residues possible and satisfying part or all of the demand. Similarly, it is important to note that the production of biofuels requires a large amount of water, while the country has problems with water supply (Godínez Madrigal et al., 2018). Within the supply chains analyzed in the country, none of them includes the water aspect or high value-added products as potential derivatives from biomass (Rendon-Sagardi et al., 2014; Santibañez-Aguilar et al., 2014). Similarly, in the case of the work presented by Espinoza-Vázquez et al. (2020), only the profit was considered as objective function, leaving out a very important aspect in the biorefinery industry such as water. For this reason, this work proposes the optimization of the supply chain for the production of biofuels and high value-added products from lignocellulosic material, taking into account the economic and water-use aspects.

2. Case Study

The case study in this work is similar to the one presented in the previous report by Espinoza-Vázquez et al. (2020). The availability of agricultural residues in Mexico is determined for four main crops: corn, sorghum, wheat and barley. Data for the distribution of such crops in all the country, and their contribution to the production of lignocellulosic residues, has been collected from the Service of Agrifood and Fisheries Information (SIAP, 2020). These raw materials are proposed to be used for the production of bioethanol and/or biobutanol, as biofuel alternatives, and levulinic acid and furfural as high-value added products. For the potential location of the facilities, the states with the highest industrial infrastructure have been selected. Finally, the production of each potential product is constrained by its national demand.

3. Mathematical Model

This model is based on the use of four raw materials (i) from the 32 states in Mexico (j). Raw materials from any state are transported to one of the five states with the highest industrial infrastructure (k), where the facilities will be located. On each facility, four products (m) can be obtained: bioethanol, biobutanol, levulinic acid and furfural. Finally, the products can be distributed to ten markets (n), selected among the 32 states because they are the locations with higher demand for gasoline. The mathematical model used in this work is as follows: the total demand of a product m in the market n , $TD(m,n)$, in mass units, is given by:

$$TD(m,n) = PD(m,n) \cdot \rho(m), \forall m, n \quad (1)$$

Where $PD(m,n)$ is the volumetric demand of the product m in the market n , and $\rho(m)$ is the density of the product m . The mass production of each product m in the k installation for each n market, $FP(m,k,n)$, should not be greater than the total demand for that product in that market:

$$\sum_k FP(m,k,n) \leq TD(m,n), \forall m, n \quad (2)$$

The amount of raw material i obtained from the source j that enters the biorefinery k to produce m , $RMAF(i,j,k,m)$, cannot be greater than the availability of raw material i in the state j , $RMA(i,j)$. This is represented as follows:

$$\sum_k \sum_m RMAF(i,j,k,m) \leq RMA(i,j), \forall i, j \quad (3)$$

The cost of transporting the biomass i from the source j to the plant k , $TSP(i,j,k)$, is represented as:

$$TSP(i,j,k) = \sum_m RMAF(i,j,k,m) \cdot D_1(i,j,k) \cdot TCSB, \forall i, j, k \quad (4)$$

$TCSB$ is the cost of transporting lignocellulosic biomass, while $D_1(i,j,k)$ is the distance of biomass i from state j to biorefinery k . In this work, $TCSB$ is taken as 0.00508 kUSD/km·kt, according to the data reported for the transport of solid biomass (OAS, 1999). The cost of transporting the product m from the installation to the market n , $CTP(m,k,n)$, is given by:

$$CTP(m,k,n) = FP(m,k,n) \cdot D_2(k,n) \cdot TCLP, \forall m, k, n \quad (5)$$

where $D_2(k,n)$ is the distance from the facility k to the market n , and $TCLP$ is the transport cost of the liquid product. In this work, $TCLP$ is taken as 0.00671 kUSD/km·kt, according to the data reported for the transportation of liquid products (OAS, 1999). A logical variable $Y(k)$ is assigned to each facility; if the variable $Y(k)$ is true, then the biorefinery must be built at location k . Similarly, a logical variable $S(m,k)$ is assigned to each product, indicating if it is produced or not. The disjunction is represented as follows:

$$\left[\begin{array}{l} Y(k) \\ TCP(k) = \sum_m \sum_n FP(m,k,n) \cdot CP(m) \\ TCRM(k) = \sum_m \sum_n \frac{FP(m,k,n)}{Yld(i,m)} \cdot RMUC(i) \\ TCL(k) = \sum_m \sum_n \frac{FP(m,k,n)}{\rho(m)} \cdot LS(m,k) \cdot LC \\ \left[\begin{array}{l} S(m,k) \\ \sum_n FP(m,k,n) = A \end{array} \right] \vee \left[\begin{array}{l} \neg S(m,k) \\ \sum_n FP(m,k,n) = 0 \end{array} \right] \end{array} \right] \vee \left[\begin{array}{l} \neg Y(k) \\ TCP(k) = 0 \\ TCRM(k) = 0 \\ TCL(k) = 0 \end{array} \right], \forall k \quad (6)$$

Where:

$$A = RMAF(i, j, k, m) \cdot Yld(i, m) \quad (7)$$

$TCP(k)$ is the total production cost of the biorefinery k , $CP(m)$ is the unit production cost for the product m , $TCRM(k)$ is the cost due to the raw material entering the facility k , $Yld(i, m)$ is the product yield obtained from the raw material i , $RMUC(i)$ is the unit cost of raw material i , $TCL(k)$ is the cost of land acquisition for installation k , $LS(m, k)$ is the required land area per liter of product, LC is the cost of land per square meter. In this work, $RMUC(i)$ has been taken as 87.5 kUSD/kt, which is an estimated selling price for agricultural residues for animal feed (Mercado Libre, 2020). LC is taken as 0.4048 kUSD/m² while $LS(m, k)$ is defined as 0.66 m²/L. The values of LC and $LS(m, k)$ are estimated from the production cost, surface and land reported for the “Francisco I. Madero” refinery, located in Ciudad Madero, Tamaulipas, Mexico (El Informador, 2019). The multi-objective optimization problem is represented as:

$$FOBJ = \min(-GN, AH) \quad (8)$$

Where the profit GN , is given by:

$$GN = \sum_m \sum_k \sum_n FP(m, k, n) \cdot SP(m) - \sum_k TCRM(k) - \sum_k TCP(k) - \sum_k TCL(k) - \sum_i \sum_j \sum_k TSP(i, j, k) - \sum_m \sum_k \sum_n CTP(m, k, n) \quad (9)$$

where $SP(m)$ is the selling price of the product m . While the water footprint AH is given by:

$$AH = \sum_m \sum_k \frac{FP(m, k, n) \cdot AW(m)}{WAP(k)} \quad (10)$$

Where $WAP(k)$ is the amount of water available for industrial processes at the facility k (Comisión Nacional del Agua, 2018). This value was obtained from the amount of water destined for industrial use in each state. On the other hand, the parameter $AW(m)$ is the amount of water required to process the product m . The minimal production of bioethanol is fixed as 10%, forced to be satisfied. The disjunctions represented by equation 6 are relaxed using the Big-M approach (TresPalacios and Grossmann, 2015) and the model was coded in the GAMS software. The resulting MILP problem consists of 887 equalities and 456 inequalities. The model has been optimized with the solver BARON, in a HP computer with a Inter Core i5-8300H CPU, 12.00 GB of RAM.

4. Results

Table 1. Limits of objective functions.

Objective	Profit	Water Footprint
Max GN	572,360.7	0.58
Min AH	0	0.02

The two previously established objectives were taken into account, generating the Pareto fronts for their analysis, forcing the code to satisfy 10% of the bioethanol demand. As a first step, the limits of the objective functions were established through their individual solution, as shown in the Table 1. Subsequently, the ϵ -constraint method is used, generating the Pareto front with about 50 points with the Hammersley technique for a

complete visualization of the data (Fu and Diwekar, 2004). It is important to mention that each point of the Pareto front-end involved 2 minutes of computing time, for a total of 100 minutes for the entire Pareto front.

Analyzing the results shown in the Figure 1, it is observed that minimizing the water footprint completely comes at the expense of a drop in supply chain profitability. While increasing the profit in very small fractions, it leads to an increase in the water footprint. Therefore, it is observed that the objective functions are competing, since the benefit of an objective function, compromises the other. It is important to mention, that the minimum water footprint is not shown, because the corresponding profit is 0, since only bioethanol is produced, and the other of the graph tend

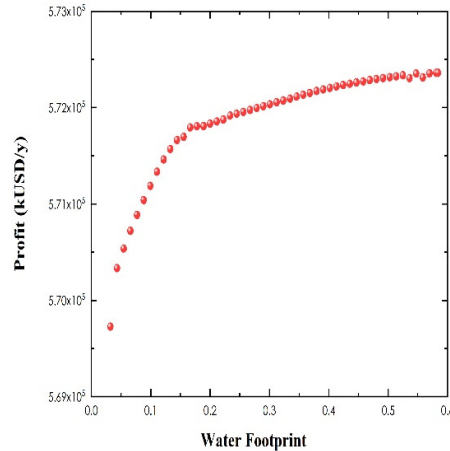


Figure 1. Pareto front: Profit Vs. Water Footprint.

to be lost due to the scale used. On the other hand, it can be seen that after the water footprint reaches a value of 0.15, the slope on the Pareto curve decreases, reducing the growing rate of the profit as the water footprint increases. Thus, a higher profit of the supply chain could be achieved if there were no restrictions on the use of water. It can be observed that the profitability tends to stabilize between $5.71 \cdot 10^5$ and $5.73 \cdot 10^5$, satisfying the required percentage of bioethanol demand and producing levulinic acid, without considerably increasing the water footprint. In terms of the supply chain, having the maximum profit, it is suggested the existence of a plant in Guanajuato, which produces bioethanol and levulinic acid, as well as a second facility located in Queretaro to produce levulinic acid. These plants fully satisfy the demand for the high value-added product and the 10% bioethanol restriction. This amount of bioethanol is only distributed to the same state where the plants are located, that is, Guanajuato and Queretaro. Finally, a greater quantity of corn waste from Aguascalientes and Jalisco is used, due to its high availability and proximity to the plants. On the other hand, the middle point, which satisfies both aspects, raises the existence of the four facilities, except Baja California. All four facilities produce bioethanol, while only Coahuila, Guanajuato and Veracruz produce levulinic acid, satisfying all the demand for the acid. As there are more plants, the amount of water available has lower impact, since as the production is distributed, it is not used in an excessive quantity. Similarly, there is a more specific distribution to the markets, for example, bioethanol produced by Coahuila goes to Nuevo Leon, while the product from Veracruz is sent the market in Puebla, and so on. This is due to the proximity of the plants to the market. In terms of the use of raw material, corn is still the most used waste, in this case obtained from Coahuila, Durango and Jalisco, distributed to the closest facilities.

5. Conclusions

A mathematical model has been developed to determine the optimal supply chain in terms of profit and water footprint for the use of lignocellulosic biomass to potentially obtain a

variety of renewable products in Mexico. This study concludes that solutions can be found that satisfy the economic and environmental aspect, producing bioethanol and levulinic acid from agricultural residues. Moreover, the facilities are located on states where the production of these biofuels and high value-added products do not have a negative impact on water use. This analysis gives us the opportunity to include more objectives within the supply chains, such as environmental and social aspects, all with the purpose of having a sustainable proposal in every aspect.

References

- S. Bhogeswararao, D. Srinivas, 2015. Catalytic conversion of furfural to industrial chemicals over supported pt and pd catalysts. *Journal of Catalysis* 327, 65–77.
- Comisión Nacional del Agua, 2018. Estadísticas del agua en México. Tech. rep., CONAGUA, México. URL http://sina.conagua.gob.mx/publicaciones/EAM_2018.pdf
- M. A. Costagliola, M. V. Prati, S. Florio, P. Scorletti, D. Terna, P. Iodice, D. Buono, A. Senatore, 2016. Performances and emissions of a 4-stroke motorcycle fuelled with ethanol/gasoline blends. *Fuel* 183, 470–477.
- M. A. Damián-Huato, A. Cruz-Leon, B. Ramirez-Valverde, O. Romero-Arenas, S. Moreno-Limón, L. ReyesMuro, 2013. Maíz, Alimentación Y Productividad: Modelo Tecnológico Para Productores De Temporal De México. *Agricultura Sociedad y Desarrollo* 10 (2), 157–176.
- A. Demirbas, 2017. Tomorrow's biofuels: Goals and hopes. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 39 (7), 673–679.
- El Informador, 2019. Serán siete las refinerías de México en 2022. URL <https://www.informador.mx/Seran-siete-las-refinerias-de-Mexico.html>
- Y. M. Espinoza-Vázquez, F. I. Gómez-Castro, J. M. Ponce-Ortega, 2020. Supply chain optimization for the production of biofuels and bioproducts from lignocellulosic biomass in Mexico. In: *Computer Aided Chemical Engineering*. Vol. 48. Elsevier, pp. 1339–1344.
- Y. Fu, U. M. Diwekar, 2004. An efficient sampling approach to multiobjective optimization. *Annals of Operations Research* 132, 109–134.
- J. Godínez Madrigal, P. v. d. Zaag, N. v. Cauwenbergh, 2018. A half-baked solution: drivers of water crises in Mexico. *Proceedings of the International Association of Hydrological Sciences* 376, 57–62.
- Mercado Libre, 2020. 3000 pacas de rastrojo de sorgo y de maíz. URL <https://articulo.mercadolibre.com.mx>
- OAS, 1999. Estudio de Integración Regional en el Transporte de Carga. Tech. rep., Fondo Financiero para el Desarrollo de la Cuenca del Plata, Uruguay. URL <http://www.oas.org/usde/publications/Unit/oea75s/begin.htm#Contents>
- F. D. Pileidis, M.-M. Titirici, 2016. Levulinic acid biorefineries: new challenges for efficient utilization of biomass. *ChemSusChem* 9 (6), 562–582.
- M. A. Rendon-Sagardi, C. Sanchez-Ramirez, G. Cortes-Robles, G. Alor-Hernandez, M. G. Cedillo-Campos, 2014. Dynamic analysis of feasibility in ethanol supply chain for biofuel production in Mexico. *Applied Energy* 123, 358–367.
- J. E. Santibañez-Aguilar, J. B. González-Campos, J. M. Ponce-Ortega, M. Serna-González, M. M. El-Halwagi, 2014. Optimal planning and site selection for distributed multiproduct biorefineries involving economic, environmental and social objectives. *Journal of Cleaner Production* 65, 270–294.
- SIAP, 2020. Anuario Estadístico de la Producción Agrícola. Tech. rep., SIAP, México. URL <https://nube.siap.gob.mx/cierreagricola/>
- F. Trespalacios, I. E. Grossmann, 2015. Improved big-m reformulation for generalized disjunctive programs. *Computers & Chemical Engineering* 76, 98–103.
- Q. Zhang, J. Dong, Y. Liu, Y. Wang, Y. Cao, 2016. Towards a green bulk-scale biobutanol from bioethanol upgrading. *Journal of Energy Chemistry* 25 (6), 907–910.

Synthesis of Combined Heat and Mass Exchange Networks Via a Trust Region Filter Optimisation Algorithm Including Detailed Unit Designs

Saif R. Kazi^a, Michael Short^{b,*}, Lorenz T. Biegler^a

^a*Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh 15213, USA*

^b*University of Surrey, Stag Hill, Guildford GU2 7XH, UK*

**m.short@surrey.ac.uk*

Abstract

Mass and heat integration are important to achieving economically and environmentally sustainable processes through increased efficiency. Typically, heat and mass exchange networks are solved separately using process integration techniques such as pinch technology or formulating nonconvex mixed-integer nonlinear programming (MINLP) problems, which are challenging to solve. To simplify the MINLP, shortcut models are employed, which can result in under/overestimation of the real network, leading to suboptimal or infeasible designs. We introduce a new optimisation algorithm for combined heat and mass exchanger network synthesis (CHAMENS), including detailed design models. The method uses shortcut models in an MINLP to find network topology, followed by a nonlinear programming (NLP) suboptimisation. The NLP allows non-isothermal and non-isocompositional mixing, uses detailed unit models of packed columns based on orthogonal collocation on finite elements (OCFE), and detailed shell and tube heat exchanger designs. We incorporate a differential-algebraic equation (DAE) based shell and tube heat exchanger design model via surrogates in a trust region filter (TRF) framework, guaranteeing optimal solutions for the detailed exchanger models are found by the surrogate models. We demonstrate the proposed approach on a case study, showcasing its performance and the need to incorporate detailed unit models in topology optimisation to find practical optimal designs.

Keywords: heat exchanger networks, mass exchanger networks, optimisation, mixed-integer nonlinear programming, process integration.

1. Introduction

Mass and heat integration increase process efficiency by reducing external heating and cooling utilities and mass separating agents (MSAs), thus maximising heat and mass recovery within the process. To solve the problem using mathematical programming techniques, mass exchanger networks (MENS) and heat exchanger networks (HENS) are commonly formulated as separate MINLPs using superstructure-based approaches such as the stage-wise superstructure (SWS) for HENS proposed by Yee and Grossmann (1990), or the approach of El-Halwagi and Manousiouthakis (1990a) for MENS. These require nonconvex MINLP formulations, which are difficult to solve, and therefore unit representations are simplified, with heat exchanger area and mass exchanger height used as sizing parameters and heat and mass transfer coefficients fixed. Despite abundant literature on solving these problems separately, mass and heat transfer coefficients are related to temperatures and velocities and therefore there is a need to integrate these

problems in combined heat and mass exchanger networks (CHAMENs). This is especially important in gas-liquid absorption, where there are benefits in operating absorption at low temperature and MSA regeneration at higher temperatures.

Srinivas and El-Halwagi (1994) formulated CHAMENs problems for the first time, using a two-step procedure, where an MINLP is first used to obtain the MEN with lowest MSA costs, followed by a HEN synthesis step. Liu et al. (2015) also solved the two networks separately using the SWS MINLP formulation to solve for network topologies before formulating an NLP to combine the models and optimise interactions between them. Recently, Kim et al. (2020) developed combined model for solving large MINLPs with MENs, HENs, and a newly developed regeneration exchanger network (REN), which incorporates different regeneration options. All these approaches, however, only use simplified representations for the unit operations and design models.

In MENS, without accounting for detailed aspects of the unit designs, the simplified formulations may result in sub-optimal or practically infeasible network designs with flooding, low mass transfer coefficients or strange column features, such as implausible L/D ratios. Isafiade and Short (2016) incorporated additional details of packed column unit design into the MINLP, such as column diameter, mass transfer coefficients, velocities, and noticed vastly different designs; however, the MINLP was difficult to converge due to the increased nonconvexity of the formulation. Short et al. (2018) developed detailed packed column models based on discretisation of the differential equations by OCFE to formulate a DAE model that can be solved via NLP. These models solve reliably, allowing for consideration of optimal diameters, heights, packing size, etc, while considering practical aspects such as column flooding. The authors then incorporated these into the MEN SWS model via a two-step optimisation algorithm, based on the work of Short et al. (2016), by using correction factors to ensure convergence of the MINLP shortcut models to the solutions of the detailed designs.

Recently, Kazi et al. (2020a) developed a shell and tube heat exchanger design optimisation model that uses efficient enumeration to determine discrete decisions on baffles, number of shells, tube passes, fluid allocation, etc., followed by a detailed DAE model. The model solves quickly and makes fewer assumptions than those based on the log mean temperature difference (LMTD). These models were incorporated into HENS, allowing for designs to incorporate detailed exchangers, including pressure drops, multiple shells, etc. (Kazi et al., 2020b). More recently, Kazi et al. (2020c) developed a novel algorithm for including these models into HENS as surrogates via a TRF algorithm from Yoshio and Biegler (2020).

In this work, we propose a two-stage approach where, in the first stage, an MINLP superstructure-based optimisation is used with simplified, shortcut unit models to find topologies for the CHAMEN. Next, an NLP is formulated with detailed optimisation unit models for the packed bed absorption columns, based on OCFE (Short et al., 2018). These models are directly incorporated in the NLP. Detailed shell-and-tube heat exchanger design models, from Kazi et al. (2020a), are incorporated into the flowsheet via a TRF algorithm. The algorithm allows for optimisation of a detailed flowsheet with first-principles models, and with superstructure decisions decided via a simultaneous optimisation procedure. The method does not increase the complexity of traditional MINLP process integration models, while providing a method of incorporating detailed unit designs. The algorithm and its application to CHAMENs demonstrates the power of the approach, highlighting the need for detailed unit models in superstructure-based process synthesis, and proposes a potential framework for future process synthesis.

2. Modelling framework

2.1. Topology optimisation

We present a modelling framework for CHAMENS that begins with solving the MEN, HEN, and REN separately as MINLPs to obtain initialisations for the full problem. Note that these formulations are very similar to common SWS approaches mentioned earlier, with a REN model following similar formulation to Kim et al. (2021). Following this, the problems are integrated through linking variables, with temperatures and flowrates of MSAs, in order to solve the combined problem as a large MINLP. These models assume constant heat and mass transfer coefficients, isothermal and isocompositional mixing, and use shortcut formulations to represent units. The shortcut heat exchanger models use areas as costing variables for the heat exchangers, ignoring the number of shells and pressure drops associated with them. Similarly, for the packed bed mass exchangers, only height is considered as a capital costing variable, with diameters and their effects on potential flooding, capital cost and mass transfer coefficients, considered constant. By doing this, we can maintain the model formulation as mostly convex and reduce complexity for the large MINLP formulation, which is challenging to solve. Parameters, along with the assumptions above, are selected to ensure that the MINLP problem underestimates the objective function, thus ensuring that this can act as a lower bound to the overall minimisation problem. If a global solution is found to the MINLP problem, this is a rigorous lower bound. However, in practice this cannot be guaranteed due to the size and non-convexity of the formulation.

2.2. Detailed network optimisation

The topology obtained from the shortcut model MINLP is used to construct an NLP, consisting of 3 linked submodels, solved simultaneously, for the REN, MEN, and HEN. The MEN and REN are formulated using the NLP optimisation model proposed by Short et al. (2018), where each selected mass exchanger is solved as a DAE model of a packed column using OCFE. This formulation allows simultaneous optimisation of the column diameters, heights, and packing sizes, using Onda's correlation for mass transfer coefficients. In Short et al. (2018), the mass balances over each column were fixed to that of the solution from the MINLP, however in our formulation, we allow for inlet and outlet concentrations of lean and rich streams to be optimised, along with column designs, providing the NLP with significant freedom. Furthermore, the model allows for non-isocompositional mixing of split streams and temperature-dependent mass transfer coefficients, linked via the Henry's Law coefficient. The MEN and REN submodels are connected via the variable flowrates for the MSAs that are regenerated.

The HEN NLP consists of the energy balance equations from the MINLP, with fixed topology via the binary variables and non-isothermal mixing. Shell and tube heat exchanger design variables that affect the overall objective function, such as number of shells, pressure drops, and areas, are included in the model via reduced order models (ROMs). These ROMs are updated at each iteration of the TRF algorithm to ensure that they are accurate within a small trust region around the previous solution, obtained from the running the detailed discretised DAE model of Kazi et al. (2020a). This strategy allows us to ensure that the solutions for the ROMs are the same as those of the rigorous models obtained from the DAE model, while also guaranteeing that we are at a stationary point for both. For more details on the TRF strategy, see the work of Yoshio and Biegler (2020). The NLP is solved at each iteration of the TRF, including all NLP subproblems, with the HEN NLP linked to the MEN and REN NLPs via temperatures and flowrates of the lean streams in the MEN, which are also the flowrates of the rich streams in the REN.

2.3. Integer-cut algorithm and implementation

Once the NLP TRF algorithm terminates, the solution includes optimal flowrates of utilities and detailed designs of individual units. These designs include optimal pressure drops, tube and shell passes, areas, fluid allocation and tube lengths for the shell and tube heat exchangers, and the optimal packing size, column heights, diameters, velocities, and stream splits for the packed columns in the MEN and REN. Additionally, the optimal temperatures for mass exchange are obtained. This solution acts as an upper bound to the overall algorithm. Integer cuts can then be generated to exclude the previously obtained topology from future iterations of the MINLP, allowing other topologies to be evaluated. Once the lower bound exceeds the upper bound, the algorithm terminates, with the upper bound solution as optimal. A graphic depiction of the algorithm is shown in Figure 1. The full strategy is implemented in Python, using Pyomo (Hart et al., 2017), with NLP problems being solved with IPOPT 3.13.3 (Wächter and Biegler, 2006) and MINLP problems solved using BARON (Tawarmalani and Sahinidis, 2005). Note that due to the size and nonconvexity of the MINLP, on certain iterations it may be necessary to relax tolerances in BARON to find feasible topologies for subsequent iterations.

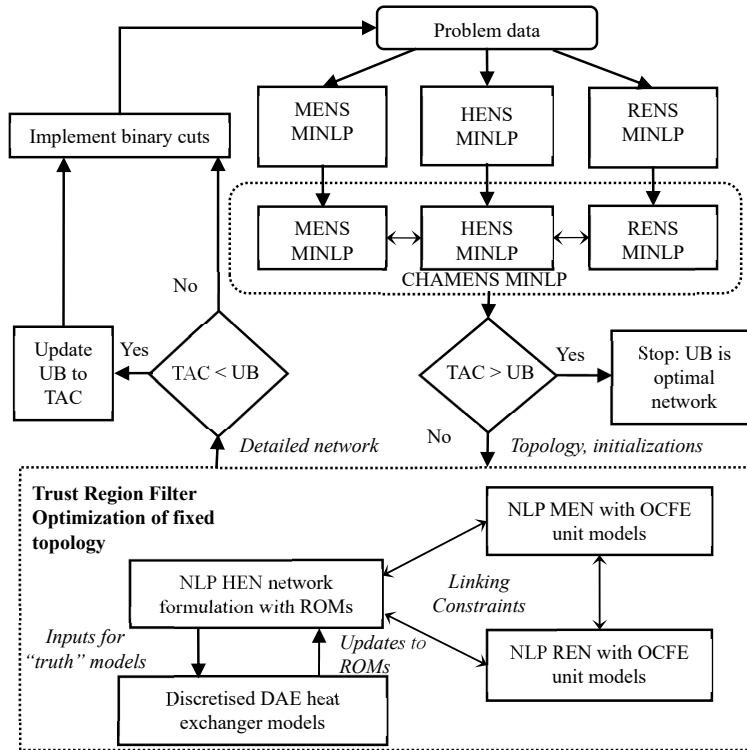


Figure 1: Overall solution algorithm

3. Case Study

We formulate an H₂S removal process, where coke-oven gas (COG) is sweetened. The problem has two rich streams, COG (R1) and tail-gas from a Claus unit (R2), both rich in H₂S, potentially absorbed by three lean streams: aqueous ammonia (L1), a process MSA,

and external MSAs (L2) 15 wt% methyl diethanolamine (MDEA), and (L3) chilled methanol regenerated via low pressure or medium pressure steam. For brevity, detailed problem data and model formulations are omitted, with results shown to demonstrate the algorithm. We use stream parameters from Kim et al. (2020), with flowrates multiplied by 10 and model formulations from aforementioned studies. For the purposes of this example we consider the temperature of L2 and flowrates of L2 and L3 as linking variables. The optimal design is shown in Figure 2. The MINLP consists of 460 constraints, 455 continuous variables and 64 discrete variables (TAC: \$26,209,870), and the NLP model consists of 11,199 constraints and 11,095 continuous variables (TAC: \$26,634,598). The full algorithm solves in 5 CPU min, with most computational effort spent in the TRF algorithm. Interestingly, the result finds that L2 and low-pressure steam are not needed in the optimal network. There are significant discrepancies in the area of the heat exchanger and column designs between the NLP and MINLP solution, showing the need for detailed designs. Interesting solution characteristics include that packed column designs are constrained by flooding, with velocities and packing sizes to ensure that each column is operating at maximum velocity to ensure high mass transfer without flooding. This generally results in smaller columns than the MINLP shortcut models, as diameters are available for optimisation. Interestingly, the NLP also redistributes mass load between the first 2 exchangers on R1 in a different way to the MINLP, using a far larger first column on R1, possibly due to the addition of mixing constraints in the NLP. This results in a slight decrease (0.02 kg/s) in methanol use and subsequent reduced costs.

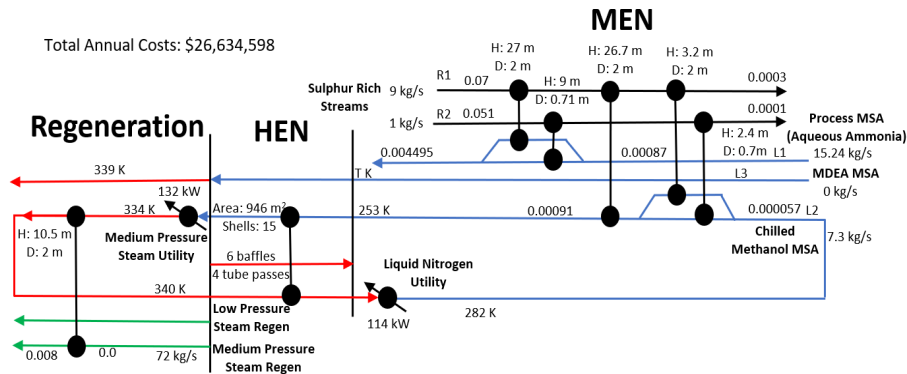


Figure 2: Optimal network identified for the Case Study

The shell and tube heat exchanger design requires many shells and high pressure drop and thus the NLP optimisation attempts to reduce the area of this exchanger, particularly since a fixed cost is introduced for numbers of shells and pressure drop is included in the objective function. The shortcut model in the MINLP underestimates the impact of this exchanger on the overall network costs without these considerations. On the second iteration of the algorithm, after the integer cut, only utilities and no process-process heat exchangers are selected and the new TAC is significantly worse due to high cooling costs.

4. Conclusions

We develop a new algorithm for solving CHAMENS that is the first to include detailed unit models as part of the optimisation formulation. The algorithm uses a MINLP model formulation with shortcut models to represent units to solve the large, nonconvex combinatorial problem first with parameters chosen to ensure an underestimation of the

objective function. This solution serves as the topology for the subsequent algorithmic steps. With topology fixed, an NLP model is then formulated. The NLP includes detailed mass exchanger designs, non-isothermal/non-isocompositional mixing constraints, and the mass and energy balances. Shell and tube heat exchangers are incorporated via surrogate models, generated from a detailed heat exchanger design algorithm, within a TRF optimisation algorithm, which guarantees that optimal solutions for the detailed exchanger models are found by the surrogate models. We demonstrate the technique on a case study involving the removal of H₂S from gaseous waste streams and find interesting designs that are different from those obtained using MINLPs alone. We present detailed heat exchanger designs and packed column designs, solving the large NLP quickly to show that the approach can be easily extended. Future work will focus on developing detailed regeneration models and including different column types that are common in MENs, in addition to a planned release of the code as an open-source heat and mass integration package. The flexible approach may also be modified to multi-component and reactive MENs in future work.

References

- El-Halwagi M.M., Manousiouthakis V., 1990a, Automatic synthesis of mass exchange networks with single component targets. *Chem. Eng. Sci.*, 45(9), 2813–2831.
- Hart W.E., Laird C.D., Watson J.-P., Woodruff D.L., Hackebeil G.A., Nicholson B.L., Siirola J.D., 2017, *Pyomo - Optimization Modeling in Python*, Springer, Cham, Switzerland.
- Liu L., Du J., Yang F., 2015, Combined mass and heat exchange network synthesis based on stage-wise superstructure model, *Chinese Journal of Chemical Engineering*, 23, 1502–1508
- S.R. Kazi, M. Short, L.T. Biegler, 2020a, Heat Exchanger Network Synthesis With Detailed Exchanger Designs - 1. A Discretized Differential Algebraic Equation (DAE) Model for Shell and Tube Heat Exchanger Design, *AIChE J.*, e17056.
- S.R. Kazi, M. Short, L.T. Biegler, 2020b, Heat Exchanger Network Synthesis With Detailed Exchanger Designs - 2. Hybrid Optimization Strategy for Synthesis of Heat Exchanger Networks with Detailed Individual Heat Exchanger Designs, *AIChE J.*, e17057.
- S.R. Kazi, M. Short, L.T. Biegler, 2020c, Heat Exchanger Network Optimization including Detailed Heat Exchanger Models using Trust Region Method, *Comp. Aid. Chem. Eng.*, 48, 1051-1056.
- S-M. Kim, A.J. Isafiade, M. Short, 2021, The synthesis of combined heat and mass exchange networks (CHAMENS) with renewables considering environmental impact. *Recent Advances in Process Intensification and Integration for Sustainable Design*, Wiley-VCH.
- M. Short, A.J. Isafiade, D.M. Fraser, Z. Kravanja, 2016, Synthesis of heat exchanger networks using mathematical programming and heuristics in a two-step optimisation procedure with detailed exchanger design. *Chem. Eng. Sci.*, 144, 372-385.
- M. Short, A.J. Isafiade, L.T. Biegler, Z. Kravanja, 2018, Synthesis of mass exchanger networks in a two-step hybrid optimization strategy. *Chem. Eng. Sci.*, 178, 118-135.
- A.J. Isafiade, M. Short, 2016, Synthesis of mass exchange networks for single and multiple periods of operations considering detailed cost functions and column performance. *Proc. Safety and Env. Prot.*, 103, 391-404.
- B. K. Srinivas and M. M. El-Halwagi, 1994, "Synthesis of combined heat and reactive mass-exchange networks," *Chem. Eng. Sci.*, 49, 2059-2074.
- Tawarmalani M., Sahinidis N.V., 2005, A polyhedral branch-and-cut approach to global optimization, *Math. Program.*, 103(2), 225-249.
- Wächter A., Biegler L., 2006, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Math. Program.*, 106, 25–57.
- T.F. Yee, I.E. Grossmann, 1990, Simultaneous Optimization Models for Heat Integration - II. Heat Exchanger Network Synthesis, *Comp. & Chem. Eng.*, 14, 1165-1184.
- N. Yoshio, L.T. Biegler, 2020, Demand-based optimization of a chlorobenzene process with high fidelity and surrogate reactor models under trust region strategies, *AIChE J.*, 17054.

Computer-Aided Refrigerant Design: New Developments

Nichakorn Kuprasertwong^a, Orakotch Padungwatanaroj^a, Arisa Robin^a, Kornkanok Udomwong^a, Anjan Kumar Tula^b, Lingyu Zhu^c, Liyang Zhou^d, Binhui Wang^e, Shuhua Wang^e, Rafiqul Gani^{f*}

^a*PSE for SPEED Company Limited, 294/65 RK Office Park, Romklao Rd., Bangkok, 10520, Thailand*

^b*College of Control Science and Engineering, Zhejiang University, Hangzhou, China*

^c*College of Chemical Engineering, Zhejiang University of Tehnology, Hangzhou, China*

^d*Juhua Group Corporation, Quzhou, China*

^e*Zhejiang Juhua Novel Material Research Institute Co. Ltd, Hangzhou, China*

^f*PSE for SPEED, Skyttemosen 6, Allerød, DK-3450, Denmark*

* *Corresponding author: rgani@pseforspeed.com*

Abstract

Refrigerants are chemical compounds with specific thermodynamic properties that affect the performance of refrigeration processes. Consequently, the properties of refrigerants play a key role in the design and/or selection of refrigerants, as well as in the improvement of refrigeration systems with respect to their operational efficiencies as well as environmental impacts. In this paper, a version of the computer-aided molecular design technique, especially adjusted for application to refrigeration selection and/or design and verification is presented. The novel features of this technique are especially developed property models suitable for the small molecules that are commonly employed as refrigerants, a database of generated small molecules and a suite of computational tools that are needed at various steps of refrigerant design and verification. Based on the new property models, databases, and process models, methods for refrigerant design (single molecules or mixtures), selection and verification have been developed. Application examples highlighting the importance of the new methods and tools are also presented.

Keywords: CAMD, Computer-aided refrigerant design, Mixture design, Refrigerant.

1. Introduction

Refrigeration systems are an essential process in industry and home applications as they perform cooling or maintain room temperature at a desired value. A cycle of refrigeration consists of heat exchange, compression and expansion with a refrigerant flowing through the units within the cycle. The refrigerants are chemical compounds having thermodynamic properties that are especially suited for a refrigeration cycle operation. They are usually small molecules having lower normal boiling points (usually less than 273 K). However, other properties related to safety (such as flammability or auto-ignition temperature) and environmental impact (such as ozone depletion or global warming) also effect their selection. Also, a set of process related properties (such as phase equilibrium properties, phase enthalpies and heats of vaporization) effect the operational performance (such as the Coefficient of Performance, COP). Therefore, properties of a refrigerant need to be considered not only for its selection and/or design, but also related to environmental-health-safety (EHS) issues as well as process performance related issues. Consequently, these problems have attracted much attention as they are ideally suited for application of

the computer aided molecular design (CAMD) technique (Churi and Achenie, 1996; Sahinidis et al., 2003; Gani et al 2017; to name a few). However, after an examination of measured data (see section 2.1) of small molecules typically used as refrigerants, the group contribution (GC) methods (for example, Marrero and Gani, 2001) that have been employed in CAMD techniques for refrigerant design and/or selection, have been found to be of questionable accuracy for some of the important design related properties (see section 3.1). Even though these GC methods give acceptable predictions for larger molecules such as solvents, their prediction errors for smaller molecules have been found to be too large for reliable design of refrigerant systems. For a more reliable design of refrigerant systems, a new set of more accurate property models, together with appropriate databases integrated with the CAMD technique is necessary.

This paper presents an especially developed database of small molecules that can be considered as a refrigerant together with their collected data. Using the available measured data, new GC models are developed for three of the important properties normal boiling point (T_b), critical temperature (T_c) and critical pressure (P_c) that define the phase behavior and related refrigerant functions. Using these GC models as a basis, the applicability of other process related property models (such as equations of state) is also verified. As the models for EHS properties are already based on small molecules, only their prediction accuracies are checked. Using the new set of property models and the associated data, methods for selection and/or design of refrigerants are adopted. The CAMD technique used in this paper is of the “generate and test” type, where, candidates are selected (or designed) and then tested for process performance through appropriate process models (for refrigeration cycle simulation). To aid in the generate and test steps, computational tools for calculating thermodynamic diagrams for refrigerants and sensitivity analysis are also employed (not shown in this paper).

Results related to property model prediction accuracy, refrigerant substitution as an example of refrigeration design, and various aspects of the refrigeration cycle operation are highlighted in this paper. More details of the databases, the developed property models, the process models, the refrigeration selection method and a special software tool can be found as an extended version of this paper in Kuprasertwong et al., (2021).

2. Methods and tools for refrigerant selection and design

2.1. Database and property models

Databases of compounds that may be suitable as refrigerants have been created with Database-1 (pure component properties), Database-2 (Vapor-Liquid Equilibrium data) and Database-3 (Binary Azeotropic data). In Database-1, the collected pure component properties are divided into 3 main sections, which are Basic Properties (such as Molecular Formula, Normal boiling point, Critical properties, etc.), Chemical Properties (GWP, ODP, LD50, etc.), Temperature Dependent Properties (density, specific heat, heat of vaporization, etc.). A total of 60-different types of properties have been collected in terms of measured and estimated data. There are a total of 1766 compounds out of which 306 are known refrigerants listed in ASHRAE (Wilson et al., 2013), NIST (NIST, 2018). The compounds are classified into different types: (hydrocarbons - paraffins, -olefins, -alkynes, -cyclohexyl, -side-chains; chlorinated; fluorinated; non-condensable gases; etc.). New GC models has been developed using measured data for T_b , T_c and P_c considering compounds with 1-7 carbon atoms. Using a new set of functional groups (considering the C-atom and its bonds with other atoms) and including an atom contribution term. The new GC model expression is similar to the MG-method (Marrero & Gani, 2001):

$$f(x) = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \quad (1)$$

Where $f(x)$ is a function of the property x to be predicted. C_i is the contribution of the atoms of type- i that occurs N_i times in a compound. D_j is the bond-contribution of the atom j that occurs M_j times and E_k is the contribution of the second and third order groups of type- k that occurs O_k times in a compound.

Measured data for 1016 small molecules having 1-7 Carbon atoms are used for regression of the model parameters for T_b , T_c , and P_c . Since there are more measured data for T_b than T_c or P_c , a correlation is also developed to predict T_c from measured T_b values. The molecular structures of each compound are represented in terms of atoms, bonds and functional groups; the contributions C_i , D_j and E_k are regressed through measured data. Table 1 gives partial model correlation results for C1-C4 compounds. For reliable refrigerant design, prediction errors should be less than 5%, but less than 1% is preferred.

Table 1: The distribution of percentage of errors for C1-C4 of estimated normal boiling point temperatures compared with measured data from regression of parameters.

Carbon Number	Percentage of compounds having			Highest %Error	Number of Compounds
	<1% error	<5% error	<10% error		
C1	85.0	100.0	100.0	1.4	20
C2	44.2	87.5	100.0	8.0	104
C3	30.3	90.3	98.7	11.1	238
C4	40.3	90.8	100.0	8.4	119
Total C1-C4	36.0	89.8	99.3	11.1	481
Total C1-C7	40.4	91.8	99.3	13.9	1016

According to the results given in Table 1, the T_b of each compound are calculated with the regressed parameters and compared with measured data. 85% of C1 compounds have less than 1 % error (relative), while the largest percentage errors for C1, C2, C3 and C4 atoms are 1.4%, 8.0%, 11.1 % and 8.4%, respectively. The model performance for T_c (correlated as a function of T_b) and P_c (GC model) are shown in Figure 1 in terms of plots of estimated property values versus measured property values together with R^2 statistics, which shows very good model accuracy.

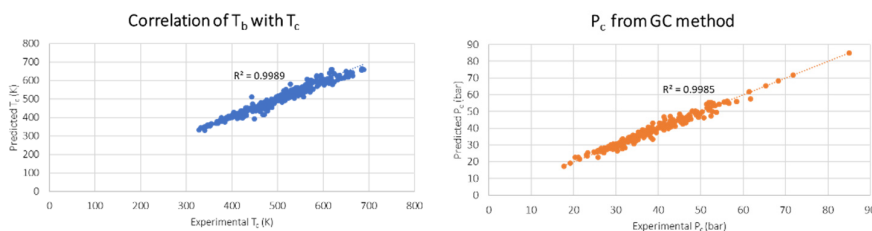


Figure 1: Plots of experimental versus estimated values for T_c (left) and P_c (right)

VLE data have been found for 214 compounds, while, binary azeotropes have been found for 258 compounds. The VLE phase equilibrium are predicted with cubic equations of state and with the PC-SAFT equation of state. The temperature and/or pressure dependent properties are calculated with in-house correlations, and/or equations of state.

2.2. Method for refrigerant selection and design

The selection method paradigm is as follows: define target properties and/or functions; find candidates that match the targets; verify selected candidates through simulation of process performance. The target properties for a refrigerant to be selected and/or designed are divided into two sets: Set-1 consists of T_b , T_c and P_c (these properties influence the phase equilibrium of the pure compound and/or mixture); Set-2 consists properties such as heat of vaporization, specific heat, density, GWP, ODP, etc. (these properties influence

the performance, sizing etc. of the process). In the selection method employed, Set-1 properties are matched as close as possible for the single refrigerant and exactly matched for mixed refrigerants. Set-2 properties are given in terms of upper and/or lower bounds and the candidate properties must be within the bounds. The operating conditions of a refrigeration process need to be consistent with the properties of a refrigerant and these are checked through the following consistency rules: T_b of the refrigerant must be lower than the room temperature (T_{room}); the maximum condenser temperature must be less than 0.9 of refrigerant T_c . Feasible candidates found through database search or CAMD are evaluated through refrigerant process simulation in terms of their performance.

Refrigerant selection - Pure molecules: It is unlikely that all target properties can be exactly matched through the properties of a single molecule. Therefore, in this case, upper and/or lower bounds for the target properties are given to find compounds that have properties within the bounds. Built-in search options in the database help to identify the candidates. Note that a database of small molecules (real and generated) has been created a priori and so new refrigerant molecules are not designed but retrieved from the database.

Refrigerant selection – mixtures: Unlike the single refrigerant molecule, mixtures can be tailored to exactly match the target values from 1 to 3 properties belonging to Set-1, while the rest of the properties (Set-2) are matched within the specified bounds. For example, as given in Table 2, the binary mixture composition is adjusted to match exactly the target T_{bT} ; ternary mixture compositions are adjusted to match exactly T_{bT} and one of T_{cT} or P_{cT} ; quaternary mixture, compositions are adjusted to match T_{bT} , T_{cT} and P_{cT} . Note that successful mixture design requires that at least one compound has a value lower than the target and one compound has a value higher than the target. The use of linear mixing rules for T_c and P_c is justified as these are also used by the cubic equations of state through which the phase equilibria are calculated. T_b is a result of the phase equilibrium at a specified pressure and thus fixes the operating condition. The value of T_b from the ideal mixing rule is used as an estimate of the operating temperature. The exact values from the phase equilibrium calculations are used later to adjust the operating temperature.

Table 2: Target properties versus compositions for refrigerant mixture design

Binary mixture	Ternary mixture	Quaternary mixture
$T_{bT} = x_1 T_{b1} + (1-x_2) T_{b2}$ (2)	$T_{bT} = x_1 T_{b1} + x_2 T_{b2} + (1-x_1-x_2) T_{b3}$ (3) $T_{cT} = x_1 T_{c1} + x_2 T_{c2} + (1-x_1-x_2) T_{c3}$ (4) Or $P_{cT} = x_1 P_{c1} + x_2 P_{c2} + (1-x_1-x_2) P_{c3}$ (5)	Match T_{bT} , T_{cT} and P_{cT} by adjusting x_1 , x_2 and x_3

2.3. Method for refrigeration cycle simulation

Refrigerators are cyclic devices where heat transfers take place from a low-temperature region to a high-temperature region and vice versa. The operations for one refrigeration cycle are shown in Figure 2 and a corresponding process model has been developed.

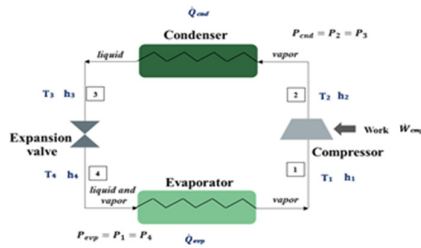


Figure 2: One cycle refrigeration process

The evaporator duty, compressor work, and condenser duty, are calculated at specified conditions of operation (defined by temperature and pressure) by using the following equations.

$$\text{Evaporator duty: } Q_{\text{evp}} = m_f(h_1 - h_4) \quad (6)$$

$$\text{Compressor Work: } W_{\text{cmp}} = m_f(h_2 - h_1) \quad (7)$$

$$\text{Condenser Duty: } Q_{\text{cnd}} = m_f(h_3 - h_2) \quad (8)$$

The performances of different refrigeration processes are compared in terms of COP:

$$\text{COP} = \frac{Q_{\text{evp}}}{W_{\text{cmp}}} = \frac{m_f(h_1 - h_4)}{m_f(h_2 - h_1)} \quad (9)$$

Note: In Eqs. 6-9, Q_{evp} is evaporator duty, W_{cmp} is isenthalpic work consumed by compressor, Q_{cnd} is condenser duty, m_r is refrigerant mass flow rate, and h_j is specific enthalpy at point j .

3. Application Examples

3.1. Consistency of data, models and CAMD solutions

To highlight the need for more accurate property models together with a need to use measured data for target properties in refrigerant selection, examples of reported feasible candidate molecules are given in Table 2 together with the target they are supposed to match as well as measured (if available) and estimated values. Only a small selection of compounds is being highlighted here (a comprehensive list can be obtained from the corresponding author). If single molecular replacements are being sought, deviations of more than 10 K from the target value may result in infeasible refrigeration cycle operation. However, if mixtures are being sought, then the target temperature can be matched exactly but the mixture compositions would not be consistent.

Table 3: Consistency check of measured-estimated property values versus target values

Compound (chemical formula)	Measured T_b (K)	Estimated T_b (K) with reported model ¹	Target value (K)
C2H6	184	GC-method not applicable	232-242
FNO	200	390	232-242
CHClO		329	242
C3H6O		316	242
C2H3FO	293	298	225
C3H4	238	260	232-242

1: Several GC-models available to the authors and including Marerro and Gani (2001)

3.2 Replacement and evaluation of R-134a

Table 4 lists selected alternative refrigerants for R-134a in terms of single, binary and ternary mixtures. The computed COP (not optimized) for the alternatives are compared with R-134a. The computations are done with the ProREFD software (Udomwong et al., 2021), where the developed property models, databases, refrigerant design and refrigeration process simulation models have been implemented.

3.3 Replacement of R-12

Another case study (Sahinidis et al., 2003) considered the replacement of R-12. A search of the database available in ProREFD (Udomwong et al., 2021) with respect to measured T_b to within 10 K of the reference gave the candidates listed in Table 5. In this problem, only compounds with at least one Carbon-atom have been considered. For each alternative, the corresponding refrigeration process has also been simulated to compute the COP at the conditions given in Sahinidis et al., (2003).

Table 4: Example of alternatives for R-134a refrigerant

Component (%wt)	Properties			Refrigeration process simulation results			
	T_b (K)	T_c (K)	P_c (bar)	Q_{evp} (kJ/h)	W_{cmp} (kJ/h)	Q_{cnd} (kJ/h)	COP
R-134a (Reference)	246.60	374.30	40.64	22,205	6,228	-28,434	3.57
Alternative of single refrigerant							
R-E170	248.35	400.10	53.70	52,624	13,695	-66,320	3.84
R-40	249.15	416.25	66.80	54,989	13,668	-68,657	4.02
Alternative of binary refrigerant							
R-12/40 (43.9/56.1)	246.60	402.49	55.57	38,710	9,911	-48,622	3.91
R-1216/E170 (36.5/63.5)	246.60	388.40	44.70	36,722	9,552	-46,275	3.84
R-1113/40 (67.1/32.9)	246.60	391.35	49.17	30,776	7,953	-38,730	3.87
Alternative of ternary refrigerant							
R-32/125/124 (7.1/32.2/60.7)	246.60	374.30	37.73	19,528	5,919	-25,448	3.30
R-32/125/124 (20.8/17.3/61.9)	246.60	376.63	40.64	23,050	6,473	-29,523	3.56

Table 5: Example of alternatives for R-12 refrigerant

Component	Properties			Calculation Result			
	Tb (K)	Tc (K)	Pc (bar)	Qevp (kJ/h)	Wcmp (kJ/h)	Qcnd (kJ/h)	COP
R-12 (reference)	243.35	384.95	41.25	11,193	7,999	-19,193	1.40
Alternative of single refrigerant							
R-717	239.72	405.65	112.80	182,237	87,459	-269,696	2.08
Cyclopropane	240.35	397.91	54.94	43,058	21,512	-64,571	2.00
R-40	249.15	416.25	66.80	46,109	22,038	-68,148	2.09

4. Conclusions

Refrigerant design and/or selection together with refrigeration process performance can be performed reliably and efficiently through the CAMD technique. However, the property models typically used for molecules that are larger than the small refrigerant molecules have been found to give unacceptable prediction errors. Therefore, a new class of GC models suitable for refrigerant molecules have been developed and validated with measured data. Another issue is the size of the database. The number of small molecules that maybe suitable for application as refrigerant and for which measured target property values are available, is not very high. This means that the number of candidates within certain ranges of temperature, for example, lower than 250 K, is not very high. To overcome this problem, CAMD has been used to generate compounds with 1-6 Carbon atoms, resulting in around 3000 additional molecules that have been found but their properties have not yet been measured. Also, structures for another 8000 molecules have been generated that do not currently exist. Truly innovative and novel refrigerants most likely will come from this list. However, novel refrigerant mixtures satisfying target properties that a single molecule cannot satisfy may also result in novel refrigerants. Much work, however, is needed to predict and validate the newly generated molecules. It should be noted that the scope and significance of methods like CAMD depends on the application range of the property models. Inclusion of data analytics and machine learning based modelling are also likely to help advance the current state of the art.

References

- N. Churi, L. E. K. Achenie, 1996, On the Use of a Mixed Integer Nonlinear Programming Model for Refrigerant Design, *I & ECR*, 35, 3788.
- R. Gani, L. Zhang, S. Kalakul, S. Cignitti, 2017, Computer-aided Molecular Design and Property Prediction in “Methods and tools for chemical product design”, *Computer Aided Chemical Engineering*, 39, 153-196.
- J. Murrero, R. Gani, 2001, Group-contribution based estimation of pure component properties, *Fluid Phase Equilibria*, 183-184, 183-208.
- N. Sahinidis, M. Tawarmalani, M. Yu, 2003, Design of alternative refrigerants via global optimization, *AIChE Journal*, 49, 1761-1775.
- N. Kuprasertwong, K. Udomwong, A. Robin, O. Padungwatanaroj, A. K. Tula, L. Zhu, L. Zhou, B. Wange, S. Wang, R. Gani, 2021, Computer-Aided Refrigerant Design, Analysis, and Verification, *Computers & Chemical Engineering* (submitted)
- K. Udomwong, A. Robin, N. Kuprasertwong, O. Padungwatanaroj, A. K. Tula, L. Zhu, L. Zhou, B. Wange, S. Wang, R. Gani, 2021, ProREFD: Tool for Automated Computer-Aided Refrigerant Design, Analysis, and Verification, *ESCAPE 31 Proceedings* (submitted).
- D. P. Wilson, D. H. Kenney, 2013, ANSI/ASHRAE Standard 34-2013 Designation and Safety Classification of Refrigerants, 46-51.
- The National Institute of Standards and Technology (NIST), 2018, NIST Standard Reference Database Number 69, <https://webbook.nist.gov/chemistry/>.

Comparative assessment of black liquor upgraded gasification in integrated kraft pulp, methanol and dimethyl ether production plants

Meire Ellen Gorete Ribeiro Domingos,^{a*} Daniel Flórez-Orrego,^b Moisés Teles dos Santos,^a Silvio de Oliveira Junior^b

^a*Polytechnic School, University of Sao Paulo, Department of Chemical Engineering, 05508-000 Sao Paulo, Brazil*

^b*Polytechnic School, University of Sao Paulo, Brazil, Department of Mechanical Engineering, 05508-030 Sao Paulo, Brazil*

Abstract

The conventional application of a pulp and paper mill, in which black liquor is typically concentrated and burnt, is compared with the upgrading gasification route for methanol (MeOH) and dimethyl ether (DME) co-production. This comparison is made in terms of exergy efficiency and environmental impact indicators, under different utilities supply scenarios. The combined energy integration and exergy analysis is used to identify the potential improvements related to the decarbonization and mitigation of the process irreversibility. The exergy efficiencies of the conventional scenario and the integrated plants average 40% and 45%, respectively, while the overall emission balances vary from 1.97 to -0.07 tCO₂/tPulp, respectively. These results reinforce the relevance of the electricity import from the Brazilian mix for pushing upwards the share of renewable energy resources in the production of traditionally fossil-based chemicals and fuels, which could be an important decision variable in favor of the exploration of forest-based integrated biorefineries.

Keywords: Black liquor, Kraft process, Gasification, Methanol, Dimethyl ether.

1. Introduction

The pulp and paper sector is classified as an energy intensive activity, accounting for about 5.6% of the total energy consumed by the industrial sector (IEA, 2017). It is also responsible for about 2% of the total yearly CO₂ emission of the industrial sectors worldwide (Leeson et al., 2017). The black liquor (BL) is a byproduct of the kraft pulping process, which contains more than half of the exergy content in the total woody biomass fed to the digester. The pressurized entrained-flow, high temperature black liquor gasification (PEHT-BLG) technology has been envisaged as a promising process to improve the energy performance in the pulp industry. The syngas produced from the gasification process of BL can be used either to generate electricity or produce chemicals and biofuels, expanding the products portfolio and matching the modern biorefinery concept. In this work, the conventional scenario of the black liquor use (i.e. concentration and combustion) is compared with the black liquor upgrading gasification process for methanol (MeOH) and dimethyl ether (DME) production, which are extensively studied as potential low carbon fuels (Fernández-Dacosta et al., 2019). The exergy destruction and the cumulative CO₂ emissions were considered as criteria for assessing the sustainability of the energy conversion systems involved.

2. Methods

2.1. Process modeling

An alternative approach to the conventional black liquor evaporation unit is adopted, namely, a mechanical vapor recompression (MVR) system. The gasification is modeled as a pressurized entrained-flow, high temperature black liquor gasification (PEHT-BLG). The produced syngas is used to produce either methanol or dimethyl ether (Figure 1). A detailed description of the kraft pulp mill, MVR and PEHT-BLG, and the syngas treatment and purification processes, as well as the data of the ultimate and proximate composition (mass basis) for BL and other industrial substances of interest, used to calculate the specific chemical exergy and lower heating value thereof, have been presented in (Domingos et al., 2021).

The methanol synthesis process is modeled using the LHHW (Langmuir-Hinshelwood Hougen-Watson) kinetic model developed by (Graaf et al., 1988). Before entering the methanol synthesis loop, the purified syngas is compressed to 90 bar. In the loop, the syngas is firstly heated up in a feed-effluent heat-exchanger (FEHE) by the reactor outlet stream. Then, the syngas is fed to a tubular plug-flow reactor (PFR) operated isothermally at 90 bar and 210°C. The reactor outlet comprises a gaseous mixture containing methanol, water and unconverted reactants (CO_x and hydrogen). This mixture is cooled and flashed twice, first to 30°C and 45 bar, and then to 30°C and 3.5 bar, in order to separate the condensable products and the non-condensable reactants (Kiss et al., 2016). The condensed stream continues to a distillation column operating at atmospheric pressure, where methanol with purity over 99% wt. is produced. The non-condensable stream is partially purged to avoid the built up of inerts, and the other part is recycled back to the methanol synthesis loop.

The DME synthesis occurs via the indirect route, i.e. via methanol dehydration over an acid zeolite catalyst. The purified syngas initially passes through the same processes described previously for methanol synthesis. Syngas is firstly compressed up to 90 bar and converted to methanol. Then, it continues to the DME synthesis loop, wherein an adiabatic DME reactor is used. The kinetic is given by (Bandiera and Naccache, 1991). The outlet stream of the reactor is a mixture of water, DME and methanol, which is fed to a first distillation column (DC), which removes DME as the distillate with purity over 98.5% wt. The bottoms stream is a mixture of water and methanol that is subsequently fed to a second DC. The methanol that exits the second column as the distillate is recycled back to the reactor, while the water exits from the bottom.

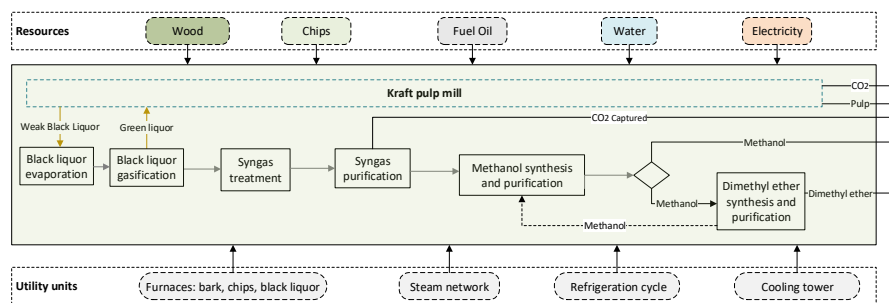


Figure 1. Integrated model used for assessing the production of methanol or dimethyl ether from black liquor upgrading gasification process.

The thermodynamic properties of the flows, the mass, energy and exergy balances of each process unit are evaluated using Aspen Plus® v8.8 software and the Peng-Robinson equation of state with Boston-Mathias modifications.

2.2. Exergy analysis indicators

Three conceptual processes are evaluated, namely (i) the standalone kraft pulp mill, based on the conventional recovery boiler, and two configurations proposed, based on the integration of black liquor gasification process for either methanol or dimethyl ether production. In turn, the two integrated chemical plants may consume either electricity from the grid (mixed mode) or only biomass as fuel import (autonomous mode).

In order to compare those configurations, two performance indicators are proposed. The rational exergy efficiency and the relative exergy efficiency. The former considers that the sum of the exergy of the two chemical products (MeOH or DME, and cellulose) and by-products (CO₂, purge gas, power) corresponds to the useful output of the chemical plant; the later compares the minimum theoretical work required to bring about only the two main chemical products of the integrated chemical plants to the overall exergy consumption in the actual production processes. The physical and chemical exergies, as well as the efficiencies are assessed using Excel add-ins embedded in Aspen Plus®.

2.3. CO₂ emissions

Two CO₂ emissions balances were performed: the overall CO₂ emissions and the net one. The overall CO₂ balance considers overall CO₂ emitted (either fossil or biogenic) minus CO₂ captured by the gas purification unit, whereas the net value subtracts the amount of CO₂ embodied by the crops, assumed as circular emissions. The indirect fossil CO₂ emitted considers the indirect emissions due to the upstream supply chains of electricity (62.09 g_{CO2}/kWh), wood (0.0043 g_{CO2}/kJ_{Wood}) and oil (0.0029 g_{CO2}/kJ_{Oil}) (Flórez-Orrego et al., 2015) The biogenic emissions comprise the emissions related to the combustion of woody components (bark and chips), based on the carbon content of the biomass.

2.4. Optimization framework and problem definition

Inasmuch as electricity can be imported from the grid, a trade-off is expected between additional fuel consumption for in-plant cogeneration and the extent of grid electricity purchase. This circumstance requires a complete redesign of the energy integration approach between the chemical units and the utility systems, so that the power and steam requirements remain satisfied.

The determination of the minimum energy requirements (MER) and the solution of the energy integration problem is handled by the OS MOSE Lua platform, developed by the IPESE group at EPFL, in Switzerland (Yoo et al., 2015). This framework allows to determine the most suitable utility systems and their operating conditions that lead to the lowest resources consumption and optimal operating cost (Maréchal and Kalitventzeff, 1998). The computational framework manages the data transfer with ASPEN Plus® software and builds the mixed integer linear programming (MILP) problem that minimizes the operating cost of the chemical plant. The utility units are modeled via equation-oriented subroutines written in Lua programming language. To this end, the additional equations required for the mass and energy balances of those units rely on the concept of layer (water, natural gas, biomass, syngas, methanol, dimethyl ether, power, carbon dioxide, heat, etc.). Finally, representative market cost for the wood (0.013 euro/kWh), chips (0.016 euro/kWh), oil (0.018 euro/kWh) and electricity (0.06 euro/kWh) consumed, as well as the selling prices of pulp (0.144 euro/kWh), methanol (0.065 euro/kWh), dimethyl ether (0.07015 euro/kWh), and CO₂ (0.0084 euro/kg) produced are taken from literature.

3. Results and Discussion

The overall exergy consumption in the conventional case achieves 42.20 GJ/tpulp, whereas for the integrated processes this value has an increment of over 14% for the mixed mode setups and of 38% for the autonomous mode configurations (Figure 2), respectively. This increase is explained by a larger amount of biomass and electricity required to drive the compressors and other ancillary equipment of the integrated chemicals production plants. The opportunity for importing ‘greener’ Brazilian electricity (i.e. in mixed mode designs) leads to a strikingly lower overall exergy consumption for MeOH and DME production compared to the autonomous cases.

The exergy efficiencies of the conventional and integrated cases for MeOH and DME, average 40%, 45% and 45%, respectively (Table 1). Thus, the integrated cases presented higher exergy efficiency in comparison to the conventional case. These results are strongly in agreement with those reported by Van Rens et al. (2011), who performed an exergetic analysis for production of MeOH and DME using wood as feedstock for gasification.

The mixed modes present the highest values of exergy efficiency (approx. 49.3% for MeOH and 48.6% for DME), whereas the autonomous cases averages 41.2% for MeOH and 40.9% DME (Table 1). This fact reinforces the idea that the electricity import, whether available, may help reducing the irreversibility in the manufacturing of bioproducts. All in all, there exists a large underexploited biomass potential that may contribute to develop new business opportunities from fuels to chemicals production. Moreover, despite the fact that an intensive industrial utilization of the Brazilian electricity mix may bringing some challenges to generation companies and other economic sectors, the advantages of the electricity import in the integrated biorefineries has been demonstrated in light of the reduction of the overall process irreversibilities.

The *Extended Exergy Consumption*, shown in Table 1, takes into account the exergy efficiency of the upstream feedstock supply chains, which ends up impairing the performance of the extended pulp and bioproducts manufacturing processes. Actually, in the case of the integrated chemical plants, when an extended control volume is observed, there is a significant increase in the exergy consumption (13.7-17.0%) compared to the exergy input to the control volume of the sole chemical plant. Although these results may not be immediately interesting for pulp and bioproducts manufacturers when evaluating the performance of the chemical plant itself, those features are certainly useful to public policies and decision-makers in environmental and benchmarking frameworks, as they permit a holistic comparison of the impact of the bioproducts sector with other industrial sectors based on other supply chains.

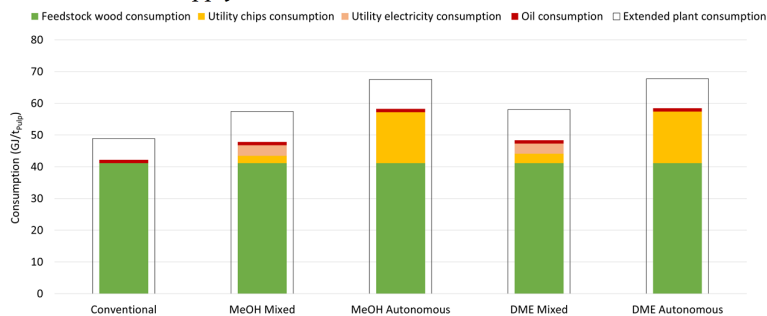


Figure 2. Plantwide and extended exergy consumption for the conventional and the integrated chemical plants (pulp and bioproducts co-production).

Table 1. Exergy destruction and efficiencies for the three studied scenarios.

Chemical plant	Conventional mill		Integrated chemical plants		
	Only pulp	Pulp + Methanol	Pulp + Dimethyl ether		
Operation mode	BL combustion	Mixed ²	Auto ³	Mixed	Auto
Rational exergy efficiency (%)	42.5	55.0	45.1	54.0	44.7
Extended rational exergy efficiency (%) ¹	36.7	45.7	38.9	45.0	38.6
Relative exergy efficiency (%)	42.3	52.8	43.3	52.0	43.0
Extended relative exergy efficiency (%) ¹	36.5	43.9	37.4	43.3	37.1
Exergy destruction (GJ/t _{Pulp})	24.3	21.5	32.0	22.2	32.3
Extended exergy destruction (GJ/t _{Pulp}) ¹	30.9	31.2	41.2	31.9	41.6

1. Overall exergy consumption increases if the cumulative efficiency of the electricity generation (55.68%), as well as of the oil (95.20%) and biomass (86.13%) supply chains are considered (Flórez-Orrego et al., 2015); 2. Mixed mode implies partial import of electricity to make up the overall power consumption in the integrated chemical plants; 3. In auto mode, in-plant cogeneration of the overall power and heat required relies only on the biomass fuel input and the steam network.

Figure 3 presents the overall and detailed CO₂ emission balance. The overall emission balances achieve 1.97 t_{CO2}/t_{Pulp} in the conventional mill; -0.07 t_{CO2}/t_{Pulp} in the combined pulp and MeOH plant, and -0.02 t_{CO2}/t_{Pulp} in the pulp and DME plant. The negative values indicate a net positive impact towards the depletion of the CO₂ present in the atmosphere. As a result, the indirect emissions from the biomass utilization are not only offset by the captured biogenic emissions, but also the import of ‘greener’ Brazilian electricity leads to a mitigation of CO₂ emissions.

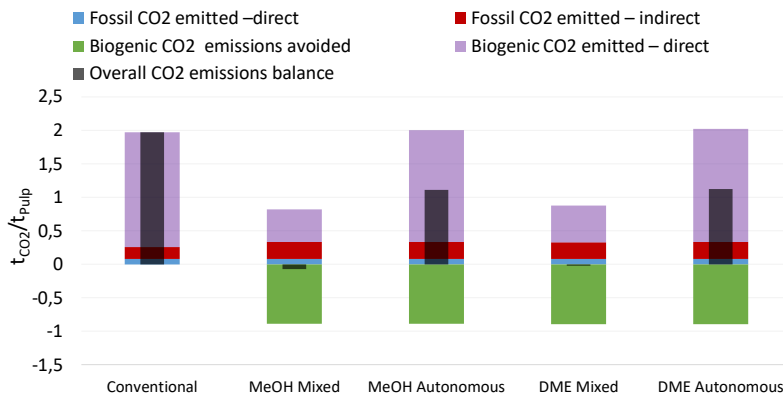


Figure 3. Overall and detailed (biogenic and fossil, directly and indirectly emitted, and avoided) CO₂ emissions for the conventional mill and the integrated kraft pulp and bioproducts manufacturing plants.

Table 2 evidences that all the integrated pulp and bioproduct production processes (mixed and autonomous) have higher operating revenues (increment between 12-15%), due to the complementary value-added CO₂ and bioproduct, not produced in the conventional kraft pulp process. These revenues, however, can be directly influenced by market prices fluctuations.

Table 2. Operating incomes, cost and revenues for the three studied scenarios.

Chemical plant	Conventional mill	Integrated chemical plants			
	Only pulp	Pulp + Methanol		Pulp + Dimethyl ether	
Operation mode	BL combustion	Mixed	Auto	Mixed	Auto
Operating Incomes (euro/ t _{pulp})	714.61	854.12	854.12	863.07	863.07
Operating Costs (euro/t _{pulp})	-153.49	-218.72	-224.89	-219.69	-225.62
Operating Revenues (euro/t _{pulp}) ¹	561.12	635.40	629.24	643.37	637.45

¹Operating revenues calculated as the difference between the gross operating incomes minus the operating cost.

4. Conclusions

The combined energy integration and exergy analyses performed allowed spotlighting the best alternatives of utility systems that ensure a competitive performance, while maximizing the recovery of the available waste heat exergy. As a result, the exergy efficiencies of the conventional and integrated cases average 40% and 45%, respectively, whereas the overall emission balance varies from 1.97 to -0.07 tCO₂/t_{pulp}, respectively. These negative values point towards the environmental benefits brought about by the production of chemicals through the use of alternative energy sources, such as biomass and electricity, in integrated chemical plants. Other environmental impacts should be also investigated.

References

- C. Fernández-Dacosta, L. Shen, W. Schakel, A. Ramirez, G. J. Kramer, 2019, Potential and challenges of low-carbon energy options: Comparative assessment of alternative fuels for the transport sector. *Applied energy*, 236, 590-606.
- J. Bandiera, C. Naccache, 1991, Kinetics of methanol dehydration on dealuminated H-mordenite: Model with acid and basic active centres. *Applied catalysis*, 69, 139-148.
- M. E. G. R. Domingos, D. Flórez-Orrego, M. Teles Dos Santos, H. Velásquez, S. De Oliveira Junior, 2021, Exergy and environmental analysis of black liquor upgrading gasification in an integrated kraft pulp and ammonia production plant. *International Journal of Exergy*, 34.
- D. Flórez-Orrego, J. A. da Silva, H. Velásquez, S. de Oliveira Jr, 2015, Renewable and non-renewable exergy costs and CO₂ emissions in the production of fuels for Brazilian transportation sector. *Energy*, 88, 18-36.
- G. Graaf, E. Stamhuis, A. Beenackers, 1988, Kinetics of low-pressure methanol synthesis. *Chemical Engineering Science*, 43, 3185-3195.
- IEA 2017. Tracking Clean Energy Progress 2017. Energy Technology Perspectives 2017 Excerpt: Informing Energy Sectors Transformations. International Energy Agency.
- A. A. Kiss, J. Pragt, H. Vos, G. Bargeman, and M. De Groot, 2016, Novel efficient process for methanol synthesis by CO₂ hydrogenation. *Chemical engineering journal*, 284, 260-269.
- D. Leeson, N. Mac Dowell, N. Shah, C. Petit, P. Fennell, 2017, A Techno-economic analysis and systematic review of carbon capture and storage (CCS) applied to the iron and steel, cement, oil refining and pulp and paper industries, as well as other high purity sources. *International Journal of Greenhouse Gas Control*, 61, 71-84.
- F. Maréchal, B. Kalitventzeff, 1998, Process integration : Selection of the optimal utility system. *Computers & Chemical Engineering*, 22, 149-156.
- G. Van Rens, G. Huisman, H. De Lathouder, R. Cornelissen, 2011, Performance and exergy analysis of biomass-to-fuel plants producing methanol, dimethylether or hydrogen. *Biomass and Bioenergy*, 35, S145-S154.
- M. Yoo, L. Lessard, M. Kermani, F. Maréchal, 2015, OS MOSE Lua: A Unified Approach to Energy Systems Integration with Life Cycle Assessment. 12th International conference PSE 2015 and 25th International conference ESCAPE 2015., 2015 Copenhagen, Denmark.

Sustainable Process Design and Synthesis for HDPE Recycling

Xiang Zhao, Fengqi You

Cornell University, Ithaca, New York 14853, USA

xz643@cornell.edu

Abstract

Plastic pollution is an ongoing problem in the U.S. This work focuses on the sustainable process design and synthesis of waste high-density polyethylene (HDPE) recycling systems to mitigate plastic pollution. A superstructure that can be implemented to manufacture various chemical products is proposed. We formulate the following superstructure optimization problem as a multi-objective mixed-integer nonlinear fractional programming (MINFP) problem, which is efficiently solved by a tailored global optimization algorithm that integrates parametric algorithm and branch-and-refine algorithm. Results reveal the economic profitability via displaying unit NPVs of optimal solutions (\$107.2 to \$151.3 per ton HDPE treated), while the unit GHG emissions of the good-choice optimal design are 0.44 ton CO₂-eq/ton HDPE treated.

Keywords: waste HDPE, superstructure optimization, global optimization algorithm

1. Introduction

Plastic pollution is still an ongoing problem around the U.S. Over the country, 76% of waste plastics, which is mainly generated by the households, go to landfill, while only 8% of the plastics are effectively recycled (EPA, 2019). Landfilled plastics damage the bio-system via being decomposed into toxic chemicals or through watering into microplastics that can be accumulated in the eco-system (Heward, 2018). In this regard, closed-loop and open-loop recycling processes are implemented into recycling these waste plastics (Williamsa and Williamsb, 1997). The closed-loop process can manufacture recovered plastics with high capital cost, while the open-loop process can be implemented into producing various useful chemicals rather than original plastics, which improves the economic viability and environmental sustainability for processing waste plastics (Fivga and Dimitriou, 2018). Although various researchers have paid their main attention to improving the economic and environmental performances for processing waste HDPE by using the open-loop recycling process, the issue of the sustainable design and synthesis of HDPE recycling process systems is still a knowledge gap.

In this work, we address this knowledge gap by introducing the sustainable design and synthesis approach to determine optimal economic and environmental performances (Gong et al. 2015). We develop by far the first superstructure that aims to convert waste HDPE into various chemicals, namely monomers, aromatic mixtures, and fuels, so that plastic pollution can be effectively alleviated. The following superstructure optimization problem is then formulated by the life cycle optimization approach to optimize the unit NPV and unit life cycle environmental impacts simultaneously. Life cycle assessment (LCA) and techno-economic analysis (TEA) are performed to provide environmental and economic parameters, respectively, and thus the superstructure optimization problem can

be formulated as a multi-objective mixed-integer fractional programming (MINFP) problem. To tackle the computational challenge caused by the combinatorial nature and the pseudo-convexity shown in fractional objective functions, we apply a tailored global optimization algorithm that integrates the inexact parametric algorithm with the branch-and-refine algorithm. The proposed framework is implemented to design an HDPE recycling processing system that can treat 295,000 ton HDPE/year in maximum in the U.S., corresponding to the estimated annual landfill amount of waste HDPE in New York State (EPA, 2019).

2. Superstructure Description of Waste HDPE Recycling

In this work, we develop by far the first superstructure of waste HDPE recycling process systems. As shown in Figure 1, the superstructure includes 1,728 processing pathways and nine sections. The corresponding flowsheet accounts for various unit processes and technologies, which are corresponding to the technology options shown in Figure 1. Only one technology or unit process can be selected for each processing pathway.

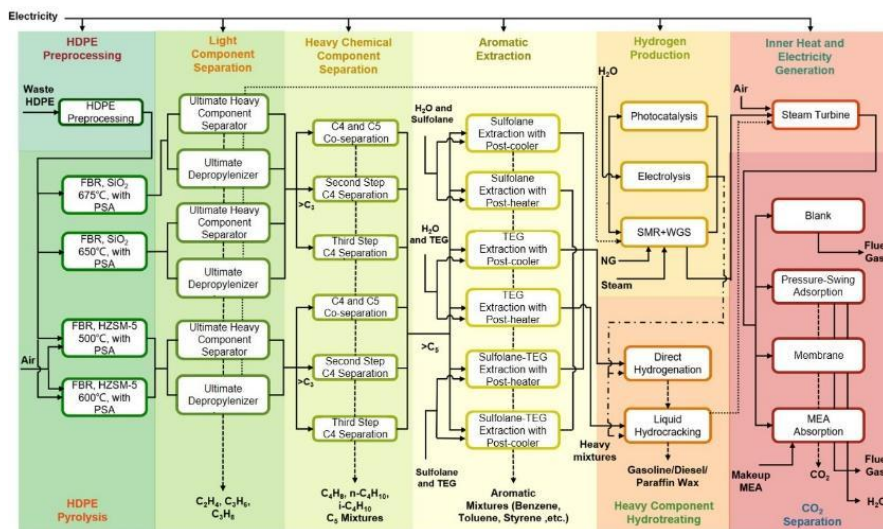


Figure 1. Overview of the proposed superstructure of Waste HDPE Recycling.

The whole superstructure starts with the HDPE preprocessing section, where the transported waste HDPE is shredded and transported to the HDPE pyrolysis section. The HDPE particles are thermally cracked into various hydrocarbon chemicals in a fluidized bed pyrolyzer (Hernández et al., 2007), where the nitrogen gas is an inert fluidized gas and circulated in the HDPE pyrolysis section. The streams with gaseous products are separated, cooled, and split into light and heavy components. In the light hydrocarbon separation section, all the technology options are considered for producing ethylene, propane, and propylene from the light hydrocarbon stream. The heavy streams from the light hydrocarbon separation and HDPE pyrolysis sections are mixed and fed into the heavy chemical component separation section, where three technologies configurations, namely second C4 separation, third C4 separation, and C4 and C5 co-separation, are considered (Yang and You, 2017). The mixture of heavy streams is split into n-butane, i-butane, butene mixture, and C5 mixture, which are important components in organic solvent and naphtha (Buekens and Froment, 1968).

The stream with heavy liquid (>C6) from the heavy chemical component separation section is sent to the aromatic extraction section to produce aromatic mixtures that are vital to produce benzene and toluene (He and You, 2014). Notably, sulfolane is selected as a solvent due to its wide usage in the typical UOP extraction process (Mayers, 2004), while the TEG is chosen because of its high selectivity to aromatics and miscibility in water (Wang et al., 1998). The raffinate from this section is directly fed into the heavy component hydrotreating section to be hydrogenated or hydrocracked. The raffinate is converted into gasoline, diesel, and wax (>C16 components), while the hydrocracking section produces fuel gas, gasoline, and diesel. The fuel gas stream is sent to the inner heat and electricity generation section to produce electricity and heat.

To produce hydrogen for the hydrogenating or hydrocracking process, we consider three technology options in the hydrogenation section, namely photocatalysis, electrolysis, and steam-methane reforming are considered. In the steam-methane reforming section, the methane stream from the light component separation section is mixed with steam and converted into hydrogen, with carbon dioxide, carbon monoxide, and unconverted methane as byproducts, which are sent to the onsite combustion section. In the onsite combustion section, all overhead gas flows are mixed and ignited in the furnace to exchange heat in selected technologies options and evaporate the precooled water into steam, which can push the steam-turbine to do the work and thus be converted into electricity in a turbine generator. The flue gas from the onsite combustion section is sent to the CO₂ separation section to split and liquefy the CO₂ for storage.

3. LCA and TEA Methodology

3.1. LCA Methodology

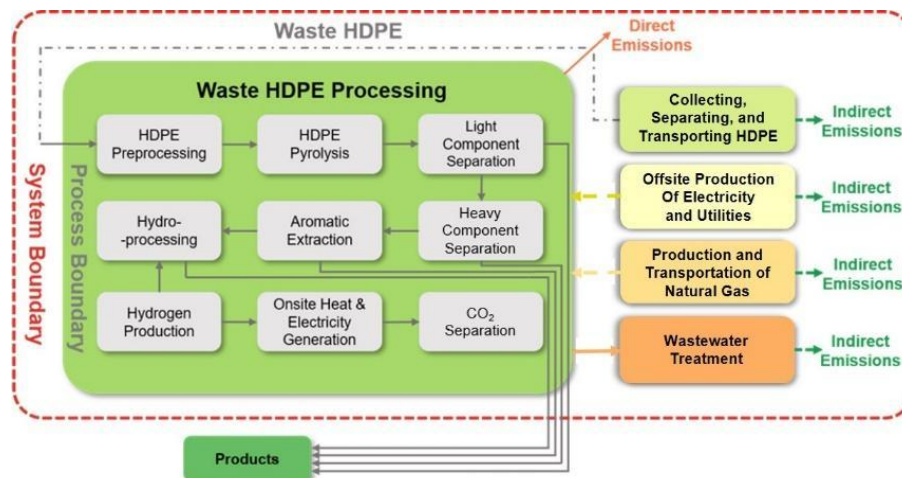


Figure 2. System boundary of a cradle-to-gate LCA for waste HDPE recycling process systems.

The LCA methodology is used for systematically quantifying the environmental impacts so that the environmental parameters can be provided to the corresponding superstructure optimization problem. The goal of this LCA is to quantify the life cycle environmental impacts from waste HDPE recycling process systems. Five life cycle stages, namely collecting, separating, and transporting HDPE, offsite production of electricity and utilities, waste HDPE processing, production and transportation of natural gas, and

wastewater treatment, are confined by the system boundary from cradle-to-gate, as shown in Figure 2. We choose the functional unit as one ton of waste HDPE processed within the waste HDPE processing system. We compile the detailed life cycle inventories (LCIs) of such five pre-defined life cycle stages based on their mass and energy balances, of which data are extracted from high-fidelity Aspen-Plus-based process simulations and the Ecoinvent V3.6 database. Notably, we extract the chemical composition of pyrolysis products from the experimental data of pyrolyzing waste HDPE. The global warming potential (GWP100) can be implemented into evaluating the greenhouse impacts over the course of 100 years relative to that of CO₂ (Hartmann et al., 2013). The GHG emissions of each life cycle stage are quantified and inputted as environmental parameters to the superstructure optimization problem. The environmental results are shown in the Pareto-optimal curve and GHG emissions breakdowns (Guillén-Gosálbez et al., 2019).

3.2. TEA Methodology

We consider the capital expenditure (CAPEX) and operating expenditure (OPEX) to provide the economic parameters to the superstructure optimization problem. The direct, indirect, and working capitals of all equipment units, as well as the land cost are accounted for calculating the CAPEX, while the OPEX includes the transportation, feedstock, utility, operations and maintenance (O&M) costs, as well as the property tax and insurance (PT&I), sales expense, and income tax. Notably, the net present value (NPV) is calculated via subtracting the CAPEX from the summation of annualized cash flow in each operating year.

4. Model Formulation

$$\begin{aligned} \max \quad & OBJ_{economic} = \frac{NPV}{HDPE \cdot yr} \\ \min \quad & GWP = \frac{(EECST + EEH + EEE + EEW + EED + EEN)_{GWP}}{HDPE} \\ \text{s.t.} \quad & \text{Mass balance constraints} \\ & \text{Energy balance constraints} \\ & \text{Environmental impacts assessment constraints} \\ & \text{Techno-economic evaluation constraints} \\ & \text{Superstructure network configuration constraints} \end{aligned}$$

The superstructure optimization problem is formulated as a multi-objective model that aims to maximize unit NPV ($OBJ_{economic}$) and minimize life cycle GHG emissions (GWP) (Thomassen et al., 2019). Notably, the model is a nonconvex MINFP problem. We introduce nonlinear terms in power functions to calculate capital costs, which are formulated as separable concave terms in the fractional objective functions. The combinatorial nature, pseudo-convexity, and separable concave terms in fractional objectives provide computational challenges for general-purpose global optimizers. Thus, a tailored global optimization algorithm that integrates inexact parametric algorithm (Zhong et al., 2014) with the branch-and-refine algorithm (You and Grossmann, 2011) is applied to tackle this computational challenge.

5. Results and Discussion

5.1. Results of the Superstructure Optimization

As shown in Figure 3, the maximum NPV is \$151.38/ton HDPE treated (corresponding NPV: \$3.4 billion) at point C, which improves by about 25% compared to that from

relevant literature. This improvement is contributed by the versatility of manufacturing multiple products from waste HDPE, heat integration, and the on-site production of high-temperature energy and electricity to reduce the cost of purchasing electricity and natural gas. However, the unit GHG emissions of solution A is 0.4 ton CO₂-eq/ton HDPE treated, which is 57% compared to that of point C. Comparatively, the good-choice solution B reveals minimum unit life cycle GHG emissions (0.44 ton CO₂-eq/ton HDPE treated), which are just 10% higher than that at point A, and a pronounced economic performance (unit NPV: \$142.9/ton HDPE treated, corresponding NPV: \$3.2 billion). The corresponding optimal HDPE recycling process selects MEA absorption as the CO₂ separation process so that more than 85% of GHGs can be removed from the flue gas, which greatly decreases the direct emissions compared to the most economically competitive design of solution C.

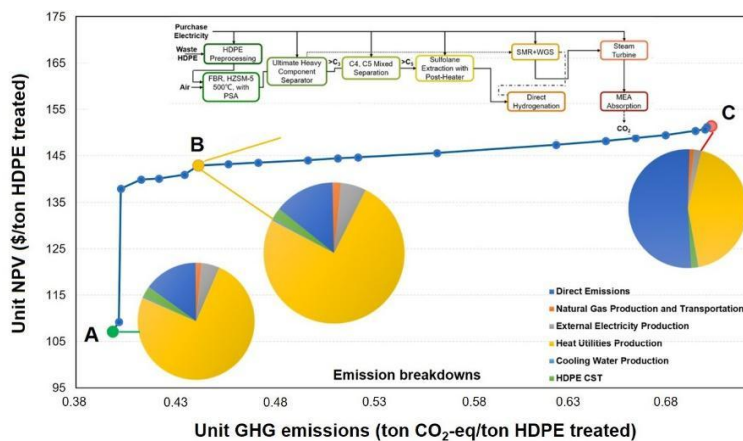


Figure 3. Pareto-optimal curve reveals the trade-offs between the unit NPV and unit life cycle GHG emissions.

5.2. Computational Performance

The superstructure optimization problem includes 104 integer variables, 63,906 continuous variables, and 67,743 constraints. A tailored global optimization algorithm is used for effectively solving this nonconvex problem. We perform the computational experiment on DELL OPTIPLEX 7040 desktop with Intel(R) Core (TM) i7-6700 CPU @ 3.40GHz and 32 GB RAM. The mathematical model and corresponding solution process are coded in GAMS 24.8.3. For all three cases corresponding to optimal solutions A to C, BARON 17.1 fails to find a global optimal solution within 44,000 seconds. Moreover, SCIP 3.2 also fails to find a global optimal solution due to the combinatorial nature and nonconvexity shown in fractional objective functions. Comparatively, all three instances are effectively solved by the tailored global optimization algorithm, which tackles the aforementioned computational challenges, within a few CPU seconds.

6. Conclusion

In this work, the sustainable design and synthesis of waste HDPE recycling process systems were addressed with optimal economic and environmental performances. We developed by far the first superstructure to convert waste HDPE into various chemicals. A tailored global optimization algorithm was applied for the effective solution of the

corresponding superstructure optimization problem. Results showed that the unit NPV of the good choice solution was \$142.9/ton, while its unit life cycle GHG emissions were 0.44 ton CO₂-eq/ton HDPE treated, which was contributed by the integration of MEA absorption. The computational results revealed the advantage of the tailored global optimization algorithm over general-purpose global optimizers.

References

- AG. Buekens, GF. Froment, 1968. Thermal cracking of propane. Kinetics and product distributions. *Industrial & Engineering Chemistry Process Design and Development*. 7(3). 435-447.
- EPA, 2019. Plastics: Material-Specific Data. <https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/plastics-material-specific-data>.
- EPA, 2019. Advancing Sustainable Materials Management: 2016 and 2017 Tables and Figures. <https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/advancing-sustainable-materials-management>
- A. Fivga, I. Dimitriou, 2018. Pyrolysis of plastic waste for production of heavy fuel substitute: A techno-economic assessment. *Energy*. 149. 865-874.Y. Brown, Year, Article or Chapter Title, etc.
- B. H. Gebreslassie, R. Waymire, F. You, 2013, Sustainable Design and Synthesis of Algae-Based Biorefinery for Simultaneous Hydrocarbon Biofuel Production and Carbon Sequestration. *AIChE Journal*, 59, 1599-1621.
- J. Gong, F. You, 2015, Sustainable design and synthesis of energy systems. *Current Opinion in Chemical Engineering*, 10, 77-86.
- J. Gong, F. You, 2018, A new superstructure optimization paradigm for process synthesis with product distribution optimization: Application to an integrated shale gas processing and chemical manufacturing process. *AIChE Journal*, 64, 123-143.
- G. Guillén-Gosálbez, F. You, A. Galán-Martín, et al., 2019, Process systems engineering thinking and tools applied to sustainability problems: current landscape and future opportunities. *Current Opinion in Chemical Engineering*, 26, 170-179.
- D. Hartmann, A. Tank, and M. Rusticucci, 2013. IPCC fifth assessment report, climate change 2013: The physical science basis. *Ipcc Ar5*, 5, 31-39.
- M. Haward, 2018. Plastic pollution of the world's seas and oceans as a contemporary challenge in ocean governance. *Nature Communications*, 9, 667.
- C. He, F. You, 2014, Shale Gas Processing Integrated with Ethylene Production: Novel Process Designs, Exergy Analysis, and Techno-Economic Analysis. *Industrial & Engineering Chemistry Research*, 53, 11442-11459.
- RA. Meyers, 2004. Handbook of petroleum refining processes. 548. McGraw-Hill New York.
- M. del Remedio Hernández, ÁN. García, A. Marcilla, 2007. Catalytic flash pyrolysis of HDPE in a fluidized bed reactor for recovery of fuel-like hydrocarbons. *Journal of analytical and applied pyrolysis*. 78(2). 272-281.
- G. Thomassen, M. Van Dael, F. You, et al., 2019, A multi-objective optimization-extended techno-economic assessment: exploring the optimal microalgal-based value chain. *Green Chemistry*, 21, 5945-5959.
- W. Wang, Z. Gou, S. Zhu, 1998. Liquid-liquid equilibria for aromatics extraction systems with tetraethylene glycol. *Journal of Chemical & Engineering Data*. 43(1):81-83.
- EA. Williams, PT. Williams, 1997. Analysis of products derived from the fast pyrolysis of plastic waste. *Journal of Analytical and Applied Pyrolysis*. 40-41. 347-363.
- M. Yang, F. You, 2017, Comparative Techno-Economic and Environmental Analysis of Ethylene and Propylene Manufacturing from Wet Shale Gas and Naphtha. *Industrial & Engineering Chemistry Research*, 56, 4038-4051.
- F. You, I. E. Grossmann, 2011, Stochastic Inventory Management for Tactical Process Planning Under Uncertainties: MINLP Models and Algorithms. *AIChE Journal*, 57, 1250-1277.
- Z. Zhong, F. You, 2014, Globally convergent exact and inexact parametric algorithms for solving large-scale mixed-integer fractional programs and applications in process systems engineering. *Computers & Chemical Engineering*, 61, 90-101.

A Novel Process Design for Waste Respirator Processing

Xiang Zhao, Fengqi You

Cornell University, Ithaca, New York 14853, USA

xz643@cornell.edu

Abstract

The ongoing COVID-19 pandemic increases the consumption of respirators. In this work, we propose a novel and effective waste respirator processing system that aims to protect public health and mitigate climate change. Respirator sterilization and pre-processing technologies are incorporated simultaneously to resist viral infection and facilitate unit processes for manufacturing and separating products, so the greenhouse gas (GHG) emission can be reduced via carbon reallocation from CO₂ to downstream products. High-fidelity process simulations are performed to extract detailed life cycle inventories used for evaluating environmental performance. Results reveal the economic viability in terms of the payback time (seven years) and the internal rate of return (21.5%). The proposed waste respirator processing system reduces GHG emissions by 59.08% compared to incineration, which reflects the potential of climate change mitigation.

Keywords: waste respirator processing, COVID-19, process design and integration.

1. Introduction

The U.S. has been trapped in the COVID-19 pandemic with no shutdown in sight. During the COVID-19 pandemic, people consume massive respirators to stem coronavirus infection, which substantially triggers the usage and disposal of respirators across the U.S (Bartoszek et al., 2020). Notably, the virus will transmit among the public if massive discarded respirators are mismanaged. In this regard, we apply the incineration process to treat these waste respirators in typical medical waste disposal sites (Nzediegwu and Chang, 2020). However, a large amount of greenhouse gases (GHGs) are emitted when incinerating respirators, which give rise to serious climate change (Klemeš et al., 2020). Even worse, the mixed-plastic component within discarded respirators can be decomposed into toxic chemicals (Bora et al., 2020), which are then digested by organics and accumulated in the food-web (Heward, 2018). In these regards, we currently call for a novel and effective waste respirator processing system to reduce the risk of viral infection and GHG emissions when fighting for the COVID-19 pandemic.

To fill in this current knowledge gap, we develop and propose a waste respirator processing system, which incorporates respirator sterilization and pre-processing technology to protect the public health and for manufacturing and separating products, so the greenhouse gas (GHG) emission can be reduced via carbon reallocation from CO₂ to downstream products. We consider seven sections in the processing system, namely respirator preprocessing, pyrolysis, light hydrocarbon separation, CO₂ separation, hydrogenation, hydrogen production, and onsite combustion. A commercially available sterilization process, which can shred, sterilize, and dehydrate the waste N95 respirators, is included in this system to disinfect respirators and triggers the thermal-cracking process

in the pyrolysis section (Nutsch and Spire, 2004) similar to plastics processing (Zhao and You, 2021). We apply a detailed life cycle assessment (LCA) approach to systematically quantify the GHG emissions from cradle-to-gate so that the potential in mitigating climate change can be evaluated. Specifically, high-fidelity process simulations of the waste respirator processing system integrated with the data from Ecoinvent V3.6 are used for extracting the detailed life cycle inventory. Techno-economic analysis (TEA) is also conducted to evaluate the economic viability by calculating the capital and operating expenses. We demonstrate the economic feasibility of establishing respirator processing systems and their potential for climate change mitigation through evaluating a proposed respirator processing system, which aims to treat 582 million waste N95 respirators that are corresponding to the HHS's recommended annual production amount in eight northeastern states in the U.S. (NY Times, 2020), namely New York, New Jersey, Pennsylvania, Massachusetts, New Hampshire, Vermont, Rhode Island, and Connecticut.

2. Process Description for the Waste Respirator Processing System

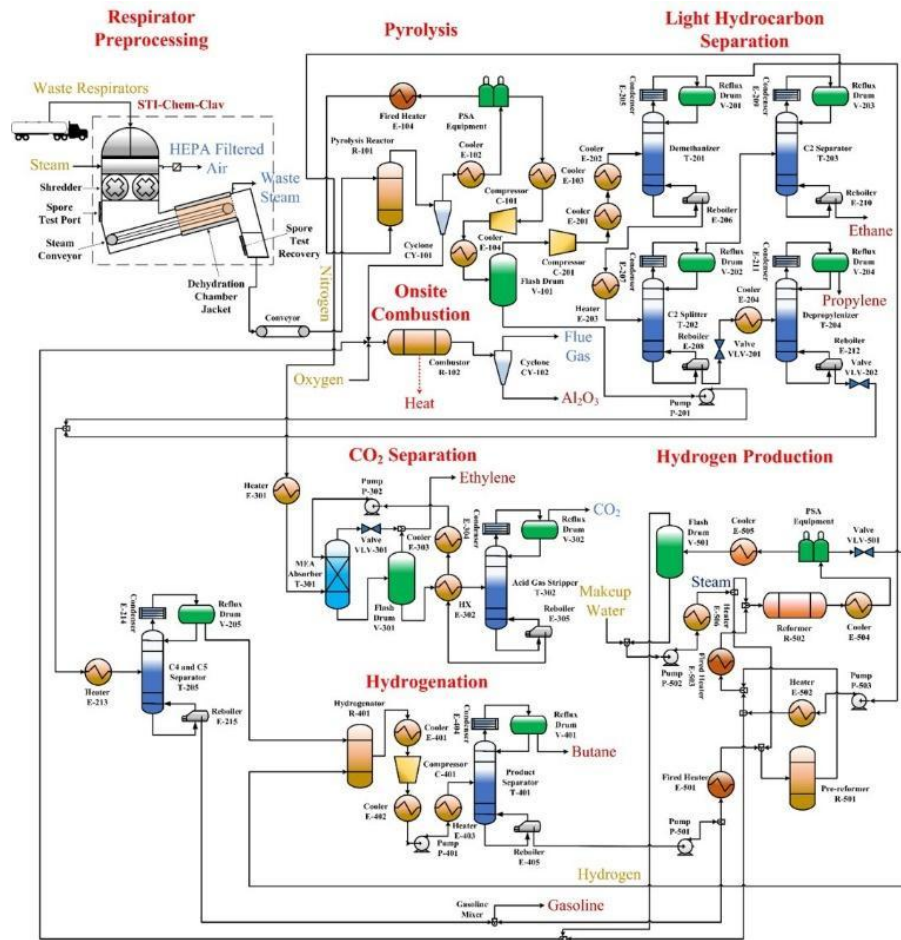


Figure 1. Process flowsheet of the proposed waste respirator processing system.

In this work, we develop and propose a novel and effective waste respirator processing system. As shown in Figure 1, the processing system integrates seven sections, namely respirator preprocessing, pyrolysis, light hydrocarbon separation, CO₂ separation, hydrogenation, hydrogen production, and onsite combustion. Detailed description of the whole processing system is shown as follows.

The whole process starts with the respirator preprocessing section, where the transported waste respirator can be shredded, sterilized, and dehydrated by steam (147.7°C, 44.61 bar) simultaneously. Notably, this section can effectively deactivate the coronavirus under the sterilization condition (95–120°C) (Nutsch and Spire, 2004). With the usage of nitrogen gas as the inert fluidized gas, the disinfected respirator particles are then thermally cracked into various inorganic and hydrocarbon chemicals in a fluidized bed pyrolyzer. The nitrogen gas is split by a pressure-swing adsorption (PSA) unit and circulated in this section. The volatile stream is then split into streams with light and heavy components in a flash tank, while the stream with char is sent to the onsite combustion section to generate high-temperature heating energy. In the light hydrocarbon separation section, the stream with light components is split into methane, ethane, ethylene, and propylene products. Notably, the methane stream is directly sent to hydrogen production or onsite combustion sections (Yang et al., 2018), while the stream with ethylene is fed into the CO₂ separation section to separate CO₂ from ethylene products. In the C₄ and C₅ Separator (He et al., 2015), the overhead liquid with C₄ components is sent to the hydrogenation section to produce butane product, while the raffinate is fed into a gasoline mixer or used for producing hydrogen (Gong et al., 2017).

Specifically, we apply NiMo catalyst (Swanson et al., 2010) in the hydrogenator to convert C₄ components into butane in the hydrogenation section. To satisfy the usage of hydrogen in this section, a hydrogen production section is implemented within the waste respirator processing system. Notably, the partial raffinate stream from the hydrogenation section is mixed with steam and produces lighter components in the pre-reformer to trigger the steam-methane reforming reaction in the downstream reformer. In the reformer, the steam is mixed with streams of methane and CO from the light hydrocarbon separation section and the mixture is converted into hydrogen following the kinetic of steam-methane reforming reaction (He et al., 2016). The hydrogen stream is separated from the PSA unit and sent to the hydrogenation section, while the steam is regenerated via pressurizing and heating the mixture of the makeup water stream and pre-cooled water stream from the flash tank. The remaining gaseous stream from the flash tank is mixed with oxygen, char, and gaseous chemical components, such as methane, and sent to the combustor to be ignited under 1000°C. The high-temperature heating energy is released from the combustor, while the solid stream (mainly Al₂O₃) is sold.

3. LCA and TEA Methodology

3.1. LCA Methodology

In this work, we apply the LCA methodology to systematically evaluate the potential of climate change mitigation via quantifying the GHG emissions from the waste respirator processing system, so the goal of this LCA is defined. Five life cycle stages, namely waste respirator transportation, waste respirator processing, offsite heating utilities production, offsite electricity production, and offsite production for inlet material, are confined by the system boundary from cradle-to-gate, as shown in Figure 2. We choose the functional unit as one thousand respirators treated within the waste HDPE processing system.

We compile the detailed life cycle inventories (LCIs) based on the mass and energy balances throughout these five life-cycle stages, of which data are extracted with the help of combining high-fidelity Aspen-Plus-based process simulations and Ecoinvent V3.6 Database. In the process simulation, we assume the chemical composition of pyrolysis products as the weighted average of chemical composition for pyrolyzing each single plastic compound (Westerhout et al., 1998).

We apply the global warming potential over the course of 100 years (GWP_{100}) to quantify the greenhouse impacts relative to that of CO_2 and thus evaluate the potential of climate change mitigation. Specifically, the GWP_{100} of methane is 28 due to the same greenhouse impacts of emitting 1 kg of methane and those of 28 kg CO_2 over the course of 100 years (Hartmann et al., 2003). The GHG emissions of each life cycle stage are quantified in this work and the results are shown as GHG emissions breakdowns.

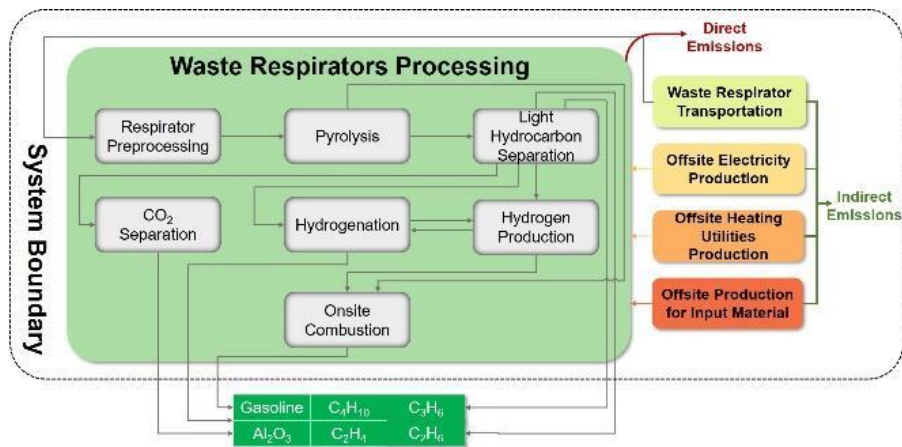


Figure 2. System boundary of a cradle-to-gate LCA for the waste respirator processing system.

3.2. TEA Methodology

We consider the capital expenditure (CAPEX) and operating expenditure (OPEX) to evaluate the economic viability of establishing the respirator processing system in terms of the net present value (NPV), payback year, and internal rate of return (IRR). CAPEX includes the direct capital, indirect capital, working capital of all equipment units, and the land cost used for setting up the processing system (Gong and You, 2018). OPEX includes the cost of transporting waste respirators, feedstock cost, utility cost, cost of operations and maintenance (O&M), property tax and insurance (PT&I), general expense, and income tax. The linear depreciation method is adopted to calculate the depreciation cost. Notably, the net present value (NPV) is calculated via subtracting the CAPEX from the summation of annualized cash flow in each operating year. The direct capital costs are extracted from Aspen-Plus Capital Cost Estimator.

4. Results and Discussion

4.1. TEA Results of the Waste Respirator Processing System

We present the breakdowns of CAPEX and OPEX of establishing the waste respirator processing system (near Citiwaste Medical Waste Disposal) in Figure 3. Notably, the total capital investment (\$16.31 million) is the major contributor to the total expenses, which

is further broken down into four categories, namely the total installation cost, total indirect capital, land cost, and working capital. The income tax is another major contributor (\$14.14 million), so that the high profitability of the waste respirator processing system is demonstrated. Specifically, the high indirect capital (\$5.82 million) is owing to the high costs for the procurement and installation of various equipment units and reactors (\$9.71 million). Owing to various distillation units and the refrigeration cycle to maintain cryogenic conditions when separating methane, the installed cost of the light hydrocarbon separation section (\$3.21 million) mainly contributes to the total installation cost, as shown in Figure 3. The installed costs for other sections are also shown in Figure 3.

The feasibility of establishing the waste respirator processing system can be illustrated in terms of the payback time of seven years, and an IRR of 21.5%. This high economic performance is mainly contributed by the various products converted from waste respirators, as well as the heat integration that minimizes the utility usage.

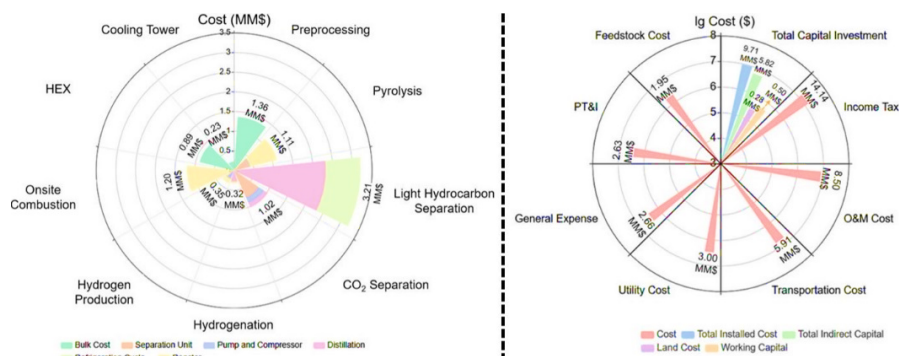


Figure 3. CAPEX, OPEX, and installed cost breakdowns.

4.2. LCA Results of the Waste Respirator Processing System

The direct emissions from the offsite combustion section and indirect emissions share 56% and 44% of total emissions, respectively. The indirect emissions can be further broken down and the major contributor to the indirect emissions is from the offsite production for inlet material (56% of indirect emissions). Moreover, GHG emissions from the offsite production of steam contribute most emissions (91.11%) among the offsite production for inlet materials, which due to the massive usage of steam in the sterilization system. Specifically, the unit GHG emissions of the respirator processing system is 12.93 kg CO₂-eq/thousand respirators, which reduces by 59.08% compared to the incineration-based system (31.60 kg CO₂-eq per thousand respirators). Hence, it is viable to establish a respirator processing system with high economic profitability and the potential to mitigate climate change.

5. Conclusion

In this work, we developed and proposed a novel and effective waste respirator processing system to protect public health and mitigate climate change. The waste respirator processing system included seven sections to convert waste respirators into various products. TEA results deciphered the economic viability for setting up the waste respirator processing system in terms of the payback time of seven years with an IRR of

21.5%. LCA results illustrated the potential of climate change mitigation by showing a reduction of GHG emissions by 59.08% compared to the incineration-based system.

References

- J. J. Bartoszko, M. Farooqi, W. Alhazzani, M. Loeb, 2020, Medical masks vs N95 respirators for preventing COVID-19 in healthcare workers. *Influenza and other respiratory viruses*, 14, 365.
- R. R. Bora, R. Wang, F. You, 2020, Waste Polypropylene Plastic Recycling toward Climate Change Mitigation and Circular Economy: Energy, Environmental, and Technoeconomic Perspectives. *ACS Sustainable Chemistry & Engineering*, 8, 16350-16363.
- J. Gong, F. You, 2015, Sustainable design and synthesis of energy systems. *Current Opinion in Chemical Engineering*, 10, 77-86.
- J. Gong, M. Yang, F. You, 2017, A systematic simulation-based process intensification method for shale gas processing and NGLs recovery process systems under uncertain feedstock compositions. *Computers & Chemical Engineering*, 105, 259-275.
- J. Gong and F. You, 2018, A new superstructure optimization paradigm for process synthesis with product distribution optimization: Application to an integrated shale gas processing and chemical manufacturing process. *AIChE Journal*, 64(1), 123-143.
- A. Nutsch and M. Spire, 2004. Carcass disposal: a comprehensive review. National Agricultural Biosecurity Center, Kansas State University, Manhattan, KS.
- D. Hartmann, A. Tank, and M. Rusticucci, 2013. IPCC fifth assessment report, climate change 2013: The physical science basis. *Ipcc Ar5*, 5, 31-39.
- J. J. Klemeš, Y. Fan, R. R. Tan, P. Jiang, 2020. Minimising the present and future plastic waste related to COVID-19. *Renewable and Sustainable Energy Reviews*, 127, 109883.
- M. Haward, 2018. Plastic pollution of the world's seas and oceans as a contemporary challenge in ocean governance. *Nature Communications*, 9, 667.
- C. He, F. You, 2014, Shale Gas Processing Integrated with Ethylene Production: Novel Process Designs, Exergy Analysis, and Techno-Economic Analysis. *Industrial & Engineering Chemistry Research*, 53, 11442-11459.
- C. He, F. You, 2016, Deciphering the true life cycle environmental impacts and costs of the megascale shale gas-to-olefins projects in the United States. *Energy & Environmental Science*, 9, 820-840.
- NYS Government, Apr. 2020. Continuing Temporary Suspension and Modification of Laws Relating to the Disaster Emergency.
- C. Nzediegwu, S. X. Chang, 2020. Improper solid waste management increases potential for COVID-19 spread in developing countries. *Resour Conserv Recycl*, 161, 104947-104947.
- NYT, Mar. 2020. Some Hospitals Are Close to Running Out of Crucial Masks for Coronavirus. <https://www.nytimes.com/2020/03/09/health/coronavirus-n95-face-masks.html>
- R. M. Swanson, A. Platon, J. A. Satrio, R. C. Brown, and D. D. Hsu, 2010. Techno-Economic Analysis of Biofuels Production Based on Gasification. National Renewable Energy Lab, Golden, CO (United States).
- R. W. J. Westerhout, M. P. Van Koningsbruggen, A. G. Van Der Ham, J. A. M. Kuipers, and W. P. M. Van Swaaij, 1998. Techno-economic evaluation of high temperature pyrolysis processes for mixed plastic waste. *Chemical Engineering Research and Design*, 76(3), 427-439.
- M. Yang, X. Tian, F. You, 2018, Manufacturing Ethylene from Wet Shale Gas and Biomass: Comparative Technoeconomic Analysis and Environmental Life Cycle Assessment. *Industrial & Engineering Chemistry Research*, 57, 5980-5998.
- M. Yang, F. You, 2018, Modular methanol manufacturing from shale gas: Techno-economic and environmental analyses of conventional large-scale production versus small-scale distributed, modular processing. *AIChE Journal*, 64, 495-510.
- X. Zhao, F. You, 2021, Waste high-density polyethylene recycling process systems for mitigating plastic pollution through a sustainable design and synthesis paradigm. *AIChE Journal*, 67, e17127.

Design and optimization of a shell-and-tube heat exchanger using the univariate marginal distribution algorithm

Oscar D. Lara Montaña^a, Fernando I. Gómez-Castro^{a*} and Claudia Gutierrez-Antonio^b

^a*Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Campus Guanajuato, Universidad de Guanajuato, Noria Alta S/N, Guanajuato, Guanajuato 36050, México*

^b*Facultad de Ingeniería, Campus Amazcala, Universidad Autónoma de Querétaro, Carretera a Chichimequillas s/n km 1, El Marqués, Querétaro, 76225, México*
fgomez@ugto.mx

Abstract

Estimation of Distribution Algorithms (EDAs) are interesting evolutionary methods; they have the characteristic of using an explicit distribution model, instead of mutation and crossover operators. In particular, Univariate Marginal Distribution Algorithms (UMDA) are some of the simplest distribution-based methods. In this work UMDA is used to design and optimize a shell-and-tube heat exchanger. To model the heat exchanger, Kern method and the Bell-Delaware method are employed. It has been found that UMDA has an inferior performance when compared with other metaheuristic algorithms. Nevertheless, other EDAs may lead to better results, which is part of future research.

Keywords: Heat exchanger, optimization, metaheuristic, estimation of distribution algorithm, Bell-Delaware.

1. Introduction

Heat exchangers are auxiliary process equipment that are necessary in almost every industry. Its main purpose is to recover energy from a process stream. Among the different available heat exchanger technologies, the shell-and-tube heat exchanger (STHE) is the most commonly used by industry. STHEs have some advantages, such as standardized design and building procedures, good heat transfer area/equipment volume ratio, and enough robustness to support high operational temperatures and pressures. Different methodologies to design STHEs can be found in open literature. Some of the most widely used are the Kern's and the Bell-Delaware methods. For both methods, many different STHEs designs can be calculated to satisfy the same energetic criteria. Therefore, it is necessary to perform an optimization routine to identify the best possible design according to an objective function.

The optimization of a STHE is not a trivial problem. Both design models contain non-convex, non-continuous, non-differentiable functions. In addition, they depend on continuous and discrete design variables. This results in a mixed-integer non-linear problem, where many local optima could exist. Deterministic procedures have been used to find optimal designs for this kind of equipment, as done by Yang et al. (2020), where a disjunctive mathematical model is proposed and considers five different heat transfer enhancement techniques. Recently, Kazi et al. (2021) developed a discretized differential algebraic equation model that can be combined with logarithmic mean temperature difference model to produce optimal STHes. Other reliable alternative is the use of evolutionary strategies to obtain optimal designs; such metaheuristic techniques does not require model adequation. Evolutionary Algorithms (EAs) are a class of optimization methods inspired in the Darwinian theory of survival of the fittest. These methods employ operators such as selection, crossover, and mutation. The different parameters of the operators must be tuned before solving an optimization problem; this represents an optimization problem in itself. To avoid the tuning procedure, EDAs were proposed as an alternative by Larrañaga (2002). In EDAs, there are neither crossover nor mutation operators. Instead, the new population (solutions) are created from an explicit probability distribution, that is built from a set of selected individuals of the previous generation. In EDAs, the interrelations of the design variables are expressed through the joint probability distribution. UMDA is a type of EDA that uses the simplest model to calculate the joint probability distribution; this model does not consider the interdependence between variables. It is reported that UMDA do not has a good performance working with non-linear problems and with problems with significant dependencies. In this work it is used as a starting point to be further compared with more complex EDAs that could work better to solve heat transfer problems.

Some works have used evolutionary strategies to optimize STHes; for example the proposal by Ponce et al. (2006) where genetic algorithms (GAs) were used, or the approach of Xiao et al. (2019) where the simultaneous design of a heat exchanger network and STHes are performed using a hybrid genetic algorithm/simulation annealing method. To the author's knowledge, distribution-based algorithms have not been used to optimize STHes. Thus, in this work a STHE is designed and optimized for a case study. Kern's and the Bell-Delaware methods are used to model the heat exchanger. As an optimization algorithm UMDA is selected, and its performance is compared with other metaheuristic methods previously used.

2. Shell-and-tube heat exchanger models

The Kern and the Bell-Delaware methods are widely used to design STHes. Kern's method gives good predictions of heat-transfer coefficients for standard designs. The prediction of the pressure drop is less accurate, because this variable is more affected by shell-side leakages, and the method does not consider it (Sinnott, 2005). On the other hand, the Bell-Delaware methodology provides a better prediction of all variables. In this method, multiple geometrical variables are calculated creating a highly non-linear search space. Also, five parameters are used to calculate the shell-side heat transfer, h_s , according to equation 1.

$$h_s = h_{id} J_c J_l J_b J_s J_r \quad (1)$$

where h_{id} is the ideal heat-transfer coefficient. J_c is the correction factor for baffle configuration. J_l is the correction factor for baffle leakage. J_b is the correction factor for bundle and pass partition bypass streams. J_s is the correction factor for baffle spacing at inlet and outlet. Finally, J_r is the correction factor for adverse temperature gradient in laminar flows.

The area of the heat exchanger is calculated for both models with equation 2. Q is the heat transfer rate, T_{LMTD} is the logarithmic mean temperature difference, F is a correction factor for T_{LMTD} , and U is the overall heat transfer coefficient computed with equation 3.

$$A = \frac{Q}{UFT_{LMTD}} \quad (2)$$

$$U = \frac{1}{\frac{1}{h_s} + R_{fs} + \frac{d_o \ln \left(\frac{d_o}{d_i} \right)}{2k_w} + R_{ft} \frac{d_o}{d_i} + \frac{1}{h_t} \frac{d_o}{d_i}} \quad (3)$$

Additionally, h_t is the tube-side heat transfer coefficient, d_i is the tube internal diameter, d_o is the tube outer diameter, R_{fs} and R_{ft} are the fouling factors for shell and tubes, respectively. Kern's method can be found in Flynn et al. (2019), and the Bell-Delaware methodology is reported by Shah and Sekulic (2003).

Commonly, the total annual cost (TAC) is used as an optimization criteria, and it depends on multiple variables and parameters as shown in equation 4 (Smith, 2005). TAC is calculated with the fixed cost, C_f , and operational cost, C_{op} , which are computed with equations 5 and 6, respectively. Also, C_p is the purchase cost of the STHE, f_m , f_p and f_c are factors to consider the construction material, operating pressure and operating temperature, respectively.

$$TAC = C_f \frac{r(1+r)^n}{(1+r)^n - 1} + C_{op} \quad (4)$$

$$C_f = C_p f_m f_p f_c \quad (5)$$

$$C_{op} = \frac{(E_s + E_t) Ec H_r}{1000} \quad (6)$$

E_s and E_t are the required pumping power (W) for shell and tube side. Ec , is the cost of the energy required for pumping. H_r is the number of operational hours per year, n is the projected lifetime, and r is the interest rate.

3. Univariate Marginal Distribution Algorithm

UMDA is a population-based evolutionary algorithm for optimization. It belongs to EDAs, which do not use operators such as mutation nor crossover to diversify the population and better explore the search space; instead, it uses an explicit model to estimate the joint probability distribution. The joint probability distribution is factored as a product of independent univariate marginal distribution, as shown in equation 7. p is the joint probability distribution, p_i is the univariate marginal frequency (Larrañaga, 2002).

$$p(X) = \prod_{i=1}^n p_i(x_i) \quad (7)$$

The selection of the sampled individuals is a very important part of this algorithm; for this work truncation is used. UMDA begins with a random population, which is ranked according to the values obtained from the objective function evaluation. Then, the best half of the population is used to build the distribution model; the population of the next generation is obtained from this model.

4. STHE optimization

The case study has been taken from Wildi-Tremblay and Gosselin (2007). It is required to design a STHE with a flow rate of 18.8 kg/s of cooling water. The inlet temperature is 33 °C and the outlet temperature is 37.2 °C. The hot fluid is naphtha with an inlet temperature of 114 °C, and 40 °C for the outlet temperature. Water is placed in tube-side because of its greater fouling resistance. The construction material for tube and shell side is stainless steel and carbon steel, respectively. Constraints are applied according to equations 8, 9 and 10, where ΔP_s and ΔP_t are the pressure drop in shell and tube side, respectively. v_t is the fluid velocity in tube-side, L is the heat exchanger length, and D_s is the diameter of the heat exchanger.

$$\Delta P_s, \Delta P_t \leq 70,000 \text{ Pa} \quad (8)$$

$$1 \text{ m/s} \leq v_t \leq 3 \text{ m/s} \quad (9)$$

$$L/D_s < 15 \quad (10)$$

The Bell-Delaware methodology depends on 11 design variables, while Kern's method on 7. Table 1 shows the established boundaries for each continuous variable. In Table 2 the allowed values for discrete variables can be found. The possibility of designs with multiple shells is not considered at this stage of the study.

Table 1. Limits of continuous design variables.

Design variable		Lower bound	Upper bound
Diameter of shell	D_s	300 mm	1,000 mm
Outer diameter of tube	d_o	15.87 mm	63.5 mm
Baffle spacing at center	L_{bc}	$0.2D_s$	0.55 mm
Baffle spacing at the inlet and outlet	L_{bo}, L_{bi}	L_{bc}	$1.6L_{bc}$
Tube-to-baffle diametrical clearance	δ_{tb}	$0.01d_o$	$0.1d_o$
Diametrical clearance of shell-to-baffle	δ_{sb}	$0.01D_s$	$0.1D_s$
Outer diameter of tube bundle	D_{otl}	0.8 ($D_s - d_{sb}$)	0.95 ($D_s - d_{sb}$)

Table 2. Allowed values used in discrete design variables.

Design variables	Allowed values
Tube pitch	P_t [1.25 d_o , 1.5 d_o]
Tube layout angle	TL [30°, 45°, 90°]
Baffle cut	B_c [25%, 30%, 40%, 45%]
Number of tube passes	s [1, 2, 4]

The equations 11 and 12 show the objective function for each method.

$$\min TAC = f(D_s, d_o, P_t, TL, N_p, B_c, L_{bc}, L_{bo}, d_{sb}, d_{tb}, D_{otl}) \quad \text{Bell-Delaware} \quad (11)$$

$$\min TAC = f(D_s, d_o, TL, N_p, L_{bc}, d_{sb}, D_{otl}) \quad \text{Kern's method} \quad (12)$$

If constraints are not satisfied a penalty is applied to the TAC; the penalized objective function is shown in equation 13. r_1, r_2, r_3 and r_4 are the penalty values, and ΔP_a is the maximum allowed pressure drop (70,000 Pa). To estimate the cost, 5,000 operational hours were assumed, an interest rate of 5%, a projected lifetime of 20 years, a pump efficiency equal to 0.85, and an electricity cost of 0.1 USD/kWh (Wildi-Tremblay and Gosselin, 2007).

$$\begin{aligned} \text{fobj} = \text{TAC} + r_1 \max[(\Delta P_s - \Delta P_a), 0] + r_2 \max[(\Delta P_t - \Delta P_a), 0] & \quad (13) \\ + r_3 [\max\{(1 - v_t)\} + \max\{(3 - v_t)\}, 0] \\ + r_4 \max[(L/D_s - 15), 0] \end{aligned}$$

5. Results

Fluid properties were computed at average temperature. For the UMDA, 50 individuals and 50 iterations were considered. The selection operator uses truncation with 50% of the total population. 30 experiments were done to obtain basic statistical information as the average and the standard deviation of the TAC. The required time to conduct one experiment for the Bell-Delaware and Kern’s method are 3.95 s and 4.01 s, respectively.

When Bell-Delaware methodology is used, the best design has a TAC of 4,093.96 USD/year, the C_{op} is 156.73 USD/year and the C_f is 3,937.23 USD/year. The TAC of the best design when Kern’s method is used is 3,613.88 USD/year, C_{op} and C_f are 286.46 USD/year and 3,327.41 USD/year, respectively. Table 3 reports the results of some important variables for both models, and its comparison with the results reported in a previous work (Lara-Montaña et al., 2020) where particle swarm optimization (PSO) and grey wolf optimization (GWO) were used. From this comparison it is evident that for both STHes models UMDA has a poor performance with respect to other metaheuristic algorithms. The optimization algorithms used in the referenced work have a better capability to avoid local optima. This can be inferred from the average TAC value that is 4,234.35 USD/year for PSO and 4,074.93 USD/year for GWO.

Table 3. Best design found and comparison with other methods (Lara-Montaña et al., 2020).

Parameter	D_s	d_o	N_t	A	L	P_t	TL	s	L_{bc}
Units	mm	mm		m^2	m	mm	°		mm
Bell-Delaware (this work)	311.83	15.87	149	34.23	4.60	19.83	90	1	79.65
Kern’s method (this work)	300.00	15.87	195	25.76	2.80	18.75	30	1	87.79
Bell-Delaware (PSO)	302.40	15.87	144	32.56	4.56	-	-	-	-
Bell-Delaware (GWO)	302.40	15.87	144	32.56	4.56	-	-	-	-
Kern’s method (PSO)	314.10	15.00	672	25.95	2.55	-	-	-	-
Kern’s method (GWO)	314.80	15.00	672	26.03	2.54	-	-	-	-

Parameter	L_{bo}	B_c	δ_{tb}	δ_{sb}	D_{atl}	C_{op}	C_f	TAC
Units	mm	%	mm	mm	mm	USD/year	USD/year	USD/year
Bell-Delaware (this work)	120.65	15	0.18	3.11	289.48	156.73	3,937.23	4,093.96
Kern’s method (this work)	-	-	-	-	239.22	286.46	3,327.41	3,613.88
Bell-Delaware (PSO)	-	-	-	-	-	160.51	3,865.99	3,966.50
Bell-Delaware (GWO)	-	-	-	-	-	160.51	3,806.01	3,966.53
Kern’s method (PSO)	-	-	-	-	-	263.34	3,262.45	3,525.80
Kern’s method (GWO)	-	-	-	-	-	257.76	3,275.97	3,525.73

After 30 experiments, an average TAC of 4,674.12 USD/year and a standard deviation of 296.68 USD/year is calculated when the Bell-Delaware model is employed with the UMDA. From these results, it can be inferred that the optimization algorithm has a poor capability to avoid local optima. The value of the standard deviation is moderate, but the difference between the best TAC and the average TAC is large. This indicates that in most of the experiments the algorithm converges in a solution far from the global

optima. Only for one of the experiments the minimum reported value of TAC was reached. For Kern's method, the average TAC and standard deviation are 3,701.81 USD/year and 44.72 USD/year, respectively. The performance of the optimization algorithm is better when Kern's method is used due its lower non-linearity and the lower number of design variables implied, compared with the Bell-Delaware method. Figure 1 shows the behavior of the TAC in each iteration for both design methods. It is evident that for Kern's method the optimization algorithm converges in less than 10 iterations. When the Bell-Delaware method is employed the UMDA method rapidly find better solutions in the first iterations, but after 25 iterations the convergence slows down, being unable to find the best design in 50 iterations.

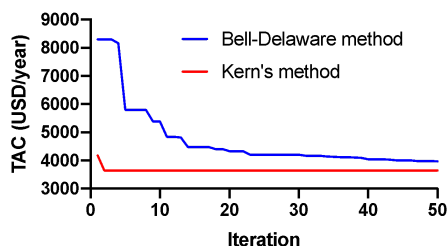


Figure 1. Behavior of the TAC for both design methods

6. Conclusion

This work explores the viability of using EDAs to solve heat transfer problems. From the results obtained it is clear that UMDA is not a good option to optimize STHes, due its poor capacity to work with non-linear models. This optimization method uses a joint probability that does not take into account dependencies between variables. Better results could be found if a more complex joint probability were employed. It is clear that the difference in the complexity of the design methods impacts on the efficiency of the optimization algorithm. When Kern's method is used, the average TAC is closer to the best TAC. This does not occur for the Bell-Delaware method that is more complex and depends on more design variables. Although the results reported in this work demonstrate that UMDA is not capable of optimizing STHes efficiently, other EDAs could have better performance by using a joint probability distribution that consider dependencies between variables.

References

- S.R. Kazi, M. Short, L.T. Biegler, 2021, Heat exchanger network synthesis with detailed exchanger designs: Part 1. A discretized differential algebraic equation model for shell and tube heat exchanger design, *AIChE Journal*, 67, 1, e17056.
- O.D. Lara-Montaña, F.I. Gómez-Castro, C. Gutiérrez-Antonio, 2020, Development of a virtual environment for the rigorous design and optimization of shell-and-tube heat exchangers, *Computer Aided Chemical Engineering*, 48, 19-24.
- P. Larrañaga, 2002, A review on estimation of distribution algorithms, in *Estimation of distribution algorithms* (Eds. P. Larrañaga, J.A. Lozano), Springer, 57-100.
- J. M. Ponce, M. Serna, V. Rico, A. Jimenez, 2006, Optimal design of shell-and-tube heat exchangers using genetic algorithms. *Computer Aided Chemical Engineering*, 21, 985-990
- R.K. Shah, D.P. Sekulic, 2003, *Fundamentals of heat exchanger design*, Wiley.
- R. Sinnott, 2005, *Chemical Engineering Design*, Coulson and Richardson's Chemical Engineering Series, Vol. 6, Elsevier.

- R. Smith, 2005, Chemical process design and integration, Wiley.
- P. Wildi-Tremblay, L. Gosselin, 2007, Minimizing shell-and-tube heat exchanger cost with genetic algorithms and considering maintenance, International Journal of Energy Research, 31, 9, 867-885.
- W. Xiao, K. Wang, X. Jiang, X. Li, X. Wu, Z. Hao, G. He, 2019, Simultaneous optimization strategies for heat exchanger network synthesis and detailed shell-and-tube heat-exchanger design involving phase changes using GA/SA, Energy, 183, 1166-1177.
- Z. Yang, Y. Ma, N. Zhang, R. Smith, 2020, Design optimization of shell and tube heat exchangers sizing with heat transfer enhancement, Computers and Chemical Engineering, 137, 106821.

Synthesis of Microgels with Minimal Final Concentration of Initiator

Falco Jung,^a Johannes M. M. Faust,^a Adel Mhamdi,^a Andrij Pich,^{b,c,d,e} and Alexander Mitsos^{a,b}

^a*Aachener Verfahrenstechnik-Process Systems Engineering, RWTH Aachen University, Aachen, Germany, amitsos@alum.mit.edu*

^b*JARA-SOFT, 52056 Aachen, Germany*

^c*DWI Leibniz Institute for Interactive Materials e.V., Aachen, Germany*

^d*Institute of Technical and Macromolecular Chemistry, RWTH Aachen University, Aachen, Germany*

^e*Aachen Maastricht Institute for Biobased Materials, Maastricht University, Brightlands Chemelot Campus, Urmonderbaan 22, 6167 RD Geleen, The Netherlands*

Abstract

Microgels are functional polymers with diverse applications synthesized by free radical polymerization using an initiator, monomers, and cross-linker as reactants. While monomers and cross-linker are incorporated into the microgel during the synthesis, initiator is left in the reactor due to its slow decomposition rate. The removal of unreacted initiator is a challenging and time-consuming post-synthesis procedure. Therefore, we use model-based dynamic optimization to determine synthesis procedures that reduce in the final concentration of unreacted initiator. The dynamic optimization requires a mechanistic model of the microgel synthesis, that needs to consider the inhibition of residual oxygen in the reactor at the beginning of the synthesis. Residual oxygen might inhibit the initiation of the synthesis when initiator concentrations are sufficiently low. We extend our existing model with the effect of an initial oxygen inhibition. We then use dynamic optimization to minimize the initial initiator concentration, while maintaining process and quality constraints. The considered product quality constraint is the hydrodynamic radius of the microgel, which must not change by more than 5 %. The determined synthesis procedure shows a 76 % reduction in the final concentration of unreacted initiator in comparison to the reference case, while maintaining the reference microgel hydrodynamic radius. The synthesis time, however, is 48 % longer than the synthesis time of the reference case. Hence, there is a trade-off between the final initiator concentration and the duration of the microgel synthesis demonstrating the value of dynamic optimization for the determination of optimal synthesis conditions valuable.

Keywords: microgels, initiator, dynamic optimization, model-based synthesis

1. Introduction

Microgel syntheses involve initiator, various monomers, cross-linker and other additives like surfactant (Pich and Richtering, 2011). Syntheses are separated in the stages of reactant preparation, the polymerization reaction, and purification of the microgel products that includes the removal of unreacted reactants in the reactor. Low amounts of unreacted monomer and cross-linker are left in the reactor after polymerization, as their

conversion is typically high (Jung et al., 2019a). The conversion of the initiator, however, is low due to high initial initiator concentrations in the reactor that are commonly used in experiments (Wu et al., 1994; Virtanen and Richtering, 2014; Ksiazkiewicz et al. 2020). The unreacted initiator has to be removed from the reactor after the reaction. The removal causes lengthy purification processes. With the perspective of producing microgels on a larger industrial scale, the surplus in initiator also causes significant costs for the large amounts of initiator used. Hence, it is desirable to determine the minimum amount of initiator required for the microgel synthesis. We propose here to use model-based dynamic optimization to determine a synthesis procedure that ensures high monomer and initiator conversion at the end of the synthesis. To analyze the effects, we define a reference experiment, which is a batch microgel syntheses based on 2,2'-Azobis(2-methylpropionamide) dihydrochloride (AMPA) as initiator ($c_i(t_0) = 1.23 \text{ mol m}^{-3}$), *N*-vinylcaprolactam (VCL) as monomer ($c_{VCL}(t_0) = 106.1 \text{ mol m}^{-3}$), and *N,N'*-Methylenebisacrylamide (BIS) as cross-linker ($c_{BIS}(t_0) = 2.59 \text{ mol m}^{-3}$), at 70 °C (Janssen et al., 2018).

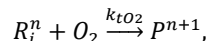
Models of microgel syntheses have been previously investigated by multiple authors, e.g., Hoare and McLean (2006), Hoare and Pelton (2007), Janssen et al. (2018), Jung et al. (2019a,b). However, these models do not take the inhibition caused by residual oxygen (Bhanu and Kishore, 1991) in the reactor into account, which may become relevant at low initial initiator concentrations. The effect of oxygen on radical polymerization reactions has been studied in the past. Two theories towards the mechanism behind oxygen inhibition have mostly been proposed: (A) oxygen terminates active radicals directly, or (B) oxygen polymerizes as co-polymer into the polymer chains and the rates of propagation and termination reactions cause the inhibition of the polymerization reaction (Schulz and Henrici, 1956; Henrici-Olivé and Olivé, 1957; Mayo, 1958; Garton and George, 1973; Bhanu and Kishore, 1991).

For microgels, only few experimental results on the effects of oxygen on the microgel synthesis have been reported. The online calorimetric and Raman spectroscopy measurements conducted by Janssen et al. (2018) reveal an inhibition phase or induction period at the beginning of the synthesis between the addition of the initiator and the onset of the polymerization reaction. They postulate that residual oxygen causes the induction period at the beginning of the synthesis. For the estimation of model parameters based on these measurements, they remove the induction period, as the mechanistic model does not include the oxygen inhibition.

Here, we include the oxygen inhibition into our previously published microgel synthesis models (Jung et al., 2019a,b). Following, we present the results of a dynamic optimization with the aim of obtaining low final concentration of unreacted initiator.

2. Model of the microgel synthesis including oxygen inhibition

We extend the previously published reaction kinetics (Jung et al., 2019a) and microgel size modeling (Jung et al., 2019b) by the oxygen inhibition. We assume that oxygen terminates active radical chains and thereby inhibits the polymerization reaction. The only reaction that is added is the termination of active radical chains with oxygen, which terminate active polymer chains and make them inactive towards propagation reactions



where k_{t,O_2} is the rate constant of the termination reaction with oxygen and r_{t,O_2} is the relative rate constant of the termination reaction with oxygen $r_{t,O_2} = \frac{k_{t,O_2}}{k_{p11}}$. Hence, we only require one kinetic rate constant to be estimated with the assumption that the rate of termination with oxygen is independent of the end group of the active radical chain.

For high values of r_{t,O_2} , an induction period develops at the beginning of the synthesis. For values above $r_{t,O_2} = 10^3$ the induction period does not change considerably, as the inhibition reaction is then limited by the number of active radical chains available; the inhibition period is much faster than the polymerization reactions. For values of r_{t,O_2} between 1 and 10^3 the inhibition reaction has approximately the same order of magnitude as the polymerization reactions. Hence, both reactions occur simultaneously causing the shape of the flow of reaction enthalpy to change. For values below $r_{t,O_2} = 10^{-1}$ no effect of the inhibiting oxygen shows in the simulation, the polymerization reactions are faster than the inhibition reaction. Hence, we set the value of the inhibition for the following simulations that the parameter is equal to $r_{t,O_2} = 10^4$.

3. Dynamic optimization

Our main aim in performing dynamic optimizations is to determine synthesis conditions that allow for low final concentration of unreacted initiator at the end of the reaction. However, we also investigate the effect of the changes in the synthesis conditions on the microgel hydrodynamic radius and the synthesis duration.

The optimization problem to be solved is

$$\min_{t_{end}, T_R(t), c_I(t_0)} c_I(t_{end}) \quad (2a)$$

$$\text{s. t. model equations, initial conditions} \quad (2b)$$

$$10^{-5} \text{ mol m}^{-3} \leq c_I(t_0) \leq 2 \text{ mol m}^{-3} \quad (2c)$$

$$60 \text{ }^\circ\text{C} \leq T_R(t) \leq 80 \text{ }^\circ\text{C}, t \in \Gamma \quad (2d)$$

$$98 \% \leq \xi_M(t_{end}) \leq 99 \% \quad (2e)$$

$$\frac{R_{\text{gel}}(t_{end}) - R_{\text{gel,ref}}}{R_{\text{gel,ref}}} \leq 0.05 \quad (2f)$$

$$500 \text{ s} \leq t_{end} \leq 10^4 \text{ s} \quad (2g)$$

Equation (2a) is the minimization of the concentration of initiator at the end of the synthesis. Equations (2b) are the modeling equations and initial conditions of the state variables that are not degrees of freedom to the optimization problem. Equations (2c) – (2d) bound the controls (degrees of freedom). The initial initiator concentration commonly used in laboratory experiments is on the order of 1 mol m^{-3} ; hence, we allow this order of magnitude and values that are up to five orders of magnitude lower, as we expect the initial initiator concentration to decrease to allow for full conversion of the initiator. Equations (2e) – (2f) are the constraints to the reactant conversions. The conversion of the monomer is set to be between 98 % and 99 % to ensure that a high conversion of monomer is achieved. Constraint (2f) forces the hydrodynamic microgel

radius to be within a range of $\pm 5\%$ around the reference value. Constraint (2g) ensures that the duration of the synthesis is between 500 s and 10^4 s.

The dynamic optimization problems are implemented and solved with the open-source software DyOS (Caspari et al., 2019). The model is integrated using the solver NIXE with absolute and relative tolerances set to 10^{-10} . The dynamic optimization problem is solved using the solver SNOPT with optimization tolerances set to 10^{-6} and feasibility tolerances set to 10^{-4} .

4. Results and discussion

To check the validity of the model extension, we compare simulation results with experimental data of microgel synthesis at two different temperatures and two different initiator concentrations (Janssen et al., 2019). The induction period for experiments run at low temperature are considerably longer than induction periods for experiments run at higher temperatures. The induction period increases for lower initiator concentrations. The agreement between model simulation and experimental data is good; especially the transition from the induction period to the polymerization is in good agreement between model and data.

We minimize the final initiator concentration; the obtained value is 0.216 mol m^{-3} . The determined initial concentration of initiator is equal to 0.283 mol m^{-3} , which translates to a final conversion of 23.9 %. Hence, the final concentration is higher, and the final conversion is lower than the values achieved in the previously presented optimization cases. The reason for this is the constraint on the microgel hydrodynamic radius. The determined synthesis procedure yields microgels with a hydrodynamic radius that is within 5 % of the hydrodynamic radius of the microgel synthesized in the reference case. The final obtained radius is at 179 nm, which is just 5 % larger than the microgels obtained in the reference case. Hence, to obtain microgels with an hydrodynamic radius that is in the same range as in the reference case, The final initiator concentration is considerably lower than the reference case. However, the synthesis duration increased considerably to 1140 s in comparison to the reference case due to the lower concentration of initiator.

Figure 1 presents selected state variables over the synthesis time. The oxygen inhibition causes the monomer conversion to increase only after an induction period of approximately 550 s. The induction period can also be observed in the simulation of the microgel radius, as the microgel radius only starts to increase considerably after the induction period has passed. The reactor temperature is kept at the lower bond of 60 °C until all oxygen has reacted with active radical chains. Then the reactor temperature is raised to the upper bond of 80 °C for the rest of the synthesis. Comparing these results to the reference case, we find that we determined a synthesis procedure that yields microgels with the same hydrodynamic radius, while we reduced the concentration of remaining initiator at the end of the synthesis from 0.93 mol m^{-3} in the reference case to 0.216 mol m^{-3} in the determined optimal synthesis, which corresponds to a 76.8 % reduction. The synthesis time, however, increases from 770 s to 1140 s, which corresponds to a 48.1 % increase in comparison to the reference case. We expect, however, that the reductions in cost due to the lower amounts of initiator used and the expected simplified purification of microgels at the end of the synthesis makes up for the

increase in synthesis time in comparison to the reference case. An additional decrease of the synthesis duration may be achievable by reducing the concentration of residual oxygen at the beginning of the synthesis. Currently, almost half of the synthesis time is used to deplete the residual oxygen in the reactor. An additional optimization showed that tightening of the constraints on the hydrodynamic microgel radius causes the initial initiator concentration to be higher and the reactor temperature to increase earlier.

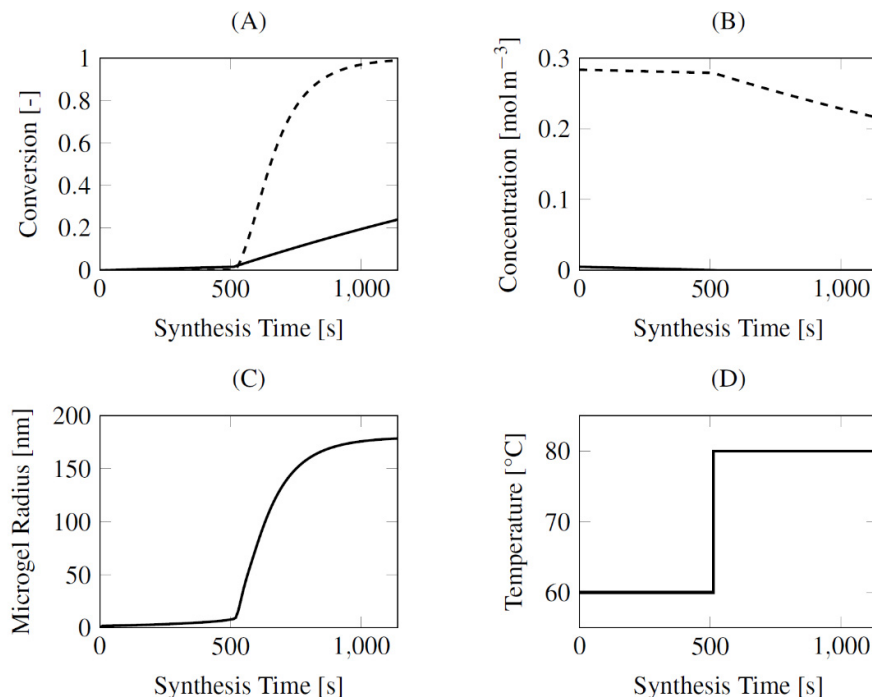


Figure 1: Dynamic optimization results considering oxygen inhibition. (A) Conversion of initiator (solid) and monomer (dashed); (B) Oxygen concentration (solid) and initiator concentration (dashed) in the reactor; (C) Hydrodynamic microgel radius; (D) Reactor temperature.

5. Conclusions

We propose to use the dynamic optimization for microgel synthesis with the aim of achieving a low concentration of unreacted initiator at the end of the reaction. In the optimization, we also consider the induction periods at the beginning of the syntheses caused by oxygen inhibition. Therefore, we extended the previously presented model of the microgel synthesis with reaction kinetics of the oxygen inhibition. We achieved good agreement with the induction periods observed in experiments run at different reactor temperatures. We minimize the final concentration of initiator in the reactor, while constraining the hydrodynamic microgel radius to be within range of the reference

synthesis. The result is a synthesis procedure that allows a 76.8 % reduction in the final concentration of initiator, while maintaining the microgel size. However, the reaction time is increased to 1140 s.

Even though the predictions of the used mechanistic model have been validated in by us for various synthesis conditions in other publications, we advise to again validate the results of the optimizations presented here. Of special interest in the experimental validation is whether the final initiator concentration is in agreement with model simulation results and whether the predicted hydrodynamic microgel radius is met.

References

- V. A. Bhanu, K. Kishore, 1991. Role of oxygen in polymerization reactions. *Chemical Reviews* 91 (2), 99–117.
- A. Caspari, A. M. Bremen, J. Faust, F. Jung, C. D. Kappatou, S. Sass, Y. Vaupel, R. Hannemann-Tamás, A. Mhamdi, A. Mitsos, 2019. Dyos - a framework for optimization of large-scale differential algebraic equation systems. In: 29th European Symposium on Computer Aided Process Engineering. Vol. 46 of *Computer Aided Chemical Engineering*. Elsevier, pp. 619–624.
- A. Garton, M. H. George, 1973. Effect of oxygen on the polymerization of vinyl chloride. i. kinetic features. *Journal of Polymer Science: Polymer Chemistry Edition* 11 (9), 2153–2167.
- G. Henrici-Olivé, S. Olivé, 1957. Reaktionskinetik der Polymerisationshemmung durch molekularen sauerstoff. *Die Makromolekulare Chemie* 24 (1), 64–75.
- T. Hoare, D. McLean, 2006. Multi-component kinetic modeling for controlling local compositions in thermosensitive polymers. *Macromolecular Theory and Simulations* 15 (8), 619–632.
- T. Hoare, R. Pelton, 2007. Functionalized microgel swelling: Comparing theory and experiment. *The Journal of Physical Chemistry. B* 111 (41), 11895–11906.
- F. A. L. Janssen, M. Kather, A. Ksiazkiewicz, A. Pich, A. Mitsos, 2019. Synthesis of poly (n - vinylcaprolactam)-based microgels by precipitation polymerization: Pseudo-bulk model for particle growth and size distribution. *ACS Omega*.
- F. A. L. Janssen, A. Ksiazkiewicz, M. Kather, L. C. Kröger, A. Mhamdi, K. Leonhard, A. Pich, A. Mitsos, 2018. Kinetic modeling of precipitation terpolymerization for functional microgels. In: 28th European Symposium on Computer Aided Process Engineering. Vol. 43 of *Computer Aided Chemical Engineering*. Elsevier, pp. 109–114.
- F. Jung, F. A. L. Janssen, A. Caspari, H. Spütz, L. Kröger, K. Leonhard, A. Mhamdi, A. Mitsos, 2019a. Dynamic optimization of a fed-batch microgel synthesis. *IFAC-PapersOnLine* 52 (1), 394–399.
- F. Jung, F. A. L. Janssen, A. Ksiazkiewicz, A. Caspari, A. Mhamdi, A. Pich, A. Mitsos, 2019b. Identifiability analysis and parameter estimation of microgel synthesis: A set-membership approach. *Industrial & Engineering Chemistry Research*.
- A. N. Ksiazkiewicz, L. Bering, F. Jung, N. A. Wolter, J. Viell, A. Mitsos, A. Pich, 2020. Closing the 1–5 μm size gap: Temperature-programmed, fed-batch synthesis of micro-m-sized microgels. *Chemical Engineering Journal* 379, 122293.
- F. R. Mayo, 1958. The oxidation of unsaturated compounds. v. the effect of oxygen pressure on the oxidation of styrene 1,2. *Journal of the American Chemical Society* 80 (10), 2465–2480.
- A. Pich, W. Richtering, 2011. Microgels by precipitation polymerization: Synthesis, characterization, and functionalization. In: A. Pich, W. Richtering (Eds.), *Chemical Design of Responsive Microgels*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–37.
- V. G. V. Schulz, G. Henrici, 1956. Reaktionskinetik der polymerisationshemmung durch molekularen sauerstoff. (versuche mit methylmethacrylat). *Die Makromolekulare Chemie* 18 (1), 437–454.
- O. L. J. Virtanen, W. Richtering, 2014. Kinetics and particle size control in non-stirred precipitation polymerization of n-isopropylacrylamide. *Colloid and Polymer Science* 292 (8), 1743–1756.
- X. Wu, R. H. Pelton, A. E. Hamielec, D. R. Woods, W. McPhee, 1994. The kinetics of poly(n-isopropylacrylamide) microgel latex formation. *Colloid & Polymer Science* 272 (4), 467–477.

Novel Dual Reactive Distillation Process for Eco-efficient Lactic Acid Recovery and Purification from Fermentation Broth

Isabel Pazmiño Mayorga,^a A.A. Kiss,^a M. Jobson^a

^a *Centre for Process Integration, Department of Chemical Engineering and Analytical Science, The University of Manchester, M13 9PL, Manchester, United Kingdom*

Abstract

Process intensification (PI) is a major area of interest for downstream processing in biorefineries, where dilute aqueous streams must be processed to separate and purify valuable components. For example, lactic acid (LA) purification faces these challenges as the feed consists of a dilute aqueous mixture of LA and other organic acids as impurities. These characteristics result in high energy requirements, large equipment, and multiple operations when considering conventional technologies. The combination of otherwise individual processing tasks is enabled by exploiting synergies following PI principles, which improves energy efficiency and reduces the processing equipment's size. This study presents a new conceptual design of an intensified dual reactive dividing wall column (R-DWC), applied to lactic acid purification. This CAPE study is the first to investigate the effect of impurities that disrupt the reaction and the separation. The approach taken in this study is a mixed methodology based on a decomposition approach, shortcut calculations, and rigorous process simulation in Aspen Plus. The intensification and the heat integration of this new realistic process led to a more eco-efficient operation in terms of material intensity (over 50 % reduction), water consumption (35 % reduction) and energy intensity (25 % reduction) compared to a 4-column flowsheet that does not account for reactive impurities and additional azeotropes. The recovery (>95 %) and purity (88 % w/w LA) targets were successfully achieved, and the acid impurity was effectively removed.

Keywords: process intensification, reactive distillation, process simulation, lactic acid

1. Introduction

Process intensification (PI) is a discipline within chemical engineering that addresses improvements in functional / phenomena / unit levels, occurring at various stages to improve process performance. For example, major enhancements have been shown by applying PI technologies to the downstream processing in biorefineries using reactive distillation (RD) (Kiss et al., 2016). Further intensification of RD attempts to overcome its inherent limitations, such as a restricted range of operating conditions and complex scaling-up. To overcome RD's limitations, exploiting additional driving forces can expand its applicability to systems that cannot be handled by RD only (Kiss et al., 2019). Therefore, the design of further intensified technologies to promote better utilization of resources (e.g. energy, water, raw materials) improves the process' eco-efficiency and reduces the equipment size, while maintaining the targets of purity and recovery, which encourages the investigation of intensified processes (Keil, 2018).

The synergy between reaction and separation in RD can be exploited by adding extra intensification features. For example, a reactive dividing wall column (R-DWC) reaches the second level of intensification by combining two already intensified technologies: DWC and RD (Weinfeld et al., 2018). R-DWC offers benefits such as increased energy efficiency, reduced number of vessels, handling of multiple outlet streams and feasible operation via decentralized control (Egger and Fieg, 2019; Mueller and Kenig, 2007). These attributes have been widely studied in academia, and the research outputs suggest that R-DWC can improve chemical processing (Kiss et al., 2019).

This paper develops the conceptual design of a novel dual R-DWC based on a decomposition approach, applied to a case study to recover and purify LA from a fermentation broth. The dual R-DWC features the esterification of the acids with methanol to facilitate the separation of esters, and the hydrolysis to recover the purified LA. For this study, a diluted mixture containing water, LA, and succinic acid was used to demonstrate the novel configuration in a flowsheet context. The kinetic data for the esterification of LA with methanol was taken from the study of Sanz et al. (2004). Also, the esterification of succinic acid with methanol and LA oligomerization were considered (Asthana et al., 2006). Amberlyst 36 ($T_{\max} = 150\text{ }^{\circ}\text{C}$) catalyst was selected for this study as previous studies indicated that Amberlyst 15 ($T_{\max} = 120\text{ }^{\circ}\text{C}$) limited the operating windows for RD (Su et al., 2013).

2. Problem statement

A primary concern of the downstream processing in biorefineries is handling large volumes of liquid containing mostly water and the target component. In addition, the removal of impurities in low concentrations requires energy-intensive processes consisting of various devices, usually of large size (Kiss et al., 2016). From the process design point of view, the development of efficient flowsheets has overseen the effects of impurities in the process performance, as process simulation research to date has not yet considered the effect of reactive impurities in the concentration of LA (Cho et al., 2008; Kim et al., 2017; Su et al., 2013). Experimental laboratory and pilot-scale studies have proven the simultaneous occurrence of the reactions of the target component and the impurities in the purification of organic acids. The evidence from these studies suggests that the impurity reaction can affect the heating and cooling needs due to the extra heat of reaction, introduce new azeotropes, require makeup of reactants, and promote operational issues such as blockage of outlet lines. To address this problem, this CAPE study is the first to investigate the effect of reactive impurities that disrupt the reaction in a novel dual R-DWC applied to purify bio-produced LA.

3. Conceptual design approach

The steps followed for the conceptual design of a novel dual R-DWC is depicted in Figure 1, and the procedure consists of a decomposition approach, shortcut models, and implementation of rigorous models.

3.1. Decomposition approach

A decomposition approach was applied to design an intensified unit that features three outlet streams (impurity, product, and excess water / methanol), and two different reactions (esterification and hydrolysis).

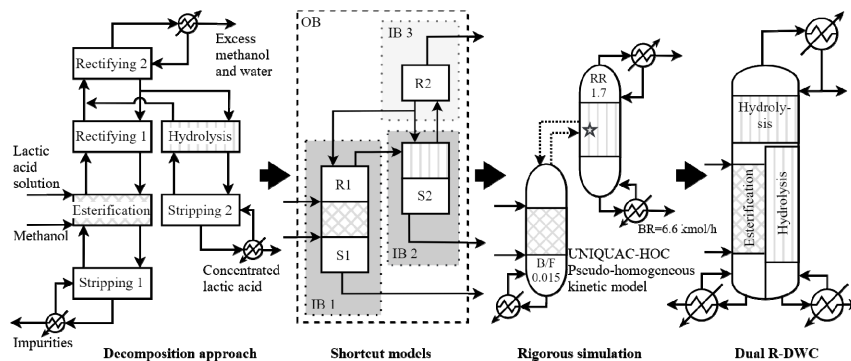


Figure 1: Methodological approach for the conceptual design of a dual R-DWC (OB: outer boundary, IN: inner boundary, B/F: bottom to feed ratio, RR: reflux ratio, BR: bottom rate)

First, the tasks and internal connections were identified: the esterification converts the organic acids into esters that are easier to separate, and the hydrolysis of the purified methyl lactate allows obtaining LA on specification. The rectifying and stripping sections remove excess compounds and reaction products to adjust the final product's composition and drive higher conversion in the reactive sections. For this case study, the heaviest cut was removed first, so the indirect sequence led to a side-stripper configuration. As the configuration proposed in this work includes two reactions, the returning stream's composition exiting the hydrolysis section depended on the vapor-liquid equilibrium (VLE) and the reaction rates. This consideration applied to the decomposition approach of dual-reactive systems differs from the one used for DWC and R-DWC, where only VLE data provides the composition of the returning stream.

3.2. Shortcut models

The tasks and interconnection identified were analyzed by defining system boundaries around the whole arrangement and grouping tasks to obtain initialization values. The grouping allowed evaluating a sub-system that contained one reaction only or a critical task that entailed a degree of freedom. For this arrangement, two inner boundaries included the esterification and hydrolysis reactions, and the top rectifying section contained the liquid split ratio. The mass balance calculations for each boundary allowed computing the interconnecting streams' values and defining operating variables such as liquid to feed ratio or reflux ratio to initialize a rigorous simulation.

3.3. Implementation of rigorous models

The simulation of the dual R-DWC was implemented in Aspen Plus v8.8 using two RadFrac modules in a thermally coupled indirect sequence leading to a partitioned side-stripper configuration. The values obtained from the shortcut calculations were used to set up the RadFrac modules separately. The observation of composition, temperature, and generation rate profiles combined with an update of the initial mass balances, accelerated a trial-and-error method to determine the initial number of reactive and non-reactive stages, location of the inlet and side-draw streams and catalyst holdup. Then, the interconnecting streams between the two RadFrac modules were implemented after updating the flowrates and compositions between iterations until reaching a match. A full flowsheet, including a preconcentrator unit and a methanol recovery column, was implemented in Aspen Plus, and the results provided key performance indicators.

4. Results and discussion

4.1. Conceptual design of a dual reactive dividing wall column

A sequential approach was used to implement the rigorous models in Aspen Plus, first building the esterification column, then implementing the hydrolysis column, and finally integrating the two with a thermally coupled arrangement.

The esterification column was initialized with the values gathered from the shortcut calculations. The base case was defined after varying variables such as the catalyst holdup, number of reactive and non-reactive stages, and feed location, so that a robust converged simulation accomplishes the performance indicators set for the esterification column: 99 % conversion of LA, low LA oligomers production, and near-to-sharp split between methyl lactate and dimethyl succinate.

The hydrolysis column was set up using the feed stream values from the simulated esterification column and the shortcut calculations' operating parameters. The number of reactive and non-reactive stages, the location of the inlet and side streams, and catalyst holdup were varied by tracking the performance indicators established for the hydrolysis section: methyl lactate conversion higher than 80 %, low methyl lactate loss in the top and bottom streams and the product specification (88 % w/w LA).

Figure 2 presents the concentration profiles in the liquid phase and the generation rates in the hydrolysis column. Figure 2a shows the initial configuration where the inlet stream entered the first top reactive stage. The liquid phase composition indicated that the upper rectifying stages were still rich in the reactants: water and methyl lactate. Then, by systematically changing the feed stream's location and observing the methyl lactate content in the top and bottom streams, Figure 2b clearly shows how the methyl lactate concentration in the liquid phase reduces towards the top and bottom ends. Figure 2c shows the generation rate profiles for lactic acid and methyl lactate with two different inlet stream locations. The most interesting aspect of this graph is how the feed stream's location can be facilitated by inspecting the column profiles.

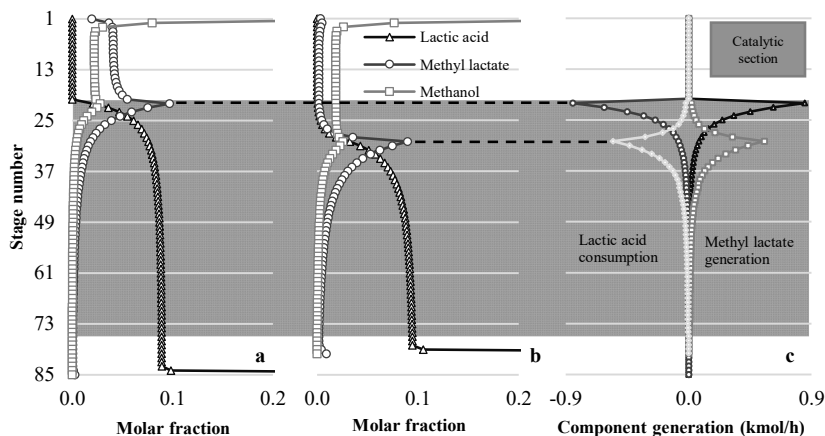


Figure 2: Liquid composition and component generation profiles in the hydrolysis column a) feed on the 1st reactive stage b) feed on the 10th reactive stage c) generation rate profiles

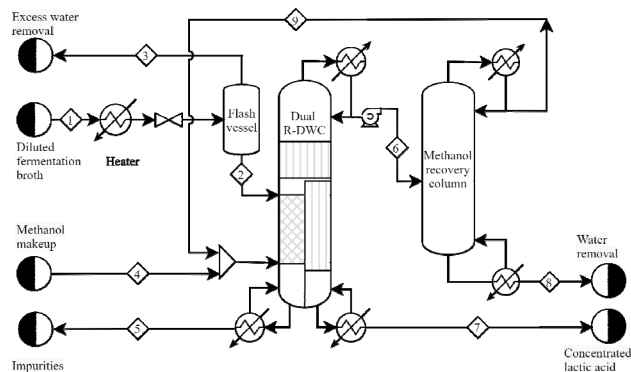


Figure 3: Process flowsheet for the pre-concentration and purification of lactic acid

An iterative procedure to match the vapor stream's composition exiting the esterification column and the returning liquid side stream into the esterification column was necessary to couple the two RadFrac modules. To facilitate a good convergence of the simulation, we generated the estimates of temperature, flowrate, and composition in the RadFrac modules, i.e., values of a previously converged simulation were used as starting values for a new run after a significant modification.

A pre-concentration operation in a flash vessel and a methanol recovery column were included to compare the novel configuration's performance with existing benchmark processes, as depicted in the flowsheet of Figure 3.

4.2. Key performance indicators

The performance of the flowsheet, including the novel dual R-DWC, was compared against two benchmark configurations that consist of a sequence of two reactive and two conventional distillation columns, and of one conventional, one reactive and one fully thermally coupled column between the hydrolysis and recovery sections (Kim et al., 2017; Su et al., 2013). The comparison is not straightforward because of the introduction of a more realistic approach by considering the esterification of succinic acid (in addition to LA) and the exploitation of additional heat recovery opportunities.

Table 1 presents the key performance indicators to evaluate the novel configuration's efficiency. Regarding the use of raw materials, considering the esterification of succinic acid resulted in an increase of the methanol makeup compared to the benchmark studies. However, the observed decrease in the methanol/LA feed molar ratio could be attributed to the solid catalyst that withstands higher temperatures, so less methanol is required to produce a quenching effect in the reactive section resulting in less methanol recirculating in the system. Also, the two benchmark studies introduced an additional water stream for the hydrolysis reaction. In contrast, the process developed in this study used the water produced in the first reaction by appropriately locating the vapor stream exiting the esterification column into the hydrolysis section, which resulted in a decrease in water consumption and material intensity. This process also achieved a higher LA recovery as milder temperatures in the hydrolysis section reduced oligomerization reactions. The direct heat integration in the dual R-DWC and the process-to-process heat recovery implemented in this study reduced the energy intensity in 25 % and 10 % compared to the benchmark technologies, respectively.

Table 1: Key performance indicators of the novel process and two benchmark processes

Parameter	Su et al. (2013)	Kim et al. (2017)	This work
Methanol makeup (kmol/h)	0.2	0.2	0.8
Water flowrate for hydrolysis (kmol/h)	14	14	0
Methanol/lactic acid feed molar ratio	4	4	3.3
Lactic acid recovery (%)	91.8	94.2	95.7
Material intensity ($\text{kg}_{\text{waste}}/\text{kg LA}$)	0.178	0.194	0.089
Water consumption ($\text{m}^3/\text{kg LA}$)	0.051	0.043	0.033
Energy intensity ($\text{MJ}/\text{kg LA}$)	14.9	12.6	11.2

5. Conclusions

This research provided insights for the conceptual design of a dual R-DWC using a combination of a decomposition approach, shortcut models, and rigorous simulations. The intensified dual R-DWC unit was successfully evaluated in a flowsheet for the pre-concentration and purification of lactic acid, including a more realistic approach by considering side reactions. The most interesting findings of this study were the effective removal of the impurity, the elimination of extra water for the hydrolysis reaction, and a more eco-efficient performance in terms of material, water and energy use.

Acknowledgments

IPM gratefully acknowledges the full fund support from SENESCYT-Ecuador. AAK gratefully acknowledges the Royal Society Wolfson Research Merit Award.

References

- Asthana, N.S., Kolah, A.K., Vu, D.T., Lira, C.T., Miller, D.J., 2006. A Kinetic Model for the Esterification of Lactic Acid and Its Oligomers. *Ind. Eng. Chem. Res.* 45, 5251–5257.
- Cho, Y., Kim, B., Kim, Dongpil, Han, Myungwan, 2008. Recovery of lactic acid by reactive dividing wall column. *Int. Conf. on Control, Automation and Systems.* pp. 2596–2599.
- Egger, L.S., Fieg, G., 2019. Process control for energy efficient operation of reactive dividing wall columns. *Chem. Eng. Res. Des.* 144, 397–404.
- Keil, F.J., 2018. Process intensification. *Rev. Chem. Eng.* 34, 135–200.
- Kim, S.Y., Kim, D.M., Lee, B., 2017. Process simulation for the recovery of lactic acid using thermally coupled distillation columns to mitigate the remixing effect. *Korean J. Chem. Eng.* 34, 1310–1318.
- Kiss, A.A., Jobson, M., Gao, X., 2019. Reactive Distillation: Stepping Up to the Next Level of Process Intensification. *Ind. Eng. Chem. Res.* 58, 5909–5918.
- Kiss, A.A., Lange, J.-P., Schuur, B., Brillman, D.W.F., van der Ham, A.G.J., Kersten, S.R.A., 2016. Separation technology—Making a difference in biorefineries. *Biomass Bioenergy* 95, 296–309.
- Mueller, I., Kenig, E.Y., 2007. Reactive Distillation in a Dividing Wall Column: Rate-Based Modeling and Simulation. *Ind. Eng. Chem. Res.* 46, 3709–3719.
- Sanz, M.T., Murga, R., Beltrán, S., Cabezas, J.L., Coca, J., 2004. Kinetic Study for the Reactive System of Lactic Acid Esterification with Methanol: Methyl Lactate Hydrolysis Reaction. *Ind. Eng. Chem. Res.* 43, 2049–2053.
- Su, C.-Y., Yu, C.-C., Chien, I.-L., Ward, J.D., 2013. Plant-Wide Economic Comparison of Lactic Acid Recovery Processes by Reactive Distillation with Different Alcohols. *Ind. Eng. Chem. Res.* 52, 11070–11083.
- Weinfeld, J.A., Owens, S.A., Eldridge, R.B., 2018. Reactive dividing wall columns: A comprehensive review. *Chem. Eng. Process. - Process Intensif.* 123, 20–33.

Exergy Analysis of Vacuum Distillation Unit

Sena Kurban^{a,b,*}, Gizem Kuşoğlu Kaya^a, Serdar Yaman^b

^a*Turkish Petroleum Refinery, 41790 Körfez, Kocaeli, Turkey*

^b*Chemical and Metallurgical Engineering Faculty, Chemical Engineering Department, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey*

Abstract

Energy efficiency studies have an important role in the petroleum industry in terms of environmental impacts and cost. In this context, distillation units are important process units in the refinery. Osuolale and Zhang (2018) pointed out they cause about 35 % of energy consumption. Energy analysis studies are useful in energy efficiency but remain incapable in some areas. Besides energy analysis in other words exergy analysis, which is related to the second law of thermodynamic, is a very powerful tool for detecting losses.

In this study, an exergy analysis study was carried out on the vacuum distillation unit located in TUPRAS Izmit Refinery. Vacuum distillation unit (VDU) processes heavy petroleum products from atmospheric distillation column. Heavy, medium, and light gas oil and vacuum residue are the products of VDU.

Vacuum distillation column was simulated in ASPEN HYSYS V10 in this study by using operational data and equipment information for exergy analysis. Energy and exergy analysis calculations were completed and case studies were carried out to examine the effect of various operational conditions on the exergy efficiency of the column in the ASPEN HYSYS. In order to see the effects of different operational conditions such as pump-around (PA) flowrate and return temperature for exergy efficiency on online platform, data was created by using of Aspen Simulation environment and an exergy prediction model was developed with machine learning models in R-Studio. Thus, support on exergy efficiency was provided with successful prediction models.

Keywords: exergy analysis, thermodynamics, vacuum distillation unit, modelling, machine learning model

1. Introduction

Energy efficiency is among the most focused topics in the industry recently. Feyzi and Beheshti (2017) stated that distillation columns, which are commonly used in chemical industry, are an energy intensive separation units. For instance, the USA alone has about 40,000 distillation columns, which consume about 6 % of the total US energy and they accounts for 40 - 60 % of the total energy use in chemical industry (Osuolale and Zhang, 2014). Therefore, designing energy efficient distillation systems and determining operational parameters that affect energy efficiency are important to minimize the energy consumption.

Rivero (2002) stated some questions such as how and where energy is degraded and how to reduce energy consumption need to be answered in order to clarify energy efficiency. To overcome this, thermodynamic analysis brings an analytic approach to the topic of energy efficiency. The First Law of Thermodynamics (FLT) is known as the law of the conservation of energy. This means that energy can be transformed, but it can be neither

created nor destroyed. The Second Law of Thermodynamics (SLT) state that the entropy of the system increases whenever any energy conversion or transfer occurs within a closed system. When comparing these laws, the first law has no knowledge of the thermodynamic process, which is related to its inability to completely convert heat to mechanical work. At this point, the concept of exergy emerges in relation to SLT and entropy (Dincer and Rosen, 2013).

Many researches focused on exergy analysis of distillation columns. Waheed and Oni (2015) improved the energy performance of crude oil distillation unit with using simulation of process techniques. As a result of these developments, both the energy and exergy efficiency of the column are increased by 4.0 and 1.6 %, respectively. Al-Muslim et al. (2003) conducted exergy analysis of single and two-stage crude oil distillation units for determining exergy losses and efficiencies. Overall exergy efficiencies of single and two-stage distillation units were found 14.0 and 31.5 %, respectively. Tarighaleslami et al. (2011) studied an atmospheric distillation unit based on the exergy analysis by changing streams and new configurations.

Simulation of process is an important tool for exergy analysis in terms of process analysis and design. The objective of this study is to present exergy analysis of vacuum distillation column in both current situation and various situation with changing operational conditions. Also, exergy analysis is supported neural network models to enhance model accuracy and reliability.

2. Exergy Analysis

The exergy of the system is described as the quantity of the ability to do work that is composed of mass, heat, work flows transfer to the system at a specified reference temperature. Exergy analysis determines the locations of the system's inefficiencies and exergy loss of system is reduced by improvements. Numerous engineers and researches prefer exergy analysis on the ground that it is an effective tool for analyzing, designing and developing processes and systems (Izyan and Shuhaimi, 2014).

The total exergy is obtained with the sum of physical and chemical exergy in Eq. (1). The physical exergy is related to difference of actual and reference conditions of enthalpy and entropy as Eq. (2).

$$Ex_{Total} = Ex_{physic} + Ex_{Chemical} \quad (1)$$

$$Ex_{physic} = (h - h_0) - T_0 \cdot (s - s_0) \quad (2)$$

Where the subscript "0" indicates the ambient conditions that are 25 °C and 101.325 kPa. Exergy efficiency in Eq. (3) gives information about the measure of approaching the ideal.

$$\psi = \frac{\sum Ex_{out}}{\sum Ex_{in}} = \frac{\text{Exergy in product outputs}}{\text{Exergy in inputs}} \quad (3)$$

3. Case Study

3.1. Process Description and Simulation

The aim of the vacuum distillation column of TUPRAS Izmit Refinery is to separate the atmospheric residue by three main streams. These are light vacuum gasoil (LVGO), a mixture of medium and heavy vacuum gas oil (MHVGO) and vacuum residue (VR).

A summary day that has exact laboratory data is chosen for the simulation. Unit operation values of summary day and equipment data sheets were used for performing the simulation. Figure 1 shows the computer-aided simulation of vacuum distillation column.

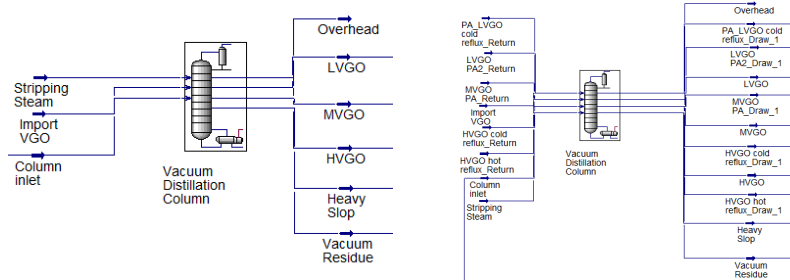


Figure 1: (a) Basic simulation of VDU; (b) Simulation of VDU with Pump-Arounds

3.2. Computer Aided Exergy Analysis

Exergy calculations were done via ASPEN HYSYS V10. The spreadsheets were created for exergy calculations in connection with the process simulation. In this study, the spreadsheets were prepared using the formula in Eq. (2) and exergy, which is directly calculated by ASPEN HYSYS. Moreover, the effect of operating parameters on exergy efficiency was examined by using “Case Studies” in ASPEN HYSYS. Lower and upper process control limit values like those given in Table 2 were entered in this case study to see the changing the exergy efficiency.

4. Machine Learning Model

Exergy efficiencies can be predicted by using the machine learning models without the need for complex exergy analysis calculations. Also, the effect of operating parameters on exergy efficiency for vacuum distillation column can be determined by means of these predictive models. Firstly, the parameters that affect exergy efficiency were determined to predict exergy efficiencies in “Case Study” of ASPEN HYSYS. Synthetic data was produced by using these parameters that input the model. In this study, Bootstrap aggregating (Bagging) and Random Forest models were used to predict exergy efficiencies in R-Studio. “Ipred” and “randomForest” packages of R-Studio were used to develop model for Bagging and Random Forest respectively. The data for both models were divided into training data (50 %), testing data (30 %) and unseen validation data (20 %). The machine learning model for exergy efficiency is of the following form:

$$\varphi = f(x_1, x_2, x_3)$$

Where φ is exergy efficiency, x_1 , x_2 , and x_3 are model inputs which are furnace exit temperature, MVGO PA temperature and HVGO cold PA temperature.

5. Results and Discussion

Exergy analysis results of VDU are summarized in Table 1. Results of Eq. (2) and exergy which is directly given by ASPEN was the same. Exergy efficiency was found to be 0.67 in current situation by the ratio of total output exergy to input exergy by using Eq. (3).

Table 1: Inlet and outlet streams exergy of vacuum distillation unit

	Stream	Mass flow rate (kg/s)	Specific exergy (kJ/kg)	Exergy rate (kJ/s)	Σ Exergy (kJ/s)
Inlet streams	Column inlet	108.944	356.442	38,832.357	
	Import VGO	25.655	34.514	885.456	39,717.814
	Stripping	0.000	1,002.207	0.000	
Outlet streams	Overhead	1.230	-8.903	-10.950	
	LVGO	10.832	23.709	256.821	
	MVGO	36.112	126.999	4,586.153	26,850.608
	HVGO	26.397	202.965	5,357.767	
	Heavy slop	13.222	263.466	3,483.599	
	Vacuum residue	46.806	281.531	13,177.218	

Table 2 shows the effect of operating parameters on exergy change between lower and upper bound. The first three parameters that affect the exergy most are shown in Table 2 are furnace exit temperature, MVGO PA temperature, HVGO cold PA temperature, respectively. These parameters were used to model input. Essentially, the change in operating PA parameters did not significantly affect the exergy efficiency. Also, Dincer and Rosen (2007) have reported that the change of temperature profile, which is controlled by the PA did not cause a significant change in exergy efficiency.

Table 2: The effect of operating parameters on exergy and energy efficiency change

Operating parameters	Lower-upper bound	Exergy efficiency change	% Absolute exergy change	Energy efficiency change	% Absolute exergy change
HVGO cold PA flow (t/h)	320-360	0.6555-0.6599	0.671	1.133-1.131	0.177
HVGO hot PA flow (t/h)	90-115	0.6560-0.6556	0.061	1.133-1.133	0.000
LVGO hot PA flow (t/h)	6.9-14.9	0.6560-0.6560	0.000	1.133-1.133	0.000
LVGO PA2 flow (t/h)	110-128	0.6552-0.6567	0.229	1.133-1.132	0.088
LVGO cold PA flow (t/h)	29.5-65.5	0.6499-0.6576	1.185	1.138-1.131	0.615
MVGO PA flow (t/h)	348-432	0.6565-0.6536	0.442	1.133-1.134	0.088
LVGO PA2 temperature (°C)	50-60	0.6570-0.6560	0.152	1.132-1.133	0.088
MVGO PA temperature (°C)	190-212	0.6431-0.6560	2.006	1.138-1.133	0.439
HVGO cold PA temperature (°C)	267-277	0.6571-0.6442	1.963	1.132-1.138	0.530
HVGO hot PA temperature (°C)	280-296	0.6529-0.6560	0.475	1.134-1.133	0.088
Furnace exit temperature (°C)	395-410	0.6655-0.6290	5.485	1.124-1.157	2.936

Figures 2 and 3 show the model performances in predicting the exergy efficiency of the vacuum distillation unit on the training, testing and unseen validation data. The sample numbers at Figures 2 and 3 consist of the case study of three parameters mentioned in the fourth section. Due to the temperature increase of the furnace output, which is one of these parameters, the exergy of column inlet stream increases depending on the increase enthalpy. Thus, the trend of the exergy efficiency in the Figures 2 and 3 decreases while the column input exergy increases considering Eq. (3).

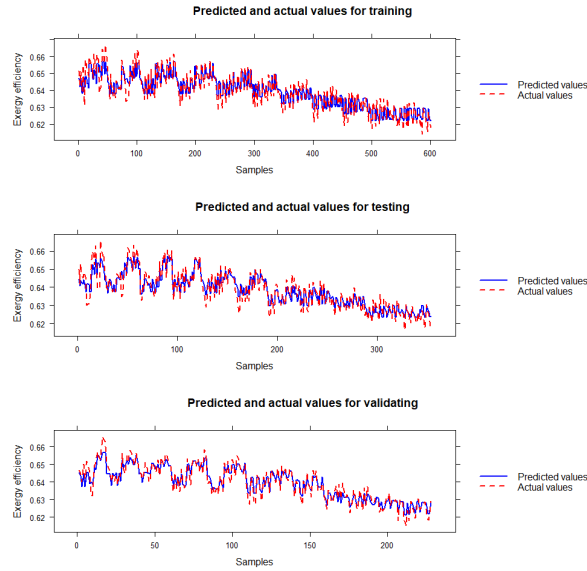


Figure 2: Actual and Bagging model predicted exergy efficiency for VDU

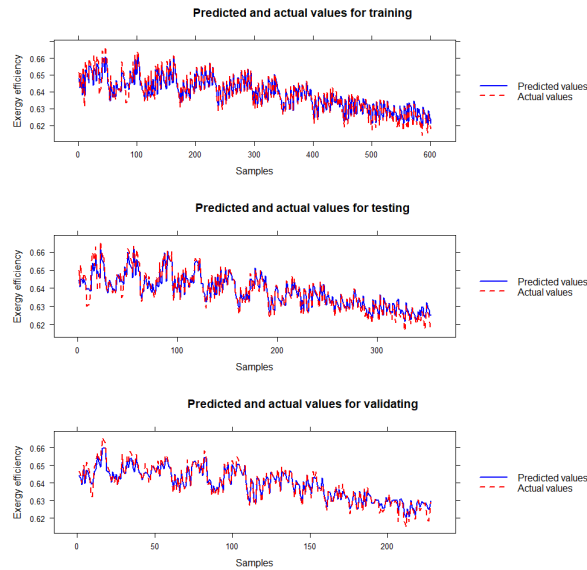


Figure 3: Actual and Random Forest model predicted exergy efficiency for VDU

The mean squared error of models can be seen in Table 3. The results in Figures 2 and 3 and Table 3 show that the machine learning models give good prediction performance, but the Random Forest model seems to give better results than the Bagging model. These models can be used to determine the exergy efficiencies for the most affecting operating parameters.

Table 3. Mean Squared Error of Models

	Training	Testing	Validation
Bagging model	0.0000126	0.0000132	0.0000080
Random forest model	0.0000060	0.0000064	0.0000047

6. Conclusions

Vacuum distillation column in TUPRAS Izmit Refinery was evaluated for exergy analysis by using parameter values under nominal conditions. Thus, an exergy analysis was made about the inefficient use of energy at which points of the process. The aim of the project is to create a machine learning model to see the effect of operational parameters on column exergy efficiency without the need for continuous running ASPEN HYSYS. The effect of various operation parameters on exergy efficiency has been examined by case studies in the ASPEN HYSYS simulation program. Changing the PA operating parameters did not lead to significant changes in exergy efficiencies. It has been seen that furnace exit temperature is the most important variables on exergy efficiency. In addition, machine learning models for exergy efficiency were improved using the process operation parameters which have the most effect on exergy efficiency. The results of two models indicated a successful performance on predicting exergy efficiency without the need for exergy-intensive calculations. From this study, we conclude that, it is necessary to prepare a substructure that requires the online working of ASPEN HYSYS continuously in order to ensure improvements through the exergy calculation. Instead of this, we obtained a model in R that can run in the online data integration tool, so we take the first step to working online platform. The reason why optimization is not considered in this study is that it is necessary to do a study for determining the optimum operating conditions by trying different values of each parameter, which is important in exergy analysis. Exergy analysis for optimization may be the subject of another study.

References

- H. Al-Muslim, I. Dincer, S. M. Zubair, 2003, Exergy analysis of single-and two-stage crude oil distillation units, *J. Energy Resour. Technol.*, 125(3), 199-207.
- I. Dincer, M.A. Rosen, 2007, Exergy analysis of crude oil distillation systems. Exergy: energy, environment and sustainable development, 1st Ed., 290-302, Oxford, USA: Elsevier
- I. Dincer, M. A. Rosen, 2013, Thermodynamic Fundamentals, Exergy: energy, environment and sustainable development, 2nd Ed., 1-20, Oxford, USA: Elsevier.
- V. Feyzi, M. Beheshti, 2017, Exergy analysis and optimization of reactive distillation column in acetic acid production process. *Chemical Engineering and Processing-Process Intensification*, 120, 161-172.
- Z. N. Izyan, M. Shuhaimi, 2014, Exergy analysis for fuel reduction strategies in crude distillation unit. *Energy*, 66, 891-897.
- F. N. Osuolale, J. Zhang, 2014, Energy efficient control and optimisation of distillation column using artificial neural network, *Chemical Engineering*, 39.
- F. N. Osuolale, J. Zhang, 2018, Exergetic optimisation of atmospheric and vacuum distillation system based on bootstrap aggregated neural network models. In *Exergy for A Better Environment and Improved Sustainability 1*, 1033-1046.
- R. Rivero, 2002, Application of the exergy concept in the petroleum refining and petrochemical industry. *Energy conversion and Management*, 43(9-12), 1199-1220.
- A.H. Tarighaleslami, M. R. Omidkhan, S. Younessi, 2011, An exergy analysis on a crude oil atmospheric distillation column, *Chemical Engineering Transactions*, 25.
- M. A. Waheed, A. O. Oni, 2015, Performance improvement of a crude oil distillation unit, *Applied Thermal Engineering*, 75, 315-324.

Computer-Aided Solvent Design Integrated with a Machine Learning-based Atom Contribution Method

Qilei Liu^a, Lei Zhang^{a,*}, Jian Du^a, Rafiqul Gani^{b,c}

^a*Institute of Chemical Process Systems Engineering, School of Chemical Engineering, Dalian University of Technology, Dalian 116024, China*

^b*PSE for SPEED, Skyttemosen 6, DK-3450 Allerød, Denmark*

^c*Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea*

Abstract

Solvents are widely applied in chemical industries. The use of efficient model-based solvent selection techniques is an option worth considering for rapid identification of solvents with desired properties. In this paper, a Machine Learning-based Atom Contribution (MLAC) method is developed for fast predictions of molecular surface charge density profiles ($p(\sigma)$) using the three-dimensional atomic descriptors. Integrating the developed MLAC method and Computer-Aided Molecular Design (CAMD) techniques, an optimization-based MLAC-CAMD framework for solvent design is established by formulating and solving a Mixed-Integer Non-Linear Programming (MINLP) model, where model complexities are managed with a decomposition-based solution strategy. Finally, a case study involving crystallization is presented to highlight the effectiveness of the MLAC-CAMD framework.

Keywords: machine learning, computer-aided molecular design, atom contribution, surface charge density profiles, decomposition-based algorithm.

1. Introduction

The selection of solvents is of importance in the optimization of product yield and purity in chemical processes. However, it is practically impossible to perform thousands of time-consuming and extensive experiments for screening solvent candidates using a trial-and-error approach. As a result, it is worth employing efficient model-based methods that can help to focus the limited experimental resources on a few promising solvent candidates. CAMD techniques are efficient model-based methods for solvent design (Zhang et al., 2016). The success of CAMD depends very much on the reliability of property prediction models used, for example, the UNIFAC models. However, a large set of experimental data are required for regression of binary interaction parameters the models need, hindering thereby wide application of UNIFAC models (e.g., extractive distillation (Chen et al., 2005)).

To overcome the above-mentioned limitation of UNIFAC models, COSMO (COnductor like Screening MOdel)-based models are developed as potential alternatives due to their powerful ability to predict solvent properties with only a small set of group-independent model parameters (Hsieh et al., 2010). However, the computationally costly Density Functional Theory (DFT) calculations needed to obtain the molecular $p(\sigma)$ in the COSMO-SAC (Segment Activity Coefficient) model limit their application in CAMD

techniques for high-throughput solvent design. Scheffczyk et al., (2017) presented a COSMO-CAMD framework, where an optimization-based genetic algorithm LEA3D and a hierarchical approach consisting of two accuracy levels of DFT calculations are combined to reduce the computational cost of the DFT calculation for novel solvent design. Dong et al., (2018) proposed a COSMO-UNIFAC method for fast activity coefficient (γ) predictions of organic liquids. Their method takes the advantages of COSMO-based models (use of pseudo-experimental data for γ) and UNIFAC models (relatively accurate prediction) to bypass the DFT calculations. Mu et al., (2007) developed a GC-COSMO method for molecular $p(\sigma)$ predictions, where the $p(\sigma)$ values are estimated by the addition of group contributions to $p(\sigma)$ in milliseconds. Although their work offers an efficient molecular $p(\sigma)$ generation for the COSMO-RS (Real Solvents) model, one limitation cannot be ignored. It is known that $p(\sigma)$ is a charge-dependent molecular property. The simple scheme of group (or descriptor) additivity is unable to describe the complex interactions among functional groups, e.g., induction effect or conjugation effect, leading to poor prediction results.

In this paper, an optimization-based MLAC-CAMD framework is established for solvent design. In Section 2, the three-dimensional atomic descriptors, weighted Atom-Centered Symmetry Functions (wACSFs), are used to develop a MLAC method for fast and accurate predictions of molecular $p(\sigma)$ for the COSMO-SAC model (Hsieh et al., 2010). Then, the MLAC method is integrated with the CAMD problem, which is formulated as an MINLP model and solved by a decomposition-based solution algorithm. In Section 3, the proposed framework is evaluated with a case study involving solvent design for the improved crystallization operation.

2. MLAC-CAMD Framework for Solvent Design

2.1. The Development of MLAC Method

A database is created before developing the MLAC method, where 1120 solvents containing H, C, N, O elements with Chemical Abstracts Service (CAS) numbers are collected from the Virginia Tech database (Mullins et al., 2006), as these elements can be found in most of the commonly used organic solvents. The Isomeric SMILES (Simplified Molecular-Input Line-Entry System) of the 1120 solvents are then collected from the PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) database using CAS numbers.

For the development of MLAC method, the SMILES representations of 1120 solvents are first converted to stereoscopic representations with cartesian coordinates using the OpenBabel tool (O'Boyle et al., 2011). Based on the stereoscopic solvents, the three-dimensional atomic descriptors, wACSFs (Gastegger et al., 2018), are calculated and selected as inputs for the establishment of High-Dimensional Neural Network (HDNN) model, which is an essential $p(\sigma)$ prediction model in the MLAC method. The wACSFs represent the local atomic environment of a centered atom i via the functions of radial (G_i^{rad}) and angular (G_i^{ang}) distributions of the surrounding atoms inside a cut-off sphere. More detailed information about the wACSFs can be found in Gastegger et al., (2018).

As the HDNN model needs atomic $p(\sigma)$ ($p_{atom}(\sigma)$) as outputs for model training, the $p_{atom}(\sigma)$ of 1120 solvents are prepared with the Gaussian 09W software (Frisch et al., 2016) and the COSMO-SAC model (Hsieh, et al., 2010). Finally, a database of $p_{atom}(\sigma)$ is established, where the number of samples for H, C, N, O atoms is 15535, 9108, 305 and 1215, respectively. With the obtained input (wACSFs) and output data ($p_{atom}(\sigma)$), the HDNN model is trained, which consists of four separate element-based (H, C, N, O) back propagation Artificial Neural Networks (ANNs). A diagrammatic sketch of HDNN model

is shown in Figure 1. For each element, atoms are randomly divided into the training, validation and test set with the ratio 6:1:1. In each ANN model, the optimizer, loss function, metrics function and activation function are selected as Adam (Kingma and Ba, 2014), Mean Squared Error (MSE), coefficient of determination R^2 and ReLU (Xavier et al., 2011), respectively. The metrics of training sample R_{train}^2 , validation sample R_{val}^2 and test sample R_{test}^2 are 0.964, 0.918, 0.907 for H element, 0.975, 0.931, 0.931 for C element, 0.950, 0.889, 0.865 for N element, 0.935, 0.867, 0.902 for O element. All these results satisfy the fitting criterion $\frac{R_{train}^2 - R_{test}^2}{R_{train}^2} < 0.1$ (note that $\frac{R_{train}^2 - R_{test}^2}{R_{train}^2} \geq 0.1$ indicates overfitting) (Zhao et al., 2020), indicating that the ANNs for H, C, N, O elements are reliable for $p_{atom}(\sigma)$ predictions.

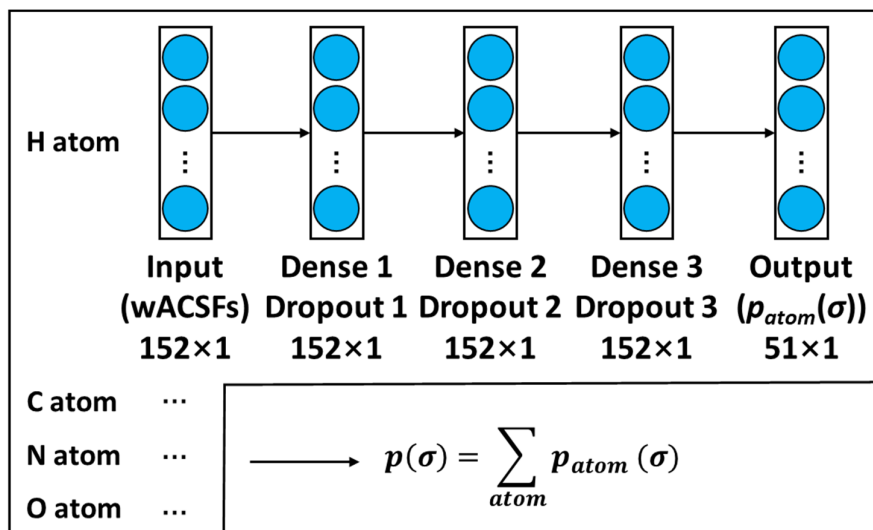


Figure 1. A diagrammatic sketch of HDNN model.

2.2. The Evaluation of MLAC Method

To further verify the feasibility and effectiveness of the HDNN model, the differences of predicted $p(\sigma)$ between the MLAC method and the DFT method (benchmark) are evaluated with the criterion R^2 . The R^2 results of the predicted $p(\sigma)$ for each discrete σ interval in the MLAC method are shown in Figure 2. Furthermore, the MLAC method is employed to determine $p(\sigma)$ for the predictions of infinite dilution activity coefficients $\gamma^\infty = f(p(\sigma), V_C)$ based on the COSMO-SAC model, where the molecular cavity volume V_C is predicted by the Group Contribution (GC) method using the MG group sets (Hukkerikar, et al., 2012) with a fitting quality of $R^2=0.9998$. The γ^∞ predictions of the MLAC method are compared with those predicted by the DFT calculated $p(\sigma)$ and V_C . Sixteen solutes (denoted as “c1~c16”) and 1120 solvents that composed of four H, C, N, O elements from the Virginia Tech database (solvents are classified in 13 categories and denoted as “s1~s13”) are selected for γ^∞ calculations. The Average Absolute Percent Error (AAPE) criterion is used to evaluate the differences of predicted γ^∞ between the DFT (benchmark) and MLAC method. Finally, the AAPEs for 16 types of solutes and 13 types of solvents using the MLAC method are shown in Figure 3.

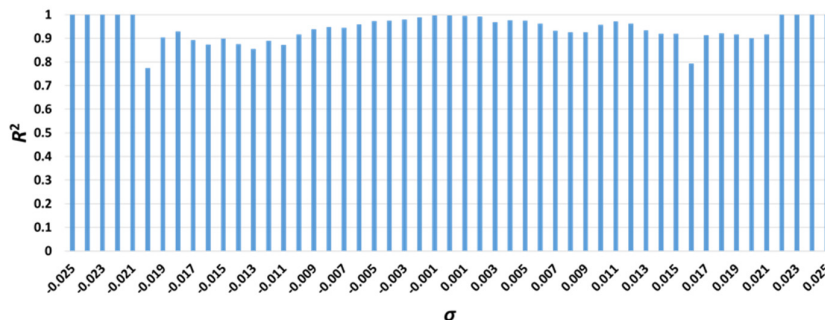


Figure 2. The R^2 results of the predicted $p(\sigma)$ for each discrete σ interval in the MLAC method.

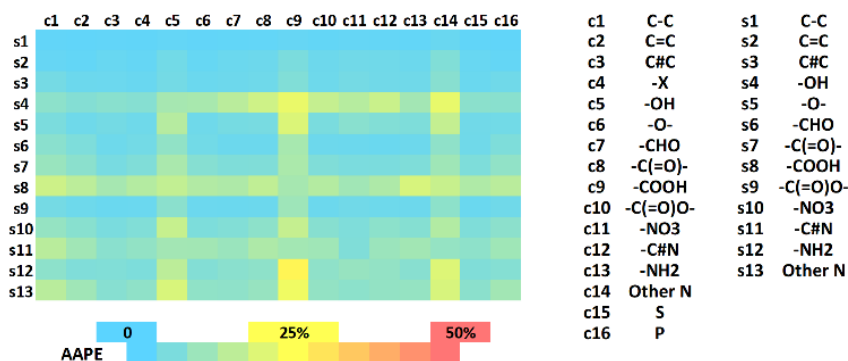


Figure 3. The heap map of AAPEs among 16 types of solutes and 13 types of solvents using the MLAC method.

It is noted that most of the AAPEs in Figure 3 are acceptable with minor prediction errors. Also, the overall AAPE for the total number of 17920 data points ($1120 \text{ solvents} \times 16 \text{ solutes}$) is calculated using the MLAC method and the result (AAPE = 6.6%) confirms that the developed MLAC method is feasible and sufficiently reliable to provide molecular $p(\sigma)$ (that is, a sum of $p_{atom}(\sigma)$ that are predicted by the HDNN model) to the COSMO-SAC model for the predictions of γ^∞ .

To highlight the computational efficiency of MLAC method, a molecule (SMILES: CCOC(C)CC(=O)OCOC(C)=O) is taken as an example for illustration. The MLAC method takes around 0.921 seconds to predict $p(\sigma)$ with SMILES on a PC laptop (Intel Core i5-8250U CPU @ 1.60 GHz), which is more efficient compared with the DFT calculation (around 51 minutes on the same laptop).

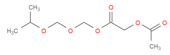
2.3. The Integration of MLAC Method and CAMD Technique

Based on the SMILES-based isomer generation algorithm (Liu et al., 2019) and GC-based property package, the ProCAPE tool (<https://www.pseforspeed.com/procape-computer-aided-property-estimation/>), the MLAC method is integrated with the CAMD technique for solvent design by formulating and solving an MINLP model through a decomposition-based algorithm (Karunanithi et al., 2005).

3. Case Study

The CAMD problem designing cooling crystallization solvents for Ibuprofen is considered. The following functional groups are selected: CH₃, CH₂, CH, C, OH, CH₃CO, CH₂CO, CHO, CH₃COO, CH₂COO, HCOO, CH₃O, CH₂O, CH-O, COOH, COO. The lower and upper bonds for structure and property constraints are given in Table 1. The objective for this case study is to design a crystallization solvent with the highest potential recovery *PR*%. More details about the objective function and constraints can be found in Karunanithi et al., (2006).

Table 1. List of structure and property constraints, as well as the top designed crystallization solvent for the Ibuprofen cooling crystallization process.

Property	Constraint	Property	Result
Number of groups	$2 \leq N_g \leq 8$	Molecule (SMILES)	CC(=O)OCC(=O)OCOCOC(C)C
Number of same groups	$N_s \leq 8$	Molecular structure	
Number of functional groups	$1 \leq N_F \leq 8$	<i>PR</i> %	95.91% (MLAC), 96.04%(DFT)
Hildebrand solubility parameter at 298 K	$17 \leq \delta \leq 19 \text{ MPa}^{1/2}$	δ (MPa ^{1/2})	18.470
Hydrogen bonding solubility parameter	$\delta_H \geq 8 \text{ MPa}^{1/2}$	δ_H (MPa ^{1/2})	11.018
Flash point	$T_f \geq 323 \text{ K}$	T_f (K)	400.172
Toxicity	$-\log(LC_{50})FM \leq 3.3 - \log(\text{mol/L})$	$-\log(LC_{50})FM$ ($-\log(\text{mol/L})$)	2.939
Normal melting point	$T_m \leq 270 \text{ K}$	T_m (K)	266.773
Normal boiling point	$T_b \geq 340 \text{ K}$	T_b (K)	523.919
Viscosity	$\mu \leq 1 \text{ cP}$	μ (cP)	0.577

Through using the decomposition-based algorithm, 272 feasible molecular candidates are designed at the first step with the constraints of molecular structures and linear properties (δ , δ_H , T_f , $-\log(LC_{50})FM$, T_m , T_b , μ), which are predicted by the GC methods (Hukkerikar, et al., 2012). Then, 6723 SMILES-based isomers are generated using the SMILES-based isomer generation algorithm. Afterwards, for the 6723 solvent candidates and solute Ibuprofen, *PR*% are individually calculated and arranged in descending order with the key property γ that is estimated by the integrated MLAC method - COSMO-SAC model. Finally, the top designed crystallization solvent is listed in Table 1. Although the DFT-based *PR*% of the top designed solvent in Table 1 (96.04%) has made a minor improvement compared with Karunanithi's solvent (94.95%), our best designed solvent is safer ($T_f = 400.172 \text{ K}$) and has lower toxicity measure ($-\log(LC_{50})FM = 2.939 - \log(\text{mol/L})$) than Karunanithi's ($T_f = 354.290 \text{ K}$ and $-\log(LC_{50})FM = 3.040 - \log(\text{mol/L})$). Further experimental verifications will be performed in the future to confirm the performance of the top designed solvent.

4. Conclusions

In this paper, an optimization-based MLAC-CAMD framework is established for solvent design, where the MLAC method is developed for fast predictions of $p(\sigma)$ using wACSFs descriptors through a HDNN model. Benefiting from the stereoscopic feature of the wACSFs descriptors and the powerful nonlinear fitting ability of the HDNN model, high prediction accuracy for $p(\sigma)$ is achieved with the MLAC method in terms of R^2 of $p(\sigma)$ and the overall AAPE of γ^∞ . A case study of designing crystallization solvents is performed, identifying a minor *PR*% improved solvent with the safer and lower toxic attributes compared with the previous work, which demonstrates the effectiveness of the MLAC-CAMD framework.

Acknowledgement

QL, LZ and JD are grateful for financial supports of NSFC (22078041, 21808025) and “the Fundamental Research Funds for the Central Universities (DUT20JC41)”.

References

- Y. Zhao, J. Chen, Q. Liu, Y. Li, 2020, Profiling the Structural Determinants of Aryl Benzamide Derivatives as Negative Allosteric Modulators of mGluR5 by In Silico Study, *Molecules*, 25, 2, 406
- Y. Dong, R. Zhu, Y. Guo, Z. Lei, 2018, A United Chemical Thermodynamic Model: COSMO-UNIFAC, *Industrial & Engineering Chemistry Research*, 57, 46, 15954-15958
- T. Mu, J. Rarey, J. Gmehling, 2007, Group contribution prediction of surface charge density profiles for COSMO-RS(OI), *AIChE Journal*, 53, 12, 3231-3240
- Q. Liu, L. Zhang, L. Liu, J. Du, A.K. Tula, M. Eden, R. Gani, 2019, OptCAMD: An optimization-based framework and tool for molecular and mixture product design, *Computers & Chemical Engineering*, 124, 285-301
- N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, 2011, Open Babel: An open chemical toolbox, *Journal of Cheminformatics*, 3, 1, 33
- M.J. Frisch, G.W. Trucks, H.B. Schlegel, et al., Gaussian 09W Rev. D.01, Gaussian, Inc., Wallingford, CT, 2016
- M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsényi, P. Marquetand, 2018, wACSF—Weighted atom-centered symmetry functions as descriptors in machine learning potentials, *The Journal of Chemical Physics*, 148, 24, 241709
- L. Zhang, D.K. Babi, R. Gani, 2016, New Vistas in Chemical Product and Process Design, *Annual Review of Chemical and Biomolecular Engineering*, 7, 1, 557-582
- J. Scheffczyk, L. Fleitmann, A. Schwarz, M. Lampe, A. Bardow, K. Leonhard, 2017, COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS, *Chemical Engineering Science*, 159, 84-92
- G. Xavier, B. Antoine, B. Yoshua, 2011, Deep Sparse Rectifier Neural Networks, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, PMLR, 315-323
- E. Mullins, R. Oldland, Y.A. Liu, S. Wang, S.I. Sandler, C.-C. Chen, M. Zwolak, K.C. Seavey, 2006, Sigma-Profile Database for Using COSMO-Based Thermodynamic Methods, *Industrial & Engineering Chemistry Research*, 45, 12, 4389-4415
- D. Kingma, J.B.J.C. Science, 2014, Adam: A Method for Stochastic Optimization, *arXiv e-prints*, arXiv:1412.6980
- C.M. Hsieh, S.I. Sandler, S.T. Lin, 2010, Improvements of COSMO-SAC for vapor–liquid and liquid–liquid equilibrium predictions, *Fluid Phase Equilibria*, 297, 1, 90-97
- B. Chen, Z. Lei, Q. Li, C. Li, 2005, Application of CAMD in separating hydrocarbons by extractive distillation, *AIChE Journal*, 51, 12, 3114-3121
- A.T. Karunanithi, L.E.K. Achenie, R. Gani, 2005, A New Decomposition-Based Computer-Aided Molecular/Mixture Design Methodology for the Design of Optimal Solvents and Solvent Mixtures, *Industrial & Engineering Chemistry Research*, 44, 13, 4785-4797
- A.T. Karunanithi, L.E.K. Achenie, R. Gani, 2006, A computer-aided molecular design framework for crystallization solvent design, *Chemical Engineering Science*, 61, 4, 1247-1260
- A.S. Hukkerikar, B. Sarup, A. Ten Kate, J. Abildskov, G. Sin, R. Gani, 2012, Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis, *Fluid Phase Equilibria*, 321, 25-43

A versatile modelling system for integrated chemical product design problems

Shiyang Chai^a, Lei Zhang^{a,*}, Jian Du^a, Anjan K Tula^b, Rafiqul Gani^{c,d}, Mario R Eden^e

^a*Institute of Chemical Process Systems Engineering, School of Chemical Engineering, Dalian University of Technology, Dalian 116024, China*

^b*College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China*

^c*Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea*

^d*PSE for SPEED, Skyttemosen 6, DK-3450 Allerød, Denmark*

^e*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*

Abstract

Chemical products can be classified as single species, mixtures and/or blends, formulated and/or functional products and chemicals-based devices. Each product class can be represented by a suite of models representing different sub-classes depending on the product function and/or application. In this work, a versatile modelling system consisting of a collection of ten sub-models (molecular structure, property, process, costing, pricing, economic analysis, quality, sustainability, environmental impact, and performance) is presented. This modelling system has been incorporated into ProCAPD, which is a chemical product design software tool together with an extended database and a library of product design templates. Application of ProCAPD with the built-in versatile modelling system is highlighted through two cases: crystallization solvent design and reaction solvent design.

Keywords: chemical product design, versatile model, ProCAPD, decomposition-based algorithm.

1. Introduction

Chemical products are widely used in modern society. Chemical product design usually has hierarchical and multi-disciplinary features. As pointed out by Ng and Gani (2019), chemical products can be classified as single species, mixtures and/or blends, formulated and/or functional products and chemicals-based devices. Each product class contains sub-classes depending on the product function and/or application. One of the difficulties in computer-aided chemical product design is that each product class and sub-class are defined in terms of different sets of product attributes, which needs multi-discipline knowledge for the modeling of the design problems.

A variety of sub-models have already been used in different problem formulations of chemical product design. For example, in solvents for CO₂ capture, Pereira et al. (2011) have employed complex thermodynamic models for prediction of CO₂ solubility. Chai et al. (2020) combined COSMO-based solubility predictions in the design of solvents for crystallization. Zhang et al. (2018) proposed a multi-objective optimization framework for the influence of government policy and consumer preference on product design. Fung

et al. (2016) proposed the use of a grand product design model including costing, pricing, sustainability, etc., together with the more common product attributes related to properties and process. Subsequently, Ng and Gani (2019) identified the technical and non-technical issues and/or problems in integrated product and process design, the relationships among which are captured in a Grand Product Design Model. Many of these recent developments have been included in teaching product and process design to prepare the new generation of chemical engineers. However, the specific models presented in their work are mainly simplified correlations, and most of the sub-models presented are conceptual equations. Therefore, a versatile modeling framework is needed for wider applications of chemical product design problems from which the needed sub-models can be retrieved and combined easily for a specific chemical product design problem.

In this paper, a versatile modeling framework suitable for chemical product design has been developed and implemented in ProCAPD (Kalakul et al., 2018). In section 2, the versatile model is formulated as a collection of ten sub-models. In section 3, the new additions to ProCAPD software are presented. In section 4, two cases studies are given for highlighting the application of ProCAPD software in chemical product design.

2. Versatile Modeling Framework

The framework of the versatile mathematical model is shown in Figure 1.

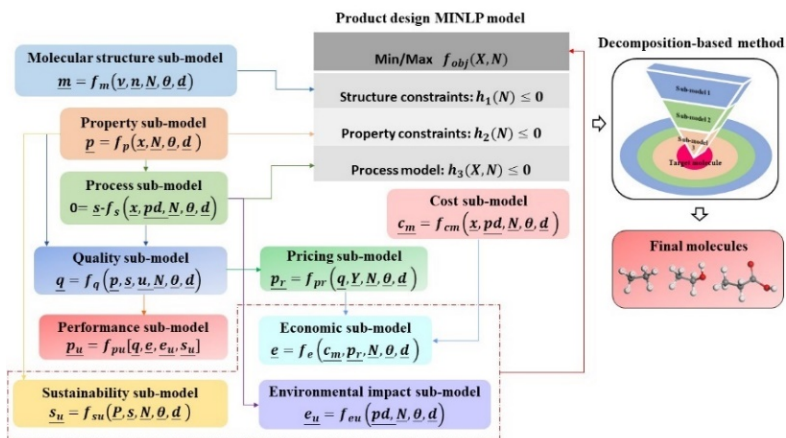


Figure 1. Versatile modelling framework for chemical product design

In Figure 1, molecular structure sub-model mainly represents the constraints related to molecular structure feasibility in terms of octet rule (v) (Odele and Macchietto, 1993) and the number of groups (n) representing the molecule. Here, N is the number of arguments, θ and d represent the supplied model parameters and specified variables, respectively. Property sub-model is used to predict the properties (p) of the compounds of the product (x). Process sub-model relates the product attributes (s) of each compound of the product (x) and process design variables (pd) through the process model equations based on conservation of mass and energy. The material properties (p), product structure attributes (s) and product use conditions (u) collectively determine the product quality (q), which is determined through the product quality sub-model. Pricing sub-model is to determine product price (p_r) and market share (Y) under the condition of given market model. Market

share (Y), in turn, affects the amount of production and the required capital investment and operating costs. Cost sub-model related to product manufacturing mainly includes the capital cost and production cost used to produce the product. All of the above models lead to an economic sub-model, which is used to evaluate the financial performance of the product being designed. There are many indicators that can be used for economic analysis, such as net present value (NPV), internal rate of return (IRR), and return on investment (ROI). For the analysis of the commercial value of products, the environmental sub-model needs to be considered in the production process, mainly due to the harmful substances generated during the production process. It is necessary to consider the recovery and treatment of these substances. So, it is necessary to consider the sustainability sub-model. Performance sub-model refers to the performance indicators in the product design process. These indicators include product quality (q), economy (e), environmental (e_u) and sustainability issues (s_u), which can be formulated as a single or multiple objective problem.

In Figure 1, each sub-model is not only a conceptual single function equation but a more detailed set of model equations. The model equations and parameters can be obtained through literature, experiment or machine learning. For each sub-model, it can be further decomposed according to its type. For example, the property sub-model can be divided into primary pure component properties, secondary pure component properties and phase equilibrium related properties. In addition, the modelling framework is continuously extended through the addition of new models in different categories. Each class of sub-models has its corresponding interface for data transfer. Also, together with an associated database and a model generation sub-system, new models can be created through data regression and machine learning based modeling. The developed modelling framework has been incorporated into a chemical product design software tool ProCAPD, which provides the software new features such as use of machine learning based models and the associated problem solution strategy.

3. Multipurpose Tool for Product design

The above versatile model is integrated into ProCAPD. The architecture of the ProCAPD tool is shown in Figure 2. It includes several other features, such as design templates for guiding the user through established workflows for specific product design problems; solution approach where different solution strategies can be applied by different problems. The aim of ProCAPD is an integrated tools which contain a suite of computer-aided tools that helps with the problem formulation, database search, properties prediction and model development.

The solution of a chemical product design problem using any model-based system requires the retrieval, calculation, use and management of a wide range of models and methods, as well as very large amounts of data. For this reason, ProCAPD is integrated with different in-house tools that are necessary to provide more efficient and flexible options for chemical product design and analysis.

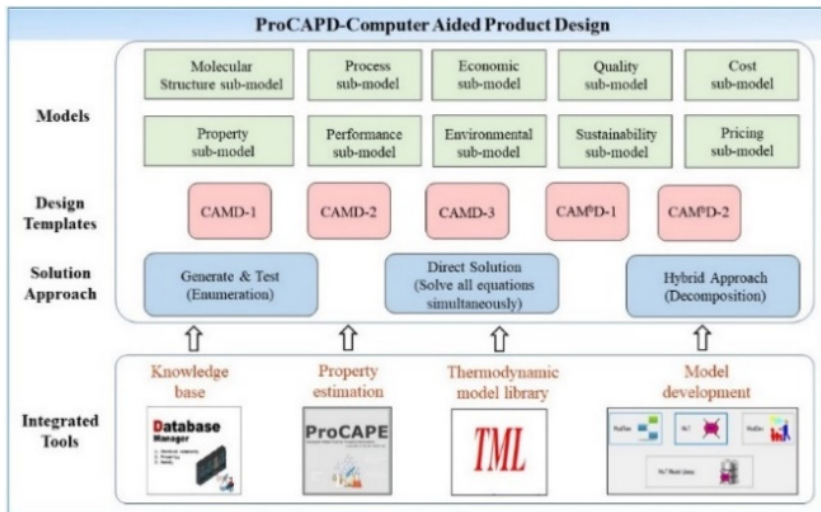


Figure 2. ProCAPD architecture for solving different product design problems

4. Case study

4.1. Crystallization solvent design

This case study is taken from Chai et al. (2020). 2-Mercaptobenzothiazole (MBT) is an important vulcanization accelerator using in the rubber industry. In the solvent-based method, it is important to select a suitable crystallization solvent to ensure the refined MBT meets the industry requirements.

In this case study, CH₃, CH₂, CH, C, aC-CH₃, aCH, aC-Cl, OH, CHCOO groups are selected. The lower and upper bonds for structure and property constraints and the optimal solvents are shown in Table 1. (A is the designed solvent that need to be kept confidential). The decomposition-based strategy (Karunanithi et al., 2005) is employed to solve the MINLP model.

Table 1. The lower and upper bonds for structure and property constraints and the candidate solvents

Objective function: $\text{Max } PR\% = \frac{1}{0.88(1-w_1)} (1 - \frac{w_2}{w_2}) \times 100\%$					
Properties	Lower bound	Upper bound	No	Name	PR%
N_G	3	8	1	Methyl lactate	97.34
N_C	-	7	2	n-Heptanol	94.61
N_F	1	6	3	A	93.73
MW (g/mol)	80	200	4	5-Methylhexan-1-ol	92.72
T_m (K)	173	310	5	2-Heptanol	92.44
T_b (K)	373	600	6	n-Hexanol	91.14
T_f (K)	273	393	7	2-Methyl-1-pentanol	90.82
$Solp$ (MPa ^{1/2})	18	21	8	2-Ethyl-1-butanol	90.69
$-\log(LC_{50})$ (-log(mol/L))	-	4.8	9	3-Methyl-1-pentanol	90.10
SLE	$\ln x_i^{sat} - \frac{\Delta H_{fus}}{RT_m} \left(1 - \frac{T_m}{T}\right) + \ln \gamma_i^{sat} = 0$		--	--	--

Note: The process sub-model also includes the COSMO-SAC model equations for activity coefficients (Chen et al., 2016), which are not listed in the table

From Table 1, the product yields (PR%) of the candidate solvents are increased to above 90%, which meet the industrial needs.

4.2. Reaction solvent design

This case study is taken from Zhang et al. (2020). TMQ is an important antioxidant and is widely used in rubber production technology. In order to further improve the quality of TMQ products, it is necessary to find a more suitable reaction solvent.

The groups (building blocks) selected in this case include CH₃, CH₂, CH, C, CH₂=CH, CH=CH, CH₂=C, CH-C, C=C, CH₃CO, CH₂CO, CH₃COO, CH₂COO, CH₃O, CH₂O, aCH, aC, aC-CH₃, aC-CH₂, aC-CH, CH₂(cyc), CH(cyc), C=C(cyc). The lower and upper bonds for structure and property constraints and the optimal solvents are shown in Table 2.

Table 2. The lower and upper bonds for structure and property constraints

Objective function: maximizing the yield of dimer: $n_d^t = \int r_d^t dt$ minimizing the energy consumption: $Q = f_q(\Delta H_{vap,298K}^i, T, T_c^i, x_i, R, F_i, \varepsilon_i)$		
Properties	Lower bound	Upper bound
N_G	4	8
N_F	1	7
T_m (K)	298	--
T_b (K)	410	440
ΔH_{vap} (kJ/mol)	0	50
$Solp$ (MPa ^{1/2})	16	22
$-\log(LC_{50})$ (-log(mol/L))	--	4.0
LLE: $x_i^I \gamma_i^I = x_i^{II} \gamma_i^{II} \quad i \in (A, B, C, D, E, F, G)$		
Reaction model: $r_i = f_r(k_{trans}^i, k_f, \gamma_i, x_i) \quad i \in (A, B, C, D, E, F, G) \quad j \in (1, 2, 3, A)$		
Distillation column model: $Q = f_q(\Delta H_{vap,298K}^i, T, T_c^i, x_i, R, F_i, \varepsilon_i) \quad i \in (A, B, D, Solvent)$		

Note: The process sub-model also includes the COSMO-SAC model equations for activity coefficients (Chen et al., 2016), which are not listed in the table

The designed solvents are listed in Table 3. From the results, different objectives lead to different solvent ranking. Among these candidate solvents, solvent T has the optimal performance in both two objective functions.

Table 3. Optimization results of the solvents for maximizing the yield of dimer and minimizing the energy consumption

No.	Solvent	Increased dimer ² (%)	Water content (%)	Solvent recovery (%)	Increased energy consumption ² (kJ)
1	T ¹	36.216	0.767	99.980	36392
2	Ethylbenzene	34.851	0.715	99.994	36615
3	Isopropyl benzene	33.850	0.683	99.997	40458
4	1-Methoxyhexane	33.759	13.660	99.988	46945
5	3-Ethyl-3-penten-2-one	33.612	5.980	99.940	41911
6	3,4-Dimethylpent-4-en-2-one	33.594	4.674	99.951	39824
7	4,4-Dimethyl-2-pentanone	33.569	4.333	99.974	39700
8	3,3-Dimethyl-2-pentanone	33.545	5.639	99.958	40604
9	4-Methyl-3-Methylidene-pentan-2-one	33.296	5.821	99.926	40605
10	1-Methoxyheptane	28.789	2.223	100.000	40625

T¹ is the solvent that need to be kept confidential

Superscript 2 indicates the added value compared to the solvent free system

5. Conclusions

A versatile modeling framework suitable for chemical product design has been developed and implemented in ProCAPD, a chemical product design software tool together with an extended database and a library of solution approaches. New features such as machine learning based models developed in previous publications, have been highlighted in this

paper in terms of their incorporation in the product design problem and the associated problem solution strategy. These new features are highlighted by two application examples (crystallization solvent design and reaction solvent design).

Acknowledgement

SC, LZ and JD are grateful for financial supports of NSFC (22078041, 21808025) and “Fundamental Research Funds for the Central Universities (DUT20JC41)”.

References

- A.T. Karunanithi, L.E.K. Achenie, R. Gani, 2005, A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. *Industrial & Engineering Chemistry Research*, 44, 4785-4797
- F. Pereira, E. Keskes, A.Galindo, G. Jackson, C. Adjiman, 2011, Integrated solvent and process design using a SAFT-VR thermodynamic description: High-pressure separation of carbon dioxide and methane, *Computers & Chemical Engineering*, 35, 474-491
- K. M. Ng, R. Gani, 2019, Chemical Product Design: Advances in and Proposed Directions for Research and Teaching, *Computers & Chemical Engineering*, 126, 147-156
- K.Y. Fung, K.M. Ng, L. Zhang, R. Gani, 2016, A grand model for chemical product design, *Computers & Chemical Engineering*, 91, 15-27
- L. Zhang, J. Pang, Y. Zhuang, L. Liu, J. Du, Z. Yuan, 2020, Integrated solvent-process design methodology based on COSMO-SAC and quantum mechanics for TMQ (2,2,4-trimethyl-1,2-H-dihydroquinoline) production. *Chemical Engineering Science*, 115894
- O. Odele, S. Macchietto, 1993, Computer-Aided Molecular Design-a Novel Method for Optimal Solvent Selection. *Fluid Phase Equilibria*, 82, 47-54
- Q. Liu, L. Zhang, L. Liu, J. Du, A. K. Tula, M. Eden, R. Gani, 2019b, OptCAMD: An Optimization-Based Framework and Tool for Molecular and Mixture Product Design, *Computers & Chemical Engineering*, 124, 285-301
- Q. Liu, L. Zhang, L. Liu, J. Du, Q. Meng, R. Gani, 2019a, Computer-Aided Reaction Solvent Design based on Transition State Theory and COSMO-SAC, *Chemical Engineering Science*, 202, 300-317
- S. Chai, Q. Liu, X. Liang, Y. Guo, S. Zhang, C. Xu, J. Du, Z. Yuan, L. Zhang, R. Gani, 2020, A grand product design model for crystallization solvent design. *Computers & Chemical Engineering*, 135, 106764
- S. Kalakul, L. Zhang, Z. Fang, H. A. Choudhury, S. Intikhab, N. Elbashir, M. R. Eden, R. Gani, 2018, Computer aided chemical product design-ProCAPD and tailor-made blended products, *Computers & Chemical Engineering*, 116, 37-55
- W. L. Chen, C. M. Hsien, L. Yang, C. C. Hsu, S. T. Lin, 2016, A Critical Evaluation on the Performance of COSMO-SAC Models for Vapor-Liquid and Liquid-Liquid Equilibrium Predictions Based on Different Quantum Chemical Calculations, *Industrial & Engineering Chemistry Research*, 55, 9312-9322
- X. Zhang, L. Zhang, K. P. Fung, G. P. Rangaiah, K. M. Ng, 2018, Product design: Impact of government policy and consumer preference on company profit and corporate social responsibility. *Computers & Chemical Engineering*, 118, 118-131

An integrated system of multi effect distillation and wind power system - Evaluation of total energy saving

O.M.A. Al-hotmani ^a, Mudhar. A. Al-Obaidi ^b, Y. M. John ^a Raj. Patel ^a, and Iqbal M. Mujtaba ^{a,*}

^a *Department of Chemical of Engineering, Faculty of Engineering and Informatics, University of Bradford, Bradford, West Yorkshire, BD7 1DP, UK*

^b *Middle Technical University, Technical Institute of Baquba, Dayala – Iraq*
I.M.Mujtaba@bradford.ac.uk

Abstract

Due to depletion of freshwater resources in arid and semi-arid regions accompanied with an increase in water demand, the focus is on improving thermal desalination technologies such as multi effect distillation (MED) that are energy intensive. Undoubtedly, this technology plays an important role in generating high quantity of freshwater from saline water resources. To maintain a sustainable seawater desalination technology, this research suggests a new hybrid system of 10 effects of MED and a farm of 100 wind turbines to be installed in Al-Shuaibah Power Plant located 103 km south of Jeddah in the Kingdom of Saudi Arabia. To carry out this proposal, a generic model for MED developed by the same authors and a set of model equations of wind power turbine are embedded in a mathematical framework. The developed framework is then used to carry out a simulation for a given set of seawater characteristics, and input design parameters at different months of the year 2019. The study investigates the average monthly saving of power for the proposed hybrid system compared to a single MED system. It is revealed that the wind turbines can save around 43% of electrical power of MED system in June 2019 based on the mean monthly wind speed.

Keywords: Seawater desalination; Multi effect distillation; Hybrid system; Wind turbine; Energy saving.

1 Introduction

Despite the abundance of water around the globe, there is a clear challenge of delivering freshwater due to scarcity of freshwater resources accompanied with a rapid growth of population. Seawater desalination has confirmed its robustness to provide freshwater especially for those countries that have coastal areas with water shortage. However, desalination plants and especially thermal one requires a considerable amount of energy to produce freshwater from seawater resources. More specifically, multi effect distillation (MED) is one of the well-known thermal desalination technologies that has been extensively used in the Gulf countries including the United Kingdom of Saudi Arabia. The most part of consumed energy in the MED system is to provide the thermal energy required to evaporate the brine besides operating the pumps (electrical energy). Due to depleting petroleum reserves and the considerable CO₂ emissions associated with these

processes, there has been progressive interest to integrate thermal desalination with a renewable source of energy. However, harnessing wind as an alternative source of power is reliant on weather and location. Interestingly, the KSA has a consistent wind energy resource as confirmed by Salah et al. [1] who stated an average wind speed of 3.5 m/s in various areas of the KSA. Thus, it would be possible to exploit wind energy as a reliable source of energy to drive the MED desalination system. The design and efficacy of a hybrid system of renewable energies and MED has been discussed in the literature. Specifically, the familiar use of wind energy of a wind turbine in the seawater desalination systems is to employ the mechanical and electrical energies. For instance, Yilmaz and [2] presented the hybrid seawater desalination system of MED and flat-plate collector (solar) and wind turbine units. They proposed to instill the hybrid system at 18 stations located at on-shore regions of Turkey. The research stated the potential of renewable energy of each location and identified the required system size of each sub-system. Also, a high-capacity wind turbine was integrated to MED system by Khalilzadeh and Nezhad [3] to exploit the wasted heat of wind turbine to produce the necessary steam for MED system. The results affirmed the merits of the proposed technique which can work adequately without any limitations during variable temperatures.

This study focuses on exploring the feasibility of a proposed hybrid renewable energy system of 100 wind turbines (Dencon Tornado type) and 10 effects MED desalination system to be installed in the Al-Shuaibah Power Plant, located 103 km south of Jeddah in the Kingdom of Saudi Arabia. The simulation-based model built in gPROMS software is carried out at fixed seawater conditions to investigate the average monthly power saving throughout 2019 that can be achieved for the hybrid system.

2 Hybrid seawater desalination system of wind turbine and MED

The proposed hybrid system of forward feed MED of 10 effects coupled to thermal vapor compression (external steam provider) and a farm of 100 wind turbines is schematically presented in Fig. 1. The wind turbine type Dencon Tornado [4] is selected to be linked with the MED system. Each effect of MED comprises an evaporator of brine, spray nozzle, demister, and feed pre-heater. The characteristics of the associated MED and design parameters of wind turbine systems are illustrated in Table 1. The purpose of installing the wind turbine as a renewable source of energy is to generate the power required to produce fresh water from the MED system.

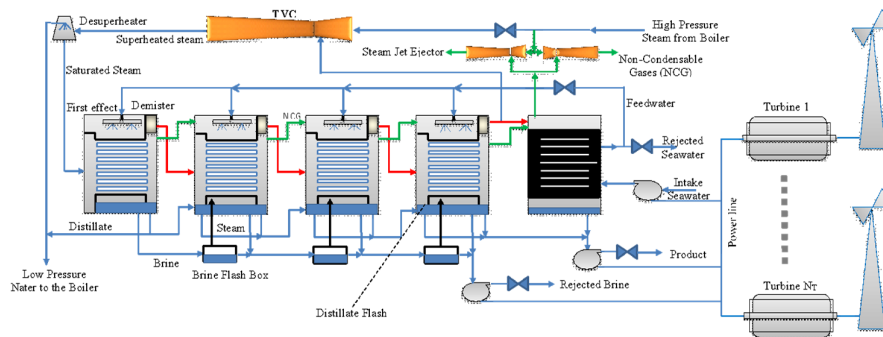


Fig. 1. Hybrid system of MED and wind turbine systems

3 Modelling of MED system

Filippini et al. [5] developed a successful model for the MED-TVC system. Some important equations of the model developed are shown in Table 2. The procedure of model developed and the associated assumptions for MED+TVC system and its validation can be found in more details in Filippini et al. [5].

Table 1. Characteristics of MED and wind turbine (Dencon Tornado) systems

Wind turbine system						
Parameter	Rated power, Pr (kW)	Height above the ground level, H (m)	Wind turbine cut-in speed, Ucut (m/s)	Wind turbine rated speed, Ur (m/s)	Wind turbine cut-off speed, Uf (m/s)	Shape parameter, k (m/s)
Value	200	32	3	12	25	2.5649
MED system						
Parameter				Unit		Value
Number of stages				(-)		10
External steam temperature, flow rate, pressure				°C, kg/s, kPa		70, 8, 1300
Brine temperature, salinity				°C, ppm		40, 6000

Table 2. Modelling of MED and TVC system (Filippini et al. [5])

Description	Equation	Unit
Feed flow rate	$Mf = \frac{Ms \lambda(Ts)}{Q_{sensible} + Q_{latent}}$	kg/s
Sensible heat in the first effect	$Q_{sensible} = Mf \int_{t1}^{t1} cp(T1, x1) dT$	kJ/s
Latent heat in the first effect	$Q_{latent} = D1 \lambda(Tv1)$	kJ/s
Temperature drop among effects (first attempt)	$\Delta T = \frac{Ts - Tb}{n}$	°C
Temperature drop among pre-heaters (first attempt)	$\Delta T = \Delta t$	°C
Feed temperature in the first effect	$t1 = tn + (n - 1)\Delta t$	°C
Temperature of vapor phase	$Tv = T - BPE(T, x)$	°C
Flow rate of flashed distillate	$D_{flashed,t} = \alpha B_{t-1}$	kg/s
Fraction of distillate by flashing	$\alpha = \frac{cp(T_{mean}, x_{mean})\Delta T}{\lambda(T_{mean})}$	-
Mean temperature	$T_{mean} = \frac{T1 + Tb}{2}$	°C
Mean salinity	$x_{mean} = \frac{xf + xb}{2}$	Ppm
Fraction of distillate by evaporation	$\beta = \frac{\alpha [xb(1 - \alpha)^n - xf]}{(xb - xf) [1 - \alpha]^n}$	-
Flow rate of evaporated distil	$D_{i,boiled} = \beta M_D$	kg/s
Total distillate	$D_i = D_{i,boiled} + D_{i,flashed}$	kg/s

Rejected brine flow rate	$B_i = B_{i-1} - D_i$	kg/s
Salinity profile in the effects	$x_i = \frac{x_{i-1}B_{i-1}}{B_i}$	Ppm
Heat load in i-th effect	$Q_i = D_{boiled,i-1} \lambda(T_{v,i-1})$	kJ/s
Description	Equation	Unit
Area of i-th effect	$\frac{Q_i}{U_{ev,i} \Delta T_{ev,i}} = A_{ev,i}$	m ²
Temperature drop in heat exchangers	$\Delta T_{ev,i} = \Delta T - BPE_{i-1}$	°C
Area of i-th pre-heater	$Mf \cdot \int_{t_{i+1}}^{t_i} cp(t, xf) dt = U_{ph,i} A_{ph,i} \Delta t_{log,i}$	m ²
Logarithmic temperature difference in pre-heaters	$\Delta T_{log,i} = \frac{\Delta T}{\log\left(\frac{T_{v_i} - t_{i+1}}{T_{v_i} - t_i}\right)}$	°C
Area of the last condenser	$Q_{COND} = U_{COND} A_{COND} \Delta T_{log,COND}$	m ²
Heat load in the last condenser	$Q_{COND} = D_n \lambda(T_{v_n})$	kJ/s
Logarithmic temperature difference in final condenser	$\Delta T_{log,COND} = \frac{tn - Tw}{\log\left(\frac{T_{v_n} - Tw}{T_{v_n} - tn}\right)}$	°C
Temperature correction factor	$TCF = 2e - 8.Tv_n^2 - 0.0006.Tv_n + 1.0047$	-
Pressure correction factor	$PCF = 2e - 7.TPm^2 - 0.0009.Pm + 1.6101$	-
Pressure at vapor temperature	$Pv = P_{crit} e^{\left(\frac{T_{crit} + 273.15}{Tv_n}\right)^{-1}} \cdot \sum_{j=1}^8 f_j$	Bar
Pressure at steam temperature	$Ps = P_{crit} e^{\left(\frac{T_{crit} + 273.15}{Ts}\right)^{-1}} \cdot \sum_{j=1}^8 f_j$	Bar
Compression ratio	$CR = \frac{Pv}{Ps}$	-
Entrainment ratio	$Ra = 0.296 \frac{Ps^{1.19} Pm^{0.015} PCF}{Pev^{1.04} Pev^{0.015} TCF}$	-
Motive steam flow rate	$Mm = Ms \frac{Ra}{1 + Ra}$	kg/s

4 Simulation of MED system

The simulation based the model developed was used to explore the required power to generate the fresh water from MED system. The simulation results based on the seawater conditions are depicted in Table 3.

5 Modelling of wind turbine energy

The generated power P_W (kW) of a single wind turbine can be calculated using the correlation of Eltamaly and Al-Shamma'a [6].

$$\begin{aligned}
 1 - P_W &= 0 && \text{if } U \leq U_{cut} \text{ and } U \geq U_f \\
 2 - P_W &= P_r \frac{U^k - U_{cut}^k}{U_r^k - U_{cut}^k} && \text{if } U_{cut} \leq U \leq U_r \\
 3 - P_W &= P_r && \text{if } U_r \leq U \leq U_f
 \end{aligned}$$

where P_r , is the rated power (kW), U the wind speed (m/s), U_{cut} the wind turbine cut-in (m/s), U_f the wind turbine cut-off (m/s), U_r the rated wind turbine (m/s) and k the rated wind and shape parameter (-), respectively.

Table 3. Seawater conditions and corresponding power of MED system

Seawater salinity	ppm	39000
Seawater temperature	°C	25
Specific energy consumption	kWhr/m ³	17.9885
Productivity	kg/s, m ³ /hr	68.327, 245.977
Power	kW	4424.760

6 Evaluation of wind turbine and MED systems at various wind speeds

The simulation of the MED system showed the productivity of about 245 m³/hr of fresh water for the design operations of the MED system and specified seawater conditions (reported in Table 3). This specifically requires generating 4424.760 kW as an average required power to produce fresh water from MED system. The variation of average monthly wind speed in the studied region (Al-Shuaibah Power Plant, Jedda in the KSA) for a period of 12 year of 2019 is shown in Table 4. Table 4 also shows the corresponding generated monthly power from a single wind turbine. The results confirmed a low generated power using only a single wind turbine system due to low average wind speed during the months of 2019 (Table 4). This in turn means the necessity of constructing a considerable number of turbines to be integrated with the MED system to generate the total required power of the MED system. However, this is not a feasible option due to a high capital cost. To solve this problem, this research proposed the deployment of a farm of 100 wind turbines to be integrated to the MED. In this respect, Table 4 provides the power saving values. Clearly, June of 2019 has the highest percentage of power saving of 42.81% compared to November 2019 of 6.67%. Also, it should be noted that the calculation of power saving is carried out based on the monthly average wind speed. Therefore, it is fair to expect a regular increase of generated power during the days that can increase the relevant energy saving. For example, 100 wind turbines and MED attains an energy saving of 63.64% at the maximum wind speed for June 2019. It is also fair to claim that the use of a larger scale of wind turbine (if it is existed) would reduce the total number of required turbines to be integrated to MED-TVC seawater desalination system. Due to the fact of variable wind speed at different time scale and the necessity of a constant source of energy to drive the desalination plant, this study therefore recommends the investigation of the integration of both solar (photovoltaic cells) and wind turbine (sources of energy) with the MED desalination system. Moreover, having a battery bank to save generated electricity from the wind turbine would be a possible solution during times of low power consumption in the plant. In this regard, the integration of solar energy to MED desalination plant (located in Aqaba, Jordan) has reduced the total thermal energy by 68% [7].

7 Conclusions

This paper has presented a proposed integrated system of wind turbine and MED thermal desalination plant to be instilled in in Al-Shuaibah Power Plant, Jedda, KSA. The simulation of the proposed hybrid system is carried out based on developed model of sub-systems. The simulation results confirmed the possibility of attaining a maximum electrical power saving of 42.81% in June 2019 based on the monthly mean wind speed for the proposed hybrid system of 10 effects MED and a farm of 100 wind turbines. The lowest energy saving was found to be in November 2019.

Table 4. Mean wind speed [8], generated electrical power, and energy saving

Month	Mean wind speed, U (m/s)	Generated power of a single wind turbine, Pw (kW)	Power saving of 100 wind turbines%
January	3.941	5.9582	13.46
February	4.887	14.676	33.16
March	4.686	12.576	28.42
April	4.500	10.755	24.30
May	3.598	3.492	7.89
June	5.260	18.944	42.81
July	4.102	7.238	16.36
August	4.081	7.067	15.97
September	3.956	6.074	13.72
October	3.828	5.107	11.54
November	3.516	2.954	6.67
December	4.535	11.089	25.06

References

- [1] Salah, M.M., Abo-khalil, A.G. and Praveen, R.P., 2019. Wind speed characteristics and energy potential for selected sites in Saudi Arabia. *Journal of King Saud University-Engineering Sciences*.
- [2] Yılmaz, İ.H. and Söylemez, M.S., 2012. Design and computer simulation on multi-effect evaporation seawater desalination system using hybrid renewable energy sources in Turkey. *Desalination*, 291, 23-40.
- [3] Khalilzadeh, S. and Nezhad, A.H., 2018. Utilization of waste heat of a high-capacity wind turbine in multi effect distillation desalination: Energy, exergy and thermoeconomic analysis. *Desalination*, 439, 119-137.
- [4] Demon. Available from: <https://en.wind-turbine-models.com/turbines/1876-dencon-tornado-200-26>.
- [5] Filippini G., Al-Obaidi M.A., Manenti F., Mujtaba I.M., 2018. Performance analysis of hybrid system of multi effect distillation and reverse osmosis for seawater desalination via modelling and simulation. *Desalination*, 448, 21–35.
- [6] Eltamaly A.M., and Al-Shamma'a A.A., 2016. Optimal configuration for isolated hybrid renewable energy systems. *Journal of Renewable and Sustainable Energy*, 8(4), 045502.
- [7] Bataineh K.M., 2016. Multi-effect desalination plant combined with thermal compressor driven by steam generated by solar energy. *Desalination*, 385, 39e52.
- [8] Wind Alert (19 January 2021). Retrieved from <https://www.windalert.com/>

Modelling and Performance evaluation of the air heated and water heated Humidification-Dehumidification desalination process using solar thermal energy

Damson Kaunga, Raj Patel, Iqbal M Mujtaba*

*Department of Chemical Engineering, Faculty of Engineering and Informatics,
University of Bradford. Bradford, West Yorkshire BD7 1DP, UK
L.M.Mujtaba@bradford.ac.uk*

Abstract

The aim of this study was to deduce the most effective desalination process by comparing performance of the water heated with the air heated Humidification-Dehumidification desalination systems. The process model was developed by applying conservation of mass and energy across the processes' major components and simulated using gPROMS ModelBuilder platform. Simulation results shows that, the water heated HDH desalination process has higher fresh water productivity than the air heated process. At the maximum inlet temperature of 60 °C the water heated and air heated HDH desalination processes have the maximum productivity of 42 kg/h and 18 kg/h respectively.

Keywords: Humidification, Dehumidification, desalination, solar thermal energy.

1. Introduction

A rapid growth in population, industrialisation and urbanisation is causing shortage of fresh water in many areas around the globe. On the other hand, sea water or underground saline water is abundant in most areas that encounter the problem of fresh-water shortage. As demand for fresh water continually increases, it becomes necessary to develop cheap desalination technologies which are powered by renewable energy (Kucera, 2014).

Multistage humidification-dehumidification (HDH) desalination using solar energy is a viable process for remote and arid areas with severe shortage of fresh water. Humidification-dehumidification is a thermal desalination cycle that operates by heating saline water using solar thermal energy, evaporating the heated water using a humidifier and finally condensing the water vapor to form fresh water in the dehumidifier. The process is viable because of its' simplicity, abundance of solar energy and moderate operating costs compared to conventional desalination processes (Al-Enezi, et al., 2006).

Various researchers have used simulation models to investigate the key parameters affecting performance of the humidification-dehumidification desalination process. On

the one hand, there is a common agreement that high yield of fresh water is favoured by increasing temperature of streams entering the humidifier. On the other hand, there is still a debate whether the maximum yield is achieved by heating the water stream (Huifang, et al., 2014) or by heating the air stream (Mohamed & El-Minshawy, 2011) before passing into the humidifier. Therefore, the purpose of this study is to use simulation tools and mathematical models to explore the performance of an emerging HDH water desalination process. Specifically, these are used to compare performance of the water heated HDH desalination system with air heated HDH desalination system to deduce system with the maximum fresh-water productivity.

2. Description of the HDH desalination models

2.1. Water heated HDH desalination cycle

In this cycle, pumped saline water from a storage vessel enters the system at the dehumidifier. As it passes in the dehumidifier it cools and dehumidifies moist air stream that comes from the humidifier as shown in Figure 1a. The enthalpy of vaporization for this condensation process is transferred from the moist air to the saline water to warm the feed stream. The condensate is collected as fresh water at the bottom. Saline water is further heated by passing through a solar heater. The hot saline water is sprayed on top of the humidifier and flows downward while a cool air stream flowing in the opposite direction is blown at the bottom. The tiny water drop formed by spraying saline water maximizes the surface area for water-gas contact which improves the evaporation rate. The humidification process occurs by transfer of heat and mass (vapor) from the hot saline water stream to the cool air stream.

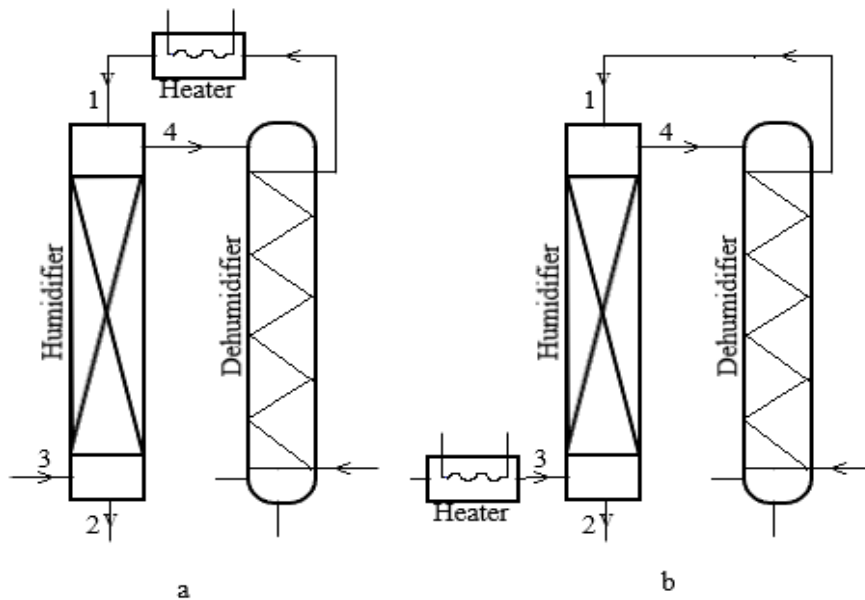


Figure 1: Schematic representation of the HDH desalination systems (a) Water heated HDH desalination cycle (b) Air heated HDH desalination cycle

2.2. Air heated HDH desalination cycle

Air heated HDH desalination cycle can be achieved by shifting the solar heater from water stream to the air stream as shown in Figure 1b. The flow of water and air through the humidifier and dehumidifier are exactly the same as described for water heated cycle, except that the heater is now placed to heat the air stream instead of heating water stream.

2.3. Modelling and simulation of the humidifier

Figure 1 shows the schematic diagram of water heated and air heated HDH desalination systems considered for developing the mathematical model in this study. They are assumed to operate at steady state. An air-water vapor mixture of known humidity ratio (ω) enters the humidifier through inlet 3 at a known pressure (p) and temperature (T). As the mixture passes through the humidifier, it comes into contact with a stream of water. If the entering mixture is not saturated, some of the water would evaporate. The pressure of the mixture is assumed to remain constant as it passes through the

humidifier. The humidity ratio ω_4 of the exiting moist air can be obtained by applying conservation of mass and conservation of energy to the humidifier of the system with an assumption that the moist air is an ideal gas mixture of dry air and water vapor. At the exit of the humidifier, air is assumed to be saturated. Moreover, the work, and changes in kinetic and potential energy are also ignored. The feed water salinity is less than 30 g/kg hence physical properties are similar to those of pure water. The enthalpy of the incoming saline water is evaluated as that of saturated liquid at the corresponding inlet temperature. Similarly, the specific enthalpy of the entering water vapor has been evaluated as that of saturated water vapor at the temperature of the incoming mixture (Moran, et al., 2011).

Conservation of mass (air and water) at steady state

$$\dot{m}_{a,3} = \dot{m}_{a,4} = \dot{m}_a \quad (1)$$

$$\dot{m}_2 = \dot{m}_1 + \dot{m}_a (\omega_3 - \omega_4) \quad (2)$$

At steady state an energy rate balance can be reduced to

$$\dot{m}_1 h_1 - \dot{m}_2 h_2 = \dot{m}_a \left((h_a + \omega h_g)_3 - (h_a + \omega h_g)_4 \right) \quad (3)$$

If enthalpy of water is approximated as $h^l \approx h^l(T^l)$, $h^g \approx h^g(T^g)$ then equation 2 and 3 can be combined and solved to get:

$$(\omega_3 - \omega_4) = \frac{\dot{m}_1 (h_f(T_1) - h_f(T_2)) + \dot{m}_a (h_{a3} - h_{a4}) + \dot{m}_a (\omega_3 h_{g3} - \omega_4 h_{g4})}{\dot{m}_a h_f(T_2)} \quad (4)$$

The dry air enthalpies of equation 4 can be calculated from the specific heat capacity and temperature of air as follows

$$h_{a3} - h_{a4} = C_{pa}(T_3 - T_4) \quad (5)$$

The evaporation rate of the system is given as:

$$\dot{m}_v = \dot{m}_a(\omega_4 - \omega_3) \quad (6)$$

The humidity ratio ω_3 at the entrance is calculated as

$$\omega_3 = 0.622 \frac{p_g(T_3)}{p - p_g(T_3)} \quad (7)$$

$$h(T) = \int_0^T C_p(T) dT \quad (8)$$

Heater energy balance equation (for both cycles) can be written as:

$$\dot{Q}_{in} = \dot{m}_f C_{p,f}(T_{f,out} - T_{f,in}) \quad (9)$$

$$\dot{Q}_{in} = \dot{m}_a C_{p,a}(T_{a,out} - T_{a,in}) \quad (10)$$

Table 1: Definition of symbols used for mathematical models

Symbols	Subscripts
C_p specific heat capacity ($\text{kJ kg}^{-1}\text{K}^{-1}$)	a air
h specific enthalpy (kJ kg^{-1})	$1,2,3,4$ inlet and exit points
\dot{m} mass flow rate (kg/s)	g,v water vapour
T temperature ($^{\circ}\text{C}$)	f liquid water
ω humidity ratio (kg_w/kg_a)	

Table 2: Parameters used for model simulation

Parameter	Value		Parameter	Value	
	System a	System b		System a	System b
C_p	1.006	1.006	hg_4	2583.2	2583.2
m_l	0.28	0.28	T_1	50	20
m_a	0.15	0.15	T_2	45	30
h_{f1}	209.33	83.96	T_3	20	50
h_{f2}	125.79	125.79	p	1	1
hg_3	2538.1	2592.1	$p_g(T_3)$	0.023	0.1235

3. Results and Discussion

In this study we used mathematical models to deduce the system with higher productivity by comparing performance of the water heated HDH desalination system to

the air heated HDH desalination system. These models were validated by comparing theoretical results obtained through model simulation with the corresponding experimental data which were previously reported by (Huifang, et al., 2014).

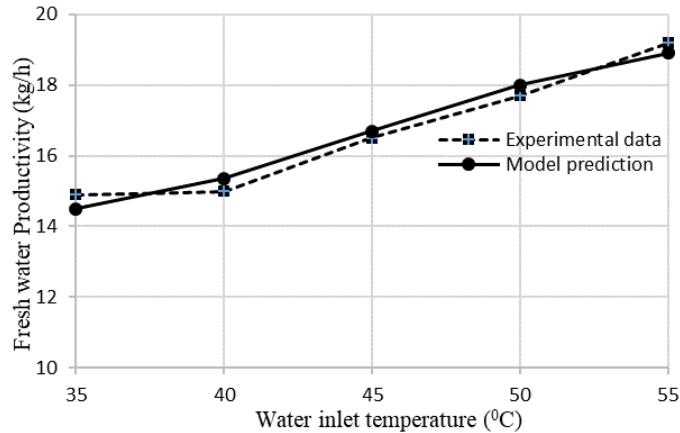


Figure 2: Comparison of the simulated and measured productivity of the system corresponding to various inlet temperature of saline water

As the base case, performance of the two HDH desalination systems were compared by changing the position of heater between the water and air streams. For water heated system the temperature of saline water and air at the inlet points of the humidifier was 50 °C and 20 °C respectively. For the air heated system the inlet air was heated to 50 °C while temperature for the inlet water was 20 °C. The flow rates for air and saline water were fixed at 0.15 kg/s and 0.28 kg/s respectively for both HDH systems. Other input parameters for both systems are shown in Table 2. The system’s mathematical model was simulated using gPROMS ModelBuilder platform for the conditions specified for both of the HDH systems. Simulations results shows that fresh water productivity for water heated system is 28 kg/h; higher than air heated system which was 15 kg/h (Figure 3).

The effect of temperature on system’s productivity is shown in Figure 3 for both water heated system and air heated system. The figure shows a significant increase of the fresh water productivity (i.e from 21 to 42 kg/h) with increase of the inlet water temperature for the water heated HDH system. On the other hand, the figure shows that, an increase of the air inlet temperature does not have significant influence on fresh water productivity. This difference of response can be attributed to the huge difference of water and air specific heat capacities which are 4.2 and 1.0 kJkg⁻¹K⁻¹ respectively. For the case of water heated system, temperature of the air stream rises as the air ascends the column. Previously, it was demonstrated by Bergman, et al (2011) that, for longer column the air temperature keeps on rising to approach the inlet water temperature. Consequently, the highest vapour carrying capacity (humidity ratio) is attained by the humid air exiting the humidifier. This is different for the air heated system whereby; the hot air stream is cooled rapidly on ascending the column. Since the heat capacity of water is 4 times higher than that of air, the water temperature is only slightly affected thus causing more condensation than evaporation as air approaches to the column’s exit

(Bergman, et al., 2011 and Sinnott, 2005). For this reason, the longer the column, the lower humidity ratio for air stream exiting the column.

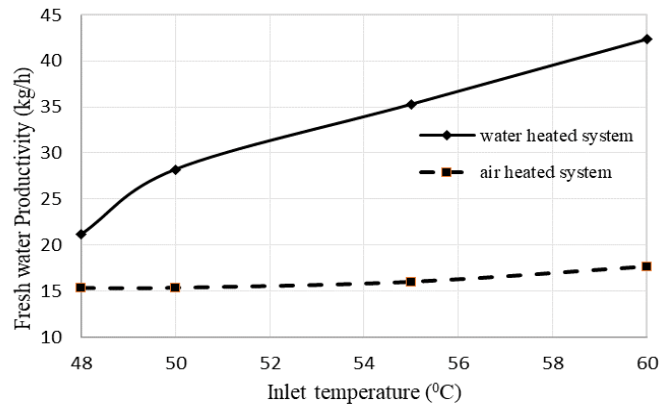


Figure 3: The effect of temperature on productivity for the water heated system and air heated system

Therefore, as recommended by (Mistry, et al., 2011) the fresh water productivity for the water heated HDH system can be maximized by increasing length of the humidifier. For the air heated system (Sharqawy, et al., 2014) recommended to reduce the length of the humidifier and/or to increase the air flow rate in order to maximise the productivity.

4. Conclusions

In this study we have developed and simulated the mathematical model of the Humidification Dehumidification desalination process. The model was used to compare performance of the water heated with the air heated HDH desalination system. Simulations results shows that, the water heated system has higher fresh water productivity than the air heated system. It was also found that the increase of water inlet temperature for the water heated system is associated with a significant increase of the fresh water productivity. A varying inlet temperature from 48 °C to 60 °C corresponded with an increase of productivity from 21 kg/h to 42 kg/h. On the other hand, the increase of air inlet temperature does not have a significant influence on the fresh water productivity for the air heated system provided that other parameters of the system are not varied. For this system, the same temperature variation raised the water productivity from 15 kg/h to 18 kg/h. Finally, it is recommended to boost the productivity by increasing length of the humidifier for the water heated HDH system and by reducing length of the humidifier and/or to increase the air flow rate for the air heated system.

References

- Al-Enezi, G., Hisham, E. & Nagla, F., 2006. Low temperature humidification dehumidification desalination process. *Energy Conversion and Management*, Volume 47, p. 470–484.
- Bergman, T., Lavine, A., Incropera, F. & Dewitt, D., 2011. *Fundamentals of Heat and Mass Transfer*. Jefferson City: John Wiley & Sons, Inc.

- Huifang, K. et al., 2014. Performance of a two-stage multi-effect desalination system based on humidification–dehumidification process. *Desalination*, Volume 344, p. 339–349.
- Kucera, J., 2014. *Desalination*. 1st ed. Massachusetts: Scrivener Publishing LLC.
- Mistry, K., Mitsos, A. & Lienhard, J., 2011. Optimal operating conditions and configurations for humidification–dehumidification desalination cycles. *International Journal of Thermal Sciences*, Volume 50, p. 779–789.
- Mohamed, A. & El-Minshawy, N., 2011. Theoretical investigation of solar humidification–dehumidification desalination system using parabolic trough concentrators. *Energy Conversion and Management*, Volume 52, p. 3112–3119.
- Moran, M., Shapiro, H., Boettner, D. & Bailey, M., 2011. *Fundamental of engineering thermodynamics*. 7 ed. Hoboken, NJ: John Wiley & Sons, Inc.
- Sharqawy, M., Antar, M., Zubair, S. & Elbashir, A., 2014. Optimum thermal design of humidification dehumidification desalination systems. *Desalination*, Volume 349, p. 10–21.
- Sinnot, R., 2005. *Chemical Engineering Design*. Fourth edition. Oxford: Elsevier Butterworth-Heinemann.

Enhancement of energy saving of reverse osmosis system of Arab Potash Company via a wind energy system

Alanood A. Alsarayreh ^a, Mudhar. A. Al-Obaidi ^b, Raj. Patel ^a, and Iqbal M. Mujtaba ^a

^a *Department of Chemical of Engineering, Faculty of Engineering and Informatics, University of Bradford, Bradford, West Yorkshire, BD7 1DP, UK*

^b *Middle Technical University, Technical Institute of Baquba, Dayala – Iraq*
I.M.Mujtaba@bradford.ac.uk

Abstract

The Reverse Osmosis (RO) system has confirmed its viability to produce high quality water from brackish water sources. However, the increase in water and electricity demands and has raised the necessity of a sustainable and integrated water treatment systems. Renewable sources of energy can be applied to reduce the environmental footprint and satisfy the energy demand. An integrated system of wind energy and large-scale RO system has been proposed in this paper to generate freshwater from brackish water. A generic model of the multistage multi pass RO system of Arab Potash Company (APC) was dedicated with a set of algebraic equations for wind turbine operation to form an accurate model to compute different performance indicators including total energy consumption. The energy analysis was carried out for each individual and the integrated systems and estimating the total generated power from the wind turbine for one year (July/2019 to July/2020). In this regard, the power for the high-pressure pumps of the RO system has been supplied from the grid and the wind energy system. The results show that the integrated wind – RO system attain a maximum energy saving during the windy months. Overall, the wind turbine system has enhanced the reduction of total required power of RO system compared to a grid powered RO system.

Keywords: Desalination; Brackish water; Reverse Osmosis; Wind turbine; Energy saving.

1 Introduction

In the last few years, membrane based separation technology has dominated the market, and consequently the Reverse Osmosis (RO) membrane technique has been increasingly implemented in industrial water treatment plants due to its better performance in terms of less energy consumption, higher water recovery, and low operational costs compared to thermal desalination processes (Qasim et al., 2019). However, several methods were implemented to reduce the energy consumption of RO process including the use of a high permeable membrane, energy recovery device, chemical demineralization, optimisation techniques, and renewable energy sources (Li, 2011).

Owing to the high prices of fossil fuel energy coupled with air pollution and global warming, it is necessary to supplement a non-renewable energy (NRE) sources (fossil fuel) with a renewable energy (RE) sources such as wind turbines and photovoltaic cells. (Nassrullah et al., 2020). Charcosset (2009) demonstrated the viability of wind turbine systems as a valuable and sustainable energy resource to be integrated with the RO process especially for coastal areas which have a high availability of wind. Wind energy offers many environmental benefits compared to conventional energy sources and a significant decrease of use of fossil fuels was noticed due to the employment of renewable energy systems. Several studies can be found in the literature that have investigated the efficiency of wind turbine in a hybrid of wind turbine and RO system. For example, Miranda and Infield (2003) tested a 2.2 kW small wind turbine to drive a RO desalination unit. A computer model was developed in order to find optimum wind turbine power that yields maximum fresh water production rate. The model has been verified and supported with experimental data. The results showed that increasing wind speed up to 11 m/s resulted in a higher water production rate. Moreover, the results showed that the specific energy consumption of RO desalination system decreased sharply as wind speed increased up to 6 m/s. Moreover, Forstmeier et al. (2007) studied a wind powered system with RO desalination system with an electrical grid connection. The results confirmed the feasibility of the hybrid system if some grid connection is available to supply energy at low specific energy consumption.

This study focuses on the development of a hybrid system of wind turbine and multistage multi-pass brackish water RO system of the APC located in Jordan. The viability of this hybrid system is tested for a selected set of inlet conditions of brackish water. To attain this goal, model simulations will be carried out to determine the total energy saving of the hybrid system. In this regard, the total number of wind turbines to appropriately provide the power of RO system will be studied for each month from July 2019 to July 2020.

2 Description of RO desalination plant powered by wind turbine system

Fig. 1 shows the layout of brackish water RO powered by wind energy system for the APC plant (capacity 1200 m³/day). It consists of two passes with permeate and retentate reprocessing designs. The 1st pass has two stages of pressure vessels arranged as (4:2). The 2nd pass has three stages of pressure vessels arranged as (2:1:1). The permeate of 1st pass is fed to the 2nd pass for further processing. The high-concentration stream of the 1st pass is sent to drain. However, the low-concentration streams of the 2nd pass are collected to constitute the high-quality water (salinity 2 ppm). The high-concentration stream of the 2nd pass is reprocessed back to the raw feed water of the 1st pass. Several wind turbines are integrated with RO system to supply the required amount of power to drive the pumps as shown in Fig. 1.

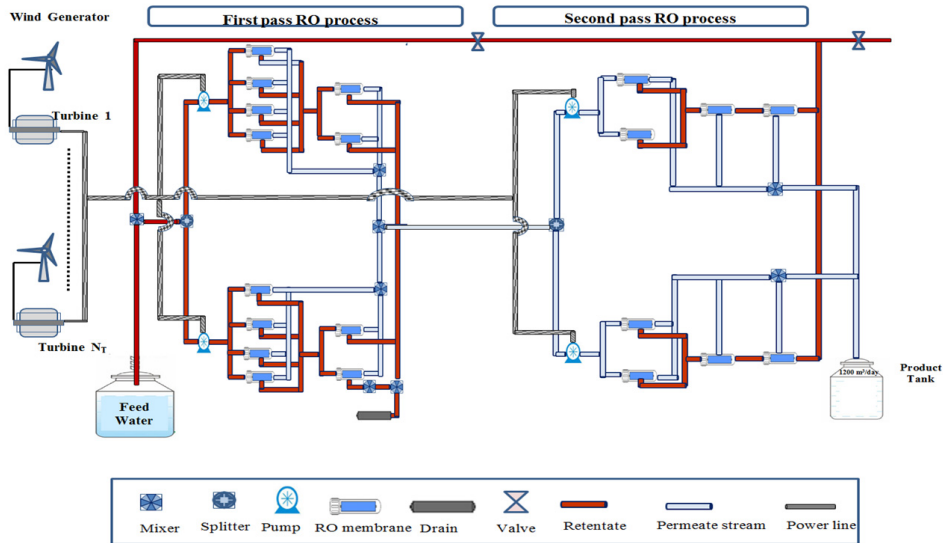


Fig. 1. Layout of brackish water RO plant of APC powered by wind turbine system

3 Modelling of RO system of APC

Al-Obaidi et al. (2018) established a steady-state model for a medium-scale brackish water RO desalination system of APC. The detailed model was validated against actual data collected from the APC and then used to study the plant performance with variable operating conditions. The mathematical model parameters and equations of RO system of APC are shown in Table 1. Recently, this model was upgraded by including a specific sub model to measure the power generated from the wind turbine as depicted in the next section.

Table 1. Mathematical model of RO system of APC

No	Model equations	Specifications
1	$Q_p = A_{w(T)} N D P_{fb} A_m$	Total water flux (m ³ /s)
2	$A_{w(T)} = A_{w(25\text{ }^\circ\text{C})} T C F_p F_f$	Water permeability constant at 25 °C (m/s atm)
3	$T C F_p = \exp \exp [0.0343 (T - 25)] < 25\text{ }^\circ\text{C}$ $T C F_p = \exp \exp [0.0307 (T - 25)] > 25\text{ }^\circ\text{C}$	Temperature correction factor of permeate
4	$N D P_{fb} = P_{fb} - P_p - \pi_b + \pi_p$	The driving pressure (atm)
5	$P_{fb} = P_f - \frac{\Delta P_{drop,E}}{2}$	The feed brine pressure (atm)
6	$\Delta P_{drop,E} = \frac{9.8692 \times 10^{-6} A^* \rho_b U_b^2 L}{2 d_h Re_b^n}$	The pressure drop along the membrane element (atm)
7	$\pi_b = 0.7994 C_b [1 + 0.003 (T - 25)]$ $\pi_p = 0.7994 C_p [1 + 0.003 (T - 25)]$	The bulk and permeate osmotic pressure (atm)
8	$C_b = \frac{C_f + C_r}{2}$	The bulk salinity (kg/m ³)
9	$Q_s = B_{s(T)} (C_w - C_p)$	The solute flux through the membrane (kg/m ² s)
10	$B_{s(T)} = B_{s(25\text{ }^\circ\text{C})} T C F_s$	The solute transport parameter at operating temperature (m/s)
11	$C_w = C_p + \left(\frac{C_f + C_r}{2} - C_p \right) \exp \exp \left(\frac{Q_p / A_m}{k} \right)$	The concentration at the membrane surface (kg/m ³)
12	$k = 0.664 k_{dc} Re_b^{0.5} Sc^{0.33} \left(\frac{D_b}{d_h} \right) \left(\frac{2 d_h}{L_f} \right)^{0.5}$	Mass transfer coefficient (dimensionless)
13	$Sc = \frac{\mu_b}{\rho_b D_b}$	Schmidt number (dimensionless)
14	$Re = \frac{\rho d_e J_w}{\mu}$	Reynolds number (dimensionless)
15	$Rec = \frac{Q_p}{Q_f} = \frac{(C_r - C_f)}{(C_r - C_p)}$	Total recovery (dimensionless)
16	$Rej = \frac{C_f - C_p}{C_f}$	Observed rejection (dimensionless)
17	$J_w = \frac{B_{s(T)} Rej_{real}}{(1 - Rej)}$	The water flux (m/s)
18	$C_p = \frac{C_f}{Rec} [1 - (1 - Rec)]^{(1 - Rej)}$	The average permeate salinity at the permeate channel (kg/m ³)
19	$C_r = C_f [1 - Rec]^{-Rej}$	The average retentate salinity at the permeate channel (kg/m ³)
20	$E1 = \frac{P_f(in)(plant) * 101325 * Q_f(Raw\ water)}{Q_p(plant) * \epsilon * p_{unp}} + \frac{P_f(Block\ 3) * 101325 * Q_f(Block\ 3)}{Q_p(plant) * \epsilon * p_{unp}}$ $\frac{36 \times 10^5}{36 \times 10^5}$	The total plant energy consumption (kWh/m ³)

4 Modelling of wind turbine system

The power generated from the wind turbine, measured in kW and based on the value of wind speed, can be calculated by the following equations;

$$Power_g(u) = 0 \quad u \leq u_c \text{ and } u \geq u_f \quad (21)$$

$$Power_g(u) = P_{rated} \left(\frac{u^k - u_c^k}{u_r^k - u_c^k} \right) \quad u_c \leq u \leq u_r \quad (22)$$

$$Power_g(u) = P_{rated} \quad u_r \leq u \leq u_f \quad (23)$$

Where P_{rated} is the rated wind turbine power (kW), u the wind speed (m/s), u_c the wind turbine cut-in (m/s), u_r the wind turbine rated (m/s), and u_f the wind turbine cut-off speed (m/s).

5 Evaluation of a hybrid system of RO and wind turbine

This section focuses on evaluating the feasibility of wind energy system to be integrated with the RO desalination system. In this regard, the necessary power to drive the pumps that satisfies the production of freshwater at a specified set of feed conditions will be outlined. More importantly, this research will determine the precise number of turbines required to supply the necessary power of the RO system. The feed characteristics of RO plant are 1098.62 ppm, 74 m³/h, 25 °C and 9.22 atm of feed water.

The simulation-based model developed by Al-Obaidi et al. (2018) based on the selected feed conditions showed that the total energy consumption of their RO system was 0.8371 kWh/m³ for water productivity rate of 48.6 m³/h. Therefore, the power required to operate the RO plant is 40.7 kW. Here, the Hummer H25 wind turbine is selected to be integrated with the RO system. Table 2 shows the features of wind turbine.

The speed values for one year (July/2019 to July/2020) of the specified location of RO system of APC were evaluated and presented in Table 3 with the associated power generated from a single wind turbine. This in turn aids to determine the number of turbines needed to run the RO plant of APC. The total numbers of turbines wind are determined through numerical simulation technique by using the gPROMS software.

Table 2. The wind turbine data (Matysik, 2020)

Type of wind turbine	Rated Power (kW)	u_c (m/s)	u_r (m/s)	u_f (m/s)	K
Hummer H25	100	2.5	10	20	2.5649

Table 3. Simulation results of RO system powered by wind energy system

Month	Mean wind speed (m/s)	Power generated from a single wind turbine (kW)	Number of wind turbines
July/2019	2.744	0.793	51
August/2019	3.187	2.541	16
September/2019	3.125	2.271	17
October/2019	3.126	2.275	18
November/2019	3.426	3.657	11
December/2019	8.822	71.699	1
January/2020	7.88	52.931	1
February/2020	6.852	36.097	1
March/2020	5.61	20.432	2
April/2020	4.4535	9.988	4
May/2020	4.197	8.163	5
June/2020	4.282	8.749	5
July/2020	3.083	2.093	19

The simulation results of Table 3 show that the highest number of turbines required to power the RO desalination plant of APC is 51 in July /2019 at the lowest wind speed of

2.744 m/s. It should be noted that this number of turbines is determined based on the power generated from a single wind turbine. Therefore, it is important to construct a farm of 51 wind turbines to satisfy the requirements of energy. However, only one wind turbine is required for the windy winter season. Thus, the excess electrical power of windy season can be used to feed the internal grid and provides a sustainable source of energy.

5 Conclusions

Nowadays, the use of wind turbine as an alternative source of energy in the RO desalination appears as a reasonable and technically mature option towards the reduction of the total required power. The feasibility of integrating the wind turbine with the RO desalination system of APC has been examined in this study. The study has confirmed the viability of this hybrid system as only one wind turbine is sufficient to provide the required power of RO desalination system of APC. However, a farm of 51 wind turbines is necessary to drive the RO system in low wind speed seasons.

References

- Charcosset, C. (2009) A review of membrane processes and renewable energies for desalination. *Desalination* 245 (1-3), 214-231.
- Forstmeier, M., Mannerheim, F., D'Amato, F., Shah, M., Liu, Y., Baldea, M. and Stella, A. (2007) Feasibility study on wind-powered desalination. *Desalination* 203 (1-3), 463-470.
- Matysik, L. B. S. (2020) wind-turbine-models. <https://en.wind-turbine-models.com/turbines/1682-hummer-h25.0-100kw> Accessed 01/10/2020.
- Miranda, M. S. and Infield, D. (2003) A wind-powered seawater reverse-osmosis system without batteries. *Desalination* 153 (1-3), 9-16.
- Nassrullah, H., Anis, S. F., Hashaikeh, R. and Hilal, N. (2020) Energy for desalination: A state-of-the-art review. *Desalination* 491, 114569.
- Qasim, M., Badrelzaman, M., Darwish, N. N., Darwish, N. A. and Hilal, N. (2019) Reverse osmosis desalination: A state-of-the-art review. *Desalination* 459, 59-104.

Multi-objective optimization for the incorporation of safety and reliability considerations in process design

Andrea P. Ortiz-Espinoza^a, Yixin Ye^b, Ignacio E. Grossmann^{b,*}, Arturo Jiménez-Gutiérrez^{a,*}

^a*Chemical Engineering Department, Instituto Tecnológico de Celaya, Celaya, Gto. 38010, México*

^b*Chemical Engineering Department, Center for Advanced Process Decision-making, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

grossmann@cmu.edu; arturo@iqcelaya.itc.mx

Abstract

In the traditional process design approach, aspects such as safety and reliability are typically left for analysis once the design phase is completed. Recently, methods have been developed to include inherent safety in the early design stages. This approach allows the generation of inherently safer designs as part of the process synthesis task. The design of reliable chemical processing plants can be enhanced by adding back-up units to prevent shutting down the plant due to failure of units. In this work, we present a multi-objective optimization model for the design of chemical process systems accounting for safety, reliability, and economics. The model is formulated within a generalized disjunctive programming framework, which is then reformulated as a mixed-integer nonlinear programming model using the ϵ -constrained method to generate pareto curves. The model includes the selection of standby units to increase the system availability and provides the optimal structure of the process flowsheet and operating parameters. The model is applied to the design of a distillation system accounting for risks due to explosion, while rating the system reliability and the process economics. The results show that the proposed optimization model yields design alternatives with optimal trade-offs among safety, reliability, and economics.

Keywords: safety assessment, reliability, multi-objective optimization.

1. Introduction

In the conventional process design safety assessment is typically left for analysis once the design phase is completed. Recently, as part of the process synthesis task, safety has been considered with a focus on generating inherently safer designs. For example, Medina-Herrera et al. (2014) presented an approach for the design of inherently safer distillation systems. In that work, quantitative risk assessment (QRA) techniques were used to incorporate safety considerations in the design of different distillation schemes. Ortiz-Espinoza et al. (2017) proposed a multi-objective assessment approach to evaluate chemical processes accounting for safety, sustainability, and economics. Within this approach, modifications to the processes were considered to increase safety performance while evaluating their impact on the remaining objectives. Nemet et al. (2018) presented

a mixed-integer nonlinear programming (MINLP) model for the synthesis of a methanol production process considering risk assessment during the synthesis stage. These works showed the conflict that generally exists between safety and economics. In this regard, Guillen-Cuevas et al. (2018) introduced a metric to reconcile safety and sustainability with economic aspects using a weighted return on investment type of metric.

Another aspect that is important to consider during process design is that of reliability. The reliability of a chemical processing plant can be enhanced by adding back-up units to prevent shutting down of the plant due to the failure of units. Recent works have addressed this topic from different perspectives. For instance, Ye et al. (2018) proposed MINLP models to determine the selection of parallel units considering the trade-off between availability and cost. Ade et al. (2018) studied the effect of inherent safety principles on the reliability of process plants with a focus on risk migration. Moreno-Sader et al. (2019) proposed an approach to incorporate safety, sustainability, reliability, and resilience during conceptual design stages.

Although safety and reliability of chemical plants are closely related, models that explicitly capture the interactions of both aspects in process synthesis and optimization have not been formally addressed before. This work presents a multi-objective optimization model for the design of chemical processes accounting for safety, reliability, and economics. The proposed optimization model includes the selection of parallel standby units for critical process stages to increase the system availability, together with the computation of frequency and consequences of possible catastrophic events to estimate risk levels. The model also considers decisions regarding selecting the structure of the process flowsheet and operating parameters.

The trade-offs among the objectives arise as the selection of more standby units benefits the availability of the systems but increases investment costs. Regarding safety, the presence of more units may increase the exposure to the risk produced by the additional units; on the other hand, increasing availability reduces the risk due to transition states such as start-ups and shut-downs. Furthermore, the selection of the flowsheet structure and operating parameters is a factor in which risk and economics may be in conflict. While extreme operating conditions may favour the economics of the process, operating under such conditions increases the consequences of the outcomes of hazardous events. The purpose of the proposed multi-objective optimization model is to synthesize process alternatives with optimal trade-offs among safety, reliability, and economics.

2. Problem Statement

Given a superstructure for a production process/system, a multi-objective modelling framework is presented to account for safety, reliability, and economics. The purpose of the model is to evaluate design alternatives that minimize risk, maximize availability, and minimize cost. For modelling purposes, the process under study is divided into K stages (e.g., compression, heating, cooling, etc.) for which design parameters and operating conditions are to be determined. Also, for certain critical stages (K_{red}), the option to install standby units is considered. When the process under study considers the selection of different flowsheet alternatives, the model can be formulated as a Generalized Disjunctive Programming (GDP) model (Grossmann and Trespalacios, 2009) of the form shown in Eq. (MO-RAC) involving Boolean variables Y_k , 0-1 variables for selecting the units in the superstructure, 0-1 variables y for selecting back-up units, and continuous variables x , c for the state, design and cost variables of the corresponding flowsheet.

$$\begin{aligned}
 & \min_{x,y} \text{Risk} \\
 & \max_y \text{Availability} \\
 & \min_{x,y} \text{Cost} \\
 \text{s.t.} \quad & g(x) \leq 0 \\
 & \left[\begin{array}{c} Y_k \\ h_k(x) \leq 0 \\ c_k = \alpha_k \\ R_k = f_R(x) \end{array} \right] \vee \left[\begin{array}{c} \neg Y_k \\ B^k x = 0 \\ c_k = 0 \\ R_k = 0 \end{array} \right] \forall k \in K \setminus K_{red} \\
 & \left[\begin{array}{c} Y_k \\ h_k(x) \leq 0 \\ c_k = \alpha_k y \\ AV_k = f_A(y) \\ R_k = f_R(x, y) \end{array} \right] \vee \left[\begin{array}{c} \neg Y_k \\ B^k x = 0 \\ c_k = 0 \\ AV_k = 1 \\ R_k = 0 \end{array} \right] \forall k \in K_{red} \quad (\text{MO-RAC}) \\
 & \Omega(Y) = \text{True} \\
 & \text{Risk} = \sum_{k \in K} R_k \\
 & \text{Availability} = A = \prod_{k \in K_{red}} AV_k \\
 & \text{Cost} = \sum_{k \in K} c_k \\
 & x \in R^n, c \geq 0, Y \in \{\text{True}, \text{False}\}^m, y \in \{0,1\}
 \end{aligned}$$

In the next section we present the equations for evaluating the availability AV_k and risk R_k that pertain to the reliability and safety of the selected flowsheet at each stage k .

3. Model formulation for reliability and safety

To increase the reliability of the process, the GDP model considers the installation of a number of potential standby units (J) in certain stages (K_{red}). The evaluation of the availability of each stage is computed through the probabilities of each unit to be available (p) (Ye et al., 2018). Eq. (1) describes the selection of at least one unit per stage, Eq. (2) is a symmetry breaking constraint for stages with identical parallel units (K_{ide}), Eq. (3) evaluates the availability of stages with only identical standby units, while the evaluation of the availability of stages with nonidentical units (K_{non}) is shown in Eq. (4). The system availability is calculated using Eq. (5).

$$\sum_{j=1}^{n_k} y_{k,j} \geq 1, \quad \forall k \in K_{red} \quad (1)$$

$$y_{k,j+1} \leq y_{k,j}, \quad \forall j \in J_k, \forall k \in K_{ide} \quad (2)$$

$$P_k = p_k \sum_{j=1}^{n_k} y_{k,j} (1 - p_k)^{j-1}, \quad \forall k \in K_{ide} \quad (3)$$

$$P_k = 1 - \prod_{j \in J_k} (1 - p_{k,j} y_{k,j}), \quad \forall k \in K_{non} \quad (4)$$

$$A = \prod_{k \in K} P_k \quad (5)$$

The cost of the system is determined considering installation and repair costs. Eq. (6) and (7) represent the cost per stage and the total cost of the system.

$$C_k = \sum_{j=1}^{n_k} y_{k,j} (c_{k,j}^{inst} + c_{k,j}^{rep}), \quad \forall k \in K_{red} \quad (6)$$

$$Cost = \sum_{k \in K} C_k \quad (7)$$

The safety performance of the processes is evaluated using QRA techniques. Within this context, risk is defined as a function of the frequency (freq) and consequences (DR) of a hazardous event. Eq. (8) shows the calculation of risk for each unit j in a stage k . The frequency element considers the failure frequency for each unit j as well as the probability of a certain hazardous event to take place (p^{haz} , p^{exp}) as shown in Eq. (9) and (10). Risk is computed per stage as shown in Eq. (11), and the risk due to transient states is calculated in Eq. (12). Finally, Eq. (13) shows how to estimate the total risk in the plant.

$$Risk_{k,j}^{Exp} = freq_{k,j}^{Exp} DR_k^{Exp}, \quad \forall k \in K, \quad \forall j \in J_k \quad (8)$$

$$freq_{k,j}^{Exp} = \lambda_k p_k^{hazf} p^{Exp} y_{k,j} (1 - p_k^i)^{j-1}, \quad \forall k \in K_{ide}, \quad \forall j \in J_k \quad (9)$$

$$freq_{k,j}^{Exp} = \lambda_{k,j} p^{haz} p^{Exp} y_{k,j} \prod_{j=1}^{j-1} (1 - p_{k,j} y_{k,j}), \quad \forall k \in K_{non}, \quad \forall j \in J_k \quad (10)$$

$$Risk_k^{Exp} = \sum_{j \in J_k} Risk_{k,j}^{Exp}, \quad \forall k \in K \quad (11)$$

$$Risk_k^{ExpT} = (-\ln A) p^{haz} p^{exp} DR_k^{Exp} \quad (12)$$

$$Risk = \sum_{k \in K} (Risk_k^{Exp} + Risk_k^{ExpT}) \quad (13)$$

The estimation of consequences is measured in this work as the damage radius resulting from an explosion event (DR). Eq. (14) shows how such a radius depends on the hazardous potential of the explosion. This hazardous potential is estimated according to energy factors and penalties. For more details on the procedure to estimate the hazardous potential the reader is referred to Rathnayaka et al. (2014).

$$DR_k^{Exp} = 0.9 (HazPot_k^{Exp})^{\frac{1}{3}} \quad (14)$$

4. Case study

To illustrate the application of the proposed multiobjective GDP model (MO-RAC), a simplified case study is considered with fixed configuration of a distillation column to separate hexane and heptane. Figure 1 shows the distillation scheme, which is based on the case proposed in (CCPS, 2000). In this case study, the installation of standby units in the condenser and reflux pump stages is considered to increase the system availability. The safety assessment considers scenarios of continuous releases of flammable material, thus leading to potential explosion incidents. Table 1 shows the parameters used in the model for the stages with potential parallel standby units.

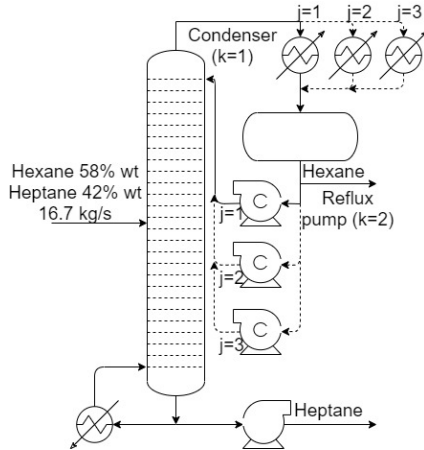


Figure 1. Distillation scheme

Table 1. Parameters for the case study

	Availability		Failure rate (fails/yr)	
	k=1	k=2	k=1	k=2
j=1	0.94	0.88	0.064	0.127
j=2	0.94	0.89	0.064	0.116
j=3	0.94	0.90	0.064	0.105
	Installation cost (k\$/yr)		Repair cost (k\$/yr)	
	k=1	k=2	k=1	k=2
j=1	$0.65(A)^{0.6}$	$26.0(\text{flow})^{0.54}$	2	3
j=2	$0.65(A)^{0.6}$	$26.8(\text{flow})^{0.54}$	2	4
j=3	$0.65(A)^{0.6}$	$27.3(\text{flow})^{0.54}$	2	5

As part of the design the model determines the reflux ratio, the pressure and temperature at the condenser, and the number of stages. The evaluation ranges of such variables are: 2 to 30 for the reflux ratio, 3 to 10 bar for pressure, 100 to 300 °C for temperatures, and 5 to 25 for number of stages.

5. Results

Figure 2 shows the Pareto curve for the multi-objective GDP optimization problem. Each point corresponds to one subproblem solved to minimize risk subject to a certain upper bound of total cost. The availability objective is implicitly considered in the optimization of risk as higher availability levels generally lead to lower risk levels (see Eq. (12)). The estimated availability for each point is shown in Figure 2. Figure 3 shows details for the following solutions: a) minimum risk (maximum cost and availability), b) maximum risk (minimum cost and availability), c) and d) intermediate scenarios. It is observed that the minimum risk solution operates at the lowest possible pressure with the installation of every possible standby unit. On the other hand, the minimum cost solution operates at the highest operating pressure considered for the column, with the installation of just one unit in each stage. The multiobjective GDP model was reformulated as an MINLP using the ϵ -constrained formulation to find the Pareto curve. The model was implemented in GAMS and solved using the commercial solver BARON requiring a total of 112.7 CPUs.

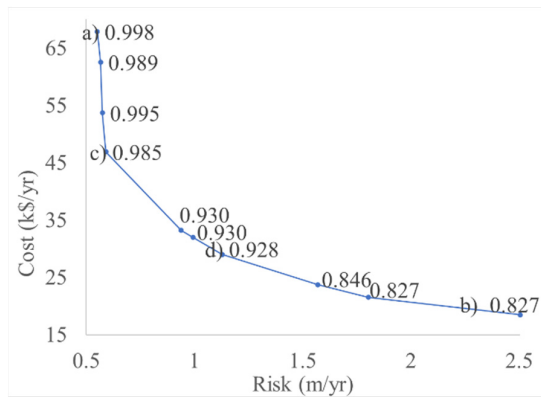


Figure 2. Pareto curve of cost vs risk for different levels of availability

6. Conclusions

A multi-objective GDP optimization model for the incorporation of safety and reliability in the design of chemical processes has been presented. The model considers the selection of standby units to increase the system availability and design decisions that impact safety performance. The application to the design of a binary distillation system showed how designs are obtained with optimal trade-offs among safety, reliability, and economics. Future work will report the application of the proposed model in the synthesis of a methanol process.

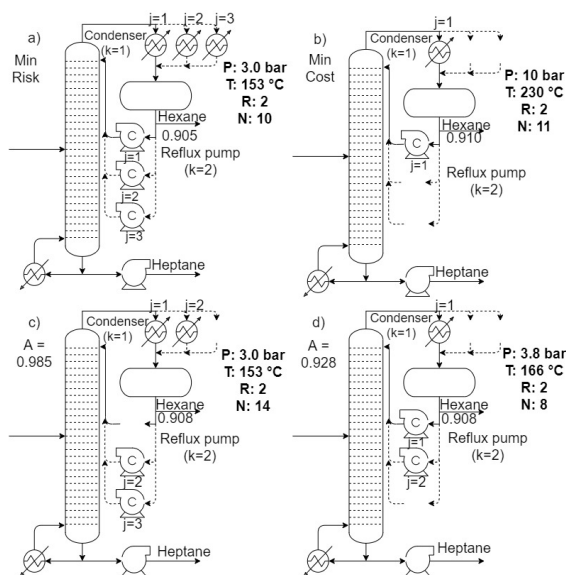


Figure 3. Details for solutions a, b, c, and d

References

- N. Ade, G. Liu, A.F. Al-Douri, M.M. El-Halwagi, M.S. Mannan, 2018, Investigating the effect of inherent safety principles on system reliability in process design, *Process Saf. Environ.*, 117, 100-110.
- AIChE, 2000, *Guidelines for chemical process quantitative risk analysis*, 2nd Ed, New York: AIChE.
- Grossmann, I.E. and F. Trespalcacios, "Systematic Modeling of Discrete-Continuous Optimization Models through Generalized Disjunctive Programming," *AIChE J.* **59**, 3276-3295 (2013).
- K. Guillen-Cuevas, A.P. Ortiz-Espinoza, E. Ozinan, A. Jiménez-Gutiérrez, N. Kazantzis, M.M. El-Halwagi, 2018, Incorporation of safety and sustainability in conceptual design via a return on investment metric, *ACS Sustain. Chem. Eng.*, 6, 1, 1411-1416.
- N. Medina-Herrera, A. Jiménez-Gutiérrez, M.S. Mannan, 2014, Development of inherently safer distillation systems, *J. Loss Prevent. Proc.*, 29, 225-239.
- K. Moreno-Sader, P. Jain, L.C. Ballesteros Tenorio, M.S. Mannan, M.M. El-Halwagi, 2019, Integrated approach of safety, sustainability, reliability, and resilience analysis via a return on investment metric, *ACS Sustain. Chem. Eng.*, 7, 19522-19536.
- A. Nemet, J.J. Klemeš, Z. Kravanja, 2018, Process synthesis with simultaneous consideration of inherent safety-inherent risk footprint, *Front. Chem. Sci. Eng.*, 12, 4, 745-762.
- A.P. Ortiz-Espinoza, A. Jiménez-Gutiérrez, M.M. El-Halwagi, 2017, Including inherent safety in the design of chemical processes, *Ind. Eng. Chem. Res.*, 56, 49, 14507-14517.
- S. Rathnayaka, F. Khan, P. Amyotte, 2014, Risk-based process plant design considering inherent safety, *Safety Sci.*, 70, 438-464.
- Y. Ye, I.E. Grossmann, J.M. Pinto, 2018, Mixed-integer nonlinear programming models for design of reliable chemical plants, *Comput. Chem. Eng.*, 116, 3-16.

Integrating Reliability and Uncertainty in Process Synthesis

Ying Chen,^a Yixin Ye,^b Ignacio E. Grossmann,^{b*} Bingzhen Chen^{a*}

^a *Department of Chemical Engineering, Tsinghua University, Beijing 100084, China*

^b *Center for Advanced Process Decision-making, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

grossmann@cmu.edu

Abstract

Plant availability and operating uncertainties have been critical considerations for the design and operation of chemical processes as they directly impact service level and economic performance. This paper proposes a two-stage stochastic Generalized Disjunctive Programming (GDP) model with reliability constraints to deal with both the exogenous and endogenous uncertainties in process synthesis, where the reliability model is incorporated into the flowsheet superstructure optimization. The proposed stochastic programming model anticipates the market uncertainties through scenarios for selecting the optimal flowsheet topology, equipment sizes and operating conditions, while considering the impact of potential parallel units for improving plant availability. An improved logic-based outer-approximation algorithm is applied to solve the resulting large-scale GDP model, which effectively avoids numerical difficulties with zero flows and provides high quality design solutions. The applicability of the proposed modeling framework and the efficiency of solution strategy is illustrated with an industrial methanol synthesis process.

Keywords: reliability-based superstructure optimization, stochastic programming, endogenous and exogenous uncertainties, logic-based outer approximation algorithm

1. Introduction

Process synthesis is the assembly and interconnection of units into a process network, involving different physical and chemical phenomena to transform raw material and energy inputs into desired products with the goal of optimizing a given objective function (Chen and Grossmann, 2017). Generalized Disjunctive Programming (GDP) is a modeling framework to explicitly represent the relationship between algebraic descriptions and the logical structure of a design problem. It is therefore well-suited to problems involving selection among discrete process alternatives with nonlinear process models (Grossmann and Trespalacios, 2013). Synthesis of process flowsheets are subjected to uncertainties, which directly impact its service level and economic performance. There are two types of uncertainties in process synthesis: exogenous, where the true parameter values are revealed independently of decisions, and endogenous, where the parameter realizations are influenced by the decisions taken (Apap and Grossmann, 2016). Exogenous uncertainties correspond typically to market uncertainties, such as product demands, product prices and utility prices. For endogenous uncertainties, decisions can influence the parameter realizations by causing alteration of the probability

distribution, or affecting the time at which we observe these realizations. Reliability-based design optimization (RBDO) captures the endogenous uncertainties from equipment failures, where the selection of redundant equipment, maintenance policy and storage sizing affect the plant availability by altering the probability distributions (Ye et al., 2018). In previous work Straub et al. (1990) and Thomaidis et al. (1994) integrated flexibility (exogenous uncertainties) and reliability in a uniform framework. However, their work only considered a quantitative measure to evaluate the proposed design alternatives, and did not consider the selection of standby units in order to improve the system availability. The major goal of this paper is to propose a novel model that integrates both exogeneous uncertainty through stochastic programming, and endogenous uncertainty through RBDO, for the synthesis of process flowsheets, where the reliability model is incorporated into the superstructure optimization. Application of the proposed model is illustrated with an industrial methanol synthesis process.

2. General model formulation

We propose the following two-stage stochastic GDP model with reliability constraints to deal with the exogenous and endogenous uncertainties in process synthesis:

$$\begin{aligned}
 & Z = \sum_i c_i + A \sum_s p_s f(x_s) \\
 \text{s.t.} \quad & g(x_s) \leq 0, \forall s \in S \tag{1} \\
 & [Y_i h_i(y_i, x_s) \leq 0 \ c_i = c_i^{fix} + c_i^{var} \times y_i + c_i^{repa}] \vee [\neg Y_i \ B^i x_s = 0 \ c_i = 0], \ \forall s \in S, i \in I_R \\
 & \left[Y_i h_i(y_i, x_s) \leq 0 \ z_{i,r+1} \leq z_{i,r} \ c_i = (c_i^{fix} + c_i^{var} \times y_i + c_i^{repa}) \sum_{r=1}^{n_r} z_{i,r} \ A_i \right. \\
 & \quad \left. = p_i \sum_{r=1}^{n_r} z_{i,r} (1 - p_i)^{r-1} \right] \vee [\neg Y_i \ B^i x_s = 0 \ c_i = 0 \ A_i = 1], \ \forall s \in S, i \\
 & \quad \in I_{iden} \\
 & \left[Y_i h_i(y_i, x_s) \leq 0 \ c_i = \sum_{r=1}^{n_r} z_{i,r} (c_{i,r}^{fix} + c_{i,r}^{var} \times y_i + c_{i,r}^{repa}) \ A_i = 1 - \prod_r (1 - p_{i,r} z_{i,r}) \right] \\
 & \quad \vee [\neg Y_i \ B^i x_s = 0 \ c_i = 0 \ A_i = 1], \ \forall s \in S, i \in I_{non} \\
 & \quad Y_i = True, \quad \forall i \notin I^D \\
 & \quad \Omega(Y) = True \\
 & \quad A_{sys} = \prod_{i \in I_{iden} \cup I_{non}} A_i \\
 & \quad y \in R^m, \ c \geq 0, \ Y \in \{True, False\}^r, \ z \in \{0, 1\}^t \\
 & \quad x_s \in R^n, \ \forall s \in S
 \end{aligned}$$

In the two-stage stochastic programming model, the first-stage variables consist of three types of design variables, the Boolean variables Y_i which determine the selection among the different process alternatives, the binary variables $z_{i,r}$ to represent whether to choose

the redundant unit r for the certain equipment i , and the continuous variables y_i related to the equipment sizes, such as reactor volume, number of trays in column, and design capacity of heat exchanger or compressor. The second-stage decisions x_s are related to the operational variables which can be adjusted to different scenarios, such as flowrates, temperatures and pressures. The objective function is to minimize the total cost of the system, including the capital cost c_i and the expected operational expenditure $A \sum_s p_s f(x_s)$. Generalized Disjunctive Programming is applied to explicitly express the logic encapsulated in the superstructure. The reliability model is integrated within disjunctions that define the selection of the units, which in turn define the structure of the selected flowsheet. Single units are given fixed probabilities of being available, and Simple Bayes Rules are used to predict the system availability A_{sys} . If a unit exists, the constraints $h_i(y_i, x_s) \leq 0$ enforce the relevant mass and energy balances, and the constraints $c_i = c_i^{fix} + c_i^{par} \times y_i + c_i^{repa}$ calculates the total cost of the unit. Otherwise, constraints $B^i x_s = 0$ describe port variable relationships when the unit is absent, and the capital cost of the unit is set to 0. The proposed two-stage stochastic programming model anticipates the market uncertainties through scenarios for selecting the optimal flowsheet topology, equipment sizes and operating conditions, while considering the impact of additional redundant units for improving plant reliability.

3. Solution Method

In this work, we propose an improved logic-based outer-approximation (LOA) algorithm to solve the resulting large-scale GDP model. LOA allows the solution of reduced space subproblems, including only constraints in selected disjuncts rather than the full-space representation. Therefore, it can effectively avoid numerical difficulties with zero flows and provides high quality design solutions. It should be noted that, when we consider the reliability, binary variables are introduced to represent the selection of parallel units. Our model becomes a hybrid GDP formulation with implicit “nested disjunctions”, which not only contain the Boolean variables to select the flowsheet structure, but also includes binary variables to select the redundant units for improving reliability. Therefore, when considering the choices of the units in the superstructure (Boolean variables) in the master problem, the reduced space subproblems in the LOA become MINLP subproblems rather than only NLP subproblems. The advanced computational tool GDPopt, provides various implementations for solving GDP problems. As an open-source platform, it incorporates recent innovations in reformulation strategies and logic-based solution algorithms (Chen and Grossmann, 2019).

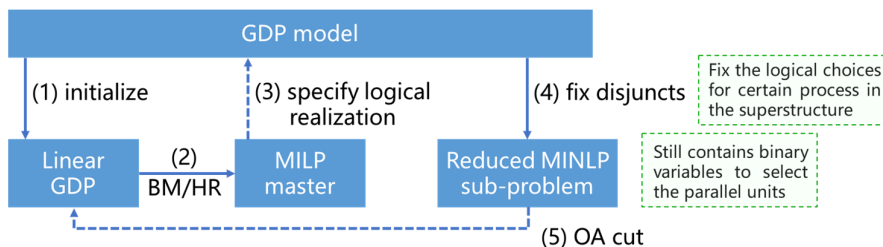


Figure 1. Improved logic-based outer approximation flow diagram for process synthesis considering reliability.

4. Industrial Case Study

In this section an industrial case study which converts syngas to methanol is presented to illustrate the application of the model (1). The methanol synthesis process was formulated by Türkay and Grossmann (1996) without reliability and uncertainty considerations. This includes the discrete decisions between two syngas feeds with different purity and cost, single and two-stage compression for both the feed and recycle streams, as well as the choice between a higher-conversion, higher-cost reactor and a cheaper variant. Our goal is to determine both design- and operational-level decisions in order to minimize the total cost of the system with both the exogenous uncertainties and plant availability taken into consideration. When considering reliability, parallel units are installed in the critical stages. The problem has four explicit disjunctions for the structural choices, involving implicit “nested disjunctions” for the selection of redundant units. The superstructure of methanol synthesis problem is shown in Figure 2.

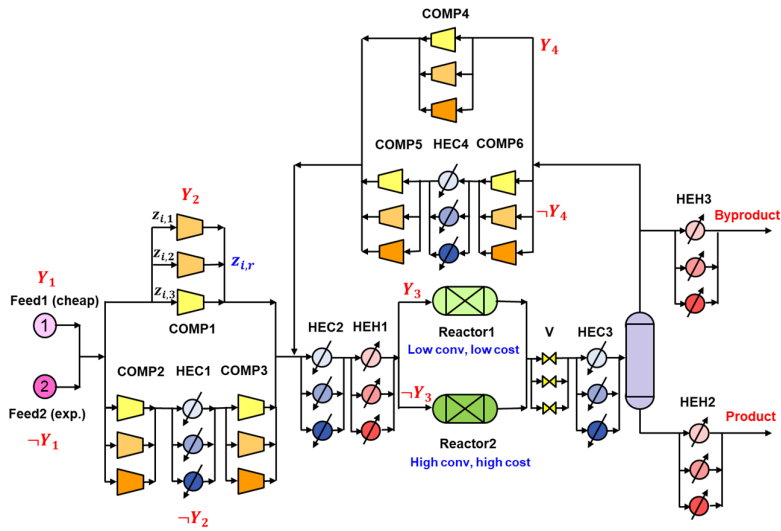


Figure 2. Superstructure of methanol synthesis problem.

The synthesis of the methanol process can be formulated as the two-stage stochastic programming problem. In the first stage the topology of network is selected, and in the second stage the process operation is selected out according to the realization of uncertain parameters. In terms of cost, the first-stage decisions represent the design decisions, including the feed selection, reactor selection, single-stage compression or two-stage compression selection, redundancy selection and design capacity of each unit, which are related to the capital expenditure; the second-stage decisions are the operational decisions, involving flowrates, temperature, pressure, and utility requirement that account for the operational expenditure. The methanol product demand and the electricity price are regarded as exogenous uncertainties. Each of them is modeled with 3 scenarios (low, medium, high) with certain discrete probability distribution, based on the historical data from the changing markets. Therefore, the two-stage stochastic programming has a total of 9 scenarios. Moreover, some critical units are given with fixed failure rates that can be regarded as endogenous uncertainties. The failure of any one of these processing stages

can result in the failure of the entire system, which will compromise its ability to meet customer demands. At each stage three potential back up units are considered.

5. Solution Results

Table 1. Solution results for different models.

Model	# of Cons	# of Cont. Vars	# of Bin. Vars	Strategy and Solver	Solution Time (s)	System Availability	Objective Profit
Deterministic Model	474	307	8	GLOA MILP-Gurobi NLP-Conopt4	35.2	0.9318	2009.16
Stochastic Programming	3986	2491	8	GLOA MILP-Gurobi NLP-Conopt4	322.8	0.9318	2156.04
Integrate Reliability and Uncertainty	4306	2537	58	GLOA MILP-Gurobi NLP-Conopt4 MINLP-Dicopt	446.1	0.9646	2174.96
VSS (\$)		146.88		VSS (%)		7.31 %	
VSS+VRS (\$)		165.80		VSS + VRS (%)		8.25 %	

To illustrate the advantages of the proposed modelling method, we compare the solution results from three different models, the deterministic model with exogenous uncertainties evaluated at their mean values, the stochastic programming model with 9 scenarios to account for exogenous uncertainties, and model (1) which integrates reliability and uncertainty to handle both exogenous and endogenous uncertainties. The first two models are standard GDP models while the last model is a hybrid GDP formulation with the introduction of binary variables to indicate the selection of the redundant units. These three models are coded in Pyomo. The logic-based outer approximation algorithm is applied to solve these three models, where GUROBI is used to solve the MILP master problems, CONOPT4 is used to solve the NLP subproblems and DICOPT is used to solve the MINLP subproblems when considering reliability. The value of stochastic solution is calculated to evaluate the profit that can be expected from implementing the stochastic solution instead of simply using the deterministic solution.

As seen in Table 1 the model that integrates reliability and uncertainty yields the best economic performance compared to the cases when either reliability is not considered, or when exogenous uncertainties are evaluated with mean values. The system availability is increased from 0.9381 to 0.9646 by adding back-up units, while the profit is increased by 8.25 % with improved operational flexibility and reliability. Figures 3-5 present the optimal designs of the corresponding process flowsheets obtained from these three models.

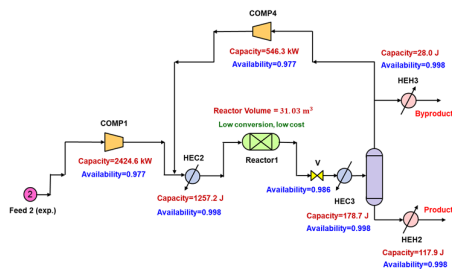


Figure 3. Deterministic Model.

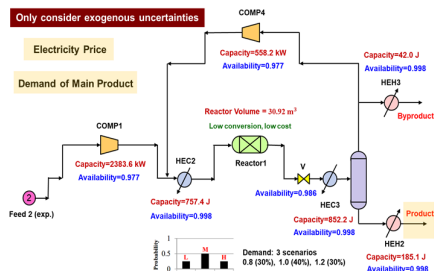


Figure 4. Stochastic Programming Model.

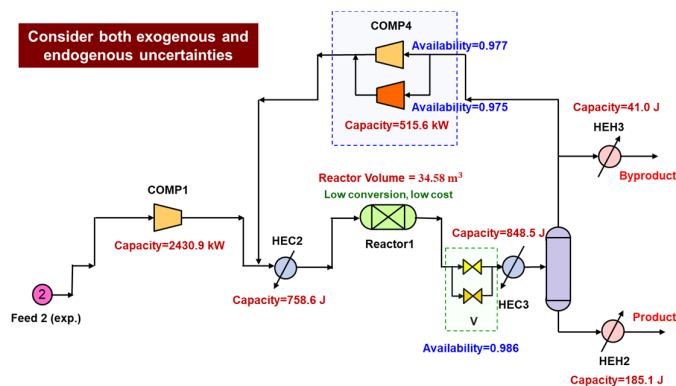


Figure 5. Integrating Reliability and Uncertainty Model.

6. Conclusions

In this paper, we have proposed a two-stage stochastic programming GDP model with reliability constraints to deal with both the exogenous and endogenous uncertainties in process synthesis. Simultaneous optimization of reliability and uncertainty in process design provides potential improvement for the operational flexibility and economic performance. An improved Logic-based Outer Approximation (LOA) algorithm was applied to the hybrid GDP model with implicit nested disjunctions, obtaining optimal solution by avoiding zero-flow singularities. The proposed methodology was illustrated with the synthesis of a methanol process to simultaneously optimize the process flowsheet and the number of stand-by units by considering exogenous uncertainties in product demand and electricity price, and endogenous uncertainties in equipment failures. It was

shown that the proposed design led to simultaneous improvement both in the profit and in the system availability.

References

- I. E. Grossmann, F. Trespalacios, 2013, Systematic Modeling of Discrete-Continuous Optimization Models through Generalized Disjunctive Programming, *AIChE Journal*, 59, 3276-3295.
- Q. Chen, I. E. Grossmann, 2017, Recent Developments and Challenges in Optimization-Based Process Synthesis, *Annual Review of Chemical & Biomolecular Engineering*, 8(1), 249.
- R. M. Apap, I. E. Grossmann, 2016, Models and Computational Strategies for Multistage Stochastic Programming under Endogenous and Exogenous Uncertainties, *Computers & Chemical Engineering*, 103, 233-274.
- Y. Ye, I. E. Grossmann, J. M. Pinto, 2018, Mixed-integer nonlinear programming models for optimal design of reliable chemical plants, *Computers & Chemical Engineering*, 116, 3-16.
- D. A. Straub and I. E. Grossmann. (1990). Integrated stochastic metric of flexibility for systems with discrete state and continuous parameter uncertainties. *Computers & Chemical Engineering*, 14(9), 967-985.
- T. V. Thomaidis and E. N. Pistikopoulos. (1994). Integration of flexibility, reliability and maintenance in process synthesis and design. *Computers & Chemical Engineering*, 18:S259–S263.
- Q. Chen, I. E. Grossmann, 2019, Modern Modeling Paradigms Using Generalized Disjunctive Programming, *Processes*, 7, 839.
- M. Türkay, I. E. Grossmann, 1996, Logic-based MINLP algorithms for the optimal synthesis of process networks, *Computers & Chemical Engineering*, 20 (8), 959–978.

Optimization of Graded Bed Reactors for Syngas to Olefin (STO) Processes

Can Ekici,^a Lorenz T. Biegler,^{a,*} Christopher Ho,^b Joseph F. DeWilde,^b Dylan Kipp,^b Paul M. Witt^b

^a*Carnegie Mellon University, Pittsburgh, PA 15213, USA*

^b*The Dow Chemical Company, Midland, MI 48640, USA*

biegler@cmu.edu

Abstract

Reactor designs with mixed catalysts play an important role to transform a multiple reactor system to single-shot reactors. In addition to savings in capital and ease of implementation, single-shot reactors are useful to break equilibrium limitations, therefore increasing the yield and selectivity of desired product. However, the nonlinear, highly exothermic and poorly conditioned nature of mixed-catalyst systems makes it difficult for commercial process simulation and optimization tools to optimize these systems. This study describes the zone and partial moving finite element approaches and their applications to o-xylene oxidation and syngas to olefin (STO) processes. 35% and 5% increases are achieved in the product yield for o-xylene oxidation and STO problems, respectively. Shown with these two examples, this tool is very promising in process intensification for other reactive systems.

Keywords: process intensification, reaction, optimization

1. Motivation and Background Information

Lower olefins (C₂ – C₄) are traditionally produced from cracking of naphtha. However, high oil prices and environmental concerns have initiated research to find other alternatives for olefin production. The two-stage syngas to olefins process is a well-studied process, in which the synthesis gas, mixture of hydrogen and carbon monoxide, reacts to methanol as the first step. Then, methanol reacts to produce olefins. However, this process has equilibrium limitations. On the other hand, olefins can be directly made in one-shot reactors by mixing the catalysts from the two-stage process [1]. Single shot reactors can break the equilibrium limitations with reduced number of unit operations.

The mixed catalyst design of a single shot reactor can be found in other graded bed problems. The ultimate objective of the study is to optimize the catalyst mixing ratio along the reactor, which makes the problem a singular control problem, where control variables appear only linearly in differential algebraic equations. In this case, the optimal control cannot be derived from the stationary condition of Hamiltonian [2]. Numerical optimization methods fail to handle the singular arcs and lead to highly oscillatory profiles, especially when a finer mesh is introduced. There have been other methods to tackle the ill-conditioned nature of the problem, but these methods mostly need a known solution structure and an expression for the singular control profile, which are difficult to determine beforehand. Therefore, a partial moving grid method with direct optimization approach is selected to solve the singular control problems [3].

In this manuscript, the detailed STO reaction mechanism and models used for o-xylene oxidation and STO problems are defined. Two separate optimization approaches are explained and their implementation and results are presented. Conclusions are drawn and future work is proposed for modeling and optimization of STO reactors and other mixed-catalyst reactive systems.

2. Reactor Modeling

Two dimensional models provide accurate representations of heat and mass transfer phenomena for packed bed reactors. In a reactor with exothermic reactions, it is crucial to capture the radial temperature profile accurately. However, two dimensional models are computationally costly to solve, especially as optimization problems. Therefore, we consider the one dimensional alpha model, which can approximate the temperature profiles of a two dimensional model. The two dimensional steady state model of a packed bed reactor in pseudo-homogeneous form, for species j and reactions i , is given below:

$$Q \frac{\partial c_j}{\partial V} = \frac{D}{R} \frac{\partial}{\partial R} \left(R \frac{\partial c_j}{\partial R} \right) + \sum_i v_{ij} r_i$$

$$\sum_j f_j c_{pj} \frac{\partial T}{\partial V} = \frac{\Lambda}{R} \frac{\partial}{\partial R} \left(R \frac{\partial T}{\partial R} \right) + \sum_i (-\Delta H_i) r_i$$

Starting with the governing equations above, the radial concentration differences are neglected by assuming the timescale of reactant consumption is much slower than that of radial dispersion. On the other hand, the radial temperature profile is non-uniform due to constant heat removal through the reactor walls. Therefore, the radial temperature profile needs to be approximated.

The total heat release H is introduced as:

$$H(c, V) = \sum_i (-\Delta H_i) r_i(c, T)$$

and a Taylor series expansion is performed around the heat-averaged temperature, \underline{T}

$$H(c, V) = \sum_i (-\Delta H_i) r_i(c, \underline{T}) e^{A(T-\underline{T})+B(T-\underline{T})^2}$$

where A and B are evaluated at $T = \underline{T}$ as:

$$A = \frac{\partial \ln \ln H}{\partial T}, \quad B = \frac{1}{2} \frac{\partial^2 \ln \ln H}{\partial T^2}$$

Hagan et al. [4] show that the heat equation can be simplified to the equation below, where α is the dimensionless heat loss parameter.

$$\sum_j f_j c_{pj} \frac{\partial T}{\partial V} = -\frac{8\alpha\Lambda}{AR_t^2} + \sum_i (-\Delta H_i) r_i$$

Here, α can be calculated using the implicit equation:

$$\frac{4\alpha}{Bi} = A(\underline{T} - T^c) + \ln \ln(1 - \alpha) - \frac{B}{3A^2}(1 - \alpha)$$

The mass balance equations are constructed in terms of change in the molar flow rates of components, f_j .

3. Optimization Formulation and Solution Strategy

3.1. Zone Algorithm

In the zone approach, the differential and algebraic equations (DAEs) are discretized and converted to nonlinear equations via orthogonal collocation on finite elements. The number of zones along the reactor bed is selected before solving the problem. Each zone is equally divided into a fixed number of finite elements and each element has a fixed number of collocation points. Three collocation points per finite element are selected for every case in this study. The control variable, the catalyst mixing ratio, is constant in each zone and only changes from one zone to another. This is consistent with the implementation in practice, because mixed catalysts can be deployed in layers within tubular reactors.

Although the number of zones is fixed for the problem, the locations of these zones are variable. This formulation increases the size and the nonlinearity of the problem but provides a quick implementation and an approximate solution to the optimal control profile.

3.2. Partial Moving Finite Element Algorithm

Partial moving grid algorithm is a novel method developed by Chen et al. [3] to find the exact solutions to singular optimal control problems. The strength of this method comes from a systematic implementation of heuristics and addresses the critical issue of size and nonlinearity for singular control problems, while enforcing the variational optimality conditions for optimal control.

Because the solution structure of the control profile is unknown, it is not possible to guess the number and location of the singular arcs in the control profile. Therefore, the break points for the singular arcs have to be freely determined by the nonlinear solver, IPOPT in this case. For that purpose, all the finite elements have to be moving elements, so that they can move “freely” along the reactor and capture the break points, and the optimal profiles. The number of finite elements must be sufficiently large, so that the profiles do not lack accuracy.

However, a higher number of moving finite elements would lead to a large problem size and an increase in the nonlinearity, and more importantly to ill-conditioning in the problem, which could easily make the problem unsolvable. Thus, a judicious selection of moving elements is crucial, and this is why this algorithm is very effective.

The partial moving grids (PMG) approach can be summarized in three main stages: constructing a fixed grid; calculating the switching function; and inserting moving elements and detecting and deleting any spikes in the control profiles. Starting from an equally spaced coarse grid and calculating the errors on non-collocation points, fixed elements are introduced at points where error constraints are violated, and moving elements used to determine optimal breakpoint locations. Along the way, any spikes in the control profile are detected and eliminated, in order to ensure that the variational optimality conditions are satisfied.

4. Results

4.1. *O*-xylene Oxidation to Phthalic Anhydride

The *o*-xylene oxidation mechanism consists of three reactions. The desired reaction is the partial oxidation of *o*-xylene to phthalic anhydride. Remaining reactions, total oxidation of *o*-xylene and total oxidation of phthalic anhydride, are side reactions. The reactions are catalyzed by a vanadium pentoxide catalyst mixed with an inert powder, and the blue plots in Figure 1 show the fraction of the active (vanadium pentoxide) catalyst. This reaction system is modelled using the alpha model. The objective is to maximize the phthalic anhydride production. The problem is solved using the zone approach, for one, three, four and six zones.

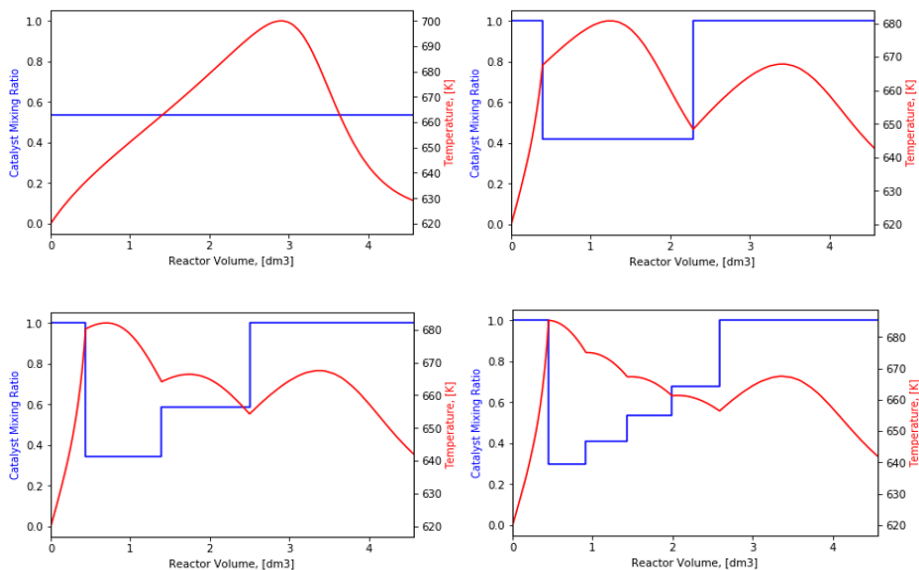
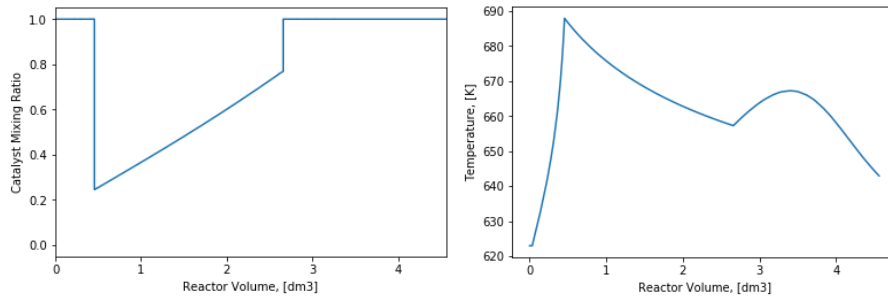


Figure 1 Control and Temperature Profiles for Alpha Model *O*-xylene Oxidation Problem with Different Number of Zones

A developing trend can be observed in the control profile as the number of zones increases. Higher phthalic anhydride production is achieved with an increasing number of catalyst zones. These results are in agreement with those of Nie et al. [5]

The same problem is then solved using the partial moving grids approach. Unlike the zone approach, where the solution depends on the number of zones selected, the partial moving grid approach can yield the exact solution to the infinite dimensional singular control problem. Compared to the solution graphs from the zone approach, both the stepwise control and kinked temperature graphs are smoothed as needed by partial moving grids and represent the exact variational solution. The NLPs for both methods have up to 15,000 variables and equations. These NLPs are solved for 1 to 6 zones using the zone approach and for the exact profiles using partial moving finite elements in under 30 CPU seconds.



Syngas to Olefin

The syngas to olefins reaction mechanism is a complex kinetic system, particularly when catalyzed by mixed catalysts. The olefin production starts with introduction of synthesis gas mixture into the reactor and CO and hydrogen react over Zn/Zr catalyst to yield methanol. Methanol reacts over SAPO catalyst to yield methanol. Moreover, as a result of Zn/Zr – SAPO catalyst mixture, the olefins can react to paraffins, which is an undesired reaction. [6]. The overall kinetic scheme is shown in the figure below.

The catalyst mixture breaks the equilibrium limitation during the methanol production step and enables higher rates, but it also promotes the hydrogenation of olefins, which is an irreversible reaction and results in loss of olefins.

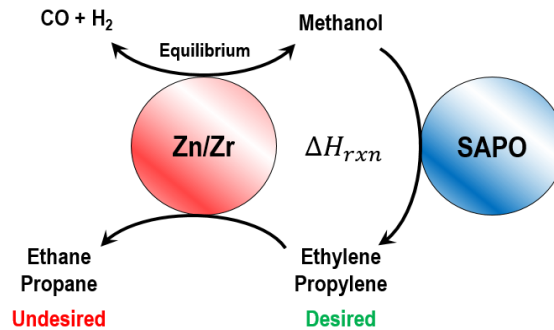
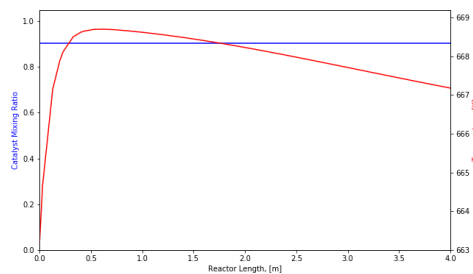
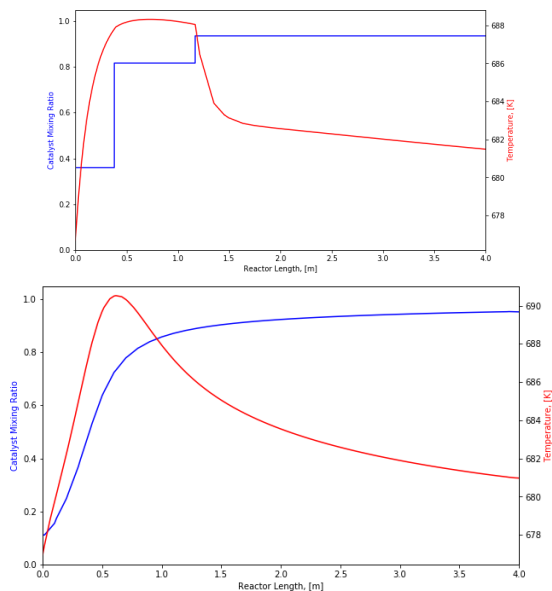


Figure 3 STO Mechanism over Zn/Zr - SAPO Catalyst Mixture





STO optimization problem is solved to maximize the lower olefin ($C_2 - C_4$) yield at the reactor exit using the zone approach. The decision variables in the optimization formulation are inlet temperature, inlet feed ratio and catalyst mixing ratio. The blue plots in Figure 4 show the Zn/Zr catalyst fraction. A 5% increase in lower olefins yield is achieved switching from one zone to three zones.

5. Conclusions

Mixed catalyst problems are singular control problems which are usually ill-conditioned and cannot be dealt with commercial optimization tools. Two different optimization techniques, the zone approach and the partial moving finite element approach, are applied to *o*-xylene oxidation and STO problems, leading to 35% and 5% increases in the product yield in 3 zones for the two problems, respectively. The zone approach is easier to implement and its results are more practical to implement in industry. However, the result will always be suboptimal, and it is not possible to know how far the solution is from the exact solution. On the other hand, it is possible to get the exact solution to the same problem with the partial moving grids (PMG) approach, which is a more complex algorithm and more difficult to transfer to practice. Thus, applications of the two different optimization algorithms to the same problem are necessary for the completeness of the study.

Also, because of the exothermic nature of the STO mechanism, thermal runaways need to be considered for a safe operation and reactor design. In addition, part of our future work will consider optimization under uncertainty. These extended optimization formulations will be implemented to address uncertainties in the reaction parameters and catalyst deactivation, using a two-stage multi-scenario approach [7]. The aim would be to maximize product yield, while respecting the conversion and selectivity constraints and avoiding thermal runaways.

References

- [1] Feng J., Li J., Pan X., Xiao J., Li H., Ma H., Wei M., Pan Y., Zhou Z., Li M., Miao S., Li J., Zhu Y., Xiao D., He T., Yang J., Qi F., Fu Q., Bao X. Selective conversion of syngas to light olefins. *Science* 351, no. 6277: 1065-1068, 2016
- [2] Kameswaran S., and Biegler L. T. Simultaneous dynamic optimization strategies: Recent advances and challenges. *Comp. & Chem. Eng.* 30, no. 10-12: 1560-1575, 2006
- [3] Chen W., Biegler L. T., A Simultaneous Approach for Singular Optimal Control Based on Partial Moving Grid. *AIChE Journal*, 2019
- [4] Hagan, P., Herskowitz, M., Pirkle, C. A Simple Approach to Highly Sensitive Tubular Reactors. *SIAM Journal on Applied Mathematics - SIAMAM.* 48. 10.1137/0148064, 1988
- [5] Nie, Y., Witt P. M., Agarwal A., and Biegler L. T. Optimal active catalyst and inert distribution in catalytic packed bed reactors: ortho-xylene oxidation. *Industrial & Engineering Chemistry Research* 52, no. 44: 15311-15320, 2013
- [6] Kirilin A., Dewilde J. F., Santos V., Chojecki A., Scieranka K., and Malek A., Conversion of Synthesis Gas to Light Olefins: Impact of Hydrogenation Activity of Methanol Synthesis Catalyst on the Hybrid Process Selectivity over Cr-Zn and Cu-Zn with SAPO-34, *Ind. Eng. Chem. Res.* 2017, 56, 13392-13401
- [7] Wang, Y. Modeling, Estimation and Optimal Control of Batch Reactors under Parameter Uncertainty. *Carnegie Mellon University Library*, 2019

Gas Lift Optimization for Optimum Oil Production from a Well Platform

Rahul Sudhanshu^{a,b}, Nitin Dutt Chaturvedi^a

^a*Department of Chemical and Biochemical Engineering,
Indian Institute of Technology Patna, Bihta, Patna – 801106 (Bihar), India*

^b*Surface Team, Mumbai High Asset,
Oil and Natural Gas Corporation Limited, Mumbai – 400051, India*

Abstract

The objective of gas-lift optimization is to allocate the available compressed gas to each well of the platform such that crude oil production is maximized from the group of wells. A wrong allocation of gas may lead to the de-optimization of those wells resulting in lesser production of fluid. Considering the cost and scarcity of compressed gas, allocating them to the group of wells in optimum quantity to enhance oil production becomes the job of utmost importance. In this paper, representative field data have been used to optimize the production using regression and mathematical linear programming. Results produced by this method have been tested at one of the well platforms of the Indian offshore oil field with satisfying outcomes.

Keywords: Artificial Lift, Gas Lift, Production Optimization, Regression, Oil and Gas

1. Introduction

As a routine practice, mostly one or two wells are put under testing and optimization thereafter to get the maximum production. On that very day, lift-gas is adjusted for production optimization to those tested wells accordingly. Since the quantity of lift-gas is not unlimited (and certainly not cheap) in the field, allocating it to all the individual wells for maximum production is even more important. Many thumb rules are being used in the field and different models have also been developed by researchers to optimize the oil production and lift-gas allocation. Posenato et al. (2012) used a Genetic Algorithm for gas lift optimization under lift-gas supply constraint due to limited compression facility. Dutta-Roy (1997) used sequential quadratic programming to study the gas lift allocation and interaction between wells flowing on a common gas lift supply header. Camponogara et al. (2006) used dynamic programming for addressing the problem of gas-lift optimization. Epelle et al (2019) optimized Net Present Value (NPV) of production system by determining lift gas allocation, well controls and other routing constraints, where NPV is combined effective value of revenue generated from oil and gas production and cost incurred on application of artificial lifts etc. Supplying an inadequate quantity of lift-gas may lead to liquid loading in the production tubing causing lesser production; on the other hand, more lift-gas into production tubing will not allow the oil to flow at the optimum rate. In this work, a simple methodology has been developed to allocate the right amount of lift gas among already tested and optimized wells for optimum production from the well platform which house group of wells.

2. Problem Statement

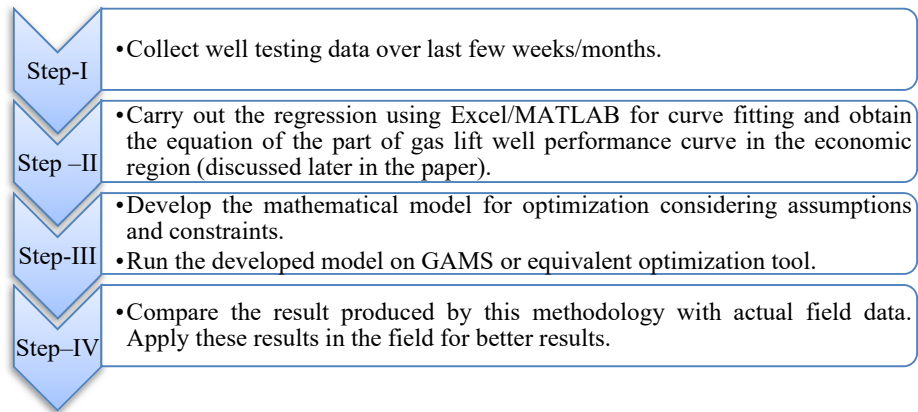
The lift-gas is being supplied to a particular well platform housing a set of wells from a common header/manifold and adjusting this lift gas to one particular well will affect the gas injection rate and pressure of other wells due to interaction among them (Dutta-Roy, 1997). Consequently, production from other wells gets de-optimized from their maximum value. Therefore, it becomes a matter of utmost importance to adjust lift-gas of all the wells such the maximum production is obtained not just from a few wells but from the entire well platform. The objective function problem can be expressed as in Eq. (1.0):

$$\text{Max } P = \sum_{i=1}^n (p_i) \quad (1.0)$$

Where, n is the set of n -number of wells designated as well- $i = \{i: i = 1, 2, \dots, n\}$, need gas injection at the rate of $q_i = \{i: i = 1, 2, \dots, n\}$ to produce liquid/oil at the rate of $p_i = \{i: i = 1, 2, \dots, n\}$.

3. Methodology

In this methodology, a simple procedure depicted in the flowchart has been followed:



Before going further following are the assumptions and declarations in this methodology of gas lift optimization:

- In this analysis, only the effect on the production due to the allocation of lift-gas to individual wells has been considered.
- the liquid produced is mostly oil (almost zero BSW), therefore, the effect of water-cut is not considered.
- Backpressure on well-head due to change in pressure in the group/common production manifold due to variation in production has been ignored.
- The effect of change in pressure at well-head due to change in lift gas rate and pressure has also been ignored.

3.1. Collect well testing data and understanding well performance curve

Figure 1 is a schematic diagram of a well testing facility in which total injection gas is coming to 6" common manifold from the gas compression facility, wherefrom it is distributed among wells. Quantity of produced oil and gas can be tested by diverting a particular well to the separator via 6" test header. The flow of oil from the well is a very complex mechanism and it depends on multiple factors like reservoir conditions, inflow

and outflow performance relationships, etc. Mayhill (1974) initially discussed the concept of gas lift performance curve (Figure 2), where point D indicates the maximum production of the well, and the line DE divides the entire plot into two areas, Area I and Area II.

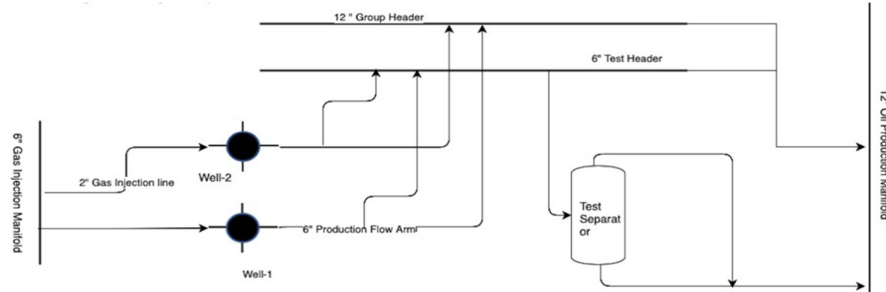


Figure 1: Well testing facility of well platform

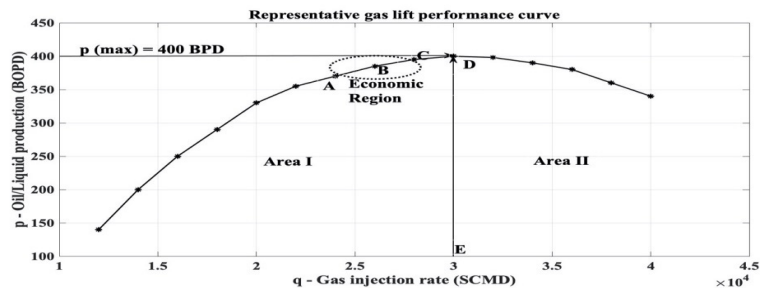


Figure 2: Representative gas lift performance curve

Production of oil is increasing function of gas injection rate till point D and it starts decreasing afterward. Upon analysis, it can be observed that near the point of maxima gas injection rate increases significantly compared to the minor improvement in oil production. Point A to C is marked with a dotted ellipse and is the most economical region for production. It can also be observed that in this small region the curve is following an almost straight-line pattern (from point A to C).

3.2. Regression analysis

When the reservoir depletes and it does not have enough pressure to flow the well on self-mode, the role of an artificial lift to flow the fluid of the well to the separator becomes pertinent. The gas lift method is the preferred way to flow them at offshore installations. Further, the study and analysis of the gas lift performance curve which shows the relationship between the oil production rate (p_i) and gas lift injection rate (q_i) help a lot in optimizing the production and in lift-gas allocation. An n-th order polynomial has been provided to correlate the relationship between gas lift injection rate (q_i) and the oil production rate (p_i) (Sun-Young Jung et al., 2016):

$$p = a_0 + a_1 q_i + a_2 q_i^2 + \dots + a_n q_i^n = \sum_{j=0}^n (a_j q_i^j) \tag{2.0}$$

where p is the rate of production of oil, q is the gas lift injection rate for i-th well and a_0, a_1, a_2, a_n are coefficients generated by the least-squares method during regression analysis. To understand the relationship between oil production rate and gas lift injection rate, well testing data have been tried to fit into the polynomial of higher orders. Though they offer a better fit to the well test data, better understanding of the

function is not developed, neither it is a good predictor. Regression analysis has always been suggested to keep the order of polynomial as low as possible. In this study, the mathematical model is considered to be of the first order as in Eq. (3.0), around the economic region where most of the well testing and optimization data lie.

$$p_i = a_0 + a_1 q_i \quad (3.0)$$

3.3. Develop the mathematical model for optimization

For further simplification and calculation of mathematical model, a curve equivalent to equation (3.0) is plotted keeping p – Oil Production on x-axis and q – Gas Injection on y-axis represented by the equation (4.0):

$$q_i = m_i p_i + c_i \quad (4.0)$$

$$\text{Maximize} \quad P = \sum_{i=1}^n (p_i) \quad (5.0)$$

$$\text{Subject to:} \quad \sum_{i=1}^n (q_i) = \sum_{i=1}^n (c_i) + \sum_{i=1}^n (m_i) * p_i \leq Q \quad (5.1)$$

$$p_i' \leq p_i \leq p_i^* \quad (5.2)$$

$$q_i, p_i \geq 0 \quad (5.3)$$

Equation ((5.0) – (5.3)) is proposed mathematical model for optimization, where P is total production from the well platform (sum of production from all individual wells), Q is the maximum injection gas available for gas lift, q_i and p_i are quantity of lift-gas and production from well i , c_i and m_i are constants generated from linear regression. p_i' is the minimum acceptable production considering the short-term optimization case and p_i^* is the maximum production that can be drawn from i -th well in the economic region.

3.4. Implementation and results

Testing data [Table 2, Annexure A] from a set of five wells of a well platform is considered for optimization under this model. Regression analysis of the data of all the wells has been done on MATLAB_R2020a and part of the well performance curve of the first order is plotted (part of the curve near economic region/optimum region is depicted in Figure 3 for well-1).

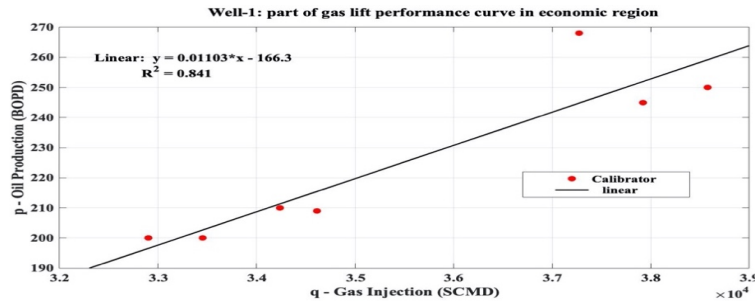


Figure 3: Part of well performance curve for well-1

As discussed during the development of the mathematical model, an equivalent curve is also plotted keeping p – Oil Production on the x-axis and q – Gas Injection on the y-axis using linear regression and the new relation between oil production and gas injected have been generated for all the wells. Figure 4 shows this relation for well-1.

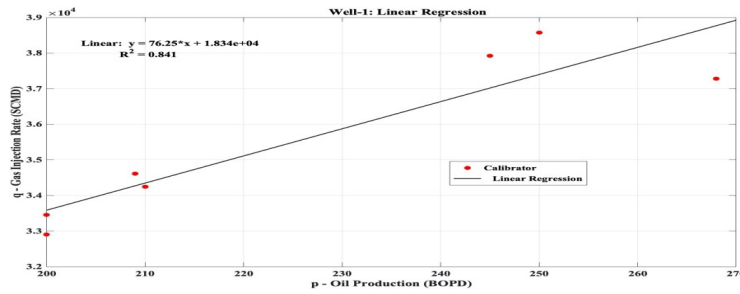


Figure 4: Relation between oil production and gas injection

Finally, the data have been fed to the mathematical model (from Eq. 5.0 to 5.3) and run on GAMS 32.1.0 with CPLEX solver. For different values of Q (190,500 – 201,500 SCMD at the interval of 500 SCMD) on each run, optimal values of variables P are generated. Figure 5 shows the relation between ‘optimized production’ vs ‘maximum gas consumed’. Equations generated by linear regression give values of q_i for all values of p_i and lift gas thus be allocated judiciously for the optimum production. Set of quantity of oil production p_i and corresponding allocated gas q_i , for maximum value of oil produced P and quantity of gas consumed Q at entire well platform is given in Table 3 [Annexure A]. Compared to the field data where oil production is in the range of (1450 – 1651) barrels of oil per day (BOPD) and total injection gas is in the range of (188,771 – 233,914) standard cubic meter per day (SCMD), this model gives the production in the range of 1780 BOPD upon the consumption of 200,000 SCMD of lift-gas from that well platform.

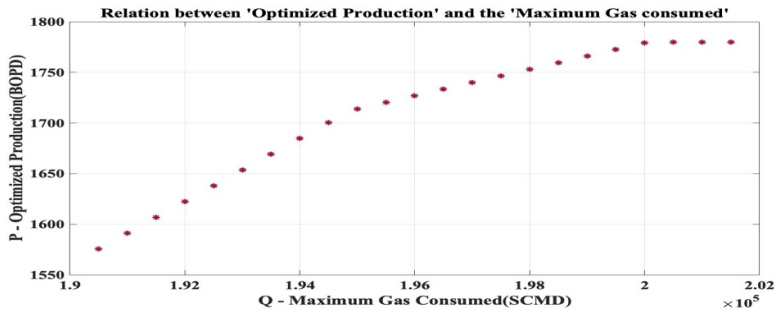


Figure 5: Optimized production Vs. Maximum gas consumed

4. Conclusion

It is obvious that result of this model is not just showing increase in oil production but approximately 30,000 SCMD of compressed gas are saved. This methodology provides wellhead engineers/managers relatively simple and quick approach to analyze and work on the available production data during field visit. Optimal lift-gas injection rate from other oil-fields can be further tested/implemented, on the wells of the field to verify actual improvement in oil production. Though results produced by this methodology will help oil field operators in day-to-day well optimization, additional constraints like impact of well head pressure upon variation in gas injection rates, changes in water cut (BSW) on oil production etc. will be considered in future work.

Acknowledgement

The authors would like to thank the Science and Engineering Research Board, India for providing the research funding for this project under the grant no. ECR/2018/000197.

Annexure A

Choke Size (X/64")	Gas Inj. Orifice / Line Size (mm")	Main line Pressure (psi)	Main line Temp (°F)	Main line Gas Inj Pr. (psi)	Separator Pressure (psi)
NIL	26.0233/2	192	180	1041	206

Well No	Well-1		Well-2		Well-3		Well-4		Well-5		All Wells	
Testing Date	Gas Inj.*	Oil Prod#	Gas Inj.	Oil Prod	Gas Inj.	Oil Prod	Gas Inj.	Oil Prod	Gas Inj.	Oil Prod	Total Inj. Gas	Total Oil
Sept-20	37277	268	40509	338	52569	260	48247	524	55312	208	233914	1598
Aug-20	33452	200	31337	356	53434	240	50703	533	53800	221	222726	1550
July-20	34239	210	31325	371	55703	240	44117	389	49517	240	214901	1450
June-20	34613	209	24354	406	55134	255	46782	423	53588	240	214471	1533
May-20	32900	200	21932	394	59199	220	46640	492	50529	251	211200	1557
Apr-20	37921	245	23628	383	54630	265	45000	400	49935	260	211114	1553
Mar-20	38577	250	22304	437	41467	270	45071	422	41352	272	188771	1651

*Unit - standard cubic meter per day (SCMD), # Unit - barrels of oil per day (BOPD)

	Well-1	Well-2	Well-3	Well-4	Well-5	Total
pi (BOPD)	268	437	270	533	272	1780
qi (SCMD)	38770.78	20247.99	46798.2	49097.11	45149.52	200063.6

Reference:

- Camponogara, E., & Nakashima, P. H. R. (2006). Solving a gas-lift optimization problem by dynamic programming. *European Journal of Operational Research*, 174(2), 1220–1246.
- Dutta-Roy, K., & Kattapuram, J. (1997). A New Approach to Gas-Lift Allocation Optimization SPE Western Regional Meeting.
- Epelle, E & Gerogiorgis, D (2019), 'Mixed-Integer Nonlinear Programming (MINLP) for production optimisation of naturally flowing and artificial lift wells with routing constraints', *Chemical Engineering Research and Design*, 152, 134-148.
- Sun-Young Jung & Jong-Se Lim (2016) Optimization of gas lift allocation for improved oil production under facilities constraints, *Geosystem Engineering*, 19:1, 39-47.
- Mayhill, T. D. (1974). Simplified Method for Gas-Lift Well Problem identification and Diagnosis. Fall Meeting of the Society of Petroleum Engineers of AIME.doi:10.2118/5151-ms
- Posenato, A., & Rosa, V. R. (2012). A Genetic Algorithm for Gas Lift Optimization With Compression Capacity Limitation. SPE Latin America and Caribbean Petroleum Engineering Conference.
- Production Operation Volume I (2019). Institute of Oil & Gas Production Technology. Oil and Natural Gas Corporation Ltd. Panvel, Navi Mumbai

Techno-economic evaluation of sorption enhanced steam gasification of PKS system for syngas using CaO for CO₂ capture

Muhammad Shahbaz,^{a,b,*} Ahmed AlNouss,^a Suzana Yusup^b, Gordon Mckay^a, Tareq-Al Ansari^a

^a*Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University (HBKU), Qatar Foundation (QF), Education City, P.O. Box 5825, Doha, Qatar.*

^b*Center for Biofuel & Biochemical Research (CBBR), Institute of Self - Sustainable Building (ISB) Chemical Engineering Department, Universiti Teknologi PETRONAS Bandar Seri Iskandar, 32610, Perak, Malaysia.*

mshahbaz@hbku.edu.qa

Abstract

Biomass Gasification is a viable process to convert biomass into valuable energy production. In this study, the techno-economic feasibility of a sorption enhanced steam gasification of palm oil waste for syngas production is developed using Aspen Plus. The sorption enhancement is achieved by utilising CaO as a medium to capture CO₂ and generate clean syngas. CaO has recently demonstrated the feasibility to capture CO₂ at reduced capital and operating investment costs compared to conventional capture processes. For this purpose, the flowsheet configuration and economic analysis has been carried out using Aspen plus. The results of the economic assessment demonstrate a total hydrogen production cost approximated at \$2.51 per kg for gasification system utilising CaO compared to \$2.57 for the base case without CO₂ capture. Moreover, a sensitivity analyses and a multi-objective optimisation has been carried out to maximise the hydrogen to carbon monoxide ratio and minimise the CO₂ emissions. The results demonstrate that by increasing the steam flowrate in the range of 500-2000 kg/h, the H₂/CO ratio, and the CO₂ emissions increase. This is mostly due to the acceleration of gasification reactions; mainly the methane reforming and water gas shift reactions. Similarly, increasing the CaO flowrates in the range of 500-2000 kg/h results in increasing the H₂/CO ratio significantly and reducing the CO₂ emissions rapidly. This CO₂ reduction with the use of CaO is experimentally validated in the literature mainly due to the carbonation reaction. Whereas, the variation of gasification temperature indicates a decreasing trend in the H₂/CO ratio and CO₂ emissions with the increase in temperature in the range of 600-800 °C. The Pareto curve generated from the multi-objective optimisation demonstrates an overall increasing trend of CO₂ emissions with the increase in the H₂/CO ratio in the produced syngas.

Keywords: PKS; Aspen Plus; CO₂ Capturing; Techno-economic analysis; CaO

1. Introduction

Biomass is a leading source of alternative energy for fossil fuels as it offers promises advantages such as its low carbon footprint due to carbon neutrality (Shahbaz et al., 2017a), and it can be considered part of a renewable and sustainable supply chain due to its abundant availability (Inayat et al., 2020a). The conversion of biomass into energy is achieved through biological and thermochemical conversion processes (Zeng et al., 2016). Thermochemical conversion processes are advantageous over biological processes due to the conversion of biomass to all forms of energy, such as gaseous, solid, and liquid fuels, in addition to (Shahbaz et al., 2020). Incidentally, gasification is the preferred thermochemical conversion process in the production of a gaseous fuels such as H₂ and syngas (AlNouss et al., 2020a). Many researchers have experimentally investigated syngas production from palm oil waste through lab and pilot-scale processing (Chan et al., 2019), whilst simulation-based modelling is also useful to inform syngas production process (Inayat et al., 2020b). The H₂ rich syngas produced from the biomass processing depends on many factors such as biomass feed, gasification agent, and most importantly, the reduction of CO₂ in the product gas (Inayta et al., 2012). In this regard, CaO based sources, such as, dolomite and limestone are used for the reduction of CO₂ in the gasification of palm oil waste through modelling and experimentation (Shahbaz et al., 2017a).

Palm oil waste is a major waste available in South East Asia and has been investigated for energy production especially through gasification (Shahbaz et al., 2017b), for which techno-economic analysis is a very important activity for the commercialisation of any technology. A process developed for the steam-gasification of empty fruit bunches (EFB) based on kinetic parameters using MATLAB was developed by Inayat et al. (2012), and Aspen PLUS was used for equilibrium modelling of palm kernel shell (PKS) (Shahbaz et al., 2017a). From the above discussion, there is no study reported for the techno-economic analysis of PKS steam gasification using the CaO for CO₂ capture. As such, the main aim of this study is to perform a techno-economic feasibility study of sorption enhanced steam gasification of palm kernel shell using Aspen Process Economic Analyser (APEA) with adsorbent regeneration. In addition, the study investigates the impact of process parameter and optimises the parameters for low H₂/CO ratio to maximize the H₂ rich syngas production. Sensitivity and, and multi curve optimisation is also developed.

2. Methodology and Process Description

For several industrial processes, such as polymers, metals, biofuels, other chemicals, modelling and simulation of solid process steps is crucial. Aspen Plus advanced process simulation capabilities allow the integrated high-performance simulation of process steps for solids and fluids with stable physical properties. In this study, palm kernel shell (PKS) undergoes steam gasification process, which is simulated using Aspen Plus software, to yield synthesis gas. The process is simulated under the assumptions of steady-state and isothermal-operation, atmospheric pressure with no pressure gradient, neglected tar formation and char is carbon. The model begins with a decomposition unit operating at 40 °C and 1 bar to convert the PKS biomass to conventional components; the attributes are summarised in Table 1 (Shahbaz et al., 2017a). The decomposition unit is linked with a calculator block to calculate the yields of the conventional components as presented by (AlNouss et al., 2020b). The converted stream enters the fluidised bed gasification unit, modelled as a Gibbs free energy reactor, along with the superheated steam to yield syngas at 800 °C and 1 bar as illustrated in Figure 1. The effluent syngas is fed to the fixed bed

gasification unit where the CaO is used to capture the generated CO₂ and yield a solid CaCO₃. A solid separator is used to remove ash and the formed CaCO₃ and keep the syngas that is cooled to further purify it from water. The CaO is then regenerated from CaCO₃ using a regeneration reactor operating at 900 °C and 1 bar and a purification step to remove vent gases from the regenerated CaO.

Table 1. Proximate and ultimate analysis of Palm kernel shell (Shahbaz et al., 2017a)

Proximate analysis		Ultimate analysis (dry mass wt% basis)	
Moisture (wt%)	9.70	C	48.78
Dry mass wt% basis		H	5.70
Volatile matter	80.81	N	1.01
Fixed carbon	14.25	S	0.21
Ash content	4.94	O (by difference)	39.36

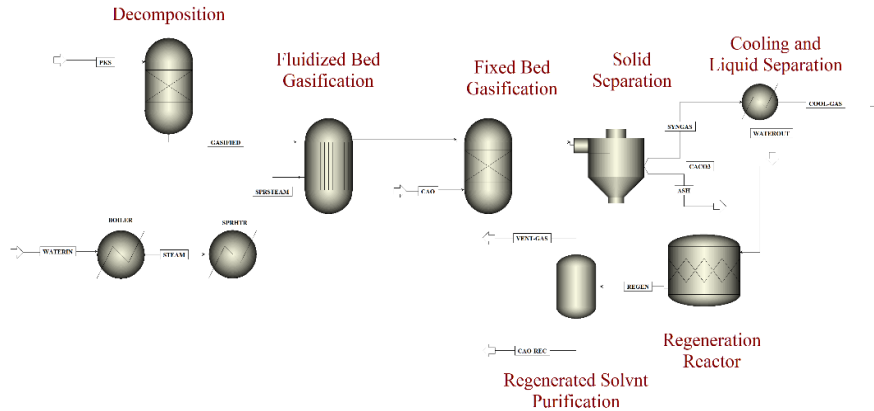


Figure 1. Process flow sheet for the sorption enhanced steam gasification of PKS.

The simulated model has then been utilised to perform a number of sensitivity cases to study the effect of changing steam flowrate, gasification temperature and CaO flowrate on the quality of syngas production. Moreover, the optimisation capabilities of Aspen Plus software has been utilised to perform singular and multi-objective optimisations to maximise the ratio on hydrogen to carbon monoxide and minimise the emissions of CO₂ in the produced syngas. Equations 1 and 2 summarise the formulation of the optimisation problem.

$$\text{Maximise} \quad \frac{y_{H_2} \cdot \text{Syngas}}{y_{CO} \cdot \text{Syngas}} \quad \text{eq. (1)}$$

$$\text{Minimise} \quad x_{CO_2} \cdot \text{Syngas} \quad \text{eq. (2)}$$

$$\text{Subject to} \quad 600^\circ\text{C} \leq T_{\text{gasification}} \leq 800^\circ\text{C} \quad \text{eq. (3)}$$

$$500 \text{ kg/h} \leq \dot{m}_{\text{Steam}} \leq 2000 \text{ kg/h} \quad \text{eq. (4)}$$

$$500 \text{ kg/h} \leq \dot{m}_{\text{CaO}} \leq 2000 \text{ kg/h} \quad \text{eq. (5)}$$

Where y_{H_2} and y_{CO} are the molar fractions of hydrogen and carbon monoxide, x_{CO_2} is the mass fraction of carbon dioxide, respectively, \dot{m} represents the mass flowrate of steam/CaO and syngas is the molar flowrate of the produced syngas. The use of CaO has also been validated by conducting a techno-economic-environmental evaluation using the built-in tools inside Aspen Plus® and utilising the equations presented earlier by (AlNouss et al., (2020c). In the study, the price of purified syngas is taken as \$24/km³.

3. Results and discussion

The results of the techno-economic-environmental evaluation based on the use of CaO to capture CO₂ demonstrates a huge decrease in CO₂ emissions compared to the base case without CO₂ capture as presented in Table 2. However, this decrease is associated with slightly higher capital and operating costs. This demonstrates the effectiveness of CaO utilisation in offsetting the high CO₂ emissions associated with the gasification process with only a slight increase in investment. Moreover, the higher hydrogen production in the CaO case has projected a lower cost of hydrogen production at \$2.51 per kg compared to \$2.57 for the base case. The process revenue for the case without CaO addition demonstrates slightly higher results approximated at \$3.772 million compared to \$3.769 for the case with CaO addition. This increase is mainly due to the increase in the CO content in the purified syngas. The addition of the regeneration unit has increased the capital and operating costs, although it resulted in an additional reduction in the CO₂ emissions, and a reduction in the costs associated with the raw material since CaO is regenerated internally. The cost of hydrogen production remained similar to the case of CaO utilisation without regeneration despite the slight increase in the total annualised cost.

Table 2: Techno-Economic analysis with and without CaO

Case	with CaO	without CaO	with CaO regeneration
Total Capital Cost [\$]	4,233,300	4,147,700	5,054,360
Total Operating Cost [\$/y]	1,427,190	1,413,370	1,435,570
Total Raw Materials Cost [\$/y]	525,600	350,400	350,400
Total Utilities Cost [\$/y]	143,318	138,912	5,054,360
Equipment Cost [\$]	207,500	159,100	259,900
Total Installed Cost [\$]	1,015,800	952,600	1,330,600
Total Annualised Cost [\$/y]	2,822,126	2,615,527	2,823,916
Revenue [\$/y]	3,769,178	3,772,167	3,769,178
H ₂ Production (kg/y)	1,123,514	1,019,029	1,123,514
(T = 700 °C, Steam = 1500 kg/h, CaO = 100 kg/h)			
Cost of H ₂ Production (\$/kg)	2.51	2.57	2.51
CO ₂ Emissions (kg CO ₂ -e/h)	535	1031	509

The results of the different analyses performed on the steam gasification of PKS are presented in terms of sensitivity relations, optimum objective functions and techno-economic comparison. Figure's 2 and 3 present the relationships between the molar compositions of the produced syngas and the changes in CaO flowrate, gasification temperature and steam flowrate, respectively. The change in CaO flowrate at constant gasification temperature of 700 °C and a constant steam flowrate of 1500 kg/h demonstrates an increase in hydrogen composition from 70% to 99% and a decrease in CO, CO₂ and CH₄ compositions from 13%, 17% and 0.27% to 1%, 1% and 0.05%, respectively. The decline in CO₂ with the addition of CaO is due to the carbonation reaction and was also reported by Shahbaz et al. (2017b).

The increase in gasification temperature at a constant CaO flowrate of 1000 kg/h and constant steam flowrate of 1500 kg/h demonstrates an increase in hydrogen and CO compositions from 76.3% and 6.5% to 77.8% and 12.3%, respectively, with the decrease in CO₂ and CH₄ compositions from 13.8% and 3.43% to 9.8% and 0.02%, respectively. The reason behind these trends is mainly due to the influence of higher temperatures on the gasification reactions leading more syngas products, where the reactions of steam-methane reforming, Boudouard and CO water shift are favoured (Shahbaz et al., 2017b).

Moreover, H₂ production trend indicates a peak at around 700 °C of temperature.

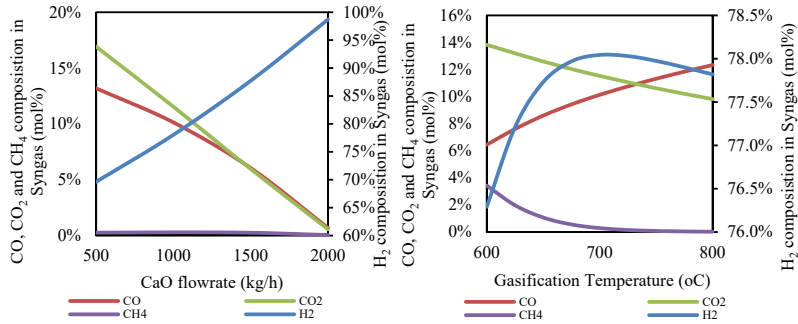


Figure 2. Impact of CaO flowrate and gasification temperature on syngas composition.

The change in steam flowrate at a constant gasification temperature of 700 °C and a constant CaO flowrate of 1000 kg/h demonstrates an increase in H₂ and CO₂ compositions from 70.5% to 78.6% and 3% to 13% respectively, with a decrease in CO and CH₄ compositions from 19% to 8% and 7.55% to 0.11%, respectively. These trends indicate an enhancement in steam-methane reforming and CO water gas shift reactions with the addition of steam, where more H atoms are converted to H₂ resulting in less CH₄ and more C atoms are converted to CO₂ resulting in less CO (Inayat et al., 2012).

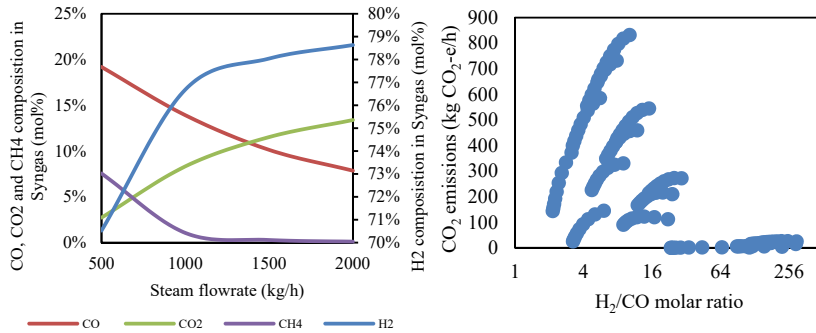


Figure 3. Impact of steam flowrate on syngas composition.

Figure 4. Pareto-front curve

Furthermore, to analyse the optimal operating parameters that will achieve an increase in the production quality and reduction in CO₂ emissions, the singular and multi-objective optimisations have reported different parameters for steam and CaO flowrates and the gasification temperature. For the singular optimisation, the solution of Eq. (1) with the constraints in Eq. (3-5) indicates an overall objective function for the H₂/CO molar ratio approximated at 297 for gasification temperature of 600 °C and 2000 kg/h flowrate for steam and CaO, individually. Whereas, the singular optimisation of Eq. (2) with the constraints in Eq. (3-5) indicates an overall objective function for the CO₂ emissions below 0.07 kg CO₂-e/h for gasification temperature of 800 °C, the steam flowrate of 500 kg/h and CaO flowrate of 2000 kg/h. The multi-objective optimisation of both Eq. (1) and (2) with the constraints in Eq. (3-5) produces the illustrated Pareto-front curve in Figure 4. This curve indicates various optimal points for the decision-makers on CO₂ emissions and H₂/CO molar ratio of the steam gasification of PKS given various operating points.

4. Conclusions

The sorption enhanced steam gasification of PKS with the utilization of CaO for CO₂ capturing developed using Aspen PLUS[®]. The techno-economic analysis performed using the tool present in Aspen PLUS[®]. The cost of H₂ production is reduced from 2.57 to 2.51 \$/kg with the use of CaO. In addition, the CO₂ emissions were significantly decreased from 1031 to 535 (kg CO₂-e/h) with the application of CaO. The H₂ production increase with the increase of gasification temperature, steam, and CaO flow rate. A significant reduction of CO₂ is observed from 17 to 1 mol% with the increase of CaO flow rate from 500-2000 kg/hr, which shows the effectiveness of CaO for CO₂ capturing. The multi-curve optimisation made and a petro curve was generated for H₂/CO molar ratio and CO₂ emissions of the steam PKS gasification for different process parameters.

Acknowledgment: The authors thank Hamad Bin Khalifa University (HBKU), Doha, Qatar & Universiti Teknologi PETRONAS, Malaysia, for providing financial and technical support.

References

- A. AlNouss, G. McKay, and T. Al-Ansari, 2020a, Enhancing waste to Hydrogen Production Through Biomass Feedstock Blending: A Techno-Economic-Environmental Evaluation. *Applied Energy* 266, 114885.
- A. AlNouss, G. McKay, and T. Al-Ansari, 2020b. Production of syngas via gasification using optimum blends of biomass. *Journal of Cleaner Production* 242, 118499.
- A. AlNouss, P. Parthasarathy, M. Shahbaz, T. Al-Ansari, H. Mackey, and G. McKay, 2020c. Techno-Economic and Sensitivity Analysis of Coconut Coir Pith-Biomass Gasification using Aspen Plus. *Applied Energy* 261, 114350.
- Y. Chan, K.W. Cheah, B.S. How, ACM. Loy, M. Shahbaz, HKG. Singh, N. Yusuf, S. Shuhaili, S. Yusup, W. Ghani, J. Rambli, Y. Kansha, H. Lam, B. Hong, and S. Ngan, 2019. An overview of Biomass Thermochemical Conversion Technologies in Malaysia. *Science of The Total Environment* 680, 105-123.
- A. Inayat, M. Ahmad, MIA. Mutalib, and S. Yusup, 2012. Process Modeling for Parametric Study on Oil Palm Empty Fruit Bunch Steam Gasification for Hydrogen Production. *Fuel Processing Technology* 93(1), 26-34.
- A. Inayat, M. Inayat, M. Shahbaz, SA. Sulaiman, M. Raza, and S. Yusup, 2020a. Parametric Analysis and Optimization for the Catalytic Air Gasification of Palm Kernel Shell using Coal Bottom Ash as Catalyst. *Renewable Energy* 145, 671-681.
- A. Inayat, M. Raza, Z. Khan, C. Ghenai, M. Aslam, M. Shahbaz, and M. Ayoub, 2020b. Flowsheet Modeling and Simulation of Biomass Steam Gasification for Hydrogen Production. *Chemical Engineering & Technology* 43(4), 649-660.
- M. Shahbaz, T. Al-Ansari, M. Aslam, Z. Khan, A. Inayat, M. Athar, SR Naqvi, M. Ahmed, and G. McKay, 2020. A State of the Art Review on Biomass Processing and Conversion Technologies to Produce Hydrogen and its Recovery Via Membrane Separation. *International Journal of Hydrogen Energy* 45(30), 15166-15195.
- M. Shahbaz, S. Yusup, A. Inayat, A., M. Ammar, DO. Patrick, A. Pratama, and SR. Naqvi, 2017a. Syngas Production from Steam Gasification of Palm Kernel Shell with Subsequent CO₂ Capture using CaO Sorbent: An Aspen Plus Modeling. *Energy & Fuels* 31(11), 12350-12357.
- M. Shahbaz, S. Yusup, A. Inayat, DO. Patrick, M. Ammar, and A. Pratama, 2017b. Cleaner Production of Hydrogen and Syngas from Catalytic Steam Palm Kernel Shell Gasification Using CaO Sorbent and Coal Bottom Ash as a Catalyst. *Energy & Fuels* 31(12), 13824-13833.
- J. Zeng, R. Xiao, D. Zeng, Y. Zhao, H. Zhang, and D. Shen, 2016. High H₂/CO Ratio Syngas Production from Chemical Looping Gasification of Sawdust in a Dual Fluidized Bed Gasifier. *Energy & Fuels* 30(3), 1764-1770.

A Robust Optimization Approach for Hydrogen Allocation Network with Parametric Uncertainties

Gaurav Shukla, Nitin Dutt Chaturvedi*

*Department of Chemical and Biochemical Engineering, Indian Institute of Technology Patna,
Bihta, Patna, 801106, Bihar, India*

nitind@iitp.ac.in

Abstract

Hydrogen is an important resource in chemical processes industries. As hydrogen demand in the refineries is increasing rapidly, hydrogen management strategies are of great interest to the refineries. Hydrogen management trades in with the optimal supply, routing, and flow allocation of hydrogen gas also called Hydrogen Allocation Network (HAN) within a refinery linking the source station to the demand station. Other than minimizing the fresh hydrogen requirement, it is extremely important to reduce the compression work, required in any HAN. In the real world, uncertainties in the parameters (e.g. flow, quality, pressure, etc.) may arise due to various reasons so that undermine the effectiveness of HAN. In this paper, to deal with these uncertainties, an optimization model based on robust linear programming formulation is proposed to minimize the compression work and fresh hydrogen as a resource within the HAN. The applicability of the robust model helps to optimize HAN with uncertain flows and quality for the application of individual sources station and demand station with the desired reliability. The resultant formulation preserves the linearity of the mathematical model and can control the degree of conservatism for every constraint. The applicability of the proposed methodology is demonstrated through an illustrative example.

Keywords: Hydrogen allocation network, Compression work, Robust Optimization, Optimization

1. Introduction

Hydrogen, a clean and efficient secondary energy source, is widely used in chemical industries to produce clean fuels. Petroleum refineries consist of many processes with complex reactions involving hydrogen consumption or production. In a hydrogen network, there are several compressor stations for supplying hydrogen producers to consumers. So hydrogen management is an important practical aspect of refineries. It aims to achieve the optimal hydrogen allocation network (HAN) of hydrogen resources to satisfy the demands of refinery processes (Brijandi et al. 2014). Operating variables play an important role in designing of HAN, such as the purity and flow rate of input and output for consumers. In the real world, uncertainties in the parameters (e.g. flow, quality pressure, etc.) may arise due to various reasons so that undermine the effectiveness of HAN (Al-Redhway et al. 2005). Therefore, the inlet or outlet variables are determined by any changing variable after optimization. Zhang et al. (2014) proposed a graphical method for targeting the minimum fresh resource consumption of hydrogen networks by considering the separation performance of purifiers. The method optimizes both the purity and the flow rates of feed and products for the purifier within a feasible operating range. Lou et al. (2015) introduced a framework to optimize the hydrogen network of refineries under uncertainty. This framework considers several scenarios representing possible future environments. Lima et al. (2018) proposed a

robust optimization model is developed to cope with the oil price and demand uncertainties in the tactical management of the downstream oil supply chain. Liu et al (2019) proposed a fuzzy optimization model for the design of a multi-component hydrogen network with parametric uncertainties is developed. A data-driven adaptive robust optimization approach is proposed by Zhao et al. (2019) to deal with the operational optimization problem of industrial steam systems under uncertainty. From the above literature review, it is found that minimizing the fresh hydrogen and compression work has rarely been studied together with the optimization of HAN. In this paper a mathematical formulation for calculating the robust target of minimizing the compression work and resource requirement in a HAN. The presented model utilizes the concept of robustness developed by Bertsimas and Sim (2004). The proposed model enables us to decide the robust targets of fresh hydrogen and compression work requirement based on budget parameters and uncertainty levels. The proposed model is explained via an illustrative example where the process data is represented through uncertain parameters. The solution set provides flexibility to the decision-maker to the trade-off between the uncertainty level with compression work or fresh hydrogen requirement.

2. Problem Definition and Mathematical Formulation

The HAN usually includes several compressor stations for supplying the hydrogen at several demand nodes at their required pressure. The problem of optimizing hydrogen allocation with uncertainty in any allocation network optimization in a plant can be stated as follows:

- A set of N_s internal sources is available with uncertain flow and quality in a bounded interval with a known deviation from its nominal value.
- A set of N_d internal demands to be satisfied are also known such that each demand accepts a flow with a quality that is less than a predetermined maximum quality.
- The unutilized flow from the internal sources is sent to an external demand, called waste, without any maximum quality or flow limits.

The objective of this study is to minimize the compression work for supplying the flow and also minimize resource requirement in a fixed desired reliability of each demand as well as each source. The model comprises of following constraints.

2.1. Flow Balance Constraints: Let f_{ij} and f_{rj} represent the flow from i^{th} source and resource to j^{th} demand. The flow balances for compressor station and demand station may be addressed as Eq. (1) and Eq. (2):

$$\sum_{j=1}^{N_d} f_{ij} + f_{iw} = F_{si} \quad \forall i \quad (1)$$

$$\sum_{j=1}^{N_s} f_{ij} + f_{rj} = F_{dj} \quad \forall j \quad (2)$$

2.2. Quality Constraints: Due to quality constraint at every demand, the hydrogen mass-load requirement for any internal demand may be mathematically expressed as follows: (Eq. 3)

$$\sum_{j=1}^{N_s} f_{ij} q_{si} + f_{rj} q_{jr} \leq F_{dj} q_{dj} \quad \forall j \quad (3)$$

Where q_{si} , q_{jr} , and q_{dj} are the quality of i^{th} source, resource, and j^{th} demand. The total fresh hydrogen requirement can be calculated using Eq. (4).

$$F_{fr} = \sum_{j=1}^{N_d} f_{frj} \tag{4}$$

2.3 Robust Mathematical Model for HAN

Adapting the robust formulation proposed by Bertsimas and Sim (2004) for dealing with uncertainty. The mathematical model can be modified as follows.

Flow uncertainty: Let us consider that the flow supply by each internal source is uncertain and can vary in the region. To include flow uncertainty Eq. (1) is modified as follows:

$$F_{st} - \sum_{j=1}^{N_d} f_{ij} - f_{iw} - \Gamma^f z^f - q^f = 0 \tag{5}$$

$$\Gamma^f - q^f \geq \hat{F}_{st} \tag{6}$$

$$\Gamma^f, q^f \geq 0 \tag{7}$$

The parameter Γ^f and q^f are introduced that controlled the degree of conservatism and budget of uncertainty in source flow. \hat{F}_{st} is the deviation magnitude of the flow.

It may be noted that after the fresh hydrogen requirement has been minimized, the fresh hydrogen source and waste flow can be treated as entities of the sources and demands sets without any loss of generality.

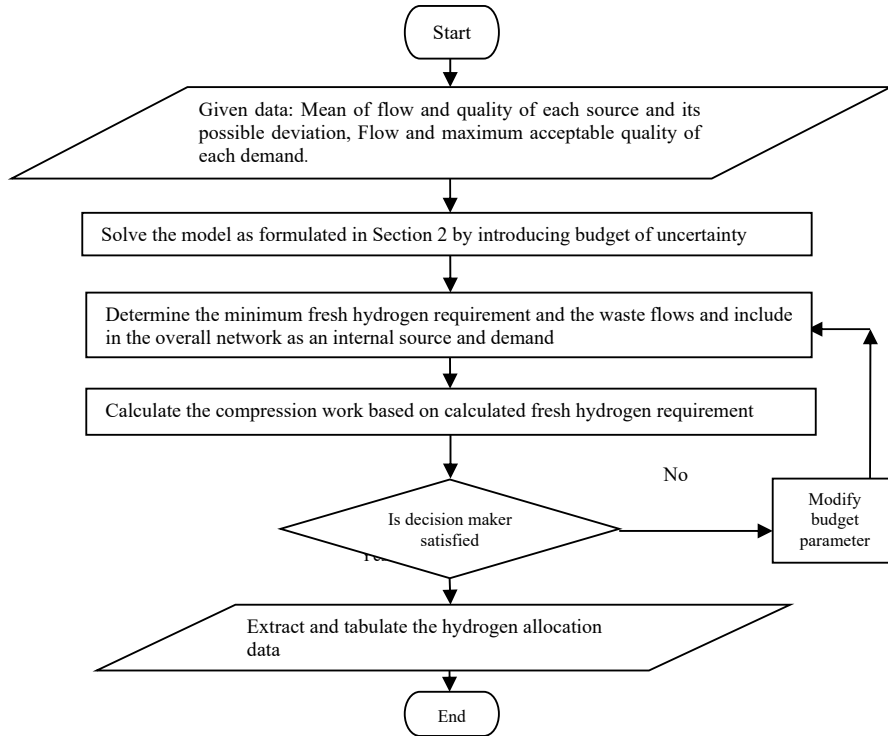


Figure 1: Flow chart showing the proposed algorithm

With these new modified sets of sources and demands with the inclusion of fresh resource and waste flows, the second part of the optimization problem is formulated. The compression work (W) required for a stream is governed by the initial (P_{in}) and final (P_{out}) states (volumetric flow (F_{in}) and pressure) as well as the process followed for compression. For isothermal compression, the work done can be expressed as Eq. (8).

$$W = P_{in} F_{in} \ln\left(\frac{P_{out}}{P_{in}}\right) \quad (8)$$

Applying the characteristic of an isothermal process;

$$PF = Const \quad (9)$$

The compression work requirement for supplying flows balance f_{ij} from various supply pressure levels (P_i) pressure level to demand at a pressure (P_j) pressure level ($P_j > P_i$) may be expressed as follows: (Shukla and Chaturvedi 2020)

$$E_j = \sum_{i=1}^{N_i} f_{ij} (\mu_j - \mu_i) \quad (10)$$

The quantity $(\mu_j - \mu_i)$ can be defined as the compression energy index.

It may be noted that supplying flow from a station having a higher pressure than the lower demand station pressure does not require any compression work. Further, it may be noted that demand pressure includes various pressure drops during flow transfer. Figure 1 shows the flow chart of the proposed algorithm.

3. Illustrative Example:

Applicability of the proposed methodology is demonstrated through illustrative examples. Data for the overall HAN from the work of Bandyopadyay et al. (2014) are given in Table 1. The source flow data are modified for illustrative purposes.

Table 1: Source and Demand Data for Illustrative Example

	Purity (%)	Quality (%)	Flow (Sm ³ /s)	Pressure (kPa)
S0 (resource)	99	1		11031.6
S1	91	9	114.7 ± 20	11031.6
S2	85	15	163.8 ± 20	16168.4
	Purity (%)	Quality (%)	Flow (Sm ³ /s)	Pressure (kPa)
D0 (waste)				
D1	92.8	7.2	131.1	10061.4
D2	87.5	12.5	196.6	15168.4

Table 1 represents the source and demand data for the example. From the deterministic approach, it is found that the minimum fresh hydrogen requirement and waste production are determined to be 58.99 and 9.79 Sm³/s, respectively and the is minimum compression power using isothermal process is found to be 1.37 MW. In the proposed example the flow is uncertain. In this case, bounded uncertainty of flow availability from the source side is given. The availability of flow parameters has ± 20 t/h variability levels for each flow. Here budget

parameter takes a value between [0, 1], its variation with objective value is calculated using the proposed methodology. The result is tabulated in Table 2.

Table 2: Solution data of example

Γ^f	0.0	0.	0.4	0.6	0.8	1.0
Fresh Hydrogen(Sm ³ /s)	58.99	60.70	65.2	73.2	81.2	89.2
Waste (Sm ³ /s)	9.79	3.506	0.363	0	0	0

It is found that for the worst case ($\Gamma^f=1$) the minimum fresh hydrogen requirement is 89.2 Sm³/s. In the presented example if it is solved considering all parameters to take their worst boundary values i.e. lower bound for flow rates using (Soyester, 1973), the resource requirement is calculated to be 89.2 Sm³/s, which increase the resource requirement by 51.21 % from the deterministic case and with maximum protection against uncertain.,This value can be considered as overestimated if the uncertainties in the parametric coefficients lie in the narrower range. Similarly, the unpredictability level in quality of each source can be included in the example and check the impact of the budget parameter in the requirement of fresh hydrogen. Next, based on the fresh hydrogen requirement and supply to demand compression station, the minimum compression work can be calculated using Eq. (10) and tabulated in Table 3.

Table 3: Minimum compression work data for an illustrative example

Γ^f	0.0	0.2	0.4	0.6	0.8	1.0
Fresh Hydrogen(Sm ³ /s)	58.99	60.70	65.2	73.2	81.2	89.2
Work (kJ/s)	1.37	1.30	1.31	1.44	1.570	1.71

The trade-off between the compression work and minimum fresh hydrogen (2 a) and uncertainty level and optimum value of compression work (2 b) is plotted.

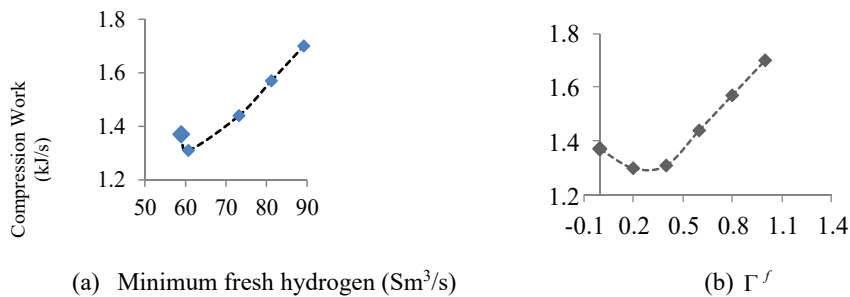


Figure 2: Trade-off between minimum hydrogen and compression work (a), Impact on the compression work with variation in the budget parameter of (b)

The proposed methodology provides a feasible solution for a relative magnitude of uncertain data in a parameter which assists decision-maker via trade-off between the feasibility tolerance and a reliability level. The model is solved using the GAMS/ CPLEX solver on the computer (Intel(R) Core(TM) i5 (3 GHz) and 4 GB RAM

4. Conclusions

In the real world for designing any process industry, it is necessary to address different parametric uncertainties. In this paper, a methodology is presented to determine the minimum compressor work required in a HAN, while satisfying the minimum external hydrogen resource requirement including the parametric uncertainties. The proposed methodology is based on robust optimization which is capable of adjusting the level of risk (worst case scenario) is applied to derive a robust optimal solution for HAN. The main benefit of using robust optimization is that it provides the trade-off between uncertainty level and an upper probability of constraint violation which helps to choose suitable functional situations for decision-makers under uncertain conditions and to do the required planning accordingly and protected the process against uncertainties to satisfy demands. In future work, the model formulation may be readily expanded to incorporate other constraints such as carbon emission, pressure uncertainty along with economical aspects such as the annual electricity consumption of compressors and the annual revenue created by fuel gas to make the problem close to the real world.

References

- Al-Redhwan, S.A., Crittenden, B.D. and Lababidi, H.M., 2005. Wastewater minimization under uncertain operational conditions. *Computers & Chemical Engineering*, 29(5), pp.1009-1021.
- Bandyopadhyay, S., Chaturvedi, N.D. and Desai, A., 2014. Targeting compression work for hydrogen allocation networks. *Industrial & Engineering Chemistry Research*, 53(48), pp.18539-18548.
- Bertsimas, D. and Sim, M., 2004. The price of robustness. *Operations research*, 52(1), pp.35-53.
- Birjandi, M.R.S., Shahraki, F. and Razzaghi, K., 2018. A New Optimization Strategy to Improve Design of Hydrogen Network Based Formulation of Hydrogen Consumers. *Chemical and Biochemical Engineering Quarterly*, 32(1), pp.91-101.
- Lima, C., Relvas, S., Barbosa-Póvoa, A.P. and Morales, J.M., 2018. Oil product distribution planning via robust optimization. In *Computer Aided Chemical Engineering (Vol. 43, pp. 949-954)*. Elsevier.
- Liu, J., Zhou, Y., Zhu, M., Deng, C. and Lee, J.Y., 2019. Optimization of Refinery Hydrogen Network with Parametric Uncertainties. In *Computer Aided Chemical Engineering (Vol. 47, pp. 77-82)*. Elsevier.
- Lou, J., Liao, Z., Jiang, B., Wang, J. and Yang, Y., 2015. A thermodynamic irreversibility based design method for multi-contaminant hydrogen networks. *International Journal of Hydrogen Energy*, 40(1), pp.435-443.
- Shukla, G. and Chaturvedi, N.D., 2020. Simultaneous Minimization of Cost and Energy in Gas Allocation Network. *Chemical Engineering Transactions*, 81, pp.1093-1098.
- Soyster A.L., 1973, *Convex Programming with Set-Inclusive Constraints and Applications to Inexact Linear Programming*, *Operations Research*, 21, 1154
- Zhao, L., Ning, C. and You, F., 2019. A Data-Driven Robust Optimization Approach to Operational Optimization of Industrial Steam Systems under Uncertainty. In *Computer Aided Chemical Engineering (Vol. 46, pp. 1399-1404)*. Elsevier.
- Zhang, Q., Liu, G., Feng, X., Chu, K.H. and Deng, C., 2014. Hydrogen networks synthesis considering separation performance of purifiers. *International journal of hydrogen energy*, 39(16), pp.8357-8373.

Process Design and Techno-economic analysis of Hydrogen Production using Green Ammonia Imported from Australia- A Korea Case Study

Malik Sajawal Akhtar ^a, Jay Liu ^{*}

^{a,*} *Department of Chemical Engineering, Pukyong National University, Busan 48513, South Korea*

jayliu@pknu.ac.kr

Abstract

Hydrogen can be a potential solution to decarbonise the transportation sector since it has a very high energy content. Its separation from water by the process of electrolysis need a vast amount of electricity and renewable energy sources like solar PV and wind electricity can play a vital role in shifting towards hydrogen economy, if they are utilised effectively. Still the problem with the hydrogen economy is the massive investment in its large scale transportation and storage infrastructure as it has a low volumetric energy density, which creates significant challenges in transporting bulk volumes. One feasible and still optimistic approach is to use ammonia as a hydrogen carrier that can be transported globally with the existing infrastructure. South Korea is evaluating the possibilities of importing green ammonia and using it for hydrogen production to meet country hydrogen demands particularly related to transportation sector. This research work presents a process design and techno-economic analysis of importing green ammonia from Australia and using it for hydrogen production in Korea at the hour of need. Processes were simulated by using commercial process simulator Aspen Plus V11. The main contributors of the economic assessment are capital and operational expenditures of electrolyser and solar photovoltaic power plant. The levelised cost of green ammonia is 870 \$/ton while that of hydrogen is 8.90 \$/kg stating that it is a viable option to use imported green ammonia to meet country hydrogen demands. Future estimates from multiple studies show that by 2030 further cost declines are possible in electrolyser and solar panels capital expenditures resulting the green ammonia cost to \$350 per ton and making this case more economically viable to meet hydrogen demands.

Keywords: Process design, green ammonia, techno-economic analysis, green hydrogen.

1. Introduction

Decarbonisation of world's industrial processes specifically the energy production is a hotspot now a days due to declining fossil fuel resources, increasing oil prices and high energy demands. Electricity and heat production are the highest contributor in carbon emissions followed by transportation. The need of transportation will be increasing with the time and more efforts would be required to decarbonise the transportation sector. Clean and green fuel is a dire need of the time in order to minimize the carbon emissions related to transportation sector. Hydrogen having the high energy content can be a potential solution to this problem since it is available in abundant quantity in the form of

water and needs only electricity for production through the process of electrolysis. One of the major challenge in implementation of hydrogen economy is its transportation in bulk volumes due to its low energy density and massive investments. To address this issue, hydrogen carriers like ammonia is under consideration. Ammonia is a well-known chemical having large existing infrastructure for transportation. Million tons of ammonia is transported every year globally through ships. In ammonia production, hydrogen is produced by reforming of natural gas which is the major cause of high carbon emissions. If it is produced through water electrolysis by using power from renewable energy sources, then decarbonisation of ammonia industry is possible. This carbon free green ammonia can be transported globally and can be used in fertiliser sector and hydrogen production through its decomposition reaction generally known as cracking.

Australia is one of the major producer of renewable electricity which is not only cost-effective but also is being produced on larger scales. The world's largest renewable energy project 'Asian Renewable Energy Hub (AREH)' of 26 GW is planned in Pilbara region of Western Australia planning to start exporting green ammonia by 2028 according to ammonia energy association. According to Sichao Kan (2018), the H₂ demand for Korea by 2022 is estimated to be at 0.47 million tons, 1.94 million tons in 2030, and 5.26 million tons in 2040, so soon new H₂ supply sources will be required. To address this issue, importing Green Ammonia from Australia and using it for H₂ production is under consideration but detailed and rigorous techno-economic analysis is required for government and industry to make a final decision.

Therefore, in this research work a process design of 300 tons/day green ammonia production plant and economic assessment of hydrogen production through thermal catalytic cracking of ammonia is studied in detail. Imported ammonia will be transported through ships from Australia in refrigerated containers normally 30,000 tons per container and will be shifted to H₂ re-fuelling stations within Korea as per requirement. Ammonia cracking units will be installed beside the re-fuelling station to ensure continuous fuel supply and avoiding extra transportation of H₂. According to Hydrogen Europe, a normal H₂ re-fuelling station has H₂ production capacity of 300 Nm³/hr and contains storage tanks, high pressure compressors, and dispensing units.

2. Methods

2.1 Description of the green ammonia production plant

The production of green ammonia considered in this work constitutes an air separation plant, green hydrogen production plant (electrolysis powered by renewable energy), and an ammonia synthesis unit. The ammonia is being produced in Australia, after production it will be compressed and stored and then shipped to Busan, South Korea. Production capacity of the plant is 746 t/day and total power requirement is about 360 MW as adapted from Morgan et al (2017) in which 335 MW (93% of total power requirement) is only for electrolysis unit. Imported green ammonia will be cracked in a cracking unit installed besides the hydrogen re-fueling station to ensure continuous supply. Aspen plus V11 was used for process designing and simulation. The heat generated throughout the process is captured by EX to generate steam. This steam is then used to run a steam turbine operating at 90% polytropic efficiency, which in turn, aids in driving the compressors to increase overall efficiency of the process. Complete process flow diagram of aspen plus simulation is shown in Figure 1.

- Air Separation Unit (ASU).

Cryogenic distillation unit was modelled which constitute multi-stage compressor having inter-stage cooling, shell and tube heat exchangers, and distillation columns. Air is at atmospheric conditions and compressed in C1 to 7 bars followed by cooling near to its liquefaction temperature. This cooled air is then introduced into a high-pressure distillation column of 25 stages and operating at 7 bar. The top and bottom products from this column are introduced in low pressure column of 20 stages operating at 1.013 bar for further separation of nitrogen. The nitrogen obtained from the top at 26.260 MT/hr is 99.999% pure and fed to the ammonia synthesis loop for ammonia production. Total power consumption of this unit is 3.24 MW.

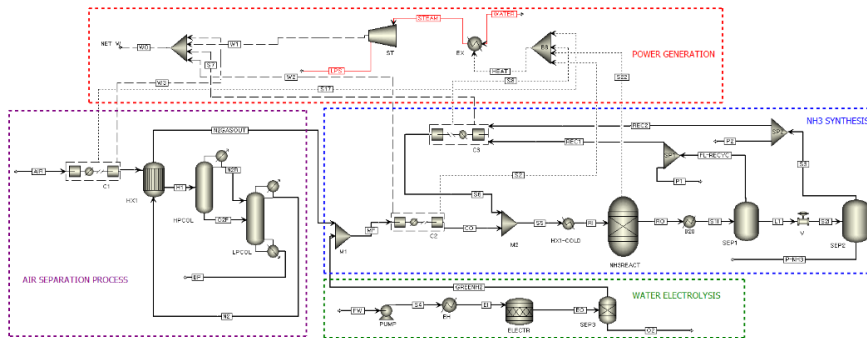


Figure 1: Aspen Plus-Process Flow Diagram (PFD) of Green Ammonia Synthesis

- Water Electrolysis.

After water purification, purified feed water is compressed in a centrifugal pump to 30 bar and heated in a heat exchanger to 80°C. Alkaline water electrolyser was modelled using RStoic reactor from model palette. Reactor outlet stream constitute 66.66% hydrogen and 33.33% oxygen, this stream was then introduced in a separator and hydrogen purity was set to 99.99% considering the purity from commercial electrolyser. Purified hydrogen gas at 5.66 MT/hr. is mixed with nitrogen and then fed to ammonia synthesis unit while total power consumption of this section is 335 MW.

- Ammonia Synthesis.

Nitrogen from air separation and hydrogen from electrolyser is first mixed adiabatically M1. This synthesis gas stream is then compressed using multi-stage (five) compressor to 150 bar pressure. This high-pressure stream is then heated to the required reaction temperature of 450°C. The reaction of ammonia synthesis is exothermic, so the heat generated during the reaction is captured by EX to generate steam which is used to power the compressors being used in the process. The reactor outlet stream is cooled to atmospheric temperature of 25°C and introduced to a flash separator Sep1 being operated adiabatically. The separator outlet stream L1 contains 97% pure ammonia, unreacted gases are recycled back to compressor train C3 and subsequently to mixer M2. Final product stream P-NH3 from Sep2 is at -18°C and 2 bar contains 99.99% pure ammonia. The power requirement for this ammonia synthesis unit is about 20 MW.

- Transportation.

In this project, transportation is considered from Dampier port of Pilbara region in Western Australia where the world biggest renewable energy project ‘Asian

Renewable Energy Hub (AREH)' is planned to Busan port, South Korea. The transport distance between the mentioned ports is 3867 nautical miles (7161 kilometre).

- Cracking Unit.

According to Hydrogen Europe, a normal hydrogen refuelling station (HRS) are operated at 300 Nm³/hr hydrogen flow rate. It is assumed that the imported ammonia from Australia is transported to hydrogen refuelling stations in Korea where cracking units are installed beside the fuelling station to ensure continuous supply. A process flow diagram of the unit is shown in Figure 2 Chung et al (2017) reported the performance of the decomposition reaction over Ru-based catalysts (Ru/La(x)-Al₂O₃) as more than 95% conversion. Reaction conditions were set at 400°C and 1.013 bar. Liquid ammonia pass through a heater to attain the reaction temperature. The outlet stream of the reactor was then introduced in pressure swing adsorption column, ultra-pure H₂ gas with 99.999% purity was collected from top while the bottom product was 99.82% pure N₂.

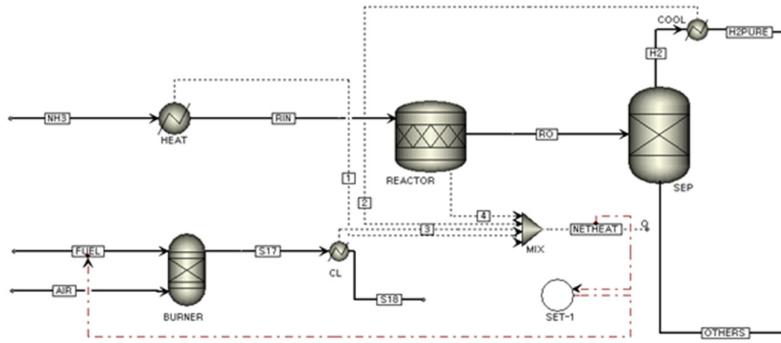


Figure 2: Aspen Plus-Process flow diagram (PFD) of Ammonia Decomposition

2.2 Economic Analysis

Lee et al (2019) compared a cracking unit for laboratory scale of 0.9 m³/hr. and for a small size refueling station of 30 Nm³/hr. Equipment cost was calculated using chemical engineering plant cost index (CEPCI) for 2020 and scaling factors obtained from Remer et al 1993 and Lee et al (2020). The capacity of our cracking unit is 300 Nm³/hr. Transportation and cost of green ammonia for Australia was obtained from Bruce et al (2018) and from green ammonia white paper by Argus media (2020). Details of cost calculations is shown in table 1 below. NREL's correlation as shown in Eq.1 was used for the calculation of levelised cost of hydrogen (LCOH) as mentioned by Short et al (1995), where TLCC is the total life cycle cost of plant including manufacturing cost, D is the discount rate, $NPV_{OTHER-PRODUCTS}$ is the net present value of the other products i.e. oxygen and nitrogen, $nH_{2_GENERATED}$ is the amount of hydrogen produced.

$$LCOH = \frac{TLCC - NPV_{OTHER-PRODUCTS}}{\sum_0^I \frac{nH_{2_GENERATED}}{(1+D)^n}} \quad (1)$$

Table 1: Economic Assessment for Cracking Unit

Parameter	Value	Reference
Assumptions for the analysis		
Depreciation year	7	-
Tax rate	30%	-
Internal Rate of Return	4.5%	-
Total Power requirement for green ammonia plant	360 MW	Morgan et al (2017)
Assessment results		
Solar PV Capex	902 \$/kwh	NREL (2019)
Green Ammonia Plant Total Cost	1.60 billion \$	-
Green Ammonia Cost (LCOA)	0.87 \$/kg	-
Ammonia required for H2	138.5 kg/hr.	-
Power requirement for cracking unit	300 KW	-
Ammonia shipping cost to Busan	0.052 \$/kg	Valentini (2020)
Cracking Catalyst cost	0.66 \$/kgH2	Lee et al (2019)
Cracking unit installed Cost	\$1,535,490	Jackson et al (2020)
Levelized Cost of Hydrogen	8.90 \$/kg	-

3. Results and discussion

The LCOH cost of hydrogen production from imported green ammonia resulted as \$ 9.12/kg. According to Jane et al (2020), the current LCOH of green hydrogen is estimated to range between \$ 2.50/kg and \$10/kg, which validate our results. So, the case of importing green ammonia from Australia for hydrogen production in Korea resulted to be a viable option. The cost of ammonia production is largely dependent on the cost of renewable electricity needed and the cost of electrolyzer. In the coming future, fall in electrolyser and renewable electricity prices is expected, so green hydrogen will be cheaper than today.

Table 2: Future estimated cost (\$/kg) of ammonia and hydrogen production

Year	2020	2025	2030	Reference
Solar PV Capex \$/kw	902	733	565	NREL (2019)
AWE Capex (including stack replacement)	1,061	849	707	Ju(2020), Matute et al., (2019)
LCOA	0.87	0.73	0.60	-
LCOH (after cracking)	8.90	8.30	7.75	-

Considering future cost reductions in capital expenditures of solar pv panels and electrolyser, we have done an analysis of future estimated prices of hydrogen production through cracking of imported ammonia excluding the expected reduction in electricity prices, as shown in table 2. Moreover, multiple research studies are also ongoing for increasing the activity and economic viability of catalysts for ammonia cracking. According to Argus media and Nayak et al (2020) estimates, by 2030 renewable electricity prices will drop to \$20-25 per MWh and LCOH of green hydrogen will eventually fall to \$2.30 to 2.40\$ per kg also the price of green ammonia will fall to \$ 310-350 per ton. Considering these future estimates, it will become more viable to import green ammonia and use it to fulfil hydrogen demands of the country and to decarbonize the ammonia and hydrogen industry. To have a fair comparison with imported case and

to reach at a point of some decision, there is also a need to do techno-economic assessment of domestic production of green hydrogen in Korea. So, our future work will consist of a comparative analysis of this case of importing green ammonia with domestic production of green hydrogen in Korea along with Monte-Carlo based sensitivity analysis.

References

- Andrea Valentini, (2020). ‘Green shift to create 1 billion tonne green ammonia market’, a white paper by argus media available at <https://view.argusmedia.com/green-ammonia.html?blog>.
- Articles on ammonia energy association at <https://www.ammoniaenergy.org/organization/asian-renewable-energy-hub/>
- Bruce S, Temminghof M, Hayward J, Schmidt E, Munnings C, Palfreyman D, Hartley P (2018) National Hydrogen Roadmap. CSIRO, Australia.
- Chung DB, Kim HY, Jeon M, Lee DH, Park HS, Choi SH, Nam SW, Jang SC, Park J-H, Lee K-Y, Yoon CW. Enhanced ammonia dehydrogenation over Ru/La(x)-Al₂O₃ (x=0-50 mol%): structural and electronic effects of La doping. *Int J Hydrogen Energy* 2017;42:1639–47.
- IRENA (2020), Renewable Power Generation Costs in 2019, International Renewable Energy Agency, Abu Dhabi
- Jane Peatie, Drew Powell. ‘Levelised costs will settle the blue-green debate’ Independent Analysis for Energy Leaders available at <https://www.petroleum-economist.com/articles/low-carbon-energy/energy-transition/2020/levelised-costs-will-settle-the-blue-green-debate>
- Jensen, J. O.; et al. Pre-Investigation of Water Electrolysis; National Laboratory, Technical University of Denmark, 2008.
- Lee, Boreum, et al. “Assessment of the economic potential: CO_x-free hydrogen production from renewables via ammonia decomposition for small-sized H₂ refueling stations.” *Renewable and Sustainable Energy Reviews* 113 (2019): 109262.
- Lee, S. H., Lim, D. H., & Park, K. (2020). Optimization and Economic Analysis for Small-Scale Movable LNG Liquefaction Process with Leakage Considerations. *Applied Sciences*, 10(15), 5391.
- Matute, G., Yusta, J. M., & Correias, L. C. (2019). Techno-economic modelling of water electrolyzers in the range of several MW to provide grid services while generating hydrogen for different applications: A case study in Spain applied to mobility with FCEVs. *International Journal of Hydrogen Energy*, 44(33), 17431–17442. <https://doi.org/10.1016/j.ijhydene.2019.05.092>
- Morgan, Eric R., James F. Manwell, and Jon G. McGowan. “Sustainable ammonia production from US offshore wind farms: a techno-economic review.” *ACS Sustainable Chemistry & Engineering* 5.11 (2017): 9554-9567.
- Nayak-Luke, Richard Michael, and René Bañares-Alcántara. “Techno-economic viability of islanded green ammonia as a carbon-free energy vector and as a substitute for conventional production.” *Energy & Environmental Science* 13.9 (2020): 2957-2966.
- NREL (National Renewable Energy Laboratory). 2019. 2019 Annual Technology Baseline. Golden, CO: National Renewable Energy Laboratory.
- Refueling stations at hydrogen europe available at <https://hydrogeneurope.eu/refueling-stations>
- Short, W., Packey, D., and Holt, T. (1995). A manual for the economic evaluation of energy efficiency and renewable energy technologies. *Renew. Energy* 95, 73–81. doi:NREL/TP-462-5173.
- Sichao Kan, “South Korea’s Hydrogen Strategy and Industrial Perspectives”, *Edito Energie*, Ifri, 25 March 2018.

Life Cycle Assessment of Hydrogen Production from Imported Green Ammonia: A Korea Case Study

Malik Sajawal Akhtar ^a, Jay Liu ^{*}

^{a,*} *Department of Chemical Engineering, Pukyong National University, Busan 48513, South Korea*

jayliu@pknu.ac.kr

Abstract

To decarbonise the transportation sector, South Korea is looking forward to evaluate possible cross-regional hydrogen supply pathways to meet the future hydrogen demand and reduce carbon emissions to ‘zero’ under Paris agreement by 2050. The use of ammonia as hydrogen carrier to meet hydrogen demands is under consideration and the goal of this paper is to evaluate environmental aspects of green hydrogen production from green ammonia imported from Australia. In the present work, a cradle to gate life cycle assessment of hydrogen production from imported green ammonia imported is carried out by using commercial life cycle assessment software Simapro. The scope of life cycle assessment includes; green ammonia production in Australia, shipping to South Korea, ammonia transportation from port to hydrogen refuelling stations in Korea, and hydrogen production through catalytic craking of ammonia. Mass and energy balance data obtained from process flow modelling are used as an input for life cycle assessment. In terms of energy consumption, hydrogen production through water electrolysis consume more than 90% of total power while ammonia synthesis only consume 5.5%. The results of life cycle assessment show that transportation of the green ammonia was the major contributor having 95% of total CO₂ emissions while hydrogen production through water electrolysis was producing about 4.5% of total carbon emissions. This study reveals that considering environmental sustainability, using imported green ammonia is not a viable option until transportation fuel is replaced with an environmental friendly fuel. However, if marine fuel is replaced by a green fuel it can further reduce the transport related emissions turning this case into a viable option to meet the hydrogen related demands.

Keywords: Life Cycle Assessment, Green Ammonia, Green Hydrogen

1. Introduction

The alarm of increasing global warming is shifting the world's focus towards clean and sustainable energy systems. The increasing concentration of greenhouse gases in the atmosphere is mainly due to the extensive burning of fossil fuels for energy generation. Energy demand is growing quickly with the rapid urbanization and industrialization of the world, and considering environmental sustainability; it is needed to offset the large dependency on fossil fuels. Renewable energy sources such as wind and solar are available in ample quantity in nature and can be used efficiently as a substitute for fossil fuels. Electricity and heat production are the highest contributors to carbon emissions, followed by transportation. Decarbonisation of the transportation and industrial sector is possible only by using clean and sustainable fuel instead of high carbon-emitting fossil

fuel. Hydrogen carrier like ammonia can be a potential candidate to solve the issue of infrastructure for the global transportation of hydrogen due to its low volumetric density. Ammonia production can be carbon-free if the hydrogen required for ammonia production is obtained from renewable energy sources through water electrolysis rather than that of the conventional process of steam methane reforming, emitting high carbon emissions. Energy hungry countries like Japan and Korea are thinking of importing green ammonia from countries like Australia to offset their large dependency on fossil fuels and meeting their hydrogen economy roadmaps. The decision to import green ammonia needs an intensive analysis on environmental sustainability assessment considering the transportation from Australia to Korea.

Currently, according to Philibert (2017), global ammonia production contributes to 1.3% of total carbon dioxide emissions, along with about 420 million tons of CO₂ being emitted into the atmosphere. Globally, ammonia production relies on fossil fuels both as an energy source and feedstock, specifically for hydrogen. The ammonia industry's decarbonisation is possible only if the source of energy and hydrogen feedstock is replaced by fossil fuel with greener and sustainable sources. Hydrogen production through water electrolysis needs electricity, and if this electricity is from renewable energy sources such as wind and solar, the produced hydrogen and then ammonia can be termed as green and sustainable. Life cycle assessment (LCA) is a well-known methodology to assess the environmental impacts associated with a product or process at each stage of its life cycle like raw materials extraction, product manufacturing, services, and final disposal. The assessment of green ammonia production's environmental impacts will give a brief insight into the process's sustainability aspects and economic indicators.

Therefore, this study presents a cradle to gate environmental assessment of hydrogen production from imported green ammonia. This assessment's system boundary includes; raw materials required for manufacturing the product, infrastructure for green ammonia production, including construction and manufacturing of electrolyzer, and transportation from Australia to Busan. This work doesn't include the construction of renewable energy power plants, compression, and hydrogen storage since it is assumed that the hydrogen generation unit is onsite and construction of hydrogen refuelling stations. It is assumed that ammonia imported from Australia will be shifted to fuelling stations in Korea as per requirement, and cracking units will be installed beside the stations to avoid hydrogen transportation and ensure continuous fuel supply.

2. Methods

A block flow diagram is shown in Figure 1 for a brief overview of the process. This work was done in three stages; process designing and simulation of hydrogen and ammonia production processes, life cycle inventory and assessment, and interpretation of results. For process design and simulation, Aspen Plus V11 was used, mass and energy balance data from process design was used as an input for life cycle assessment. SimaPro 9.1.1.1 was used for life cycle assessment, and method CML-IA baseline V3.06 was selected for evaluation of economic indicators.

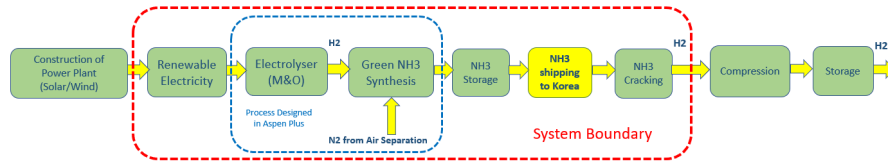


Figure 1: Block diagram of the case study

2.1 Process design and simulation

Ammonia here is produced by the combination of nitrogen (from air separation unit) and hydrogen (water electrolysis) at high temperatures (450°C) and pressure (150bar) in the presence of Fe_2O_3 as catalyst. Aspen plus process flow diagram of green ammonia synthesis is shown in Figure 2. Electricity from renewable energy sources is used as an input of power source of electrolyzer and production process of ammonia of 746 tons/day with 5.66 ton/hr. of hydrogen from electrolyzer and 26.260 ton/hr. nitrogen from air separation requires 360 MW in total, according to Morgan et al. (2017). Power consumption of each unit is; air separation 3.24 MW, ammonia synthesis unit about 20 MW, and water electrolysis about 335 MW. This green ammonia will then be shipped to Korea and then transported from port to hydrogen refuelling stations as per requirement. It is assumed that cracking units are installed beside the fuelling station for onsite production of hydrogen and to avoid hydrogen transportation. A process flow diagram of the unit is shown in Figure 3, reaction conditions were set at 400°C and 1.013 bar. Liquid ammonia pass through a heater to attain the reaction temperature. The outlet stream of the reactor was then introduced in pressure swing adsorption column, ultra-pure H_2 gas with 99.999% purity was collected from top while the bottom product was 99.82% pure N_2 . A normal hydrogen fuelling station operates at a 300 Nm^3/hr hydrogen flow rate. High-pressure compressors (750-1000 bar) are used for hydrogen at the fuelling station. However, hydrogen compression and storage is not considered in the environmental assessment.

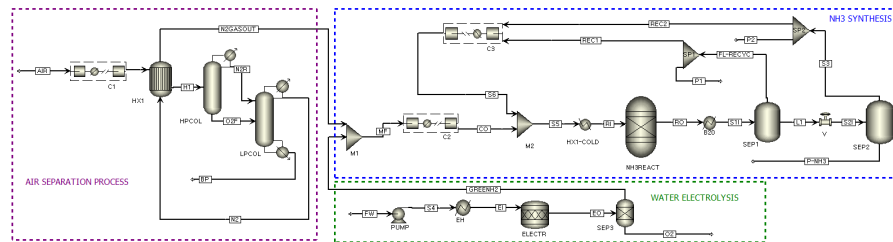


Figure 2: Process flow diagram of green ammonia synthesis

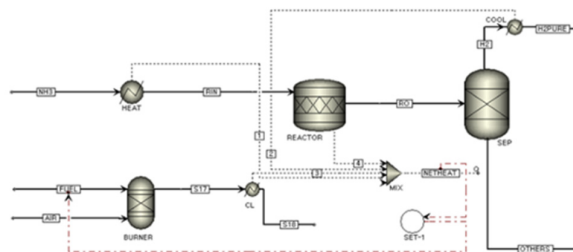


Figure 3: Process flow diagram of cracking of green ammonia

2.2 Life Cycle Impact Assessment

Life cycle assessment is a universal and mature tool to assess the environmental impacts from manufacturing, packaging, storing, transportation, and final disposal. Life cycle inventory is calculated by using mass and energy balance data obtained from process design modelled in Aspen Plus V11. The system boundary for this assessment includes; raw materials required for manufacturing the product, infrastructure for green ammonia production which includes construction and manufacturing of electrolyser, and transportation from Australia to Busan. This work doesn't include construction of renewable energy power plants, hydrogen re-fuelling stations, and compression and storage of hydrogen since it is assumed that hydrogen generation unit is onsite. Australian energy statistics (2020) published by the department of industry, science, energy, and resources reported that electricity generation by renewable energy sources (wind, hydro, and solar) 21% in 2019 in which wind and solar contributed equally at 6%. Hydrogen production from electrolysis needs a huge amount of electricity, about 54.3 kwh per kg of hydrogen (Koj et al., 2017). So, it is the major source of carbon emissions in the green hydrogen production process. In contrast, the electrolyzer construction has very little contribution to the emissions, as shown in Figure 4 below.

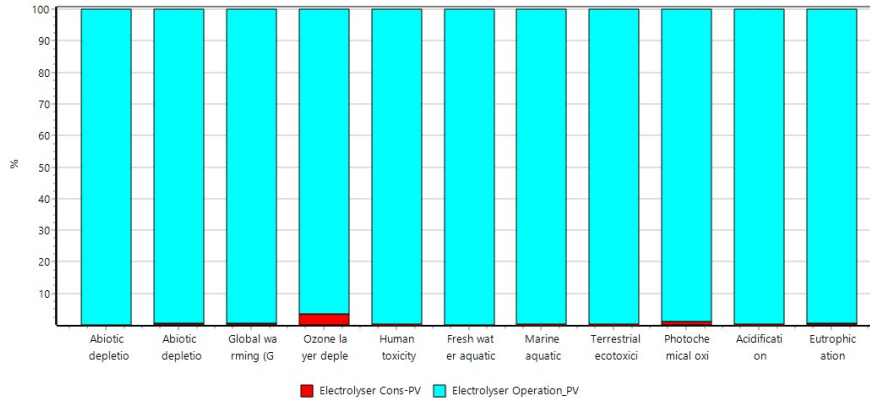


Figure 4: LCA results of hydrogen production from electrolysis

The transportation from Australia to Korea emits high CO₂ and SO_x emission since no renewable fuel is used in marine transport now. These results are shown as relative contribution, since ammonia is being transported from Australia to Busan having distance of 7161 km so it is contributing maximum in the emissions. There is a probability of using green ammonia as a marine fuel, and if marine fuel is replaced with green energy, then the high emissions related to transport can also be mitigated. The results of environmental indicators after LCA of green ammonia production is shown in Figure 5. Following the ammonia transportation, hydrogen production presents the second highest emissions throughout the process. Hydrogen production through cracking of green ammonia has the lowest contribution since very less power, about 300 kwh for hydrogen production of 300 Nm³/hr. is required for the process.

2.3 Interpretation of results

The results of the complete life cycle assessment of the case study as shown in figure 5, reveal that more than 90% of GWP related emissions come from NH₃ shipping while

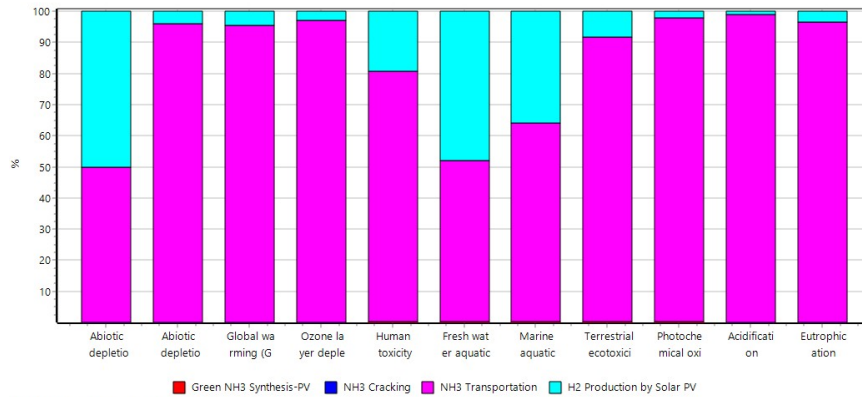


Figure 5: LCA results of hydrogen production from imported green ammonia

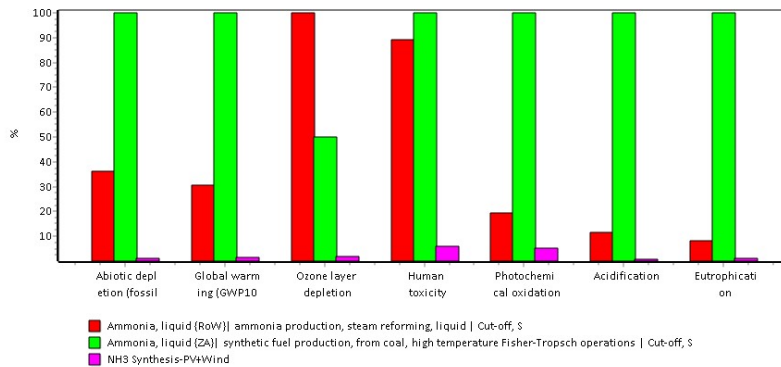


Figure 6: Comparison of LCA of ammonia production routes

hydrogen production by water electrolysis emits about 5% of total emissions. Ammonia synthesis unit results into only 0.12% while ammonia cracking process presents almost negligible amount of carbon related emissions as compared to other stages of process. Selected environmental indicators for life cycle impact assessment are explained below in table 1 along with the sources. Anthropogenic CO₂ emissions in global warming cause climate change, which in turn results in negative effects on human health and ecosystem quality. Similarly, other indicators such as the depletion of minerals and fossil fuels that come under abiotic depletion are also evaluated. Eutrophication, along with other economic indicators, was also evaluated in the analysis. CH₄-methane; CFC_s-chlorofluorocarbons; HCFC_s-hydro-chlorofluorocarbons; CH₃Br-bromomethane; PM10-particulate matter 10 μm; SO₂-sulphur dioxide; NMVOC-non-methane volatile organic compound.

3. Conclusion

Production of ammonia is a very mature process commercially, but source of hydrogen production is through steam reforming of natural gas (fossil fuel), which is a major source of high carbon emissions. We have compared our green ammonia production process with that of conventional ammonia production processes in Figure 6. According to Luke et al. (2020), ammonia production contributes to 1.3% of the world's total carbon dioxide emissions. Green ammonia production contributes only 4% of carbon dioxide emissions

as compared to conventional ammonia production routes and that 4% are due to the raw materials required for electricity production. Steam reforming uses natural gas route use natural gas and ammonia production as synthetic fuel by Fischer-Tropsch process use coal, so it has the highest contributions to the emissions. Results indicate that importing green ammonia for meeting hydrogen demands is not environment friendly now due to huge amount of emissions related to NH₃ shipping, but it can be viable in the coming future. If marine transportation is replaced by ammonia as fuel by using fuel cells or ammonia fuelled engines (2 stroke engines), which is also under research work (MAN 2019), then transportation of green ammonia will be more beneficial for the environment.

Table 1: Economic Indicators, definition, and contribution sources (Chisalita et al. 2020)

Impact Category (Units per kg emission)	Definition	Relevant LCI Data
Abiotic Depletion (ADP) (kg Sb eq.)	Collective quantification of the impact caused by extraction of minerals due to inputs in the system	Extraction of mineral resources
Abiotic Depletion (fossil fuels) (AFFDP)(MJ)	Surplus energy (lower heating value) per extracted MJ, kg, or m ³ fossil fuel, as a result of lower quality resources; unavailable for use by future generations	Extraction of fossil fuel resources
Global Warming (GWP 100a) (kg CO ₂ eq.)	GWP potential for time horizon 100 years: The impact caused by emissions of greenhouse gases	CO ₂ , NO ₂ , CH ₄ , CFCS, HCFCS, CH ₃ BR
Ozone Layer Depletion (ODP) (kg CFC-11 eq.)	Thinning of the stratospheric ozone layer due to anthropogenic emissions due to greenhouse gases	CFCS, HCFCS, CH ₃ BR, Halons
Human Toxicity (HTP) (1,4-dichlorobenzene eq.)	Potential impacts of toxic substances on human health present in the environment	Human toxic substances
Photochemical oxidation (kg C ₂ H ₄ eq.)	Formation of reactive chemical compounds like ozone, by the action of sunlight on emissions of air pollutants	PM ₁₀ , NH ₃ , SO ₂ , NO _x , and NMVOC
Acidification (kg SO ₂ eq.)	Formation of acidic compounds as a result of the manufacturing process	SO _x
Eutrophication (kg PO ₄ eq.)	Collective quantification of formation of phosphorus compounds	Nitrogen and Phosphorus compounds

References

- Chisalita, D. A., Petrescu, L., & Cormos, C. C. (2020). Environmental evaluation of european ammonia production considering various hydrogen supply chains. *Renewable and Sustainable Energy Reviews*, 130, 109964.
- De Rosa, L., & Castro, R. (2020). Forecasting and assessment of the 2030 Australian electricity mix paths towards energy transition. *Energy*, 118020.
- Koj, J. C., Wulf, C., Schreiber, A., & Zapp, P. (2017). Site-dependent environmental impacts of industrial hydrogen production by alkaline water electrolysis. *Energies*, 10(7), 860.
- MAN, ES Technical paper, (November 2019). Engineering the future two-stroke green-ammonia engine.
- Morgan, Eric R., James F. Manwell, and Jon G. McGowan. "Sustainable ammonia production from US offshore wind farms: a techno-economic review." *ACS Sustainable Chemistry & Engineering* 5.11 (2017): 9554-9567.
- Nayak-Luke, R. M., & Bañares-Alcántara, R. (2020). Techno-economic viability of islanded green ammonia as a carbon-free energy vector and as a substitute for conventional production. *Energy & Environmental Science*, 13(9), 2957-2966.
- Philibert, C. (2017). Producing ammonia and fertilizers: new opportunities from renewables. *International Energy Agency*.

Optimal Design of Distillation Columns Using a Combined Optimisation Approach

Dian Ning Chia^a, Fanyi Duanmu^a and Eva Sorensen^{a,*}

^a*Department of Chemical Engineering, UCL, Torrington Place, London WC1E 7JE, UK*
^{*}*e.sorensen@ucl.ac.uk*

Abstract

Distillation is the most important separation process in the chemical industry despite being very energy-intensive, hence any attempt to optimise its design may potentially have a significant impact. In this work, a combined optimisation strategy based on Genetic Algorithm (GA, stochastic method) and an outer approximation method (OAERAP, deterministic method) is proposed and demonstrated to reliably and efficiently optimise different distillation processes by achieving significant energy and/or capital savings. Three different case studies are presented to compare the combined strategy to the stand-alone GA and OAERAP methods. The combined optimisation strategy shows excellent results, with the total annualised costs obtained from the combined strategy being similar or better than for the stand-alone methods. More importantly, the combined strategy requires much shorter CPU time, as well as significantly less manual effort, thus greatly increasing the time efficiency when compared to the OAERAP method.

Keywords: Optimisation, Genetic Algorithm, Outer-approximation, Distillation

1. Introduction

The optimisation of many chemical processing units, including distillation columns, is a Mixed Integer Nonlinear Programming (MINLP) problem, as well as a highly non-convex problem, leading to the existence of potentially multiple local optima (Javaloyes-Antón et al., 2013). Current optimisation methods used to solve these problems can be categorised into deterministic and stochastic methods, where the former requires the translation of the problem into a convex model to find the global optimum (Grossmann et al., 2005), an approach that can be quite challenging (Bonami et al., 2008). Stochastic methods do not require the model to be transformed, but will instead typically require a long time to converge. In this work, a combined optimisation strategy is proposed that exploits the ability of the stochastic method to easily identify the proximity of the global optimum for a non-convex problem, and the quick convergence of the deterministic method once close to that optimum. Several case studies will be used to illustrate the strategy, including a two column sequence, a dividing wall column (DWC), and a hybrid separation process.

2. Methodology

The proposed strategy uses a stochastic method in the form of Genetic Algorithm (GA) to find an approximate, preliminary, optimal design in the first step. In the second step, and based on the preliminary design from the first step, a deterministic method based on outer approximation (OAERAP) is used to find the global optimum. For the combined

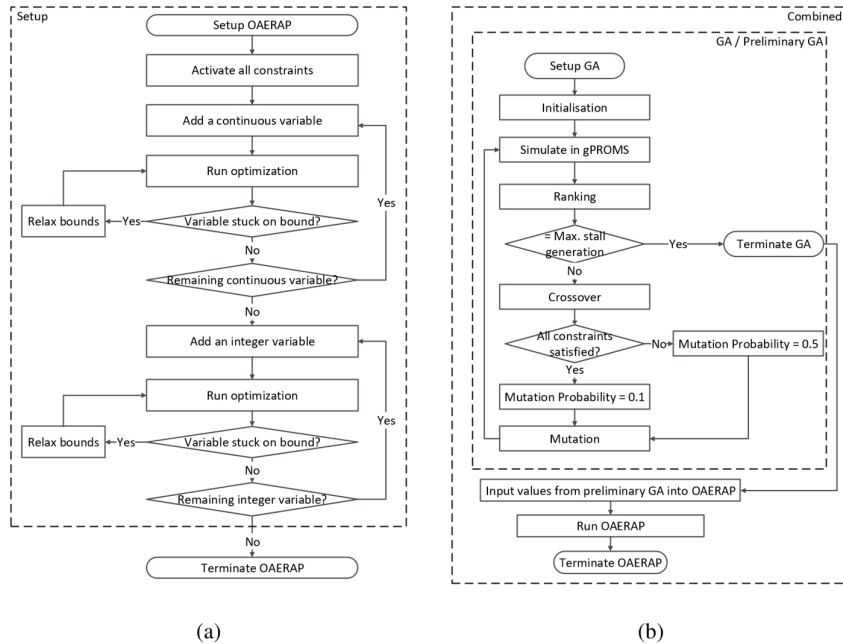


Figure 1. Flowcharts of (a) deterministic (OAERAP) method, (b) stochastic (GA) method and combined strategy.

approach, the values returned from the GA algorithm represent a valid solution to the model and the outer approximation initialisation is therefore non-problematic. All the case studies are performed in a process simulator, gPROMS ProcessBuilder, and the same distillation process model is used for all methods. gO:MATLAB is used to exchange data with MATLAB within which the external GA optimiser is implemented.

The OAERAP algorithm may easily fail to converge if the optimisation problem is highly non-convex and complex. Thus, a setup procedure as shown in Figure 1a is introduced to overcome the convergence issue of initialising the OAERAP algorithm. The OAERAP used in this work is built-in in gPROMS ProcessBuilder, and the procedure can be found in the gPROMS documentation (Process Systems Enterprise, 2020). The optimisation convergence is highly dependent on the variable bounds and the initial values provided, and the optimisation problem is therefore typically implemented in two stages. First, the continuous variables are added one-by-one and the optimisation problem is solved after each. Next, the integer/discrete variables are similarly included, and finally the whole problem optimised. It should be noted that for certain problems some of the variables may be linked, e.g. number of stages and feed locations for DWC, and these variables should then be added together. During optimisation, some of the variables may be stuck on bounds, which must then be relaxed and the problem re-optimised before adding another variable into the problem. To improve the convergence, the search space may subsequently be narrowed again before adding another variable. This manual procedure is clearly time consuming but unfortunately often necessary.

The procedure of the GA method is shown in the inner box in Figure 1b. In this work, the GA method was developed in MATLAB and was validated against several classical

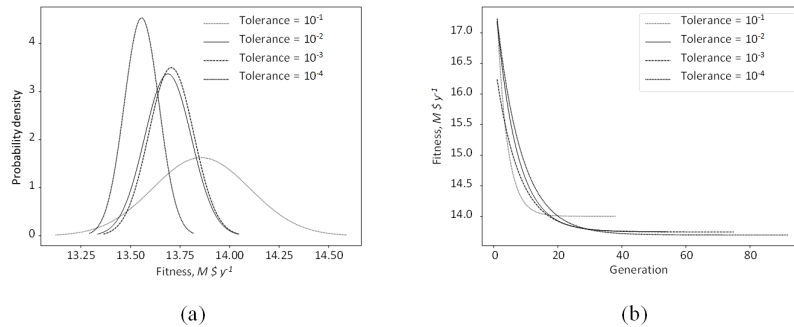


Figure 2. (a) Probability density function of fitness, and (b) Exponential regression for TAC-Generation, at different values of fitness tolerance (Case study 1).

mathematical functions (e.g. Schwefel function and constrained Rosenbrock function) and literature (Deep et al., 2009; Deb, 2000) (not shown). Based on this validation, the population size in this work is chosen as five times the number of optimisation variables, and the GA is stopped when the best fitness value remains constant (within the fitness tolerance) for 20 consecutive generations. The discrete crossover is used for all case studies (Umbarkar and Sheth, 2015). The top 50 % of the ranked chromosomes are chosen to participate in the crossover, and the elitism percentage is chosen as 10 %. To improve the performance of the GA, the mutation probability is set as 50 % initially and the value is changed to 10 % after a feasible (all constraints are met) design is found. Furthermore, the penalty function proposed by Deb (2000) is used to avoid guessing the R value. To improve the optimisation convergence, the following approaches were applied: 1) dynamic bounds were introduced for some variables (for example, the feed location should always be lower than the number of stages in a column, thus the upper bound of the feed location should be dynamically equal to the number of stages in its chromosome); and 2) the model was initialised using a feasible simulation with the number of stages close to the lower bound before each simulation. By applying these strategies, the percentage of infeasible simulation dropped from 90 % to 10 % in each generation.

The procedure of the combined strategy proposed in this work is shown in the outer box in Figure 1b. First, the GA method with a looser tolerance, and thereby a faster convergence, is used to locate a preliminary global optimum design. Figure 2 (based on a minimisation problem) shows that with a tighter fitness tolerance, the fitness is better but more generations are required. By trading off between fitness and generations (CPU time), a tolerance of 10^{-2} is chosen as this value provides good results but needs only half the generations compared to a 10^{-4} tolerance, which is the default value normally used in OAERAP and GA. Apart from the change of fitness tolerance, all other GA settings are kept the same. Once the preliminary GA design is obtained, this design becomes the initial values in the OAERAP problem and the entire optimisation is solved in one step without applying the setup procedure for the OAERAP.

3. Case Studies

The optimisation strategy is illustrated using three case studies. For each case, several optimisations are performed for each optimisation methods with almost identical solutions. In the following, the solution with the most common fitness and CPU time (referring to the peak in the probability density function) is presented. All simulations

Table 1. Case study 1 – Two column direct sequence: Comparison between GA, OAERAP, and the combined strategy.

Items	GA	OAERAP	Combined		Units
			Prelim. GA	Final	
Column 1					
Total stages	29	28	28	30	-
Feed stage	15	14	13	13	-
Distillate flowrate	334	335	332	335	$kmol h^{-1}$
Reflux ratio	1.42	1.41	1.49	1.38	$mol mol^{-1}$
Column 2					
Total stages	24	25	30	29	-
Feed stage	13	14	18	16	-
Distillate flowrate	332	331	333	331	$kmol h^{-1}$
Reflux ratio	1.21	1.21	1.21	1.15	$mol mol^{-1}$
Fitness and Time					
TAC	13.5112	13.4429	13.6970	13.2803	$M \$ y^{-1}$
Single CPU time	767	89	306	57	s
Total CPU time	767	-	363		s

were performed using a laptop with an Intel® Core™ i7-9850H CPU with 2.6 GHz and 16 GB memory.

3.1. Description

Case study 1 is the separation of an equi-molar benzene/toluene/o-xylene mixture separated in a two-column direct sequence. The feed is a saturated liquid at 0.37 bar with flow rate of $500 kmol h^{-1}$. The column operates at 0.37 bar with no pressure drop, and is modelled in gPROMS ProcessBuilder using “Distillation_column_gML”, which is an equilibrium model. For case study 2, the same separation task takes place in a dividing wall column (DWC), with all other information remaining the same as for case study 1. As there is no library model for a dividing wall column (DWC), a thermodynamically equivalent Petlyuk structure is considered instead. Case study 3 is a hybrid distillation-*per*vaporation column separating an azeotropic water-ethanol mixture. The feed is an equi-molar saturated liquid at 1.1 bar with flow rate of $500 kmol h^{-1}$. The column pressure is 1.1 bar with no pressure drop. The membrane model is a user-defined model coded in gPROMS ProcessBuilder which was validated against the membrane model by Luyben (2009). For all case studies, UNIQUAC is the thermodynamic model.

The objective function is to minimise the total annualised cost (TAC), including capital and operating costs, with the assumption of 8 years of payback period with an annual operating hour of $8400 h y^{-1}$, and a membrane lifetime of 2 years. The column sizing equations are obtained from Seider et al. (2016) and the cost equations from Sinnott and Towler (2020).

3.2. Results and Discussion

To evaluate the performance of each optimisation method, TAC and CPU time is chosen as the main indicators. Tables 1 to 3 show all the optimised variables, TAC, and CPU time for each case study. The tables show that the preliminary design obtained from GA in the proposed combined strategy has a slightly higher TAC but much lower CPU time when compared to the stand-alone GA method. The final optimal designs for the combined optimisation strategy (preliminary GA + OAERAP), show better TAC (up to

Table 2. Case study 2 – Dividing Wall Column (DWC): Comparison between GA, OAERAP, and the combined strategy.

Items	GA	OAERAP	Combined		Units
			Prelim. GA	Final	
Prefractionator					
Total stages	23	25	25	28	-
Feed stage	12	16	17	17	-
Main Column					
Total stages	44	46	45	47	-
Feed stages	9/33	8/34	9/35	6/35	-
Liq. sidedraw stages	18/9	19/8	18/9	18/6	-
Vap. sidedraw stage	33	34	35	35	-
Liq. side flowrates	332/166	328/178	332/185	330/161	$kmol h^{-1}$
Vap. side flowrate	601	640	643	598	$kmol h^{-1}$
Distillate flowrate	333	336	332	335	$kmol h^{-1}$
Reflux ratio	2.26	2.19	2.27	2.19	$mol mol^{-1}$
Fitness and Time					
TAC	9.5699	9.4990	9.6110	9.4931	$M \$ y^{-1}$
Single CPU time	13586	608	7100	319	s
Total CPU time	13586	-	7419	-	s

2% reduction) compared to the stand-alone GA method. Due to the fast convergence speed of the single OAERAP optimisation in the second step, the total CPU time of the combined strategy can be reduced by 19-53 % depending on the complexity of the system.

In general, the optimal TAC obtained from the combined strategy is close to the TAC obtained from the OAERAP method. However, the CPU time of OAERAP within the combined strategy (labelled as “Final” in the tables) was greatly reduced (36-76 %) when using the preliminary GA design as initial values. It should be noted that the CPU time for the standard OAERAP method is misleading as this is only the solution time and therefore does *not* take into account the considerable time required to set up the problem, and the trial and error needed to adjust the bounds etc. This time is therefore highly dependent on other factors such as the optimisation experience of the user, initial guesses for the values, and/or short-cut design method used. In our experience, and especially for complex systems, the OAERAP optimisation often fails to converge, gives values stuck on bounds, runs “infinitely” (e.g. for DWC more than 30 hours without yielding any results) with poor initial values, etc., which lead to a much higher time consumption compared with the combined strategy. Most of the optimisation time in the combined strategy is used in the preliminary GA which is far more robust and unlikely to fail hence not needing constant monitoring.

4. Conclusion

A hybrid optimisation strategy combining Genetic Algorithm (stochastic method) and OAERAP (deterministic method) is proposed in this paper and illustrated using three distillation-based case studies. The case studies show that using only the deterministic method may yield an infeasible solution if the bounds of the optimisation variables are too wide, and that to narrow these bounds, significant manual trial and error is required. In addition, the optimisation variables must typically be added manually and gradually to achieve convergence. For the stochastic method, the GA can yield a preliminary optimal

Table 3. Case study 3 – Hybrid process: Comparison between GA, OAERAP, and the combined strategy.

Items	GA	OAERAP	Combined		Units
			Prelim. GA	Final	
Column					
Total stages	35	38	35	37	-
Feed stages	34/31	37/34	34/32	36/33	-
Distillate flowrate	305	304	303	304	$kmol h^{-1}$
Reflux ratio	1.26	1.24	1.33	1.25	$mol mol^{-1}$
Membrane (Single Module)					
Mem. temp.	388	390	389	390	K
Area	406	357	363	357	m^2
Fitness and Time					
TAC	9.0639	8.8507	9.0468	8.8561	$M \$ y^{-1}$
Single CPU time	4291	17	3486	4	s
Total CPU time	4291	-	3490	-	s

design in a moderate amount of time, however, significantly more time is needed to converge to the actual optimum. For all the case studies, the proposed combined strategy shows a greatly reduced overall optimisation time. Although the combined strategy requires more computational effort compared to the stand-alone GA and OAERAP, the significant saving in time and manual effort, especially for complex systems, overwhelms the limitation of the combined strategy. To conclude, the proposed combined strategy makes optimisation easier, requiring less manual effort, and can yield a better global optimum in a shorter time. The combined strategy can also be used for other types of optimisation problems, e.g. based on a superstructure.

References

- P. Bonami, L. T. Biegler, A. R. Conn, G. Cornuéjols, I. E. Grossmann, C. D. Laird, J. Lee, A. Lodi, F. Margot, N. Sawaya, A. Wächter, 2008. An Algorithmic Framework for Convex Mixed Integer Nonlinear Programs. *Discrete Optimization* 5 (2), 186–204.
- K. Deb, 2000. An Efficient Constraint Handling Method for Genetic Algorithms. *Computer Methods in Applied Mechanics and Engineering* 186 (2–4), 311–338.
- K. Deep, K. P. Singh, M. L. Kansal, C. Mohan, 2009. A Real Coded Genetic Algorithm for Solving Integer and Mixed Integer Optimization Problems. *Applied Mathematics and Computation* 212 (2), 505–518.
- I. E. Grossmann, P. A. Aguirre, M. Barttfeld, 2005. Optimal Synthesis of Complex Distillation Columns Using Rigorous Models. *Computers & Chemical Engineering* 29 (6), 1203–1215.
- J. Javaloyes-Antón, R. Ruiz-Femenia, J. A. Caballero, 2013. Rigorous Design of Complex Distillation Columns Using Process Simulators and The Particle Swarm Optimization Algorithm. *Industrial and Engineering Chemistry Research* 52 (44), 15621–15634.
- W. L. Luyben, 2009. Control of A Column/Pervaporation Process for Separating The Ethanol/Water Azeotrope. *Industrial and Engineering Chemistry Research* 48 (7), 3484–3495.
- Process Systems Enterprise, 2020. gPROMS ProcessBuilder version 1.4.
URL <https://www.psenterprise.com/products/gproms/processbuilder>
- W. D. Seider, D. R. Lewin, J. D. Seader, S. Widagdo, R. Gani, K. M. Ng, 2016. *Product and Process Design Principles: Synthesis, Analysis and Evaluation*, 4th Edition. Wiley.
- R. Sinnott, G. Towler, 2020. *Chemical Engineering Design*, 6th Edition. Elsevier.
- A. Umbarkar, P. Sheth, 2015. Crossover Operators in Genetic Algorithms: A Review. *ICTACT Journal on Soft Computing* 6 (1), 1083–1092.

Process and product design for the simultaneous synthesis of xylitol and sorbitol from biomass

Guillermo Galán,^a Mariano Martín,^{a,*} Ignacio E. Grossmann.^b

^a*Department of Chemical Engineering. University of Salamanca. Plz. 1-5, 37008, Salamanca, Spain*

^b*Department of Chemical Engineering. Carnegie Mellon University, 5000 Forbes Ave, 15213, Pittsburgh PA, USA*

* mariano.m3@usal.es

Abstract

This work deals with the design of integrated facilities for the simultaneous production of xylitol and sorbitol from lignocellulosic biomass. A superstructure of alternatives is formulated. Different pretreatments, such as dilute acid or AFEX for the fractionation of the biomass are considered. Next, after hydrolysis, the C5 and C6 sugars are processed separately, for which either a catalytic or a fermentation stage are evaluated. Finally, the purification of any of the products is carried by crystallization in a multi-state evaporator system. Surrogate models based on experimental data for the different stages of the process including the pre-treatment and the synthesis and first principles models for crystallization are developed to formulate the superstructure MINLP optimization model. The optimal results suggest the use of dilute acid and the catalytic system, and a system of three crystallizers is selected. This optimal flowsheet is also used to evaluate the performance of different raw materials into the product portfolio as a function of the biomass composition as well as the design of the optimal biomass.

Keywords: Integrated processes, Biomass, added valued products, Optimization.

1. Introduction

The chemical industry is undergoing a transformation towards a more sustainable future starting from the use of renewable instead of fossil resources, which constitutes the 7th principle of green chemistry (Erythropel et al., 2018). Biomass has emerged as a rich raw material towards the production of energy and chemicals (Keshwani and Chen 2009). The diversification of the use of biomass results in the need to use it for products beyond fuels and energy. In particular xylitol and sorbitol are considered in the production of dietetic foods as well as in pharmaceutical applications (mainly as a carrier), cosmetics industry (as an emulsion stabilizer), moisturizer, texturizer and softener having being listed on the top 12 high value-added building block intermediate chemicals that can be produced from renewable biomass resources (Werpy et al., 2004). Xylitol can be obtained from xylose via fermentation or catalytic hydrogenation (Rafiqul and Sakinah, 2013). Sorbitol is obtained from glucose via catalytic hydrogenation or from fructose via fermentation (van Gorp et al., 1999; Chun and Rogers, 1988). Fructose can be obtained from glucose via

isomerization (Illanes et al., 1992). In spite of previous efforts in the design of production processes, no attempt has been made on the design of an integrated process for the simultaneous production of both from lignocellulosic biomass.

2. Process description

Biomass starts with a size reduction step before pretreatment. There are many alternative pretreatments (Sung and Cheng, 2002). However, among them the ones that have reached commercial exploitation are (1) dilute acid (H_2SO_4) pretreatment, and (2) ammonia fibre explosion (AFEX). Once the physical structure of the switchgrass is broken, we separate cellulose from hemicelluloses sugars. Cellulose is hydrolyzed at 45-50°C for 3 days to obtain glucose (Piccolo and Bezzo, 2009). Glucose can be isomerised into fructose for its fermentation into sorbitol or catalytically hydrogenated. Next, the sorbitol is recovered by crystallization in a three-effect evaporator. Xylose follows a similar path into xylitol. It can be either fermented or catalytically hydrogenated into it. The superstructure of alternatives is presented in Figure 1.

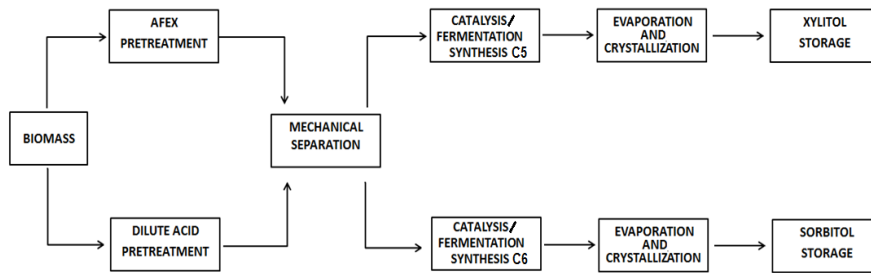


Figure 1.-Superstructure for the production of xylitol and sorbitol

3. Modelling approach

Surrogate models based on detailed simulations and/or experimental data are developed for particular units such as the pretreatments, the ammonia recovery column, and the crystallizers. The model for the superstructure is expressed in terms of total mass flows, component mass flows, component mass fractions, and temperatures of the streams in the network.

Pretreatment: Both pretreatments, AFEX and dilute acid, are modelled using response surface models developed experimental data from the literature to estimate the yield and process simulators for the recovery of ammonia. For the acid pre-treatment the yield is a function of the operating temperature, T , the acid concentration, %, the enzyme added, Enz , and the residence time (Shi et al., 2011).

$$Y_{cellu} = f(T, \%, Enz, t)$$

$$Y_{Hemi} = f(T, \%, Enz, t) \quad (1)$$

In the case of the AFEX, the yield to sugars depends on the water and ammonia added per kg of biomass, $NH_{3,added}$ or Wa_{added} , the residence time and the operating temperature, T (Garlock et al., 2012).

$$Y_{cellu} = f(T, NH_{3,added}, Wa_{added}, t) \quad (2)$$

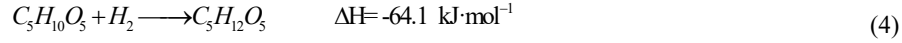
A surrogate model is developed to evaluate the effect of the feed composition on the operating conditions of the column that recovers ammonia.

Xylitol production:

Via fermentation the reaction consumes ammonia and hydrogen as follows, operating at 30°C and 1 bar with a conversion of 92% (Mountraki et al., 2017).



The catalytic synthesis reaction takes place at 100-120 °C and 40-60 bar for 3-4 h. A model to estimate the conversion is developed from the kinetics results in the literature (Ngoc Pham et al., 2016).

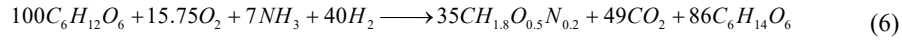


Sorbitol production:

Via fermentation an additional step is required to isomerise the glucose into fructose



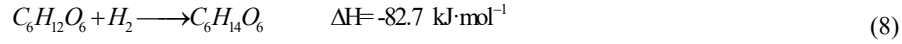
That is next fermented into sorbitol as follows:



Based on experimental data, the conversion of the fructose into sorbitol can be predicted by the following correlations

$$X_{(\text{fructose})} = -1.40 \cdot 10^{-6} (\text{concentration}(\text{g}\cdot\text{L}^{-1}))^2 + 8.700000 \cdot 10^{-4} (\text{concentration}(\text{g}\cdot\text{L}^{-1})) + 0.819 \quad (7)$$

The catalytic hydrogenation of glucose is carried out at 100-140 °C, 40-60 bar



Crystallization: Each crystallizer is modelled using mass and energy balances. Note that the pressure at the evaporation chamber will depend on the steam generated in the previous one and the composition of the mixture to concentrate. The solubility as a function of the temperature has been correlated from the literature data (Zambé et al, 2001) as well as the ebulloscopic effects. S represents steam, s condensed stream, F feed, E, evaporated water, L liquid product and C crystals.

$$F_j \cdot x_{j,f,i} = C_{j,i} + L_j x_{j,l,i} \quad \forall j \in \{\text{effects}\}, i \in \{X, S\}$$

$$H_s + H_f = H_s + H_e + H_l + H_c \quad (9)$$

$$H_i = \sum_i f_i \cdot \Delta h_{i,f} + F \int_{T_{ref}}^T c_p(X, T) dT$$

4. Optimization procedure

The superstructure is decomposed into four flowsheet alternatives: one per pretreatment and one for each of the synthetic routes, either fermentation or catalysis. An NLP problem is solved for each one where the objective function consists of a simplified production costs given by eq. (10)

$$Z = P_{Xylitol} m_{Xylitol} + P_{Sorbitol} m_{Sorbitol} - P_{Steam} \sum_i \frac{Q_i}{\lambda} - P_{H_2} \quad (10)$$

Subject to the models described in section 3. Each NLP consists of around 2700 eqs and 3800 variables and it was solved with GAMS requiring 12.5 s of CPU-time. After the optimization, a heat exchanger network is designed to reduce energy consumption. Finally, a detailed economic evaluation of the alternatives is performed to compute the

production and investment costs of the facility using the cost correlations in Martín and Grossmann, (2011) and the procedure described in that work.

5. Results

In this section the major operating results as well as the economics of the different alternatives are presented. In addition, a subsection commenting on biomass design has also been included.

5.1. Process analysis

The superstructure of alternatives has been decomposed to evaluate the performance of each of the alternative processes including two pre-treatments and two synthetic lines. Table 1 shows the major results of each of the optimized paths.

Table 1.- Major yields of the alternative production paths

	Dilute Acid Catalysis Hydrogenation	AFEX Catalysis Hydrogenation	Dilute Acid Fermentation	AFEX Fermentation
kg H ₂ / Kg Biomass	0.007	0.005	0.001	0.001
kg H ₂ SO ₄ / Kg Biomass	0.024	-	0.017	-
kg NH ₃ / Kg Biomass	-	1.052	-	1.052
kg Xylitol _{crystallized} / Kg Biomass	0.259	0.187	0.166	0.120
kg Sorbitol _{crystallized} / Kg Biomass	0.282	0.231	0.142	0.102
kg Steam/ Kg Biomass	0.505	0.587	2.013	1.374

Two main results can be highlighted. Dilute acid pretreatment and catalytic hydrogenation show the largest yield to products. In general, fermentation requires more energy than catalytic synthesis due to the diluted products and operating conditions, while dilute acid pretreatment consumes also more steam than AFEX, since it operates at milder temperature. Note that the isomerisation of the glucose is endothermic, increasing the need for utilities. The best objective function corresponds to the use of dilute acid pretreatment followed by catalytic hydrogenation. Even though further results of the four paths will be presented, the focus of the analysis will be that this best process. But for the case of the fermentation-based processes, three effect evaporation systems are used operating under vacuum.

5.2. Process economics

Table 2 shows the investment and production cost of the four alternatives. The fermentation-based processes require additional investment due to the need for isomerisation of the glucose and the high residence times in both the isomerisation and the fermentation itself, that doubles the one when catalytic hydrogenation is considered. In addition, the larger consumption of steam in the case of fermentation also increases the production costs which double the ones using catalytic synthesis. Figure 2 shows the detail of the breakdown for the production costs, Figure 2a, and investment costs, Figure 2b. With regards to production costs, 70% of them correspond to the raw material so that the more efficient process in using it will yield better economics. That is the reason for the catalytic hydrogenation to be the best synthetic path and the dilute acid the selected pretreatment. The best process reports an investment cost of 121 M € and a production cost of 18.74 M€/yr. In terms of investment, biomass pretreatment towards sugar production represents over 50% of the investment cost. That is the reason for biomass-

based processes to show a barrier to enter the market versus the more established use of fossil alternatives. The second major contribution corresponds to the heat exchanger network responsible for reducing steam use at the cost of the investment required in the network.

5.3. Biomass design

The inverse problem is solved here. Instead of using a fixed biomass composition, belonging to a lignocellulosic species such as switchgrass, the optimal flowsheet is used to determine the best composition within the typical ranges of hemicelluloses, cellulose and lignin for the simultaneous production. The resulting biomass composition corresponds to 15% water, 20% cellulose, 40% hemicellulose, 15% lignin and 5% ash. The higher cost of the xylitol results in a biomass with larger fraction of hemicelluloses. The cellulose represents a secondary role due to lower market price of the sorbitol and higher amount of steam used respect to xylitol production.

Table 2.- Summary of economic analysis

	Dilute Acid Catalysis Hydrogenation	AFEX Catalysis Hydrogenation	Dilute Acid Fermentation	AFEX Fermentation
Investment (MM€)	120.78	146.44	323.74	273.45
Production Cost (MM€)	18.74	22.94	45.80	37.80

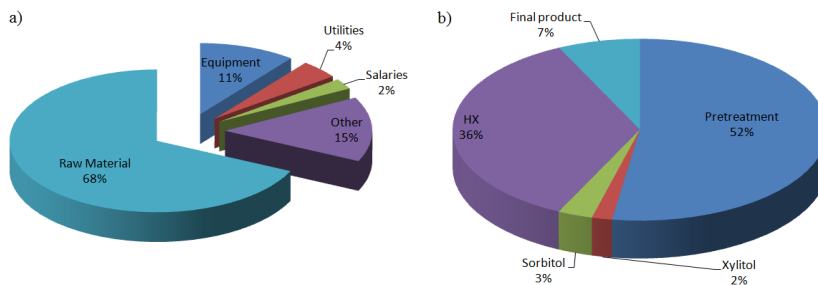


Figure 2.- Breakdown of a) production costs b) investment cost.

Among typical biomasses such as wheat straw, corn stover, switchgrass, hybrid poplar, forest residues from pine, birch, spruce and sugarcane bagasse the one that fits the best is sargassum algae (*sargassaceae*) with a composition of 20.48% cellulose and 43.19% hemicellulose and similar lignin content as the optimization results (Rabemanolontsoa and Saka, 2013). Further sensitivity analysis on the effect of the product prices is to be carried out. Working with the optimal biomass the production cost and investment decrease almost 0.5 M€ each one but there is a reduction in the production cost of xylitol almost two times, from 0.59 €/kg to 0.37 €/kg. (Table 3)

Table 3.- Economic parameters comparison

	Investment (M€)	Production Cost (M€/yr)
Switchgrass	120.78	18.74
Best biomass	120.50	18.37

6. Conclusions

The best alternative obtained to produce xylitol and sorbitol from biomass is the catalytic synthesis process with a dilute acid pretreatment due to a highest $\text{kg}_{\text{xylitol}}/\text{kg}_{\text{biomass}}$ and $\text{kg}_{\text{sorbitol}}/\text{kg}_{\text{biomass}}$ and low use of steam, reaching the lowest production and investment cost. Following this optimization, the biomass design was made to determinate the better biomass composition, reaching a key percent of 40% hemicellulose and 20% cellulose, finding this composition in the sargassum algae (*sargassaceae*), highly available in almost of the world.

The use of the best biomass would yield a saving of around 0.5 M€ in the production and investment costs, and reduction of xylitol production cost from 0.59 €/kg to 0.37 €/kg.

Acknowledgement

The authors thank the PSEM3 GIR at USAL, the CAPD at CMU and the funding received from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 778168. GG appreciates the FPU PhD fellowship from the Spanish Government

References

- U.H. Chun, P. L. Rogers. 1988 Sorbitol and gluconic acid production Appl Microbiol Biotechnol. 29:19-24
- H.C. Erythropel, J.B. Zimmerman, T. M. de Winter, L. P., F. Melnikov, C. H. Lam, A. W. Lounsbury, K. E. Mellor, N. Z. Janković, Q. Tu, L. N. Pincus, M. M. Falinski, W. Shi, P. Coish, D. L. Plata P. T. Anastas 2018. The Green ChemisTREE: 20 years after taking root with the 12 principles Green Chem. 20, 1929–1961.
- K. van Gorp, E. Boerman, C.V. Cavenaghi, P.H. Berben, 1999. Catalytic hydrogenation of fine chemicals: sorbitol production. Catalysis Today 52, 349–361.
- D.R. Keshwani, J.J. Cheng, 2009. Switchgrass for bioethanol and other value-added applications: a review. *Bioresour Technol.*, 100, 1515–1523
- R.J, Garlock, V. Balan, B.E. Dale, 2012. Optimization of AFEX™ pretreatment conditions and enzyme mixtures to maximize sugar release from upland and lowland switchgrass. *Biores Technol.* 104, 757–768.
- A. Illanes, M.E. Zúñiga, S. Contreras, A. Guerrero, 1992. Reactor design for the enzymatic isomerization of glucose to fructose. *Bioprocess Engineering*, 7(5), 199–204.
- M. Martín, Grossmann, I.E. 2011. Energy Optimization of Bioethanol Production via Gasification of Switchgrass. *AIChE J.* 57, 12, 3408-3428.
- A.D. Mountraki. K.R. Koutsospyros, B.B. Mlaya, A.C. Kokossis 2017. Selection of Biorefinery Routes: The Case of Xylitol and its Integration with an Organosolv Process. *Waste Biomass Valor*, 8, 2283-2300.
- T. Ngoc Pham, A Samikannu, AR. Rautio, K.L. Juhasz, Z. Konya, J. Wärna, K. Kordas, J.P. Mikkola 2016. Catalytic hydrogenation of D-xylose over Ru decorated carbon foam catalyst in a SpinChem rotating bed reactor. *Topic in Catalysis*, 59, 1165-1177.
- H. Rabemanolontsoa, S. Saka. 2013, Comparative study on chemical composition of various biomass species. *RSC Advances*, 12, 3757- 4094.
- S. M. Rafiqul & A. M. Mimi Sakinah. 2013. Processes for the Production of Xylitol—A Review, *Food Reviews International*, 29:2, 127-156
- T. Werpy, G. Petersen, A. Aden, J. Bozell, J. Holladay, A. Manheim, D. Eliot, L. Lasue . S. Jones Top Value-Added Chemicals from Biomass, vol. 1, U. S. Department of energy, Oak Ridge, TN (2004)
- C. Piccolo, F. Bezzo F. 2009 A techno-economic comparison between two technologies for bioethanol production from lignocelluloses. *Biomass Bioener* 33, 478 – 91.

- J. Shi, M.A. Ebrik, C.E. Wyman 2011. Sugar yields from dilute sulfuric acid and sulfur dioxide pretreatments and subsequent enzymatic hydrolysis of switchgrass. *Biores Technol*, 102: 8930–8.
- Y. Sung, J. Cheng. 2002. Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour Technol*, 83: 1-11.
- A. Zambé, A. Lee, D. Storey. 2001. Polyols in confectionery: the route to sugar-free, reduced sugar and reduced calorie confectionery. *British Journal of Nutrition* 85, S31-S45.

An approach for simultaneous computer-aided solvent design and process design for CO₂ chemical absorption processes

Ye Seol Lee^a, Amparo Galindo^a, George Jackson^a and Claire S. Adjiman^{a,*}

^a*Dept. of Chemical Engineering, Centre for Process Systems Engineering and Institute for Molecular Science and Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK*
c.adjiman@imperial.ac.uk

Abstract

In the field of Computer-Aided Molecular and Process Design (CAMPD), a variety of solution methods have been developed to handle the complexities associated with the non-convexity and non-linearity of molecular structure-property and process models. However, most algorithms are prone to failing to generate feasible solutions when the integrated solvent-process model renders a significant portion of the search space infeasible. In this work, we propose a solution approach for the integrated design of an optimal chemical absorption process in which tailored feasibility tests are incorporated into a process optimisation problem. The solution approach allows the exploration of a design space without unnecessary difficulties by recognising infeasibilities. The effectiveness of the approach is demonstrated on an aqueous amine solvent-based CO₂ capture process.

Keywords: CAMPD, carbon dioxide capture, amine-based solvent, MINLP

1. Introduction

Amine-based chemical absorption processes are regarded as some of the most effective and mature technologies for post-combustion carbon dioxide (CO₂) removal from large point sources (Rochelle, 2009). The conventional solvents used for this chemical absorption are monoethanolamine (MEA) and methyldiethanolamine (MDEA). However, the use of such solvents requires high energy consumption for solvent regeneration, consequently introducing significant economic penalties that are a concern in many contexts such as flue gas cleaning in power plants (Lee et al., 2013). In order to overcome this adverse effect, substantial research efforts have been directed towards the development of novel solvents.

A promising research direction to accelerate the identification of new solvents is the development of Computer-Aided Molecular and Process Design (CAMPD) techniques. CAMPD offers a systematic methodology to explore a very large molecular domain in which the overall performance of a process is evaluated by capturing mutual interaction between molecular properties and process performance (Adjiman et al., 2014). While a variety of strategies for the solution of CAMPD have been developed, many algorithms resort to decomposition to overcome numerical challenges. These arise mainly because the integrated solvent-process model exhibits a number of infeasible regions, thereby making it challenging to generate a feasible solution with some solvent structures. This strategy may however result in convergence to a suboptimal solution. Thus, the development of robust algorithms that make it possible to avoid infeasibilities during the exploration of a large search space is an important enabling step for CAMPD.

In this work, a robust optimisation framework for the integrated design of an optimal aqueous solvent and the associated CO₂ chemical absorption process is presented. The focus is on enabling the solution of such problems when a large set of potential molecules is entailed. New feasibility tests are introduced as an extension of the approach proposed by (Gopinath et al., 2016) for physical absorption, to ensure applicability to the more complex setting of chemisorption. The design of the tailored feasibility tests is focused on the screening of infeasible solvent candidates and the identification of feasible ranges of process conditions by

investigating: 1) physicochemical properties of the solvent mixture; 2) environmental impact arising from solvent loss; and 3) thermodynamic feasibility of satisfying separation targets. The efficiency of the proposed algorithm is demonstrated on a case study.

2. Problem Formulation

2.1. Process and cost model

The integrated design of an aqueous solvent and a CO₂ chemical absorption process is considered to determine an optimal solvent structure and process conditions. This CAMPD problem is often posed as Mixed-Integer Nonlinear Programming (MINLP) problem. The generic mathematical formulation of the design problem can be defined by:

$$\begin{aligned}
 \min_{\mathbf{x}, \mathbf{n}} \quad & f(\mathbf{x}, \mathbf{n}), \\
 \text{s.t.} \quad & \mathbf{g}(\mathbf{x}, \mathbf{n}) \leq \mathbf{0} \\
 & \mathbf{h}(\mathbf{x}, \mathbf{n}) = \mathbf{0} \\
 & \mathbf{C}\mathbf{n} \leq \mathbf{d} \\
 & \mathbf{x} \in \mathbb{R}^n, \mathbf{n} \in \mathbb{N} \subset \mathbb{Z}^q
 \end{aligned} \tag{1}$$

where \mathbf{x} is a n -dimensional vector of continuous process variables, \mathbf{n} is a q -dimensional vector of integer variables that define the molecular design space, $\mathbf{g}(\mathbf{x}, \mathbf{n})$ is a vector of inequality constraints that represent design and feasibility constraints, $\mathbf{h}(\mathbf{x}, \mathbf{n})$ is a vector of equality constraints that include structure-property models and process models and a set of linear equations $\mathbf{C}\mathbf{n} \leq \mathbf{d}$ represents molecular feasibility constraints and bounds on vector \mathbf{n} .

A schematic of the process configuration considered is shown in Figure 1. The process and cost estimation models are based on Alhajaj et al. (2016), who adopted an equilibrium-stage model combined with a lumped hydrodynamic model for process optimisation. Total annualised cost (TAC) is selected as an objective function to measure the overall economic performance for given molecules. We consider three key process design variables: the temperature of the solvent entering the absorber (T_0), the lean solvent loading (θ_0) and the desorber pressure ($P_{N,d}$). For a detailed description of the modelling strategy and assumptions used, the reader is referred to Alhajaj et al. (2016).

2.2. Molecular design space

The molecular design space is constructed by selecting groups present in typical CO₂ capture solvents and taking into account the availability of group contribution values. It is generated from the following 13 functional groups: NH₂CH₂, NH₂CH, NH₂C, NHCH₃, NHCH₂, NHCH, NCH₃, NCH₂, CH₃, CH₂, CH, C, OH. The property prediction methods of Hukkerikar et al. (2012) are used for the normal melting temperature (T_{mp}), the auto-ignition temperature (T_{AIT}), the flash point (T_{fp}), and the LC50 toxicity ($LC_{50,mgL}$). The method of Hsu et al. (2002) is used to predict surface tension (σ) and viscosity (μ). The SAFT- γ Mie group contribution equation of state (Papaioannou et al., 2014; Perdomo et al., 2020) is used to predict the fluid-phase behaviour of water, solvent and CO₂ mixtures. To make use of this thermodynamic model, the original functional groups are translated into the following SAFT- γ Mie groups: NH₂, NH, N, CH₂, CH, C, CH₂OH. The total number of functional groups in the molecule is limited by bounds $[n_t^L, n_t^U] = [3, 12]$. The total number of groups with amine and hydroxyl functionality are constrained by $[n_a^L, n_a^U] = [1, 2]$ and $n_{OH}^U = 3$, respectively. Because the two sets of functional groups are not exactly equivalent, the set of potential molecule is defined by the intersection of the two sets. This means for instance, that OH can only appear in the designed solvent when connected to CH₂. An additional

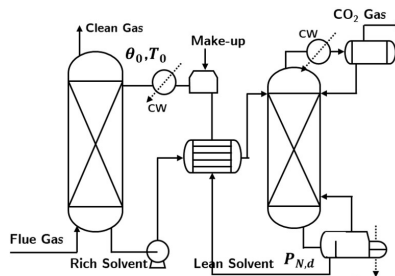


Figure 1: Schematic of the CO₂ chemisorption process

constraint, $n_{\text{CH}_2} \geq n_{\text{OH}}$, is imposed to enforce this.

3. Feasibility Tests

In this section, we describe four feasibility tests that are applicable to CAMPD for chemical absorption, as described in section 2.

3.1. Test 1: Solvent property feasibility

Test 1 is employed to evaluate whether the properties of the pure candidate solvent or those of the aqueous solution are within suitable ranges. Test 1 is formulated with explicit constraints on eight essential properties that are independent of the optimal process conditions. These include T_{bp} , $LC_{50,\text{mgL}}$, T_{AIT} , T_{fp} , T_{mp} , μ at lowest possible solvent handling temperature and bubble point (T_{bb}) at atmospheric pressure. The resulting formulation is as follows:

$$\begin{aligned} T_{sh}^U - T_{\text{bp}}(P = 1\text{atm}, \mathbf{n}) &\leq 0 & T_{\text{mp}}(P = 1\text{atm}, \mathbf{z}(\mathbf{n})) - T_{sh}^L &\leq 0 \\ LC_{50,\text{mgL}}(\mathbf{n}) - LC_{50,\text{mgL}}^U &\leq 0 & \mu(T_{sh}^L, P = 1\text{atm}, \mathbf{z}(\mathbf{n})) - \mu^U &\leq 0 \\ T_{sh}^U - T_{\text{fp}}(P = 1\text{atm}, \mathbf{n}) &\leq 0 & T_{\text{bb}}(P = 1\text{atm}, \mathbf{z}(\mathbf{n})) - T_s^U &\leq 0 \\ T_{op}^U - T_{\text{AIT}}(P = 1\text{atm}, \mathbf{n}) &\leq 0 \end{aligned} \quad (2)$$

where $\mathbf{z}(\mathbf{n})$ is a composition of aqueous solvent mixture, T_{sh}^L and T_{sh}^U are lower and upper bounds on solvent handling temperature, and T_{op}^U is an upper bound on the process operating temperature to avoid potential risks of solvent degradation in the reboiler.

3.2. Test 2: Solvent phase stability

Biphasic solvents are receiving increasing attention as alternatives to conventional solvents. They undergo liquid-liquid separation upon heating or CO₂, and only the CO₂-rich phase is sent to the desorber for regeneration, thereby reducing energy requirements and equipment costs (Papadopoulos et al., 2020). However, it is necessary for the aqueous solvent to form a homogeneous liquid phase for the successful operation of the absorber. Test 2 is designed to examine whether the absorber can operate outside the liquid-liquid region across its operating temperature range with the selected solvent. This is done by identifying the highest solvent concentration (W_{solvent}^*) at which only one liquid phase can form, at conditions corresponding to the absorber pressure (P_{N_a}) and the maximum temperature of the solvent leaving the absorber ($T_{N_a}^U$). If the highest solvent concentration obtained is less than the upper bound on solvent concentration, i.e. $W_{\text{solvent}}^* < W_{\text{solvent}}^U$, the molecule is discarded from the search space. Here, the upper bound on the solvent concentration is set to 0.4 kg/kg to prevent corrosion and foaming.

3.3. Test 3: Absorption capacity feasibility

Test 3 is introduced to eliminate operating conditions or solvents that cannot achieve the required absorption capacity. Test 3 is used to determine an upper bound on the lean solvent loading (θ_0^U) based on thermodynamic feasibility on the bottom stage (N_a). A constraint proposed by Gopinath et al. (2016) is also included in the problem to guide the solution in the desired direction of separation:

$$\begin{aligned} \theta_0^U &= \max_{T_{N_a}, \mathbf{z}} \theta_0 \\ \text{s.t.} \quad & [\mathbf{y}_{N_a} \ \mathbf{x}_{N_a} \ \alpha \ \beta]^T = \mathbf{f}_{\text{ext}}(\mathbf{z}, T_{N_a}, P_{N_a}, \mathbf{n}) \\ & \alpha \geq \epsilon_{ph}, \quad \beta \geq \epsilon_{ph} \\ & x_{N_a, \text{CO}_2} - y_{N_a+1, \text{CO}_2} - \frac{(y_{N_a+1, \text{CO}_2} - x_{0, \text{CO}_2})}{(y_{N_a+1, N_2} - x_{0, N_2})} (x_{N_a, N_2} - y_{N_a+1, N_2}) > 0 \\ & x_{N_a, N_2} < y_{N_a+1, N_2}, \quad x_{0, \text{CO}_2} \leq x_{N_a, \text{CO}_2} \\ & W_{N_a, \text{solvent}} \leq W_{0, \text{solvent}} + \epsilon_s \\ & 0 < z_i < 1 \quad \forall i \in \mathcal{NC}, \quad \sum_{i \in \mathcal{NC}} z_i = 1 \\ & \max(T_{mp}(\mathbf{n}) + 10, T_0^L) \leq T_{N_a} \leq T_{op}^U \end{aligned} \quad (3)$$

where the subscript a denotes the absorber, the set $\mathcal{NC} = \{ \text{H}_2\text{O}, \text{solvent}, \text{CO}_2, \text{N}_2 \}$ and \mathbf{y}_{N_a} , \mathbf{x}_{N_a} are the compositions of the vapour and liquid phases at equilibrium, $W_{N_a, \text{solvent}}$ is liquid phase solvent concentration in weight fraction at the bottom stage of the absorber, $W_{0, \text{solvent}}$ is solvent concentration at the absorber inlet, α and β are the vapour and liquid phase fractions, fl_{ext} an isobaric-isothermal flash calculation, ϵ_{ph} is a positive value to ensure the coexistence of vapour and liquid phases at stage N_a and ϵ_s is a positive value to avoid the excessive loss of water in aqueous solvents. In this study, ϵ_{ph} is set to 10^{-3} and ϵ_s is set to 10^{-1} respectively.

Equivalently, a lower bound on the lean loading (θ_0^L) is obtained by investigating the minimum achievable loading on the bottom stage (N_d) of the desorber with respect to global composition \mathbf{z} and desorber operating temperature (T_{N_d}).

$$\begin{aligned}
 \theta_0^L = & \min_{\theta_0, T_{N_d}, \mathbf{z}} \theta_0 \\
 \text{s.t.} & [\mathbf{y}_{N_d} \ \mathbf{x}_{N_d} \ \alpha \ \beta]^T = fl_{ext}(\mathbf{z}, T_{N_d}, P_{N_d}, \mathbf{n}) \\
 & \alpha \geq \epsilon_{ph}, \quad \beta \geq \epsilon_{ph} \\
 & W_{N_d, \text{solvent}} \geq W_{0, \text{solvent}} - \epsilon_s \\
 & 0 < z_i < 1 \quad \forall i \in \mathcal{NC} \\
 & \sum_{i \in \mathcal{NC}} z_i = 1 \\
 & \max(T_{mp}(\mathbf{n}) + 10, T_{N_d}^L) \leq T_{N_d} \leq T_{op}^U \\
 & P_{N_d}^L \leq T_{N_d} \leq \min(P_b(\mathbf{n}), P_{N_d}^U)
 \end{aligned} \tag{4}$$

where subscript d denotes the desorber and $W_{N_d, \text{solvent}}$ is liquid phase solvent concentration in weight fraction at the bottom stage (N_d) of the desorber.

If the bounds on lean loading violate the condition $\theta_0^L \leq \theta_0^U$, the candidate solvent is discarded from the molecular design space.

3.4. Test 4: Operating pressure

Given the lean loading range $[\theta_0^L, \theta_0^U]$ generated from Test 3, Test 4 is designed to provide tighter pressure bounds on the desorber, $P_{N_d}^U$. This is done by finding the highest pressure at which: 1) a vapour-liquid phase split occurs in the bottom stage of the desorber; and 2) there exists a lean solvent composition that satisfies the target lean loading. An additional constraint on the solvent concentration is imposed in order to prevent the optimisation algorithm from converging to a trivial solution. The resulting formulation is as follows:

$$\begin{aligned}
 P_{N_d}^U = & \max_{P_{N_d}, T_{N_d}, \mathbf{z}} P_{N_d} \\
 \text{s.t.} & [\mathbf{y}_{N_d} \ \mathbf{x}_{N_d} \ \alpha \ \beta]^T = fl_{ext}(\mathbf{z}, T_{N_d}, P_{N_d}, \mathbf{n}) \\
 & \alpha \geq \epsilon_{ph}, \quad \beta \geq \epsilon_{ph} \\
 & 0 < z_i < 1 \quad \forall i \in \mathcal{NC} \\
 & \sum_{i \in \mathcal{NC}} z_i = 1 \\
 & P_{N_d}^L \leq P_{N_d} \leq P_{N_d}^U \\
 & \max(T_{mp}(\mathbf{n}) + 10, T_{N_d}^L) \leq T_{N_d} \leq T_{op}^U \\
 & W_{N_d, \text{solvent}} \geq W_{0, \text{solvent}} \\
 & \theta_0^L \leq \theta_0 \leq \theta_0^U
 \end{aligned} \tag{5}$$

Table 1: Key input parameters and specification of the CO₂ capture case study

Parameter	Symbol	Value	Parameter	Symbol	Value
Gas flow rate Nm ³ /s	m_f	5,000	Degree of CO ₂ captured, %	-	85
Gas temperature, K	T_f	323	Condenser temperature, K	$T_{N_d}^L$	333
Gas pressure, MPa	P_{N_a}	0.1	Min. temperature approach, K	ΔT_{min}	20
Gas molar composition of H ₂ O	y_{N_a+1, H_2O}	0.12	Bounds on desorber pressure, MPa	$[P_{N_d}^L, P_{N_d}^U]$	[0.1, 5]
Gas molar composition of CO ₂	y_{N_a+1, CO_2}	0.05	Bounds on handling temperature, K	$[T_{N_d}^L, T_{N_d}^U]$	[303, 333]
Gas molar composition of N ₂	y_{N_a+1, N_2}	0.73	Bounds on lean solvent temperature, K	$[T_0^L, T_0^U]$	[313, 353]
Max. operating temperature, K	T_{op}^U	413	Bounds on lean loading, mol/mol	$[\theta_0^L, \theta_0^U]$	[0, 2]

4. Case Study and Results

CO₂ capture from an exhaust gas from a 400 MW CCGT power plant (Alhajaj et al., 2016) is considered as a case study to assess the performance of the proposed feasibility tests. The key input parameters and the pre-specified bounds on the design variables are provided in Table 1. An automated implementation of the sequence of tests and solution of the primal problem (process optimisation with fixed solvent structure) is developed in C++. Individual tests and the primal problem are implemented in gPROMS ModelBuilder 6.0.2 with the gSAFT Foreign Object interface for the SAFT- γ -Mie calculations.

A set of 4,179 solvent candidates is generated in accordance with the molecular feasibility constraints described in section 2.2, and the investigation of the effectiveness of the feasibility tests is carried out over this entire molecular design space. The overall results of applying the feasibility tests to these alkanolamines and alkylamines are summarised in Table 2. Within the 4,179 candidate molecules, only 910 amine-based solvents pass all feasibility tests. As can be seen in Table 2, the majority of the alkylamines are found to be infeasible in Tests 1 and 2, mainly due to their low water miscibility as well as low safety and environmental performance. For example, the maximum solvent concentration (W_{solvent}^*) of trimethylamine (TMA) that can ensure a homogeneous liquid phase at absorber operating conditions is identified as 2×10^{-2} kg/kg. It indicates that the constraint $W_{\text{solvent}}^U \leq W_{\text{solvent}}^*$ is violated and TMA should be eliminated from the design space. The performance of Tests 3 and 4 is examined for the solvents that pass Tests 1 and 2. Tests 3 and 4 lead to the tighter bounds on the lean loading and desorber pressure and their average values are presented in the Table 2. For the case of MDEA, the minimum requirement on lean solvent purity, $\theta_0^L = 0.25$, is calculated as a result of Test 3. This implies that if the initial guess on the lean loading were set to be greater than 0.25, the initialisation step of the process model would lead to numerical failure, making it impossible to evaluate the performance of this solvent. Instead, the initial point can be chosen to be below 0.25 and the primal problem can be solved successfully.

To exemplify the solution of primal problems, the impact of the feasibility tests and updated bounds are investigated for MEA and 2-amino-2-methyl-1,3-propanediol (AMPD) and the results are summarised in Table 3 and Figure 2. The results are generated based on five runs using different initial guesses for each solvents. They highlight that the process optimisation problems converge successfully and the computational cost is reduced by 12.6% and 72.2% for MEA and AMPD, respectively, with the introduction of feasibility tests. Fewer search steps are taken during the course of the optimisation, mainly as a consequence of the reduction in the ranges of the design variables achieved in Tests 3 and 4. This robustness and efficiency are likely to be critically important in molecular and process design problems where many candidate solvents are infeasible or where the evaluation of the primal problems is computationally expensive.

5. Conclusions

In this paper, we have proposed a robust optimisation approach for the design of an optimal aqueous solvent and CO₂ chemical absorption processes. New feasibility tests have been combined with the primal problem to provide a reliable way to converge to an optimal so-

Table 2: Overall results of feasibility tests over the complete list of candidate molecules

	Single Amine		Diamine	
	Alkanolamine	Alkylamine	Alkanolamine	Akylamine
Number of molecules tested	765	416	1918	1080
Number of molecules screened by Test 1	194	198	463	262
Number of molecules screened by Test 2	0	218	559	313
Number of molecules screened by Test 3	432	-	165	485
Number of molecules screened by Test 4	0	-	0	0
Feasible molecules	139	0	731	20
Average value of updated bound on θ_0^U	0.4036	-	0.9031	0.9446
Average value of updated bound on θ_0^L	0.0038	-	0.0038	2.10×10^{-6}
Average value of updated bound on $P_{N_d}^U$	1.1279	-	0.9519	0.9332

Table 3: Outcome of the solution of the tests and the primal problems for MEA and AMPD for five runs. Only the smallest objective function is presented as an optimal solution

	Unit	MEA	AMPD
Min. lean loading, θ_0^l	mol/mol	0.4052	0.4102
Max. lean loading, θ_0^r	mol/mol	10^{-6}	10^{-6}
Max. desorber pressure $P_{N_d}^U$	MPa	0.3782	0.4
TAC	\$M/year	29.22	23.39
Total energy consumption	GJ/tonCO ₂	9.73	7.03
Lean loading	mol/mol	0.310	0.259
Lean solvent temperature	K	313	345
Desorber pressure	MPa	0.21	0.22
Solvent circulation rate	mol/hr	14.25	17.67
Avg. CPU time with tests	s	551	696
Avg. CPU time without tests	s	631	2509

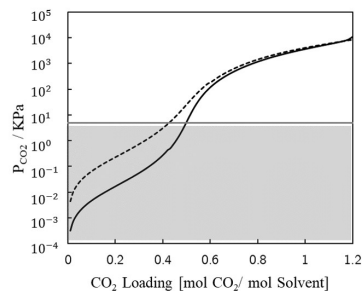


Figure 2: CO₂ partial pressure (P_{CO_2}) versus solvent loading (mol CO₂/mol solvent) for AMPD (---) and MEA (—) at the optimal process conditions

lution by removing infeasible process conditions and molecular structures from the search space. This has also led to a reduction in the computational cost by enabling the ranges of the operating conditions to be reduced automatically. The efficiency of the proposed algorithm has been highlighted through a case study on the design of solvents for CO₂ chemical absorption processes. Future work involves the full integration of the approach within a mixed-integer nonlinear optimisation algorithm in order to identify the best performing solvent and it could be directed at testing the performance of the feasibility tests on more case studies to access general conclusions of their performance on the design of optimal solvents and CO₂ chemical absorption processes.

Acknowledgements

The authors gratefully acknowledge financial support from from British Federation of Women Graduates via funds for women graduates, and Centre of Process Systems Engineering Research Committee of Imperial College London via Roger Sargent scholarship.

References

- C. S. Adjiman, A. Galindo, G. Jackson, 2014. Molecules matter: the expanding envelope of process design. In: Computer Aided Chemical Engineering. Vol. 34. pp. 55–64.
- A. Alhajaj, N. Mac Dowell, N. Shah, 2016. A techno-economic analysis of post-combustion CO₂ capture and compression applied to a combined cycle gas turbine: Part i. a parametric study of the key technical performance indicators. International Journal of Greenhouse Gas Control 44, 26–41.
- S. Gopinath, G. Jackson, A. Galindo, C. S. Adjiman, 2016. Outer approximation algorithm with physical domain reduction for computer-aided molecular and separation process design. AIChE Journal 62 (9), 3484–3504.
- H.-C. Hsu, Y.-W. Sheu, C.-H. Tu, 2002. Viscosity estimation at low temperatures ($T_r < 0.75$) for organic liquids from group contributions. Chemical Engineering Journal 88, 27–35.
- A. S. Hukkerikar, B. Sarup, A. Ten Kate, J. Abildskov, G. Sin, R. Gani, 2012. Group-contribution + (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. Fluid Phase Equilibria 321, 25–43.
- A. S. Lee, J. C. Eslick, D. C. Miller, J. R. Kitchin, 2013. Comparisons of amine solvents for post-combustion CO₂ capture: A multi-objective analysis approach. International Journal of Greenhouse Gas Control 18, 68–74.
- A. I. Papadopoulos, G. Shavaliyeva, S. Papadokonstantakis, P. Seferlis, F. A. Perdomo, A. Galindo, G. Jackson, C. S. Adjiman, 2020. An approach for simultaneous computer-aided molecular design with holistic sustainability assessment: Application to phase-change CO₂ capture solvents. Computers & Chemical Engineering 135, 106769.
- V. Papaioannou, T. Lafitte, C. Avendaño, C. S. Adjiman, G. Jackson, E. A. Müller, A. Galindo, 2014. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. Journal of Chemical Physics 140 (5).
- F. Perdomo, S. Khalit, C. Adjiman, A. Galindo, G. Jackson, 2020. Description of the thermodynamic properties and fluid-phase behaviour of aqueous solution of linear, branched and cyclic amines. Submitted to AIChE Journal.
- G. T. Rochelle, 2009. Amine scrubbing for CO₂ capture. Science 325 (5948), 1652–1654.

Simulation-based Derivative-free Optimization for Hybrid Separation Design

Shuang Xu^a, Selen Cremaschi^a, Mario R. Eden^a, and Anjan K. Tula^{b*}

^a*Department of Chemical Engineering, Auburn University, Auburn, AL, USA*

^b*College of Control Science & Engineering, Zhejiang University, Hangzhou, China*

anjantula@zju.edu.cn

Abstract

Distillation is the most widely applied separation technique in the chemical process industry. Due to its high energy consumption and low thermal efficiency, researchers have investigated several approaches to reduce utility requirements, such as hybrid separation and using intensified equipment. In most prior examples, hybrid separations have been studied to separate azeotropes and/or close boiling mixtures. In this work, two hybrid separation design strategies to replace any distillation separation task are developed. These strategies are shown to improve thermal efficiency and thus save energy in the range of 10-20%. Additionally, by implementing hybrid separation schemes, the throughput of the distillation process can be increased. In this paper, two solution strategies have been developed based on simulation-based derivative-free optimization (DFO) algorithms and rigorous mathematical optimization. Both methods are applied to a case study of methanol/water separation, and the results show that the proposed hybrid separation processes yield at least 20% energy reduction and 20% capacity improvement.

Keywords: Distillation, Hybrid Separation, Optimization.

1. Introduction

With the chemical process industry continuing to move in a more efficient and sustainable direction, researchers have investigated several approaches to reduce energy consumption, improve thermal efficiency and sustainability, including new methods for process synthesis, hybrid separation, and design of intensified equipment. Tula et al. (2017) pointed out that after applying hybrid distillation-membrane techniques, 15-20% energy reduction of the retrofitted process can be achieved without significant additional capital investment. In their approach, distillation was used to purify the distillate until a certain composition (switching composition) and then sent to a membrane system to reach the purification target. O'Connell et al. (2019) further evaluated the influence of the switching composition in hybrid distillation-membrane separations through case studies involving the separation of C6 isomers and ethylbenzene/styrene. Through these studies, O'Connell et al. (2019) established the strong causal relationship between switching composition and the economic feasibility of retrofitted processes. Figure 1 shows the relationship between normalized separation driving force as a function of feed concentration. The driving force is defined as the concentration difference between the two phases, thus representing the ease of separation. As the distillation/membrane maximum driving forces are located at different positions in the diagram/composition space, the separation can ideally be divided into two regions where Region I is more efficient for distillation, and Region II is more efficient for membrane separation.

Therefore, determining the location of these two regions and the corresponding operating/design parameters is an important problem to be solved. One feasible solution method is sensitivity analysis, which analyses the impact of different variables and then chooses the best combination. A more rigorous approach is to formulate an optimization problem and solve it mathematically. Caballero et al. (2009) presented an algorithm based on a shortcut/rigorous distillation model to identify the optimum hybrid separation configuration. In our work, a mathematical optimization model and a simulation-based derivative (DFO) free optimization model are developed to determine the hybrid distillation design/operating parameters.

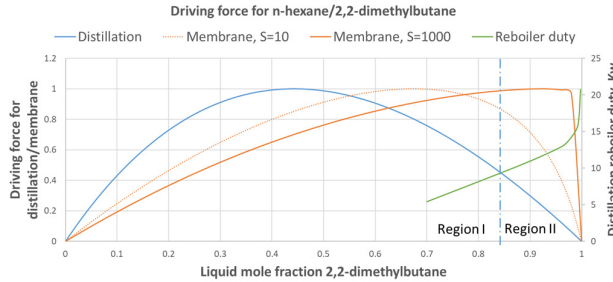


Figure 1. Driving force for n-hexane/2,2-dimethylbutane (distillation and membrane separation).

2. Solution Strategies

2.1. Simulation-based Optimization

The simulation-based optimization approach uses DFO algorithms to select operating/design variables. This approach incorporates the inner process simulation results with an outer DFO algorithm. In the inner-loop, process simulations are performed, and the results are used to calculate the objective function corresponding to the design decisions. The results from the inner-loop (operating/sizing variables) are sent to the outer-loop, where a suitable solver based on a gradient-free optimization algorithm is employed to obtain new/improved values of the decision variables/design parameters. This iterative calculation continues until the maximum number of iterations is reached or the optimal solution is found. Simulation-based optimization takes advantage of the process simulator, which represents implicit equations, making the problem relatively easy to solve. In this study, three open-source DFO algorithms (RBFOpt, GPyOpt, ZOOpt) are applied to all the case studies. The algorithms are applied to find the optimum design and operating variables for methanol/water separation process using distillation-membrane and distillation-adsorption approaches. A generalized form of the optimization problem solved by the DFO algorithms is given in Equations (1)-(7).

$$\min: TAC = \frac{i(i+1)^n}{(i+1)^n - 1} \times IC + AUC \quad (1)$$

$$\text{st. } IC = \sum_j Cost_j(q_{s,j}) \quad (2)$$

$$AUC = 24 \times 300 \times \sum_j Utility_j(q_{o,j}) \quad (3)$$

$$q = \theta(x, z, p(x)) \quad (4)$$

$$\sum_j Utility_j \leq Utility_0 \quad (5)$$

$$x_{product} \geq purity \quad (6)$$

$$x \in X, \quad z \in Z \text{ integer} \quad (7)$$

In the model, TAC is the total annualized cost, IC is the investment cost, AUC is the annualized utility cost, i is interest, n is plant life, $Cost_j$ is the investment cost of equipment j (Turton et al., (2008)). $Utility_j$ is the utility cost of equipment j (Kalakul et al., (2014)). $Utility_0$ is the utility cost of the initial single distillation column, q_{sj} is the sizing variable for equipment j , q_{oj} is the operating variable for equipment j , $p(x)$ is the membrane estimation model (Sommer and Melin, (2005)). $\Theta(x, z, p(x))$ is the process simulation model, x is the switching composition, and z is the number of membrane/adsorption stages.

2.2. Mathematical Optimization

For mathematical optimization, the distillation-membrane separation problem is considered with only one membrane system following the distillation separation. Since the final purification is achieved by the membrane, the original distillation column will underperform at the new conditions. The objective is to validate if column capacity can be increased while maintaining operation at feasible normal conditions. The retrofitting distillation model and hydraulic analysis (Gadalla et al. (2003)) are included in the optimization. To investigate the relationship between capacity improvement and energy cost reduction, a generalized bi-objective nonlinear programming problem (NLP) is formulated, which is shown in Equations (8)-(13).

$$\min_{x,y} F_1 = f_1(x, y, q) \quad (8)$$

$$\max_{x,y} F_2 = f_2(x, y, q) \quad (9)$$

$$st. h(x, y, q) = 0 \quad (10)$$

$$g(x, y, q) \leq 0 \quad (11)$$

$$q = q(x, y) \quad (12)$$

$$x, y \in R^+ \quad (13)$$

Here, f_1 is the total annualized cost, f_2 is the profit, $h(x, y, q)$ is a simple regressed thermodynamic model and retrofitting distillation model, $g(x, y, q)$ is the hydraulic and purity constraints, $q(x, y)$ is the dependent variable, which is determined by three decision variables x (switching composition of the distillate and bottom) and y (feed flowrate).

3. Case Study

Both solution strategies are applied to a case study of methanol/water separation. The base case methanol/water distillation column has 32 stages with stage efficiency of 60%, and is fed to stage 12. The inlet conditions are: 2.5 atm, 107°C, 1342 kmol/h, 76.5 mol% water, 23.5 mol% methanol. The base case distillation column's reboiler duty is 36.9 GJ/h, and the product requires 99.9 mol% purity of both water and methanol.

3.1. Distillation-membrane (DFO)

Okamoto et al. (2001) proposed a Zeolite NaA vapor permeation membrane for methanol/water separation, which shows high selectivity for water permeation. Sommer and Melin (2005) developed a model to estimate the membrane permeability at different operating conditions. In our work, the hybrid distillation-membrane is considered, and the process scheme is shown in Figure 2. The binary mixture of water and methanol is first fed into a distillation column (B1) for separation (Region I). Then, the distillate stream passes through a membrane system (B4 and B7) for further purification (Region

II). To maintain the same outlet conditions of the products as the original distillation column, the permeate and retentate streams are heated/compressed. The top product is methanol, and the bottom is water. Membrane selectivity and permeability depend on the inlet stream conditions, thus the number of membrane stages should be determined. Here, the maximum number of membrane stages is two because of the high selectivity. In this problem, three decision variables: distillation top/bottom composition (switching composition) and the number of membrane stages, need to be determined. Note that the driving force of the membrane separation is the vapor pressure difference of the permeate compound between the permeate and feed side. In this membrane system, either a vacuum pump or sweeping gas is required to enhance the permeation. In our case, we specify the total pressure at the permeate side as near vacuum (2 kPa).

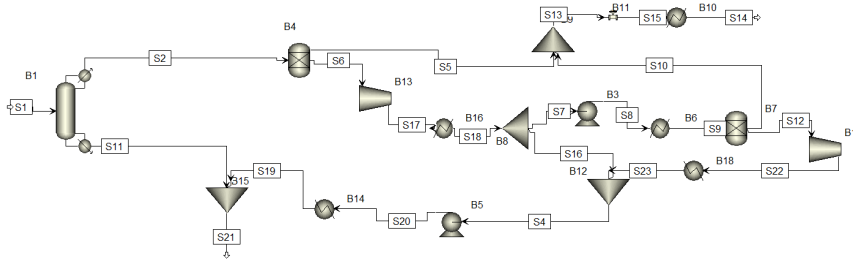


Figure 2. Process scheme of the distillation-membrane separation process.

All algorithms yield similar results (Table 1). The optimal top switching composition is 88 mol% of methanol, and the system requires one membrane stage after the distillation separation, resulting in 36.8% in energy savings and a payback time of three years.

Table 1. DFO results for distillation-membrane separation.

	RBFOpt	GPyOpt	ZOOpt
Top switching (methanol), mol%	88.4%	88.8%	88.3%
Bottom switching (water), mol%	99.90%	99.91%	99.94%
Stages	1	1	1
Annualized cost (\$)	1,950,848	1,952,675	1,952,774

Table 2. DFO results for distillation-adsorption separation at different selectivities.

	S = 1,000			S = 500		
	RBFOpt	GPyOpt	ZOOpt	RBFOpt	GPyOpt	ZOOpt
Top switching (methanol), mol%	96.68%	98.16%	98.22%	97.58%	97.78%	98.20%
Bottom switching (water), mol%	99.95%	99.94%	99.94%	99.91%	99.92%	99.90%
Stages		1			2	
Annualized cost (\$)	1,488,569	1,492,973	1,494,281	1,542,266	1,539,384	1,535,048
Energy savings		33%			31.5%	
Payback time (year)		0.3			0.4	

3.2. Distillation-adsorption (DFO)

In this case, the process scheme and DFO model are similar to the distillation-membrane system. The adsorption of water is a widely used technology, and Gabruš et al. (2015) proposed a Langmuir-Freundlich model for adsorption capacity estimation. In the process model, the membrane is replaced with an adsorption system that is integrated with distillation. An analogous set of three decision variables, as described above, are

identified for the distillation-adsorption process. Due to the lack of specific selectivity data for the adsorbent system, two different selectivities, 500 and 1,000, are used for optimization. The results are shown in Table 2. All three algorithms give similar results, with over 30% in energy savings. For the lower selectivity, two adsorption stages are required, compared to only one at the higher selectivity. To estimate the energy cost of the adsorption/desorption system, a constant desorption energy 45.95 kJ/mol is used in the calculations (Gabruś et al., 2015).

3.3. Mathematical Optimization

A schematic depiction of the distillation-membrane process is shown in Figure 3. The binary mixture is first fed into an existing distillation column for separation. Then, a membrane system is used to purify the distillate. In the process, a compressor is used to keep the product at the same conditions as the original distillation column.

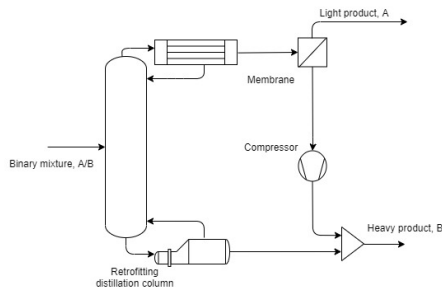


Figure 3. Distillation-membrane process scheme for mathematical optimization.

The ϵ -Constraint method developed by Deb (2001) is used to solve the bi-objective problem. The resulting Pareto front is shown in Figure 4. The annualized cost and profit are varied along with switching composition and column capacity. The optimization results indicate that the objective function is very sensitive to column capacity only when the annualized cost is lower than \$3.2 million/year. However, when the annualized cost is higher than \$3.2 million/year, both switching composition and column capacity influence the objective function. The profit increases faster when the annualized cost is lower than \$3.2 million/year. That is because this region has higher energy efficiency, resulting in lower product cost, i.e. in this region, it is easier to improve the profit by varying the column capacity. A switching composition of 92.4 mol% is suggested for this hybrid operation, which can achieve 20% capacity improvement and energy savings.

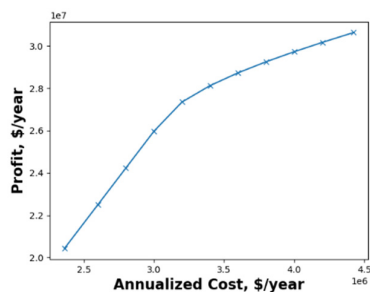


Figure 4. Pareto front of the bi-objective optimization results.

4. Conclusions

Distillation is an efficient separation technique until a certain top product purity; however, the separation task becomes thermally inefficient when moving beyond this point. Replacing the less favorable region of distillation with a different separation technique can yield significant energy savings. In this work, two different optimization strategies (DFO and mathematical NLP model) are developed to design hybrid distillation schemes involving adsorption and membranes. The DFO strategy utilizes process simulation tools, which are readily available. Although the DFO algorithm cannot guarantee the global optimum, increasing the number of iterations can generate results that are close to the global optimum. So, the DFO strategy is very efficient for preliminary design. The second strategy uses a more rigorous model, and this strategy can be used for obtaining the final design parameters as it can guarantee the global optimum.

Using the simulation-based DFO strategy, the best switching composition for vapor permeation is around 88 mol% methanol, while for adsorption, the optimal switching composition is around 96-98 mol% methanol. It is also concluded that for the adsorption selectivities evaluated, the hybrid distillation adsorption is more economical than the hybrid distillation membrane process. It has also been shown that the retrofitted hybrid columns can be used to increase the distillation processing capacity. The distillation column can achieve a maximum of 20% energy savings and 20% capacity improvement. Future work will focus on comparing the sustainability, safety, and controllability between the distillation and hybrid processes.

References

- Caballero, J. A., Grossmann, I. E., Keyvani, M., & Lenz, E. S. (2009). Design of hybrid distillation-vapor membrane separation systems. *Industrial & Engineering Chemistry Research*, 48(20), 9151-9162.
- Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms* (Vol. 16). John Wiley & Sons.
- Gabruś, E., Nastaj, J., Tabero, P., & Aleksandrak, T. (2015). Experimental studies on 3A and 4A zeolite molecular sieves regeneration in TSA process: Aliphatic alcohols dewatering-water desorption. *Chemical Engineering Journal*, 259, 232-242.
- Gadalla, M., Jobson, M., & Smith, R. (2003). Increase capacity and decrease energy for existing refinery distillation columns. *Chemical Engineering Progress*, 99(4), 44-50.
- Kalakul, S., Malakul, P., Siemanond, K., & Gani, R. (2014). Integration of life cycle assessment software with tools for economic and sustainability analyses and process simulation for sustainable process design. *Journal of Cleaner Production*, 71, 98-109.
- Okamoto, K. I., Kita, H., Horii, K., & Kondo, K. T. (2001). Zeolite NaA membrane: preparation, single-gas permeation, and pervaporation and vapor permeation of water/organic liquid mixtures. *Industrial & Engineering Chemistry Research*, 40(1), 163-175.
- O'Connell, J. P., Eden, M. R., Tula, A. K., & Gani, R. (2019). Retrofitting distillation columns with membranes. *Chemical Engineering Progress*, 115(12), 41-49.
- Sommer, S., & Melin, T. (2005). Influence of operation parameters on the separation of mixtures by pervaporation and vapor permeation with inorganic membranes. Part 1: Dehydration of solvents. *Chemical Engineering Science*, 60(16), 4509-4523.
- Tula, A. K., Befort, B., Garg, N., Camarda, K. V., & Gani, R. (2017). Sustainable process design & analysis of hybrid separations. *Computers & Chemical Engineering*, 105, 96-104.
- Turton, R., Bailie, R. C., Whiting, W. B., & Shaeiwitz, J. A. (2008). *Analysis, synthesis and design of chemical processes*. Pearson Education.

Material Screening for Thermochemical Energy Storage in Solar Power Systems

Ishan Bajaj^a, Xinyue Peng^a, Christos T. Maravelias^{b,c,*}

^aDepartment of Chemical and Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706, USA

^bDepartment of Chemical and Biological Engineering, Princeton University, 50-70 Olden Street, Princeton, NJ 08540, USA

^cAndlinger Center for Energy and Environment, Princeton University, 86 Olden Street, Princeton, NJ 08540, USA
maravelias@princeton.edu

Abstract

A framework is proposed to systematically identify promising solid-gas reaction candidates for thermochemical energy storage (TCES) in concentrating solar plants (CSP). It involves four steps: (i) generation of a list of reaction candidates, (ii) screening reactions based on thermodynamic criteria, (iii) solution of a process synthesis model to compute levelized cost of electricity (LCOE) and thermal energy storage (TES) costs, and (iv) selection of candidate reactions. The analyses identify ten reactions that can be economically and energetically superior to the state-of-the-art two-tank molten salt storage system.

Keywords: solar energy, process synthesis and optimization, reaction screening.

1. Introduction

Thermal energy storage is a promising way to manage intermittency of solar energy in concentrated solar power (CSP) plants. Commercially, CSP plants are typically integrated with a two-tank molten-salt sensible heat storage unit. There are two major drawbacks of this system. First, heat is delivered to the power cycle at low temperature (565 °C), resulting in low solar-to-electricity efficiency. Second, due to low energy density of molten-salt, large quantity of material is needed, which can become too costly for large-scale systems. TCES systems, on the other hand, are a more promising alternative because of their high energy density and ability to deliver heat at high temperature.

TCES is based on a reversible reaction, wherein heat is required for the forward reaction and thus, the reaction enthalpy is stored in the products. Today, TCES is at an early stage of development. Fluid-phase and solid-gas reactions are the two classes of TCES systems that have been studied. Recent analyses of fluid-phase reactions show that they suffer from low energy efficiency and high cost due to the need to store reacting gases (Peng et al., 2017). However, solid-gas reactions are more promising because they require less gas storage and the products can be easily separated (expressed as: $A(s) + \Delta H_R \leftrightarrow \nu_B B(s) + \nu_C C(g)$, where ΔH_R is reaction enthalpy).

Figure 1 shows a CSP plant consisting of four components: collector, receiver, TCES system and power cycle. During the day, the collector focuses sunlight onto a receiver, where photons are absorbed and converted into heat. The flow of heat transfer fluid (HTF) is split such that a part of it flows through the reactor (R1) to provide heat to an

endothermic reaction resulting in the decomposition of A and the remaining flows through the power block to provide heat to the working fluid (WF). The products of the endothermic reaction (B and C) are stored and transferred to R2 during the night, where exothermic reaction occurs and provides heat to WF.

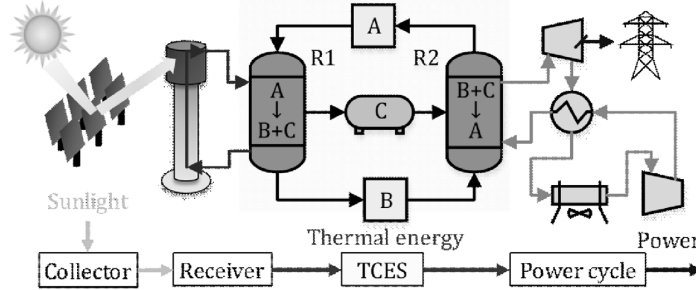


Figure 1. Schematic diagram of a CSP plant with solid-gas TCES system.

Previously, most of the studies had focused on improving the performance of a handful of reaction systems and investigating their applicability. However, given the diversity of reactions applicable for TCES, there is a need for a systematic screening approach to find novel reaction candidates. While some work has been done in reaction screening (André et al., 2016; Pardo et al., 2014) based on equilibrium temperature and energy density, it still remains unclear which reactions are competitive in terms of LCOE and TES costs. To this end, we develop a framework to identify reactions based on energetic and economic performance.

2. Screening Methodology

In this section, the screening framework to identify promising TCES systems is outlined. The framework involves generating reaction candidates, eliminating reactions based on thermodynamic criteria, computing LCOE and TES costs for the remaining reactions and selecting the top performing reaction candidates.

2.1. Generating reaction candidates

The following four classes of reactions are considered: (i) oxide/oxide ($A=M_aO_c$, $B=M_bO_d$), (ii) oxide/metal ($A=M_aO_c$, $B=M$), (iii) hydroxide/oxide ($A=M_a(OH)_c$, $B=M_bO_d$), and (iv) carbonate/oxide ($A=M_a(CO_3)_c$, $B=M_bO_d$). Reaction candidates are created such that both A and B are inorganic materials. Additionally, only one gaseous component is allowed to avoid the need for gas separation. Aspen Plus is used to identify applicable materials, and their properties including density, melting point, specific heat capacity, enthalpy, and Gibbs energy (G) are stored.

An algorithm is designed to automatically identify all possible reactions for the collected materials. A total of 364 reactions comprising of 166 oxide/oxide, 162 oxide/metal, 19 hydroxide/oxide, and 17 carbonate/oxide, are identified.

2.2. Eliminating reactions based on thermodynamic criteria

For each of the reactions, equilibrium temperature (T_{eq}) is estimated:

$$T_{eq} = -\frac{\Delta G_R}{R \ln \ln p_C^{v_C}}$$

where ΔG_R , R and p_c are the Gibbs energy of reaction, universal gas constant, and partial pressure of gas C , respectively. The value of p_c is taken to be 0.21 bar for oxide/oxide and oxide/metal reactions, and it is fixed to 1 bar for hydroxide/oxide and carbonate/oxide reactions. For a reaction to be considered suitable for TCES system, its T_{eq} needs to meet two criteria: (i) it should lie within the range 25-2000 °C and (ii) it should be lower than the melting points of A and B so that no phase change occurs. Based on this criteria, 305 reactions are eliminated. The chosen reactions include 34 oxide/oxide, 6 oxide/metal, 7 hydroxide/oxide, and 12 carbonate/oxide. Key properties of the selected reactions are shown in Figure 2. The material prices are taken from the United States Geological Survey mineral commodities database. T_{eq} impacts the overall plant efficiency and ΔH_R affects the storage costs and efficiency. For high efficiency, T_{eq} needs to be in the range 500-1200 °C and ΔH_R needs to be high. It can be observed that there are several reactions with appealing properties, however, it is not clear how these properties impact the energetic and economic performance of the integrated plant. To this end, we developed a process model for an integrated CSP-TCES plant, which is outlined in the next section.

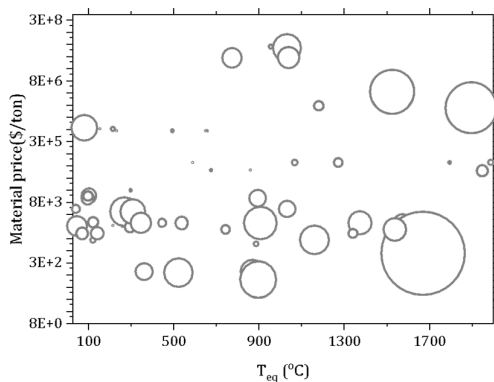


Figure 2. Key properties of reactions obtained after satisfying thermodynamic criteria. Each bubble represents a candidate reaction and its size represents reaction enthalpy.

2.3. Stochastic programming-based process synthesis model

An optimization model is developed (shown in Figure 3) for the design of CSP plants with solid-gas TCES systems (Peng et al., 2020, 2019). The inputs required for the model include material properties (e.g., material price, density, specific heat capacity), reaction characteristics (e.g., reaction enthalpy, equilibrium temperature), plant specifications (e.g., plant location and nameplate capacity), weather data and cost data. In this study, we consider a 100 MW CSP plant with power tower configuration located in Daggett, California (34.87°N, 116.78°W) and its weather data is obtained from the National Solar Radiation Data Base. The assumptions for economic analysis are given in Peng et al. (Peng et al., 2020).

The optimization model considers seasonal and daily variability. We use scenarios to represent seasonal variability and modes to represent daily variability. The optimization model is formulated as a two-stage stochastic programming (SP) model with multiple modes embedded within each scenario. The first stage variables are design variables that include solar field area, receiver area, and TCES equipment sizes. The second stage are operational variables for each scenario and mode including storage hours, flow rates, and operating temperatures. To consider the seasonal

variability in solar irradiance in our analyses, we generate six representative scenarios (i.e., days) from the annual data using centroid clustering algorithm and approximate daily solar irradiance using two modes. While it is possible to consider a more accurate dynamic model with finer temporal granularity, this approach would be unnecessary and infeasible, from a computational standpoint because we aim to screen hundreds of reactions. Furthermore, it was demonstrated in Peng et al. (Peng et al., 2019) that the multimode SP approach is computationally efficient and provides reasonably accurate solutions.

The objective of the model is to minimize the overall LCOE, which is defined as the sum of annualized costs of collector, receiver, thermal storage, and power block divided by the annual electricity generated. The constraints arise from the physical model of plant components, unit mass and energy balances, equipment sizing, and cost calculations. The collector efficiency is fixed to 0.6. The receiver efficiency is a function of temperature such that increasing temperature results in lower efficiency due to radiation and convection losses. Both the reactors are modeled as fluidized bed reactors. We assume that the conversion is 100%, no side products are formed, and the conversion rates and yields remain the same with cycles. For oxide/oxide and oxide/metal reactions, air is used as the source of O_2 and therefore, no gas storage vessels are required. Storage tanks for storing CO_2 and H_2O are required for carbonate/oxide and hydroxide/oxide reactions, respectively. Furthermore, a compressor is also needed for carbonate/oxide reactions because of low boiling point of CO_2 . Supercritical CO_2 Brayton cycle is used for power generation and a model is developed to evaluate its performance at different turbine inlet temperatures (Peng et al., 2020). The optimization model is a nonlinear programming model, formulated in GAMS, and solved using a novel solution technique (Section 2.4).

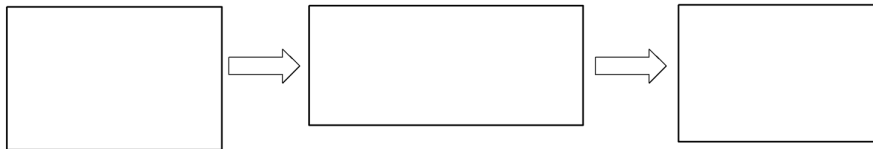


Figure 3: Overview of optimization-based CSP-TCES process model.

2.4. Solution method

The method is based on three ideas. First, tight lower and upper bounds of the first-stage and second-stage variables are computed through constraint propagation based on the initial bounds on a set of key variables including the temperature of R1 (T_{R1}), R2 (T_{R2}), and HTF (T_{HTF}). Second, the space of these variables is divided into small subregions and the model is solved in each of the regions to global optimality using ANTIGONE (Misener and Floudas, 2014). Third, to accelerate the convergence, we use two strategies. First, when solving the model in a new subregion, we add a cut based on the best objective function value obtained in the explored subregions so far, thereby enabling pruning of nonpromising subregions. Second, if any of the decision variables lie at the bounds of the current subregion, the subregion that shares the decision variable is selected as the next region to be explored, leading to the identification of good solutions quickly. Computational comparisons with other methods suggest that our solution method finds the global minimum faster for several problem instances.

3. Results

The solar-to-electricity efficiency (η^{s-e}), which depends on the efficiencies of the four plant components, is used as the primary energetic metric. Receiver efficiency decreases with increase in receiver temperature (T_r) and power cycle efficiency increases with increasing turbine inlet temperature (T_p). Due to heat transfer limitations, the following relationship holds between the operating temperatures:

$$T_r = T_{HTF} > T_{R1} > T_{eq} > T_{R2}$$

T_p lies between T_{R1} and T_{R2} , and moves closer to T_{R2} as the fraction of sensible energy storage decreases. Since T_r is always higher than T_p , it is impossible to have low T_r and high T_p . For high η^{s-e} , T_r should be in the range of 500-1200 °C and the difference between T_r and T_p ($T_r - T_p$) should be small to reduce exergy losses. However, if $T_r - T_p$ is small, less sensible heat is stored, which leads to high storage costs. The optimization algorithm balances these trade-offs such that design and operational decisions are obtained resulting in minimum LCOE.

The top ten reactions identified based on LCOE and TES costs are shown in Figure 4. LCOE for the selected reactions range from 11.2-12.9 ¢/kWh and TES costs range from 8.2-23.1 \$/kWh_{th}. Importantly, all the selected reactions have the potential to lead to lower LCOE and TES costs than the two-tank molten salt storage system. Note that seven of the ten promising reactions in terms of LCOE belong to oxide/oxide class, whereas two belong to hydroxide/oxide class, and one belongs to carbonate/oxide class. In terms of TES cost, seven reactions belong to oxide/oxide class, two belong to hydroxide/oxide class, and one belongs to oxide/metal class. Only two reactions are found to be promising in terms of both LCOE and TES costs, which implies that low TES costs do not imply low LCOE and underscores the importance of simultaneously considering all the components of a CSP plant.

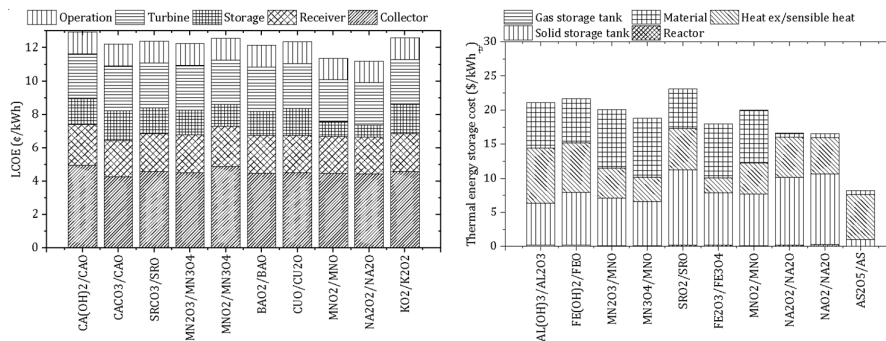


Figure 4. Top ten reactions selected based on (a) LCOE and (b) TES costs.

4. Conclusions

We developed a framework to systematically screen reactions for thermochemical energy storage. First, we identify a total of 374 reactions belonging to oxide/oxide, oxide/metal, hydroxide/oxide, and carbonate/oxide classes. Two thermodynamic criteria are then applied to eliminate 305 reactions. A stochastic programming-based optimization model is solved using a novel solution strategy for the remaining reactions to compute LCOE

and TES costs. Our analyses identify ten reactions that can be energetically and economically superior to the two-tank molten salt storage system.

Table 1. Key reaction properties and the overall efficiency for the selected reactions.

Reactions	T_{eq} (°C)	Material price (\$/ton)	ΔH_R (kJ/kg)	η^{s-e}
Ca(OH) ₂ /CaO	521.0	150.0	1341.9	0.19
CaCO ₃ /CaO	895.9	100	1664.9	0.22
SrCO ₃ /SrO	1159.0	954.4	1349.3	0.21
Mn ₂ O ₃ /Mn ₃ O ₄	901.0	2535.2	203.6	0.21
MnO ₂ /Mn ₃ O ₄	536.3	2535.2	646.2	0.20
BaO ₂ /BaO	742.2	1738.8	475.5	0.22
CuO/Cu ₂ O	1031.8	5650.0	816.4	0.21
MnO ₂ /MnO	905.6	2535.2	1509.5	0.22
Na ₂ O ₂ /Na ₂ O	868.1	157.0	1117.2	0.22
KO ₂ /K ₂ O ₂	885.7	776.0	314.3	0.21

Key properties and η^{s-e} of the top reactions selected based on LCOE are shown in Table 1. As expected, for a reaction system to be competitive, η^{s-e} needs to be high. For the selected reactions, η^{s-e} range from 0.19-0.22 and the corresponding T_{eq} range from 521-1159 °C. Mn₂O₃/Mn₃O₄ and MnO₂/MnO reaction systems have similar T_{eq} , but the latter has higher reaction enthalpy, which leads to low $T_r - T_p$, resulting in higher efficiency. Higher reaction enthalpy of MnO₂/MnO system compared to Mn₂O₃/Mn₃O₄ leads to lower storage costs for MnO₂/MnO system. The material price of CuO/Cu₂O is 37 times more than Ca(OH)₂/CaO, but LCOE is lower for the former due to its marginal higher η^{s-e} . This illustrates how small improvements in efficiency can impact the process economics.

References

- André, L., Abanades, S., Flamant, G., 2016. Screening of thermochemical systems based on solid-gas reversible reactions for high temperature solar thermal energy storage. *Renew. Sustain. Energy Rev.* 64, 703–715.
- Deutsch, M., Müller, D., Aumeyr, C., Jordan, C., Gierl-Mayer, C., Weinberger, P., Winter, F., Werner, A., 2016. Systematic search algorithm for potential thermochemical energy storage systems. *Appl. Energy* 183, 113–120.
- Misener, R., Floudas, C.A., 2014. ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations. *J. Glob. Optim.* 59, 503–526.
- Pardo, P., Deydier, A., Anxionnaz-Minvielle, Z., Rougé, S., Cabassud, M., Cognet, P., 2014. A review on high temperature thermochemical heat energy storage. *Renew. Sustain. Energy Rev.* 32, 591–610.
- Peng, X., Root, T.W., Maravelias, C.T., 2019. Optimization-based process synthesis under seasonal and daily variability: Application to concentrating solar power. *AIChE J.* 65, 1–16.
- Peng, X., Root, T.W., Maravelias, C.T., 2017. Storing Solar Energy with Chemistry: The Role of Thermochemical Storage in Concentrating Solar Power. *Green Chem.* 2427–2438.
- Peng, X., Yao, M., Root, T.W., Maravelias, C.T., 2020. Design and Analysis of Concentrating Solar Power Plants with Fixed-bed Reactors for Thermochemical Energy Storage. *Appl. Energy* 262, 114543.

Network design optimization of waste management systems: the case of plastics

Rafael Castro-Amoedo,^a Alina Dahmen,^b Ana Barbosa-Povoa,^b François Maréchal^a

^a *Industrial Process and Energy Systems Engineering group, École Polytechnique Fédérale de Lausanne, Rue de l'Industrie 17, 1950 Sion, Switzerland*

^b *Centre for Management Studies, Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais, 1049 Lisbon, Portugal*

Abstract

Failure in appropriately managing complex municipal solid waste systems in general and waste plastics in particular negatively affects public health and environment.

A decision-making tool incorporating concepts of logistics and supply chain, considering different types of waste was developed. A multi-objective, multi-period mixed-integer linear programming (MILP) formulation, focusing on the plastic recycling network, incorporating features of multimodal transportation was implemented. The model aims at minimizing net present cost (NPC) and environmental impacts, illustrated by a case study of a Swiss waste treatment cluster. A reference situation was used for benchmarking and 7 new scenarios considered - corresponding to different plastics collection configurations.

From the solutions space, designs arise that lead to a 2-fold decrease of NPC at the expense of increasing by 20 % current emissions, as well as solutions that provide small net gains in environmental aspects - close to 3 %, albeit a 4-fold increase in NPC due to large investments in recycling facilities and increased transportation costs.

This work sheds light on the impact of supply chain and logistics aspects on the separate collection and treatment of waste plastics, in the context of a real municipal waste treatment policy, where a thermodynamic model to deal with the conversion processes is integrated into the general supply chain decision. As a result, important insights to guide decision-makers were obtained.

Keywords: Waste plastics, MSW, Recycling, Thermovalorization, Supply chain

1. Introduction

As reported in (Bhada-Tata et al., 2018) due to rapid urbanization and population growth, an increase in *per capita* generation of waste - and plastic waste in particular - is foreseen. If not properly managed, waste might have a significant impact on population health, environment, and economy. Designing an integrated waste management system is a strategic Supply Chain (SC) problem as it refers to long-term decisions on a combination of functional elements in the waste chain, from generation, storage, collection, separation and transportation to treatment, distribution, and disposal.

Management of plastic waste presents additional challenges, such as low transport efficiencies (due to low material density), as well as frequent contamination with other types of plastics as described in (Bing et al., 2013). Moreover, plastics have a high energy content - valuable from a thermovalorization point of view: if segregated, municipal solid waste (MSW) loses an important share of its energetic content. Thus, recycling and energy recovery present themselves as competitors for treating this

particular kind of waste contrasting, according to (Groot et al., 2014), with the general belief of routing waste plastic for recycling facilities whenever possible.

This work is a necessary step to a generalized superstructure of plastic waste treatment and conversion routes. Logistics and supply chain considerations are core concepts that have proved to be relevant from an economic perspective, as recently confirmed by (Colvero et al., 2020), and might condition tactical, operational and investment decisions. Furthermore, heat integration between units explores synergies and trade-offs between recycling and thermovalorization of different waste fractions.

2. Problem statement and mathematical formulation

The problem is generically formulated as a multi-echelon, multi-period, multi-commodity and multi-stakeholders. The goal is to determine the set of optimal configurations along the Pareto front and characterize each solution by the cost and associated emissions; each solution yields the optimal allocation of waste flows, the number of products and by-products, and the transport structure used. System boundaries include waste collection, transportation, and processing of the different waste flows, with respective costs and emissions associated. Unlike many similar studies, it was considered that the introduction of a separate collection of plastics affects not only quantities but also composition and heating value of the combustible waste.

The model is divided into two phases: parameter estimation and mathematical optimization. As it is often not possible to get specific collection costs by municipalities, a model based on the work of (Groot et al., 2014) was used for the first task. The optimization problem was formulated using mixed-integer linear programming.

The economical objective - Net present cost (NPC) - is given by the sum of operating and investment costs (Eq. 1). Binary variables ($\Psi_{i,t}$ and Ψ_i) account for the existence or not of a given unit (waste conversion units, sorting centres and recycling facilities, as well as transport units), while continuous variables ($f_{i,t}$ and f_i) account for the size of each unit. t_t^{op} represents the operating time of each time step $t \in T$.

Transport and supply chain nodes are represented by units i belonging to the set L , containing sub-sets Lu and Ls , for which investment is required and waste is sourced, respectively. Units have a fixed (c_i^{inv1}) and variable (c_i^{inv2}) investment, fixed ($c_{i,t}^{op1}$) and variable ($c_{i,t}^{op2}$) operating costs as well as fixed ($c_{i,t}^{imp1}$) and variable ($c_{i,t}^{imp2}$) environmental impacts. Each unit contains relevant information to define mass and energy requirements. Investment costs were annualised using an interest rate (ar) and project lifetime (n), while operating costs were discounted with a monthly interest rate (mr).

$$\min NPC = \sum_{t \in T} \frac{\sum_{i \in L} (c_{i,t}^{op1} \cdot \Psi_{i,t} + c_{i,t}^{op2} \cdot f_{i,t})}{(1 - mr)^t} \cdot t_t^{op} + \frac{i(1 + ar)^n}{(1 + ar)^n - 1} \cdot \sum_{i \in Lu} (c_i^{inv1} \cdot \Psi_i + c_i^{inv2} \cdot f_i) \quad Eq. 1$$

Multi-objective optimization was used to address conflicting objectives, particularly the trade-off between costs and emissions. Environmental impact - Eq. 2 - was used for (ε -constraints) formulation, using the GWP 100a method, corresponding to the IPCC 2013 global warming potential impact method, as well as the ecological scarcity method (ESM). The widespread use of the former and the tailor-made criteria for Switzerland of the latter motivated the choice. Values were retrieved from Ecoinvent database v.3.6.

$$ENV = \sum_{t \in T} \sum_{i \in L} (c_{i,t}^{imp1} \cdot \Psi_{i,t} + c_{i,t}^{imp2} \cdot f_{i,t}) \cdot t_t^{op} \leq \varepsilon \quad Eq. 2$$

Constraints were added to ensure full waste collection (Eq. 3), with $\dot{M}_{w,i,j,t}$ the mass of waste $w \in W$, received in each SC node and $gen_{w,i,t}$ the waste type generated per origin. Minimum (f_i^{min}) and maximum (f_i^{max}) facilities capacities (Eq. 4), as well as resource consumption (Eq. 5, Eq. 6), with $\dot{m}_{r,i,t}^+$ the reference quantity of resource $\forall r \in R$ needed in each SC node were also considered. Mass balances are closed (Eq. 7), for each SC node. Heat cascade formulation (Eq. 8, Eq. 9) was added, following the formulation developed by (Maréchal and Kalitventzeff, 1998), in which residual heat ($\dot{R}_{t,k}$) is transferred according to the 2nd law of thermodynamics, from higher ($k + 1$) to lower temperature levels (k).

$$\sum_{j \in L \setminus L_s} \dot{M}_{w,i,j,t} = gen_{w,i,t}, \forall w \in W, \forall i \in L_s, \forall t \in T \quad Eq. 3$$

$$f_i^{min} \cdot \Psi_{i,t} \leq f_{i,t} \leq f_i^{max} \cdot \Psi_{i,t}, \quad \forall i \in L, \quad \forall t \in T \quad Eq. 4$$

$$\dot{M}_{r,i,t}^+ = f_{i,t} \cdot \dot{m}_{r,i,t}^+, \quad \forall r \in R, \forall i \in L, \forall t \in T \quad Eq. 5$$

$$\dot{M}_{r,i,t}^- = f_{i,t} \cdot \dot{m}_{r,i,t}^-, \quad \forall r \in R, \forall i \in L, \forall t \in T \quad Eq. 6$$

$$\sum_{i \in L} \dot{M}_{r,i,t}^- = \sum_{i \in L} \dot{M}_{r,i,t}^+, \quad \forall r \in R, \forall t \in T \quad Eq. 7$$

$$\sum_{i \in L} \dot{Q}_{i,t,k} \cdot f_{i,t} + \dot{R}_{t,k+1} - \dot{R}_{t,k} = 0, \quad \forall t \in T, \quad \forall k \in K, \text{ with } T_{k+1} \geq T_k, \quad Eq. 8$$

$$\dot{R}_{t,k} \geq 0, \quad \dot{R}_{t,k_{max}+1} = \dot{R}_{t,1} = 0, \quad \forall t \in T \quad Eq. 9$$

The MILP problem is written in AMPL and solved by CPLEX (v12.7.0.0). Every experiment was run on a Microsoft Windows v10.0.18363 machine equipped with a 2.4 GHz Intel(R) Xeon (R) 8 core processor with 16GB of RAM.

3. Case Study

3.1. Structure

A waste management system in western Switzerland was used as case-study and comprises: (i) **Waste Source Nodes** - Municipalities, Waste Dumping Centers and Industries, with the source of waste assumed to be the centre of each municipality; (ii) **Recycling Chain** - Plastic waste from municipalities can be routed to different sorting centres, where they are separated and washed from contaminants. Plastics are then routed to a recycling centre; recycled polymers are absorbed by the market while by-products are routed to thermovalorization units; (iii) **Thermovalorization Centers** - Waste from municipalities, industries and separation centres (contaminants and by-products) go to thermovalorization units for heat and electricity recovery or cement plants to be used as fuel; (iv) **Transshipment Centers** - Only train stations are considered based on their ability for handling intermodal transport; (v) **Markets** - Recycled polymers, heat and electricity are absorbed by market units, mimicking real demand.

The waste streams considered are MSW (comprises domestic and industrial from municipalities), Polyethylene Terephthalate (PET) bottles, Polyethylene (PE) bottles (high and low density), and mixed plastics (MP) - which comprises other types of

plastics and small fractions of PE and PET. A tax on waste bags (0.058 \$/L) was considered as well as its variation due to separate collection of plastics.

3.2. Collection scenarios

Scenario analysis was used to study distinct collection policies. The optimization algorithm is then free to select the set of best paths for each type of waste at hand, with respect to compatibility constraints.

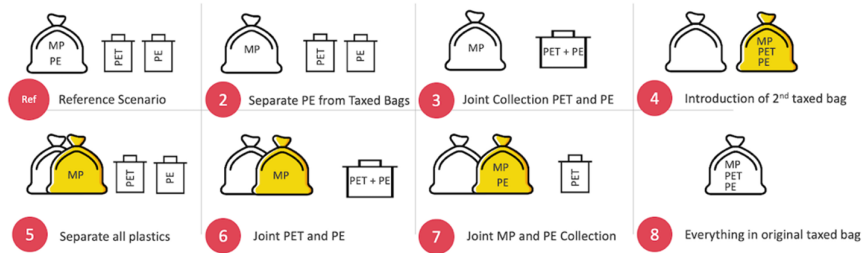


Figure 1 – Plastic collection scenarios - white bags refer to current implemented taxed bags and include MSW for all scenarios; yellow bags refer to taxed plastic bags.

Selected scenarios, defined in collaboration with the industrial partners (Figure 1), comprise the current state (Ref) used for benchmarking - on which PET and small fractions of PE are separately collected and recycled, with the remaining MSW routed for thermovalorization - as well as different configurations for plastics collection. Scenarios 4 to 7, on which a plastic taxed bag (0.035 \$/L) was considered, were also investigated. An extreme scenario (8) without any sorting was also studied.

4. Results and Discussion

For comparison, results were normalized by the amount of waste. illustrates the solution generation and dispersion for both environmental impacts assessed. Negative values indicate profitable systems and net environmental benefits.

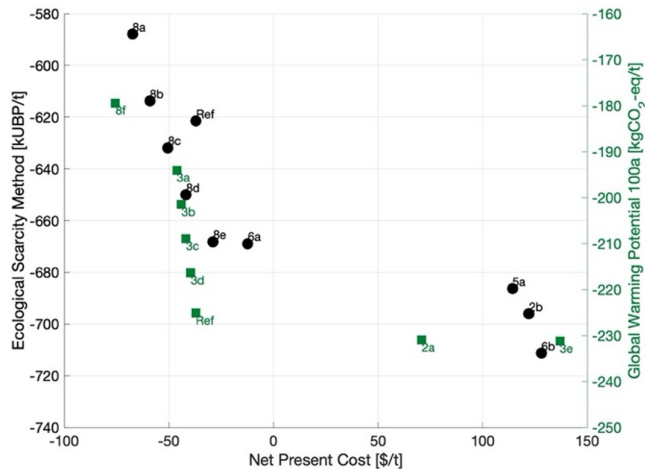


Figure 2 - Pareto points; circles refer to ESM and squares to the GWP 100a method; numbers next to symbols represent different collection scenarios (Figure 1) and letters indicate different configurations within each scenario; Ref identifies the reference collection scenario.

Regardless of the impact method, several system configurations achieve better economic and environmental metrics than the reference situation. For GWP, several configurations for collection strategy 3 (3a to 3d), in which PET and PE are collected together, can achieve moderately better (up to 25 %) economic performance compared with the reference system, though degrading environmental performance (up to 13 %). Solution for scenario 8, on which all wastes are collected in one undifferentiated bag, shows the lowest cost of all configurations due to reduced logistics costs - achieving a reduction of 100 % (equivalent to duplicate profit). However, environmental impact is degraded by 20 %. On the opposite spectrum of the plot, solution 2a and 3e perform slightly (up to 3 %) better environmentally but to the cost of considerable investment on new recycling units, which were chosen to address the increased collection of PE - converting thus the system to a non-profitable one - for the time horizon considered.

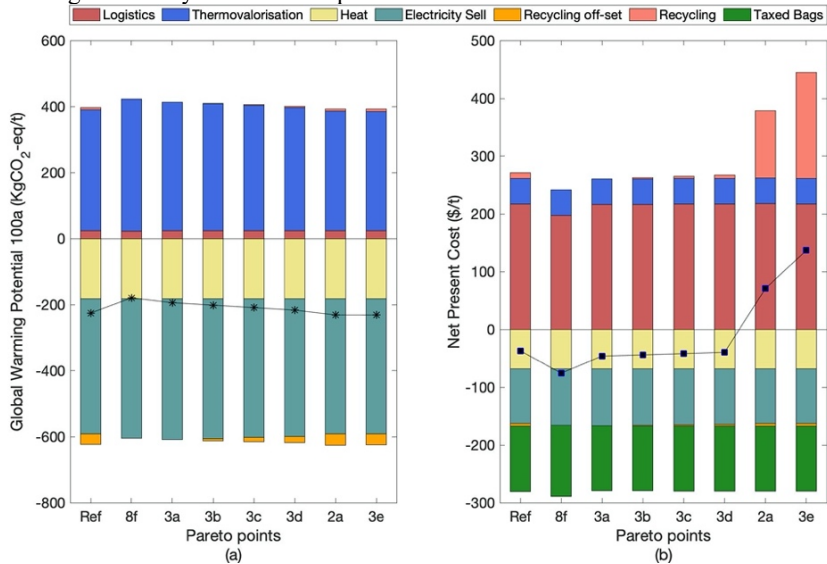


Figure 3 - Environmental impact (a) and NPC (b) partition by activity

ESM shows the benefit of using a more comprehensive indicator. Contrarily to GWP, the reference scenario does not belong to the Pareto curve, which implies that more interesting solutions are sure to be found. Similarly to GWP, solutions in which waste is collected all together (scenario 8) yield more profitable configurations to the price of degrading environmental indicators (8a and 8b); however, if natural gas co-burn is allowed, better environmental indicators are achievable (8c, 8d and 8e) associated with the off-set of heat and electricity. Solutions on the end spectrum (5a, 2a and 6b) perform better (up to 25 %) environmentally to the expense of heavy investment.

Cost and environmental partition by type of activity were studied in more detail for Pareto solutions using GWP indicator - Figure 3. Logistics have a considerable economic impact - close to 40 %, contrasting with very reduced (ca. 2 %) environmental impacts. Conversely, thermovalorization emissions account for the largest share of environmental impact albeit at a reduced cost. However, those emissions are compensated by the off-set of heat and electricity production for market demand. Recycling shows an environmental off-set larger than the associated emissions, although costs do not follow the same pattern - the recycling process can be more expensive than the production of the corresponding plastics. Besides, for solutions 2a and 3e, investment in new recycling

facilities surpasses the operational expenditures with thermovalorization and are comparable with the logistics annual costs; furthermore, the environmental benefit it brings comes at an expense of 18 and 28 \$/kgCO₂-eq, respectively.

The allocation of flows for each of the Pareto configurations shows that the total segregation of plastics (scenarios 4,5,6,7) is not part of the optimal set of solutions for the current set of waste flows and market prices. It seems reasonable to assume that the plastic content currently collected with the MSW taxed bag brings added value to the system and does not create a substantial environmental burden.

5. Conclusions

A superstructure-based approach comprising thermovalorization and recycling units embedded on a SC network was developed to study the collection and management of plastic waste. Cost and environmental indicators show that better economic performance is linked with scenarios with no plastic segregation, keeping the energetic potential of plastic within the thermovalorization perimeter and reducing logistics costs; however, it comes with higher environmental burdens. Reduction (3 % on average) of emissions is possible, for configurations that promote a separate collection of plastic waste and invest in new recycling facilities. Nevertheless, the cost associated is substantial.

Given the current set of assumptions, plastic waste currently collected with MSW taxed bag brings added value to the system, as a consequence of efficient energy recovery in thermovalorization units, not posing a major environmental threat. Solutions are heavily dependent on the nominal waste flows as well as environmental and economic assumptions. An extensive uncertainty analysis to first identify hotspots on the input parameters is desirable, followed by uncertainty propagation. On the technology side, if carbon capture, usage and storage (CCUS) technologies are added, different flow allocations might be expected, due to increasing environmental performance of thermovalorization units.

References

- Bhada-Tata, P., Kaza, S., Yao, L., Van Woerden, F., 2018. What a Waste 2.0 A Global Snapshot of Solid Waste Management to 2050, Urban Development Series.
- Bing, X., Groot, J.J., Bloemhof-Ruwaard, J.M., van der Vorst, J.G.A.J., 2013. Multimodal network design for sustainable household plastic recycling. *Int. J. Phys. Distrib. Logist. Manag.* 43, 452–477. <https://doi.org/10.1108/IJPDLM-04-2012-01134>
- Colvero, D.A., Ramalho, J., Gomes, A.P.D., Matos, M.A.A. de, Tarelho, L.A. da C., 2020. Economic analysis of a shared municipal solid waste management facility in a metropolitan region. *Waste Manag.* 102, 823–837. <https://doi.org/10.1016/j.wasman.2019.11.033>
- Groot, J., Bing, X., Bos-Brouwers, H., Bloemhof-Ruwaard, J., 2014. A comprehensive waste collection cost model applied to post-consumer plastic packaging waste. *Resour. Conserv. Recycl., SI:Packaging Waste Recycling* 85, 79–87. <https://doi.org/10.1016/j.resconrec.2013.10.019>
- Maréchal, F., Kalitventzeff, B., 1998. Energy integration of industrial sites: tools, methodology and application. *Appl. Therm. Eng.* 18, 921–933. [https://doi.org/10.1016/S1359-4311\(98\)00018-0](https://doi.org/10.1016/S1359-4311(98)00018-0)

Acknowledgements

The authors acknowledge the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754354, as well as FCT and P2020 under the project PTDC/EGEOGE/28071/2017, Lisboa -01.0145-Feder-28071 SFRH/BD/148499/2019 and SFRH/BD/143538/2019.

Rational Design of Ionic Liquid Phase-Change Material for Efficient Thermal Energy Storage

Teng Zhou,^{a,b,*} Huaiwei Shi,^a Kai Sundmacher^{a,b}

^a *Process Systems Engineering, Otto-von-Guericke University Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany*

^b *Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany*

* *Corresponding author: zhout@mpi-magdeburg.mpg.de*

Abstract

The selection of phase change material (PCM) is essential for the development of efficient thermal energy storage (TES) processes. Due to their benign properties, ionic liquids (ILs) have been demonstrated to be a new type of promising PCM. Unfortunately, there is a lack of systematic study on the optimal selection or molecular design of IL-PCMs for TES. In this work, a computer-aided molecular design (CAMD) method is proposed to systematically design IL-PCMs for a practical TES process. The influences of different IL properties, such as latent heat, heat capacity, and thermal conductivity, are simultaneously captured with a detailed TES process model. Optimal ILs holding a best compromise of all the properties are identified by solving an optimization-based CAMD problem where the TES performance of the process is maximized. It is found that our optimally designed IL-PCMs show much higher TES performance than the industrial state-of-the-art PCMs.

Keywords: phase change material, heat storage, ionic liquid, computer-aided molecular design

1. Introduction

Thermal energy is widely available and easy to access, which can be stored in the form of latent heat, sensible heat, or both. Generally, latent heat storage is more attractive than sensible heat storage because of its high energy density (Abhat 1983). Among all the latent heat storage approaches, the solid-liquid phase transition is most attractive for large-scale use due to its small volume change. In such a system, a phase change material (PCM) absorbs heat from a high-temperature heat transfer fluid (HTF) and melts whereas it releases heat to the cold HTF when it solidifies (Mehling and Cabeza, 2008).

PCM selection is essential for developing highly efficient TES systems. Organic PCMs usually have low thermal conductivity and they are normally volatile and flammable; inorganic salts suffer a lot from corrosion and supercooling (Kenisarin and Mahkamov, 2007). These drawbacks promote the development of new high-performing TES materials. Organic salts, which are also known as ionic liquids (ILs), consist of an organic cation and an organic or inorganic anion. This new type of chemicals is thermally stable, nonflammable and nonvolatile (Zhang et al., 2006). Moreover, the wide range of melting temperature spanning from negative up to 100 °C (Plechkova and Seddon, 2008) and the ability to tailor the properties (Zhou et al., 2012) by changing the cation, anion, and/or the substituents make ILs very promising PCMs. Despite that, except a few scattered

experimental studies (Terasawa et al., 2010; Vijayraghavan et al., 2013), there is a lack of systematic work on the optimal selection or molecular design of ILs for TES.

So far, most of the previous work is focused on the heat of fusion when selecting PCMs. However, one should note that in reality, rarely are these materials being used only during the phase change period. In fact, the latent heat storage is always combined with sensible heat storage. In this context, in addition to melting point and latent heat, heat capacity is also essential for selecting PCMs. Moreover, thermal conductivity needs to be considered as well because it can greatly affect the heat transfer rate. The impact of all the different properties can be captured only when a real TES process is studied and modelled.

In this work, a computer-aided molecular design (CAMD) based method (Austin et al., 2016; Song et al., 2018; Zhou et al., 2019; Zhou et al., 2021) is proposed to systematically design IL-PCMs for a practical TES process. Different group contribution-based models are employed to predict the key properties of ILs that affect the TES process performance. Combining these property models with a detailed TES process model, an optimization-based CAMD problem is finally formulated and solved to identify the best IL-PCM possessing a maximal TES performance.

2. Modeling

The TES device is depicted in Figure 1. As shown, the spherical PCM capsules with a diameter of 0.04 m are packed in a cylindrical TES tank ($D = 1.0\text{ m}$ and $l = 1.52\text{ m}$) where the high-temperature HTF water at 343.15 K flows through the tank with a mass flow rate $\dot{m} = 0.0796\text{ kg/s}$. Initially, all the PCM spheres have a temperature of 323.15 K and the space between PCM capsules in the tank is already filled with water at the same temperature. When the thermal charging process is started, hot water flows in from the top of the tank and meanwhile the solid PCMs start to absorb heat. The PCM starts to melt when the temperature reaches its melting point and the TES process is completely finished when all the PCMs reach 343.15 K . However, this process can be terminated earlier because the thermal charging rate decreases over time.

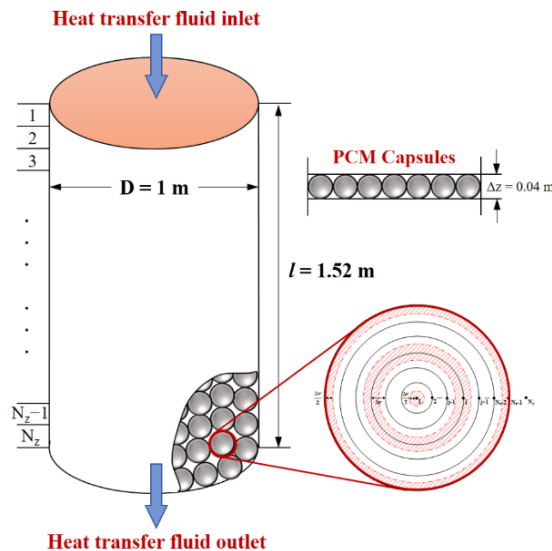


Figure 1. Layout and details of the thermal storage tank

In order to evaluate the TES performance of an IL-PCM for the thermal storage process illustrated in Figure 1, two types of models are generally required. They are IL property models that link IL structures to their physical properties and process models that predict the TES performance based on the IL properties.

2.1. Process Modeling

It is assumed that the HTF water has a one-dimensional axial flow and temperature distribution. Only heat convection is considered in this flow direction. For the PCM spheres, one-dimensional radial heat conduction is considered. The density, thermal conductivity, and heat capacity of the PCM are assumed constant during the whole TES process. Moreover, the heat conduction between different PCM spheres are neglected and the ambient heat losses and radiation heat transfer are also neglected. As indicated in Figure 1, the cylindrical tank is discretized into 38 equal sections ($N_z = 38$) in the fluid flow direction. The height of each section equals 0.04 m, making one tank section containing exactly a single layer of PCM capsules. The entire PCM capsule is discretized into 13 control volumes ($N_r = 14$) along the radial direction and the discretization index of N_r represents the position where water locates. Such a TES system is governed by a set of heat balances in the form of partial differential equations, which are discretized and solved as algebraic equations using the explicit finite difference method (Shi et al., 2021).

2.2. Property Modeling

As indicated by the process model, five thermo-physical properties of PCM, i.e., melting point T_m , latent heat L , density ρ , thermal conductivity λ , and heat capacity C_p , are required to quantify the TES performance of a PCM. Group contribution (GC) methods (Zhou et al., 2018; Song et al., 2020) are widely used in CAMD to predict physical properties of compounds (Austin et al., 2016), due to their easy incorporation within mathematical optimizations and high qualitatively correct estimations. Different GC models are employed to predict the five properties of ILs. They have been elaborated in Shi et al. (2021).

2.3. TES Performance Evaluation

The thermal storage power of the system represents the TES performance of an IL-PCM. However, one should note that the thermal storage power is not a constant during the charging process. In the beginning, heat transfer is fast because the temperature gradient between PCM and water is high. However, the thermal storage power decreases gradually as the PCM temperature gets closer to the water temperature. Therefore, an average TES power, defined as the total stored heat over the thermal charging time, is used to represent the overall TES performance of the material.

3. Results and Discussion

3.1. Effect of IL Properties on the TES Power

As evidenced by the process models, five IL properties including the melting point T_m , latent heat L , density ρ , thermal conductivity λ , and heat capacity C_p are required to quantify the TES power of the IL-PCM. When the T_m of the IL is low, the temperature gradient on the solid-liquid interface (i.e., between the outer-layer melted PCM and the inner-layer melting PCM) is large, which makes the latent heat absorption faster and consequently leads to a higher TES power. Except T_m , all the other properties show positive effects on the TES performance of the material and the reason can be easily understood from the physical knowledge about the TES system. Additionally, the

significance of the property effect can be quantified by a parameter sensitivity study with the process models.

3.2. Computer-Aided IL Design

The CAMD method is used to design the optimal IL-PCM in this work. In total, 59 IL building groups including 54 cation-anion-pair groups and 5 substituent groups are considered for IL design. In CAMD, a molecule is represented by a vector of group numbers. Typically, several equality and inequality constraints on the number of groups must be considered in the CAMD program to ensure that the generated ILs are structurally reasonable. For the complete list of the 59 IL groups and all the necessary structural constraints, one could refer to Shi et al. (2021) where the solution strategy of the CAMD problem is also elaborated.

In order to determine the average thermal storage power of an IL-PCM in the TES system (see Figure 1), a total thermal charging time needs to be pre-specified. In this work, we assume that the hot water at 70 °C is generated by solar heat and thus the thermal charging time depends on the total sunshine duration. Figure 2 shows the average monthly sunshine hours in Germany from June 2019 to May 2020 (Statista website). As indicated, the annually average sunshine time is about 5.5 hours/day. Thus, 5.5 hours is set as our thermal charging time for the IL-PCM design.

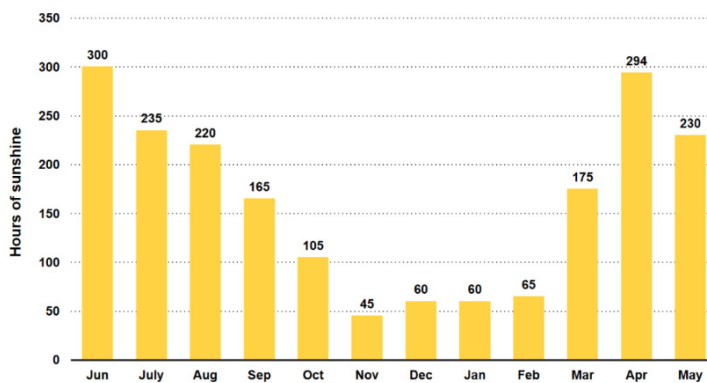


Figure 2: Average sunshine time in Germany from June 2019 to May 2020

3.3. Design Results

The group combinations and molecular structures of top nine ILs are listed in Table 1 and plotted in Figure 3, respectively. The five key physical properties and corresponding average TES power (P_{ow}) of the PCMs are also summarized in Table 1. In order to demonstrate the advantage of model-based PCM design, the TES performance of the designed IL-PCMs is compared with that of an industrial benchmark PCM paraffin wax # 60. As indicated, all of the top nine ILs show a substantially higher TES power than the wax. When comparing them, it is found that the influence or contribution of the heat capacity and thermal conductivity, indicated by the product of the property difference and its corresponding sensitivity value, on the thermal storage power is very small. Even though the wax has a higher latent heat that can increase its thermal storage power, wax's higher melting point (note T_m has a large negative effect) and the much lower density make this material worse than the designed ILs.

Table 1. Top 9 IL-PCMs identified from CAMD for 5.5-hour thermal charging time
(Pow in W, T_m in K, C_p in J/(kg·K), ρ in kg/m³, L in kJ/kg, λ in W/(m·K))

Rank	Group combinations	Pow	T_m	C_p	ρ	L	λ
1	2 CH ₂ , 1 OH, 1 [MPy][TfO]	5277	323.8	2171	1426	134.3	0.158
2	2 CH ₂ , 1 OCH ₃ , 1 [MPy][TfO]	5225	324.0	1758	1333	151.6	0.152
3	1 CH ₂ , 1 OCH ₂ , 1 CH ₃ , 1 [MPy][TfO]	4995	324.0	1758	1333	115.8	0.152
4	1 CH ₂ , 1 OCH ₃ , 1 [MIm][TfO]	4958	323.2	1644	1387	94.60	0.160
5	1 CH ₂ , 1 OH, 1 [MPy][TfO]	4760	327.6	2166	1474	137.3	0.162
6	1 CH ₂ , 1 OCH ₂ , 1 OCH ₃ , 1 [MIm][I]	4719	323.4	1812	1450	71.70	0.159
7	1 CH ₂ , 1 OCH ₃ , 1 [MIm][NO ₃]	4684	325.7	1538	1221	129.7	0.196
8	2 OCH ₃ , 1 [Im13][BF ₄]	4648	325.3	1661	1314	100.7	0.183
9	1 CH ₂ , 1 OCH ₃ , 1 [MPy][TfO]	4644	327.7	1793	1374	155.3	0.156
BM	Paraffin Wax # 60	3403	333.1	2150	850	190.0	0.230

[MPy][TfO] is found to be the most favorable IL group, and OH and OCH₃ are the preferred functional groups for maximizing the TES power of the ILs. It is also found that two CH₂ groups are most beneficial for the ILs [MPy(CH₂)_nOH][TfO] and [MPy(CH₂)_nOCH₃][TfO]. Increasing the alkyl-chain length results in a lower melting point, leading to a higher TES power due to the large negative effect of T_m . However, when the number of CH₂ group is larger than 2, the melting point becomes lower than 323.15 K, which results in a significantly lower TES power due to the absence of latent heat storage contribution (note that the initial temperature of the PCM is 323.15 K). Moreover, when comparing the first with the second ILs and comparing the fifth with the ninth ILs, it is found that OH is slightly better than OCH₃ when [MPy][TfO] is selected as the main group. The reason is that even though OCH₃ contributes more to the latent heat, the OH group has a larger contribution to both the heat capacity and density, which determines its superiority over OCH₃ for increasing the TES power of the IL-PCM.

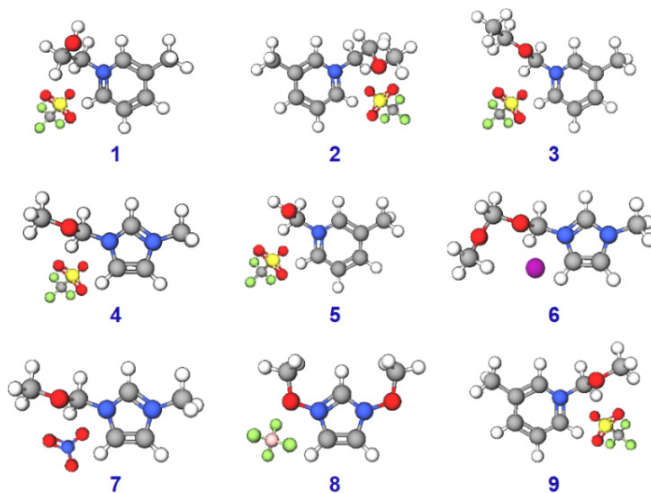


Figure 3: Molecular structures of the top 9 ILs identified from CAMD

4. Conclusions

The CAMD method is employed to design ILs as PCMs for a practical TES process. The influence of different IL properties on the TES performance is simultaneously captured with a detailed TES process model. The optimal IL structure that possesses a best compromise of all the properties is identified by solving an optimization-based CAMD problem where the average TES power of the IL-based thermal storage system is maximized. It is found that our designed ILs perform better than the conventional PCM in terms of the average TES power.

References

- A. Abhat, 1983, Low temperature latent heat thermal energy storage: heat storage materials, *Sol Energy*, 30, 4, 313-332
- H. Mehling, L.F. Cabeza, *Heat and Cold Storage with PCM*, Springer, 2008
- M. Kenisarin, K. Mahkamov, 2007, Solar energy storage using phase change materials, *Renew Sust Energ Rev*, 11, 9, 1913-1965
- S. Zhang, N. Sun, X. He, X. Lu, X. Zhang, 2006, Physical properties of ionic liquids: database and evaluation, *J Phys Chem Ref Data*, 35, 4, 1475-1517
- N.V. Plechkova, K.R. Seddon, 2008, Applications of ionic liquids in the chemical industry, *Chem Soc Rev*, 37, 1, 123-150
- T. Zhou, L. Chen, Y. Ye, L. Chen, Z. Qi, H. Freund, K. Sundmacher, 2012, An overview of mutual solubility of ionic liquids and water predicted by COSMO-RS, *Ind Eng Chem Res*, 51, 17, 6256-6264
- N. Terasawa, S. Tsuzuki, T. Umecky, Y. Saito, H. Matsumoto, 2010, Alkoxy chains in ionic liquid anions; effect of introducing ether oxygen into perfluoroalkylborate on physical and thermal properties, *Chem Commun*, 46, 10, 1730-1732
- R. Vijayraghavan, U.A. Rana, G.D. Elliott, D.R. MacFarlane, 2013, Protic ionic solids and liquids based on the guanidinium cation as phase-change energy-storage materials, *Energy Technol*, 1, 10, 609-612
- N.D. Austin, N.V. Sahinidis, D.W. Trahan, 2016, Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques, *Chem Eng Res Des*, 116, 2-26
- Z. Song, C. Zhang, Z. Qi, T. Zhou, K. Sundmacher, 2018, Computer-aided design of ionic liquids as solvents for extractive desulfurization, *AIChE J*, 64, 3, 1013-1025
- T. Zhou, Z. Song, X. Zhang, R. Gani, K. Sundmacher, 2019, Optimal solvent design for extractive distillation processes: a multiobjective optimization-based hierarchical framework, *Ind Eng Chem Res*, 58, 15, 5777-5786
- T. Zhou, H. Shi, X. Ding, Y. Zhou, 2021, Thermodynamic modeling and rational design of ionic liquids for pre-combustion carbon capture, *Chem Eng Sci*, 229, 116076
- H. Shi, X. Zhang, K. Sundmacher, T. Zhou, 2021, Model-based optimal design of phase change ionic liquids for efficient thermal energy storage, *Green Energy Environ*, <https://doi.org/10.1016/j.gee.2020.12.017>
- T. Zhou, S. Jhamb, X. Liang, K. Sundmacher, R. Gani, 2018, Prediction of acid dissociation constants of organic compounds using group contribution methods, *Chem Eng Sci*, 183, 95-105
- Z. Song, H. Shi, X. Zhang, T. Zhou, 2020, Prediction of CO₂ solubility in ionic liquids using machine learning methods, *Chem Eng Sci*, 223, 115752
- Statista, Average monthly sunshine hours in Germany, <https://www.statista.com/statistics/982758/average-sunshine-hours-germany/>, 2020.

Tailor-made solvents by integrated design of molecules and CO₂ absorption processes

Johannes Schilling,^{a,b} Madlen Hopp,^c Joachim Gross,^c André Bardow,^{a,b,d,*}

^a*Energy and Process Systems Engineering, Department of Mechanical and Process Engineering, ETH Zurich, Tannenstrasse 3, 8092 Zurich, Switzerland*

^b*Institute of Technical Thermodynamics, RWTH Aachen University, Schinkelstrasse 8, 52062 Aachen, Germany*

^c*Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany*

^d*Institute of Energy and Climate Research - Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Strasse, 52425 Jülich, Germany*

abardow@ethz.ch

Abstract

Solvents are key to many chemical processes. To find optimal combinations of solvents and processes, solvent and process design have to be integrated. In this work, we present a method for the integrated thermo-economic design of separation processes and solvents that achieves a high level of the desired integration. The method is based on our 1-stage Continuous-Molecular Targeting-Computer-aided Molecular Design method. The PC-SAFT equation of state is used as thermodynamic model to calculate both equilibrium and transport properties. Computer-aided molecular design allows designing the solvent as a degree of freedom of the process optimization. To allow for ease of use, we implemented the method in the process flowsheeting software gPROMS ProcessBuilder, enabling the convenient definition of the flowsheet. We present the method for the integrated design of a physical absorption process to separate CO₂ from power plant flue gases.

Keywords: CAMPD, PC-SAFT, CO₂ absorption, carbon capture.

1. Introduction

Identifying optimal solvents for separations is an essential task in the chemical industry. Economically efficient separations require an optimal combination of process and solvent. For this purpose, solvent design must be integrated into process design (Papadopoulos et al., 2018). However, the integration of solvent design adds many discrete degrees of freedom to the process optimization leading to mixed-integer nonlinear programs (MINLP) (Gani, 2004). To solve the resulting MINLP problems, systematic solution approaches have been developed in recent years (Papadopoulos et al., 2018).

Many solution methods employ equilibrium thermodynamic models. In this case, process performance can be assessed by a thermodynamic objective function, as, e.g., demonstrated for the design of CO₂ absorption processes by some of the authors (Lampe et al., 2015; Stavrou et al., 2014). However, thermodynamically optimal solvents can still be undesirable in terms of economics since investment costs are neglected.

To quantify investment cost, equipment design needs to be integrated into the process and molecular design. Equipment sizing additionally requires the transport properties of the designed molecule. In pioneering work, Pereira et al. (2011) directly integrated a CAMD formulation into process design for physical CO₂ absorption. Here, transport properties

and equipment sizing were based on empirical correlations. While limited to the design of n-alkanes, the work demonstrates the advantages of the integrated thermo-economic design. The group extended their work, e.g., by domain reduction (Gopinath et al., 2016) to remove infeasible process and molecular design regions. However, the lack of consistent models for transport properties with strong predictive power has enforced equipment sizing based on heuristics, which may not accurately capture trade-offs between processes, equipment, and molecule. Furthermore, the application of such design methods is usually challenging and limited to expert software, limiting their wider adoption.

In this work, we present the integrated thermo-economic design of solvents and separation processes based on the 1-stage Continuous-Molecular Targeting-Computer-Aided Molecular Design (1-stage CoMT-CAMD) method (Schilling et al., 2017a; Schilling et al., 2017b). In 1-stage CoMT-CAMD, a consistent thermodynamic model for equilibrium and transport properties is used, the PC-SAFT equation of state. The integration into the process flowsheeting software gPROMS ProcessBuilder enables a straightforward application of the integrated design. So far, 1-stage CoMT-CAMD has only been applied for Organic Rankine Cycles. Here, we extend the method for the integrated thermo-economic design of solvents and separation processes. For this purpose, diffusion coefficients and mass transfer need to be predicted. We demonstrate the method for a physical CO₂ absorption process.

2. The framework of 1-stage CoMT-CAMD

1-stage CoMT-CAMD aims at maximizing the economic benefits offered in the molecular design space (Schilling et al., 2017b). The resulting problem is formulated as MINLP, Problem (1), based on the generic formulation introduced by (Gani, 2004):

$$\begin{array}{ll}
 \min_{x, y^s} & f(x, \theta, \kappa) \\
 \text{s.t.} & \left. \begin{array}{l}
 g_1(x, \theta, \kappa) = 0 \\
 g_2(x, \theta, \kappa) \leq 0 \\
 \kappa = k(x, \theta, y^s) \\
 p_1(x, \theta) = 0 \\
 p_2(x, \theta) \leq 0 \\
 \theta = t(x, y^s) \\
 F_1 \cdot y^s = d \\
 F_2 \cdot y^s \leq d \\
 x_{\text{lb}} \leq x \leq x_{\text{ub}} \in \mathbb{R}^m \\
 y_{\text{lb}}^s \leq y^s \leq y_{\text{ub}}^s \in \mathbb{Z}^l
 \end{array} \right\} \begin{array}{l}
 \text{equipment model} \\
 \text{PC-SAFT (transport)} \\
 \text{process model} \\
 \text{PC-SAFT (equilibrium)} \\
 \text{CAMD} \\
 \text{process degrees of freedom} \\
 \text{molecular structure}
 \end{array} \quad (1)
 \end{array}$$

In Problem (1), a thermo-economic objective function f is optimized (e.g., the net present value), depending on process and equipment variables x (e.g., pressure levels), equilibrium properties θ (e.g., enthalpies), and transport properties κ (e.g., viscosities). The process and equipment models encompass equality constraints g_1, p_1 (e.g., energy balances or mass transfer correlations) and inequality constraints g_2, p_2 (e.g., limitations of pressure levels or flooding conditions), respectively. The equilibrium properties θ and transport properties κ of the molecules are calculated using the PC-SAFT equation of state (Gross and Sadowski, 2001) (for details, see Section 3). A CAMD formulation is used to integrate the molecular design into the process design. Here, the designed molecule is represented by the integer vector y^s , which contains the number of a particular group of the molecular structure. The structural feasibility of the molecular structure is ensured by equality and inequality CAMD constraints (F_1, F_2) (Struebing, 2011).

The resulting degrees of freedom of the full MINLP problem are the process and equipment variables x and the molecular structure of the molecule y^S . To allow for an easy and fast application of the method, we have integrated the 1-stage CoMT-CAMD method into the commercial process modeling software gPROMS ProcessBuilder (Process Systems Enterprise, 1997-2021) in the work of Schilling et al. (2020). Thereby, the process flow-sheet can be quickly developed and adapted using the equipment model libraries of ProcessBuilder. The CAMD model and equipment sizing models are directly integrated as libraries into ProcessBuilder. Thus, the overall integrated design problem can be set up based on drag-and-drop flowsheeting. The MINLP can be directly defined and solved in gPROMS ProcessBuilder, allowing for an integrated design with one user-friendly tool. The MINLP is solved by the standard local MINLP solver OAERAP (Outer Approximation/Equality Relaxation/Augmented Penalty) of gPROMS ProcessBuilder. OAERAP combines an outer-approximation formulation with a relaxation strategy. Initially, the integer variables are relaxed to identify a hypothetical, optimal molecule, the target. Subsequently, an outer-approximation formulation identifies the optimal real molecule jointly with the corresponding optimal process settings and equipment sizes. Integer-cuts are used to obtain a ranking of optimal molecules. Here, the MINLP is solved repeatedly and previous solutions are excluded from the design space using integer-cut constraints.

3. Equilibrium and transport properties based on PC-SAFT

The key to the proposed method is the detailed molecular picture underlying the modern thermodynamic model used to calculate the equilibrium and transport properties of the designed molecules: the perturbed-chain statistical associating fluid theory (PC-SAFT) (Gross and Sadowski, 2001). In PC-SAFT, each molecule is described by a set of typically 3 to 7 pure component parameters. The physical background of these parameters makes the overall model based on PC-SAFT particularly suited for relaxation strategies in optimization algorithms. PC-SAFT is linked to the CAMD formulation by the homosegmented group-contribution (GC) approach of PC-SAFT (Sauer et al., 2014), which calculates the pure component parameters from the molecular structure y^S .

PC-SAFT is used to calculate both equilibrium and transport properties in every state of the process. The model is based on the residual Helmholtz energy. To obtain absolute equilibrium properties, we calculate the heat capacity of the ideal gas as reference property from the molecular structure using a GC approach (Joback and Reid, 1987). The transport properties κ are calculated from PC-SAFT based on Rosenfeld's entropy-scaling. Therein, the transport properties were found to depend on the residual entropy univariately. Using the dependency, GC approaches have been developed by some of the authors for viscosities η (Lötgering-Lin et al., 2018), thermal conductivities λ (Hopp and Gross, 2019), and self-diffusion coefficients D^S (Hopp et al., 2018) based on entropy-scaling and PC-SAFT. From the self-diffusion coefficient D_k^S of component k , we calculate the diffusion coefficient D_{ik}^∞ in infinite dilution of component i in component k using the empirical correlation:

$$D_{ik}^\infty = D_k^S \left(\frac{m_k^{0.34} \sigma_k}{0.5(m_i^{0.34} \sigma_i + m_k^{0.34} \sigma_k)} \right)^{2.97}, \quad (2)$$

where the segment number m and diameter σ are pure component parameters of PC-SAFT. The diffusion coefficients in infinite dilution are transformed into Maxwell-Stefan diffusion coefficients using the predictive Darken equation (Liu et al., 2011). From the Maxwell-Stefan diffusion coefficients, we calculate the required Fick diffusion coefficients using the thermodynamic factor (Taylor and Krishna, 1993) from PC-SAFT.

In this work, we consider functional groups of branched alkanes, cyclic alkanes, and aromatics. We constrain the molecular design space of the cyclic molecules to single ring structures. Further functional groups can be easily considered as soon as the group contributions are fitted to measurement data or the polar contribution of PC-SAFT is integrated into gPROMS ProcessBuilder.

4. Case study: Design of a carbon capture system with physical absorption

We exemplify 1-stage CoMT-CAMD for the integrated design of the solvent and process of a carbon capture system with physical absorption. We investigate precombustion carbon dioxide capture from the syngas stream of a coal-fired integrated gasification combined cycle power plant. The syngas feed contains hydrogen, carbon dioxide, and water. We notably account for water in the system due to its significant impact on the phase equilibria. The process topology includes the primary unit operations absorption, desorption, and carbon dioxide compression (Figure 1). Following Stavrou et al. (2014), the absorber has a fixed number of 7 equilibrium stages with a constant CO₂ capture rate of 90%. The desorption proceeds in two stages: a first flash stage, which feeds a hydrogen-rich gas back to the absorber, followed by a desorption flash. The process parameters are taken from Stavrou et al. (2014). The heat exchangers are modeled as shell-and-tube heat exchangers in counter-flow, assuming single-phase forced convection (Gnielinski, 1975).

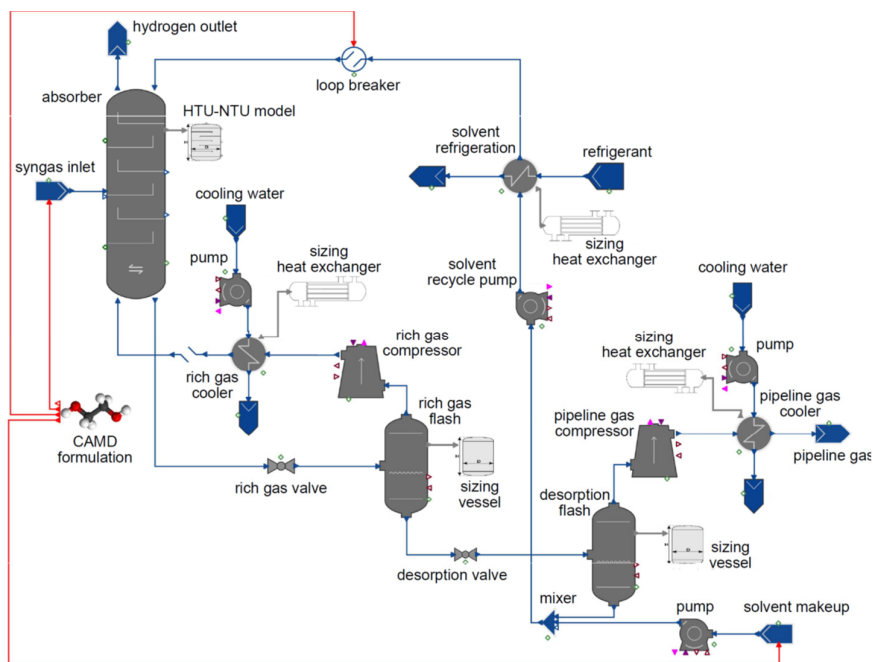


Figure 1: Flowsheet for CO₂ capture by physical absorption. The CAMD model and the sizing models of the equipment are directly linked to the process flowsheet. The flowsheet is created using gPROMS ProcessBuilder 1.2.

The absorption column is sized based on the height and number of transfer units using the so-called HTU-NTU method (Sinnott, 2005). The height of a transfer unit is sized

using the mass transfer correlation of Onda et al. (1968) depending on the mixture's viscosity and diffusion coefficients. We consider binary diffusion coefficients of hydrogen/carbon dioxide (gas phase) and solvent/carbon dioxide (liquid phase). The column diameter is obtained from the flooding conditions (Sinnott, 2005).

The resulting MINLP optimization problem consists of 5 continuous process degrees of freedom (e.g., mass flow rates or pressure levels) and 33 binaries representing the molecular structure of the solvent. The objective function f_{tot} is defined as the annual costs per mass of captured CO₂ calculated from investment cost and operational cost for electricity, cooling, solvent loss, and hydrogen loss. We assume a depreciation period of $n = 5$ a and an interest rate of $i = 0.08$. Initially, the relaxation problem is solved, leading to a hypothetical, optimal solvent, the target, with a target value of 15.99 €/t_{CO₂}. A ranking of the top 5 solvents is calculated using 1-stage CoMT-CAMD, identifying branched C10-C12 alkanes as the best solvents (Table 1). The optimal identified solvent is n-undecane, with costs of 16.0 €/t_{CO₂}, which is very close to the target, indicating no further room for improvement in the considered molecular space. In practice, alkanes are not used for physical CO₂ absorption. Thus, while the results demonstrate the advantages of an integrated design, future work should integrate further promising functional groups for CO₂ absorption, e.g., ethers. The identified solvents show an optimal trade-off between operational and investment costs. Solvents with longer chains reduce the operational cost f_{op} but lead to higher investment costs f_{inv} due to higher viscosities. If only the operational cost f_{op} without investment cost f_{inv} is considered as the objective function, branched C12-C13 alkanes, and thus solvents with longer chains are identified. The trade-off between economics, process, and molecule is captured within our integrated design framework leading to overall optimal solutions. Identifying a ranking of the top 5 solvents requires ca. 22 h using an Intel-Xeon CPU with 3.0 GHz and 64 GB RAM.

Table 1: Target and top 5 identified solvents minimizing the total annual cost per mass of captured CO₂ f_{tot} , and the annual operational (f_{op}) and investment (f_{inv}) costs per mass of captured CO₂.

Rank	Name	f_{tot} in €/t _{CO₂}	f_{op} in €/t _{CO₂}	f_{inv} in €/t _{CO₂}
-	target	15.99	12.70	3.29
1	n-undecane	16.00	12.71	3.29
2	n-dodecane	16.08	12.67	3.41
3	methyldecane	16.13	12.78	3.35
4	dimethylnonane	16.32	12.88	3.44
5	n-decane	16.47	13.29	3.18

5. Conclusion

In this work, we present a method for the integrated thermo-economic design of solvents and separation processes in chemical engineering. For this purpose, modern physically-based thermodynamic models based on PC-SAFT were linked to a CAMD formulation and rigorous models of the separation processes. Our 1-stage CoMT-CAMD method enables the solvent design based on thermo-economic objectives using detailed models to size the equipment. We implemented the method into the commercial process modeling software gPROMS ProcessBuilder to allow for a straightforward definition of the integrated design problem. For a physical absorption process to separate CO₂ from power plant flue gases, our method systematically identifies the thermo-economically optimal solvent and the corresponding optimal process. The integrated design of solvents and separation processes allows reducing the cost for CO₂ capture.

References

- R. Gani, 2004. Chemical product design: Challenges and opportunities. *Comp. Chem. Eng.* 28 (12), 2441–2457.
- V. Gnielinski, 1975. New equations for heat and mass transfer in the turbulent flow in pipes and channels. NASA STI/Recon Technical Report A 75, 8–16.
- S. Gopinath, et al., 2016. Outer approximation algorithm with physical domain reduction for computer-aided molecular and separation process design. *AIChE J* 62 (9), 3484–3504.
- J. Gross, G. Sadowski, 2001. Perturbed-Chain SAFT: An equation of state based on a perturbation theory for chain molecules. *Ind. Eng. Chem. Res.* 40 (4), 1244–1260.
- M. Hopp, J. Gross, 2019. Thermal conductivity from entropy scaling: A group-contribution method. *Ind. Eng. Chem. Res.* 58 (44), 20441–20449.
- M. Hopp, J. Mele, J. Gross, 2018. Self-diffusion coefficients from entropy scaling using the PCP-SAFT equation of state. *Ind. Eng. Chem. Res.* 57 (38), 12942–12950.
- K.G. Joback, R.C. Reid, 1987. Estimation of pure-component properties from group-contributions. *Chem. Eng. Comm.* 57 (1-6), 233–243.
- M. Lampe, et al., 2015. Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT. *Comp. Chem. Eng.* 81, 278–287.
- X. Liu, T.J.H. Vlught, A. Bardow, 2011. Predictive Darken equation for Maxwell-Stefan diffusivities in multicomponent mixtures. *Ind. Eng. Chem. Res.* 50 (17), 10350–10358.
- O. Lötgering-Lin, et al., 2018. Pure substance and mixture viscosities based on entropy scaling and an analytic equation of state. *Ind. Eng. Chem. Res.* 57 (11), 4095–4114.
- K. Onda, H. Takeuchi, Y. Okumoto, 1968. Mass transfer coefficients between gas and liquid phases in packed columns. *J. Chem. Eng. Jpn.* 1 (1), 56–62.
- A.I. Papadopoulos, et al., 2018. Computer-aided molecular design: Fundamentals, methods, and applications, in: Reedijk, J. (Ed.), Reference module in chemistry, molecular sciences and chemical engineering. Elsevier, pp. 1–76.
- F.E. Pereira, et al., 2011. Integrated solvent and process design using a SAFT-VR thermodynamic description: High-pressure separation of carbon dioxide and methane. *Comp. Chem. Eng.* 35 (3), 474–491.
- Process Systems Enterprise, 1997–2021. gPROMS.
- E. Sauer, M. Stavrou, J. Gross, 2014. Comparison between a homo- and a heterosegmented group contribution approach based on the perturbed-chain polar statistical associating fluid theory equation of state. *Ind. Eng. Chem. Res.* 53 (38), 14854–14864.
- J. Schilling, et al., 2017a. 1-stage CoMT-CAMD: An approach for integrated design of ORC process and working fluid using PC-SAFT. *Chem. Eng. Sci.* 159, 217–230.
- J. Schilling, et al., 2017b. From molecules to dollars: Integrating molecular design into thermo-economic process design using consistent thermodynamic modeling. *Mol. Syst. Des. Eng.* 2 (3), 301–320.
- J. Schilling, C. Horend, A. Bardow, 2020. Integrating superstructure-based design of molecules, processes and flowsheets. *AIChE J* 66 (5), e16903.
- R. Sinnott, 2005. *Chemical Engineering Design: Chemical Engineering Volume 6*. Elsevier.
- M. Stavrou, et al., 2014. Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for CO₂ capture. *Ind. Eng. Chem. Res.* 53 (46), 18029–18041.
- H. Struebing, 2011. Identifying optimal solvents for reactions using quantum mechanics and computer-aided molecular design. PhD thesis. Department of Chemical Engineering, London, UK.
- R. Taylor, R. Krishna, 1993. *Multicomponent mass transfer*. John Wiley & Sons.

Small-scale GTL Applications with Heat Integration in Reforming and Fischer-Tropsch Stages

Ligia B. Dias^{a,b}, Frederico S. Coelho^b, Fernando G. Martins^c, José Carlos B. Lopes^{a,b}, Mariana G. Domingos^{a,b,}*

^aCoLAB NET4CO₂, Rua Júlio de Matos, 828-882, 4200-355 Porto, Portugal

^bLSRE-LCM - Laboratory of Separation and Reaction Engineering - Laboratory of Catalysis and Materials, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

^cLEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal

mariana.domingos@net4co2.pt

Abstract

The associated natural gas produced on crude oil offshore platforms is often flared. The potential of two small-scale Gas-to-Liquids processes to monetise this gas whilst reducing carbon dioxide emissions is investigated in this work. Compact NETmix microreactors are considered for the reactor stages, and the process energy expenditure is reduced through heat integration of endothermic reforming stages, where syngas (a mixture of hydrogen and carbon monoxide) is produced, with exothermic Fischer-Tropsch stages, which convert syngas into liquid fuels. The base concept of the two Gas-to-Liquids processes considered differs mainly on the reforming stage: Steam Methane Reforming or Dry Methane Reforming. The two Gas-to-Liquids routes proposed are compared in terms of capital expenditure, operational expenditure and their emissions reduction potential compared to flaring.

Keywords: Gas-to-Liquids, Steam Methane Reforming, Dry Methane Reforming, Heat Integration, Aspen

1. Introduction

According to the World Bank's Global Gas Flaring Reduction Partnership (GGFR), gas flaring releases about 400 million tons of CO₂ equivalent emissions each year (GGFR, 2020). Remote offshore crude oil platforms and floating vessels are amongst the main contributors to these emissions. Given their limited access to a gas market, when associated gas cannot be reinjected to increase oil production, or used onsite, it ends up being flared or vented to the atmosphere (IEA, 2020). Flaring reduction initiatives, like Zero Routine Flaring by 2030 from the World Bank, and tighter environmental regulations are driving a growth in the Gas-to-Liquids (GTL) market.

Gas-to-Liquids processes comprise two main steps: 1) conversion of gas, such as natural gas (NG) into a mixture of carbon monoxide (CO) and hydrogen (H₂), called synthesis gas or syngas, followed by 2) conversion of this syngas into a mixture of long liquid hydrocarbon (HC) chains by Fischer-Tropsch (FT) reactions. The most common industrial method for syngas production is Steam Methane Reforming (SMR). However, given the potential to re-use part of the carbon dioxide present in some reserves, Dry Methane Reforming (DMR) is also receiving increased attention (Wang et al., 2018).

In this work, the potential of both routes to monetise NG flared at a crude oil production facility is assessed. For GTL processes to present an attractive alternative for monetisation of associated gas and remote/stranded natural gas reserves, where conventional means of transportation are not practical or economical, it is essential to develop technologies for small scales (100-1000 bpd) with a fraction of the capital costs of conventional large-scale plants (Chen and Grossman, 2019). The study considers small-scale facilities, with 500 bpd production scale, and performs process simulation calculations, heat integration and economic estimations using AspenTech software packages. To reduce the installation size and improve the energy efficiency of the process, compact microreactors based on the NETmix technology are assumed for all reaction steps. These reactors are characterized by very high heat transfer coefficients (Costa et al., 2017), which together with heat integration of the processes mitigate the two major challenges associated with small-scale GTL.

2. Methodology and Process Description

The processes were simulated in Aspen Plus[®] using the Peng-Robinson property model, recommended when hydrocarbons and light gases are present, and the ENRTL-RK model for process sections containing electrolytes. The process flowsheets from Figure 1 comprise a heat exchanger network, capable of integrating process heat and minimizing energy consumption, which was developed in Aspen Energy Analyzer[™] through Pinch analysis. Equipment costs were estimated with Aspen Process Economic Analyzer[™], with the exception of the reactor costs, whose costs were obtained from the NET4CO₂ Collaboratory Laboratory, the entity with expertise in building these type of reactors (Lopes et al., 2008). Capital expenditure (CAPEX), operational expenditure (OPEX) and carbon footprint were assessed for both processes.

Figure 1a shows the Aspen flowsheet of the GTL process via SMR. It includes four major steps: NG steam reforming, CO₂ capture with amine scrubbing, FT reaction and FT syncrude purification. A mixture of natural gas and steam is pre-heated and fed to the SMR NETmix reformer R-01, at 860 °C and 20 bar with kinetic expressions based on the work of Tonkovich et al. (2007), and with a steam-to-carbon molar ratio of 2.3. This gives a global hydrocarbon conversion of 79 % and a final H₂/CO ratio of 4. The CO₂-rich syngas is pressurised to 33 bar (to minimise stage requirements, absorbent flow rate and overall equipment size) and is scrubbed with a 40 wt. % Methyl-diethanolamine (MDEA) solution. 89 % of CO₂ is removed from the syngas with a 98 wt. % purity.

The CO₂-lean syngas is sent to the Pressure Swing Adsorption (PSA) system T-03 with 80 % molar H₂ recovery. A fraction of the H₂ produced is fed back to the process to correct the H₂/CO molar ratio to 2. Although it is not depicted in Figure 1a, the remaining portion of H₂, with 99.5 wt. % purity, is diluted with other streams containing light gases from the purification stage (streams 53 and 59) and it is fed to the furnace for energy recovery purposes. The Fischer-Tropsch reaction occurs in R-02, the NETmix FT reactor,

operating at 250 °C and 30 bar, with kinetic model obtained from Almeida et al. (2013), leading to a CO conversion of 88 %. Finally, the FT syncrude produced is separated from lighter hydrocarbons in flash vessels V-02 and V-04 and from water in decanter V-03.

The GTL via DMR differs mainly in the feedstocks, processing steps (NG dry reforming and Water-Gas Shift instead of steam reforming) and heat integration network. In Figure 1b, a natural gas stream containing 70 % (v:v) CO₂ is pre-heated before entering the NETmix dry reforming reactor R-01, with kinetic expressions based on the work of Richardson and Paripatyadar (1990), operating at 900 °C and 4 bar, and with 98 % conversion. The resulting syngas with a H₂/CO ratio of 0.6 is sent to the Water-Gas Shift (WGS) reactor R-02, operating at 180°C and 3 bar and simulated with kinetic expressions from Amadeo and Laborde (1995). R-02 is responsible for raising the H₂/CO ratio from 0.59 to 2. After WGS, the processing steps are the same of GTL via SMR: syngas is scrubbed with the same solution of MDEA, giving a 96 % recovery and 98 wt. % purity; FT reactor R-03 is simulated at the same operating conditions with 87 % CO conversion; and the final purification step is also performed with two flash vessels and one decanter.

The heat integration of the GTL via SMR process resulted in a network of 11 heat exchangers, while the GTL via DMR has a network of 13 heat exchangers (these values do not include reboiler and condenser from the stripper). Note that to achieve the same production of 500 bpd, DMR uses less NG. This reduction of NG per barrel of product is larger than the increase in the heat of reaction for the DMR case, relative to the SMR. This contributes to significantly less heat to be integrated for the DMR case in comparison to the SMR, 18 MW against 30 MW. In addition, DMR is a more heat-efficient process, integrating 39 % of available process heat, compared to 32 % via SMR. Both processes have a similar energy conversion efficiency of around 65 %.

3. Results & Discussion

Table 1 compares the CAPEX and OPEX associated with both SMR and DMR routes. It is possible to conclude that SMR has the highest capital investment, being the compressors, the 4 packed towers of the PSA system and the heat exchanger network the most expensive equipment. Regarding OPEX, cooling water costs have the highest impact for SMR, while electricity is the main expense for DMR (mainly for C-01 compressor, which departs from 3 bar instead of 20 bar as in SMR). Nevertheless, the CAPEX + OPEX cost estimated for FT syncrude, for both routes, is lower than the average Brent crude oil price in 2019, \$64.36/bbl (MWV, 2020) and is about 1/5 of target cost for synthetic e-fuels (Bosch, 2020).

GTL via SMR leads to a 28 % reduction of carbon emissions when compared to the flaring of the same natural gas, while GTL via DMR increases emissions by 10 %. SMR takes advantage of H₂ combustion in the furnace, avoiding the need of burning extra natural gas to meet heating needs. Both proposed GTL plants are within the limits of 1700 m² topside area and 3130 Mtonne weight estimated for an offshore 5000 bpd modular GTL concept (Kwon et al., 2015). For a 500 bpd capacity, the estimated topside area and weight are one and more than one order of magnitude below the limits, respectively.

A sensitivity analysis on the impact of interest rate, operational lifetime and NG cost on the CAPEX and OPEX for DMR case is presented in Figure 2a-c. The lowest CAPEX is achieved for around 20 years of operation. A 10 % raise on the interest rate increases

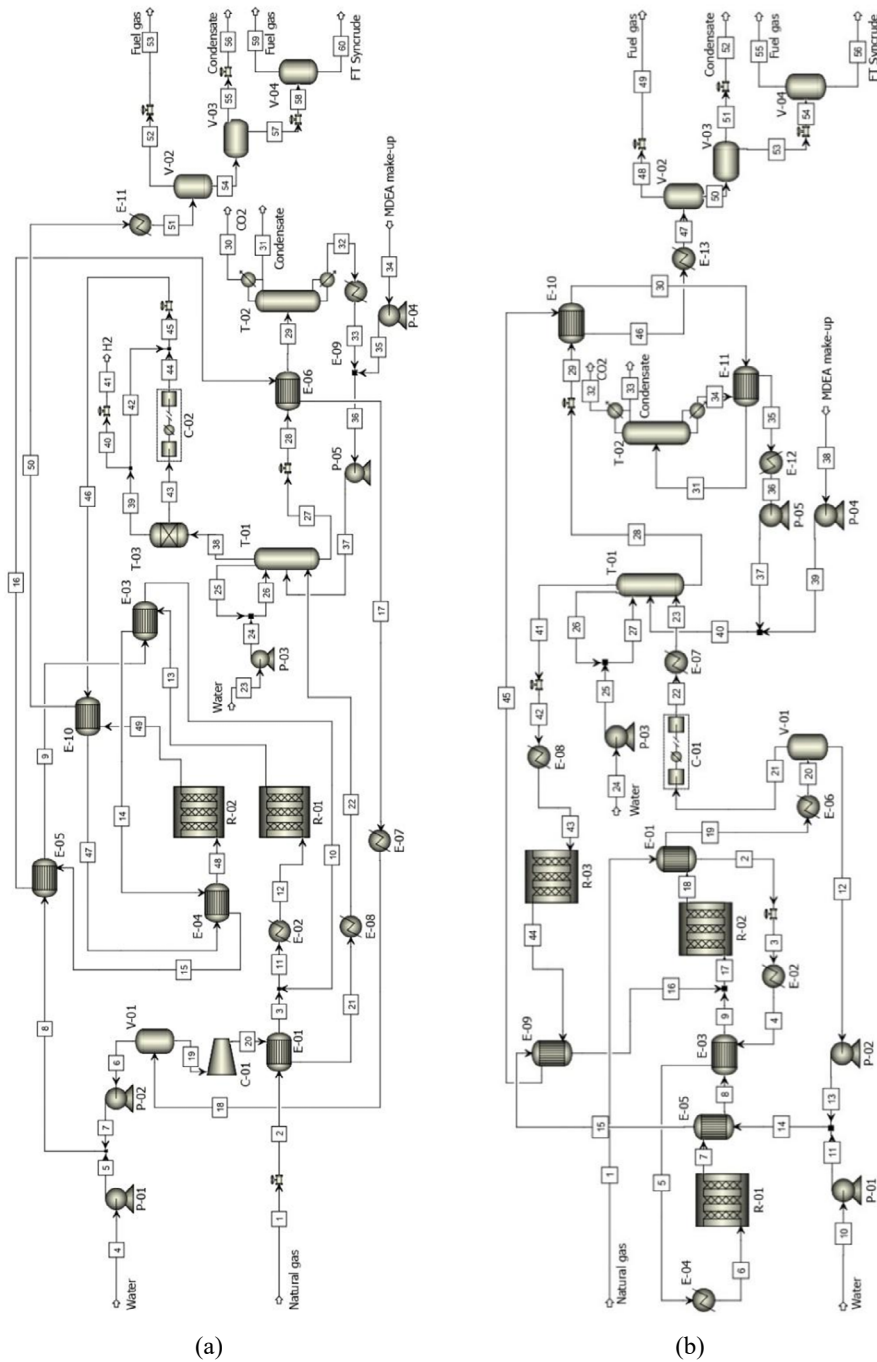


Figure 1. Aspen Plus® flowsheets: (a) GTL via SMR, (b) GTL via DMR.

CAPEX in 550 %, demonstrating the importance of the interest rate on the economically feasibility of a GTL project.

Table 1. Estimated capital and operating costs for small-scale GTL via SMR and DMR, for 20 years operation lifetime, 4.5 % interest rate, and \$0.01/kg of NG.

Reforming	SMR	DMR	Reforming	SMR	DMR
Capital	(10 ⁶ \$)	(10 ⁶ \$)	Operation	(\$/bbl)	(\$/bbl)
Compressors	5.4	3.1	Natural gas	4.4	7.7
Pumps	0.65	0.35	Water	0.062	0.031
Drums	0.51	0.49	Cooling water	8.74	7.9
PSA towers	4.0	-	Electricity	5.54	14
Absorber	0.47	0.45	OPEX	19	30
Stripper	0.97	0.86	Carbon footprint (kgCO ₂ e/bbl)	820	1450
SMR reactor	1.5	-	Total area* (m ²)	178	113
DMR reactor	-	1.0	Weight* (Mtonne)	2.6	0.9
WGS reactor	-	2.5	*preliminary estimates without detailed equipment design or plant layout. Pumps, compressors, boilers, furnaces and piping were not considered.		
FT reactor	1.1	0.87			
Catalysts	0.053	0.050			
Furnace	2.5	1.7			
Heat exchanger network	3.2	2.5			
Cooling tower	0.65	0.65			
Total capital	27	19			
CAPEX (\$/bbl)	20	14			

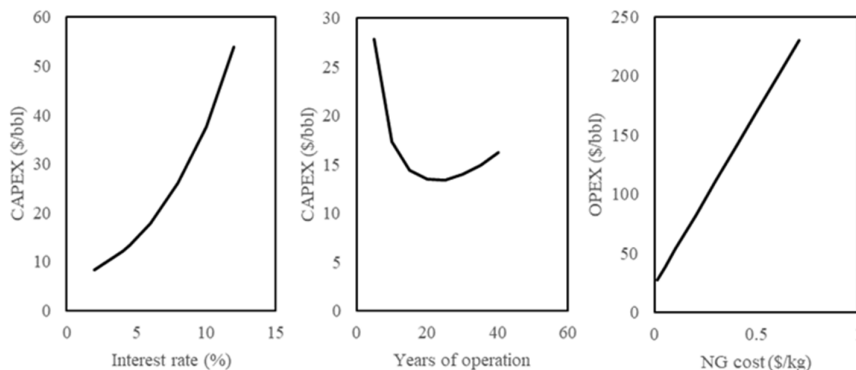


Figure 2. Impact of (a) interest rate on CAPEX, (b) operation lifetime on CAPEX, assuming an interest rate of 4.5 %, and (c) NG cost on OPEX for DMR.

4. Conclusions

Comparing the two small-scale GTL routes, the SMR is the most economically favourable, with \$20/bbl of CAPEX and \$19/bbl of OPEX and enables a 28 % reduction on carbon emissions when compared to flaring, which is advantageous particularly in a scenario of rising CO₂ emission taxes. Nevertheless, the choice between SMR and DMR

must be made carefully, since the decision on the route to follow is highly dependent on NG field size, composition, CO₂ availability, taxes, oil markets, among other factors.

5. Acknowledgements

Financial support from projects POCI-01-0145-FEDER-006984, financed by FCT – Fundação para a Ciência e Tecnologia; NORTE-01-0145-FEDER-000006, financed by Comissão de Coordenação e Desenvolvimento Regional do Norte is gratefully acknowledge; and by base funding with the reference UIDB/00511/2020 of the Laboratory for Process Engineering, Environment, Biotechnology and Energy – LEPABE - funded by national funds through the FCT/MCTES (PIDDAC).

References

- A. L. Y. Tonkovich, B. Yang, S. T. Perry, S. P. Fitzgerald, Y. Wang, 2007, From seconds to milliseconds to microseconds through tailored microchannel reactor design of a steam methane reformer, *Catalysis Today*, 120(1), 21-29.
- Bosch, 2020, Synthetic fuels – The next revolution. <https://www.bosch.com/stories/synthetic-fuels/>. Last accessed 06/11/2020.
- D. A. Wood, C. Nwaoha, B. F. Towler, 2012, Gas-to-liquids (GTL): A review of an industry offering several routes for monetizing natural gas, *Journal of Natural Gas Science and Engineering*, 9, 196-208.
- D. Förtsch, K. Pabst, E. Groß-Hardt, 2015, The product distribution in Fischer–Tropsch synthesis: An extension of the ASF model to describe common deviations, *Chemical Engineering Science*, 138, 333-346.
- GGFR, 2020, Global Gas Flaring Tracker Report. World Bank. USA. <http://pubdocs.worldbank.org/en/503141595343850009/WB-GGFR-Report-July2020.pdf>
- H. Kwon, D. Choi, Y. Moon, S. Ahn, N. Jang, H. Lim, W. Kim, D. Denholm, T. Park, 2015, GTL FPSO & Modular GTL as Potential Solutions for Developing Offshore Oil & Gas Fields, OTC Brasil, Rio de Janeiro, Brazil.
- IEA, 2019, World Energy Outlook 2019, IEA, Paris <https://www.iea.org/reports/world-energy-outlook-2019>
- IEA, 2020, Flaring Emissions, IEA, Paris. <https://www.iea.org/reports/flaring-emissions>
- J. C. B. Lopes, P. E. M. S. C. Laranjeira, M. M. G. Q. Dias, A. A. A. Martins, 2008, Network Mixer and Related Mixing Process, Patent No. EP172643 B1.
- J. T. Richardson, S. A. Paripatyadar, 1990, Carbon dioxide reforming of methane with supported rhodium, *Applied Catalysis*, 61(1), 293-309.
- L. C. Almeida, O. Sanz, D. Merino, G. Arzamendi, L. M. Gandía, M. Montes, 2013, Kinetic analysis and microstructured reactors modeling for the Fischer–Tropsch synthesis over a Co–Re/Al₂O₃ catalyst, *Catalysis Today*, 215, 103-111.
- M. F. Costa, C. M. Fonte, M. M. Dias, J. C. Lopes, 2017, Heat Transfer Performance of NETmix – A Novel Micro-Meso Structured Mixer and Reactor, *AIChE Journal*, 63, 2496-2508.
- MWV, 2020, Average annual Brent crude oil price from 1976 to 2020 (in U.S. dollars per barrel). In Statista. <https://www.statista.com/statistics/262860/uk-brent-crude-oil-price-changes-since-1976/>. Last accessed 06/11/2020.
- N. E. Amadeo, M. A. Laborde, 1995, Hydrogen production from the low-temperature water-gas shift reaction: Kinetics and simulation of the industrial reactor, *International Journal of Hydrogen Energy*, 20(12), 949-956.
- Q. Chen, I. E. Grossman, 2019, Economies of Numbers for A Modular Stranded Gas Processing Network: Modeling And Optimization, *Computer Aided Chemical Engineering*, 47, 257-262.
- Technavio, 2019, Global Gas-to-Liquid (GTL) Market 2019-2023.
- Y. Wang, L. Yao, S. Wang, D. Mao, C. Hu, 2018, Low-temperature catalytic CO₂ dry reforming of methane on Ni-based catalysts: A review, *Fuel Processing Technology*, 169, 199-206.

Automated Process Synthesis Using Reinforcement Learning

Quirin Göttl,^a Dominik Grimm,^{b,c,d} Jakob Burger^{a,*}

^a*Technical University of Munich, Laboratory of Chemical Process Engineering,
Schulgasse 16, 94315 Straubing, Germany*

^b*Technical University of Munich, Campus Straubing for Biotechnology and
Sustainability, Bioinformatics, Schulgasse 22, 94315 Straubing, Germany*

^c*Weihenstephan-Triesdorf University of Applied Sciences, Petersgasse 18, 94315
Straubing, Germany*

^d*Technical University of Munich, Department of Informatics, Boltzmannstr. 3, 85748
Garching, Germany*

burger@tum.de

Abstract

A novel method for automated flowsheet synthesis based on reinforcement learning (RL) is presented. Using the interaction with a process simulator as the learning environment, an agent is trained to solve the task of synthesizing process flowsheets without any heuristics or prior knowledge. The developed RL method models the task as a competitive two-player game that the agent plays against itself during training. The concept is proven to work along an example with a quaternary mixture that is processed using a reactor or distillation units.

Keywords: process synthesis, machine learning, reinforcement learning, automated method.

1. Introduction

Automated process synthesis is a central research field of process systems engineering. It can be grouped into three categories (Siirola, 1996): superstructure optimization, evolutionary modification, and systematic generation. In the latter one, a flowsheet is created sequentially by adding more and more process units from a predefined set. The decision process is usually done with heuristics that are based on prior knowledge. Prominent examples are the well-known expert systems. In the present work, a novel machine-learning (ML)-based method in the category systematic generation is introduced. As ML and artificial intelligence (AI) are rapidly expanding fields, a lot of research focuses on applying these kind of techniques in computer-aided process engineering. In the area of process synthesis, ML is for example applied to create surrogate models for reducing computational cost in simulation (Eason and Cremaschi, 2014). AI offers however more potential (Dimiduk et al., 2018), including recognizing system structures and components and making recommendations for the final design. The type of ML techniques, that could address these kind of problems, seems to be reinforcement learning (RL). The objective of RL is to teach an agent, which could be for example a neural net, to master a given task through repeated interactions with its

environment. This concept is already applied in process engineering, however almost exclusively in process control. Among rare exceptions are Zhou et al. (2017), who employed RL to set experimental conditions for the optimization of chemical reactions. Khan and Lapkin (2020) demonstrated that it is possible to identify promising processing routes in hydrogen production by using a RL approach. Given one of the open streams in the process, the agent was trained to choose a process unit, which is then connected to this stream.

The present work shows a RL based method, which provides the whole flowsheet as input for the agent. The decision is hierarchical: at first, the agent chooses one of the open streams to work with and afterwards a suitable process unit. This approach is an extension of the work of Göttl et al. (2021), in which the agent chooses an open stream and the corresponding process unit at once from a flat, non-hierarchical action space. The agent is trained without the usage of prior knowledge or heuristics. Through repeated process simulation during the training phase of RL, the agent develops artificial process engineering intuition. After briefly explaining the methodology, an example proves the concept to work.

2. Methodology

Framework. The present work is an extension of the method of Göttl et al. (2021), where we refer to for a description of the details. The framework for RL is depicted in Figure 1. The environment is a steady-state process simulator that contains all physical knowledge, i.e. models of thermodynamic properties and a given set of standard process units. The agent does not have any prior knowledge and learns only through interaction with the environment. The agent takes actions on the environment by sequentially adding process units to a flowsheet. The agent observes the stream table that is obtained as result of the process simulation performed in the environment. After the agent decides that the flowsheet synthesis is complete, it obtains a reward from the environment, e.g. some monetary cost function of the process simulation results like the net present value.

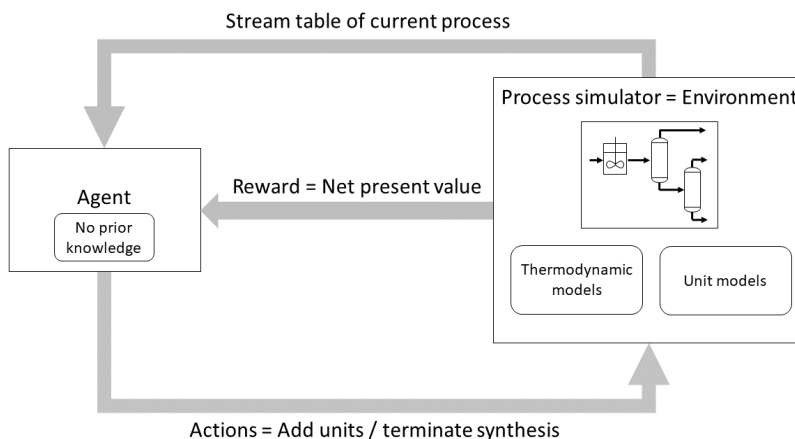


Figure 1: Framework for formulating flowsheet synthesis as reinforcement learning problem.

The presented framework is quite general and is further specialized to yield a proof of concept. A major assumption/limitation in the present work is that the agent takes only concrete actions, e.g. choose an open stream and afterwards place a process unit. The agent does not specify any continuous operation parameters of the apparatus. The reason for this limitation is two-fold: on the one hand, a robust simulation environment can be employed (cf. below). On the other hand, the used RL methods are native in discrete decision spaces and would require computationally expensive extensions for continuous parameters. Despite the limitation to discrete decisions, interesting design problems can be considered as the most important decisions in conceptual design are discrete.

Process Example. As example process in the environment, we chose a chemical model system that is zeotropic and consists of four compounds A (low-boiler), B, C and D (high-boiler). In a reactor with heterogeneous catalyst, the following, kinetically controlled reaction $A + B \rightarrow C + D$ happens. The following process unit are available to the agent:

- Place a distillation column with one of the following sharp split: D1 (A - BCD), D2 (AB - CD), D3 (ABC - D).
- Place a reactor (denoted R) with a fixed residence time. The kinetic is of first order in A and B: $\dot{n}_A^{\text{in}} - \dot{n}_A^{\text{out}} = 5 \text{ kmol/h} \cdot x_A^{\text{out}} x_B^{\text{out}}$
- Place a mixer (denoted M) to combine two streams.

D1, D2, D3 and R are applied to any single open stream, whereas M requires two open streams as input. The net present value is used to evaluate the obtained processes in the case studies. Since the complexity of its calculation is not relevant for the presented methodology and the process models are rather simple, a rather simple scheme is used: the output streams of the process yield a positive value when they are pure C or D. If an output stream consists of a mixture, it is assigned a negative value depending on its composition (the components A and B have a lower value than C and D). Costs for process units are assumed flat and annualized over 10 years. Energy costs are neglected.

SynGameZero method of reinforcement learning. Numerous off-the-shelf RL methods could be tried out on the problem shown in Figure 1. We have identified two major challenges: a) It is not trivial to reward the agent after every action immediately. For example, think of a multi-step separation sequence that only works successfully after a recycle has been closed. After placing the first process unit of the sequence, a constructive reward is hard to determine. b) For the same reasons, the problem is prone to local optima. Powerful exploration schemes are thus indispensable.

In the present work, both challenges are tackled simultaneously using a novel method called *SynGameZero* - Flowsheet Synthesis in a Game Environment with Zero Knowledge (Göttl et al. 2021). The task of creating a profitable flowsheet is modelled as a two-player game. Two competing players create a flowsheet for the same task. The game is turn-based and at each turn, a player can add a process unit in its own flowsheet or terminate the flowsheet synthesis. Both players are able to see their own and the opponents flowsheet. The game ends, when both players have completed their flowsheets. The game winner is the player who yielded a higher net present value. If the net present value is tied, the player who completed the flowsheet first wins the game. The winner obtains the reward $r = 1$, the loser $r = -1$. The agent is trained by playing many games against itself to win the game. Such a game-like setup allows us to use a modified version

of the efficient techniques proposed by Silver et al. (2017) that have been successfully used to master the board game of Go.

Agent setup. A state of the environment consists of the present flowsheet structure and the stream table (from simulation). This information is stored together in a flowsheet matrix. The input for the agent consists of two of those matrices – one for the flowsheet of the player whose turn it is, one for the flowsheet if its opponent. The agent consists of an artificial neural net (ANN) and a tree search. As mentioned before, the agent first chooses an open stream in the flowsheet to work with and afterwards a process unit, which is placed at the chosen stream. Therefore the ANN is also divided into two (connected) parts, the first one being responsible for choosing a stream, the second one being responsible for choosing a process unit. The first part of the ANN takes the flowsheet matrices as input and outputs a vector \mathbf{p}_1 and a scalar v_1 . The vector \mathbf{p}_1 represents a probability distribution (its entries are in the range [0; 1] and sum up to 1) and has one entry for every stream in the flowsheet and an additional entry for an action called T, which terminates the flowsheet synthesis. The vector suggests, which open stream should be taken (for adding a process unit). The scalar v_1 is an estimate of the reward (range [-1; 1]) at the end of the game for the current player. The second part of the ANN is connected to the last hidden layer of the first part and takes as additional input a vector, which indicates which stream was chosen. This part of the ANN outputs a vector \mathbf{p}_2 and a scalar v_2 . The vector \mathbf{p}_2 has one entry for every available process unit and similarly as \mathbf{p}_1 it represents a probability distribution. v_2 is another estimate of the reward. To improve its performance, the agent does not directly select the actions with the highest probabilities in \mathbf{p}_i . Instead, all outputs of the ANN are used to guide a tree search to plan several actions in advance. To avoid extensive computations, the tree search is adaptive in depth and does not use a full enumeration of all actions. Only promising actions are explored, where \mathbf{p}_i and v_i quantify the word promising. The action taken by the agent is ultimately decided based on the statistics of the tree search (basically it is counted how often each possible action was taken during the tree search). The ANN is trained with these statistics to give better suggestions for actions and estimates of the reward in the future. The tree search is described in detail in Göttl et al. (2021).

3. Example case study and results

To prove the concept to work, the same flowsheet problems as approached by Göttl et al. (2021) are considered. At the beginning of each game during training, one of the following feed situations is chosen randomly: in situation 1 two feed streams of the types $[\dot{n}_A; \dot{n}_B; 0; 0]$ and $[0; 0; \dot{n}_C; \dot{n}_D]$ are considered. In situation 2 two feed streams of the types $[\dot{n}_A; 0; \dot{n}_C; 0]$ and $[0; \dot{n}_B; 0; \dot{n}_D]$ are considered. In situation 3 one feed stream of the type $[\dot{n}_A; 0; \dot{n}_C; \dot{n}_D]$ is considered. In situation 4 one feed stream of the type $[0; \dot{n}_B; \dot{n}_C; \dot{n}_D]$ is considered. The non-zero entries are sampled randomly out of the interval [0.2; 1.2] kmol/hr before each game. Both players obtain the same flowsheet problem (identical feed streams) and try to create a process, which yields a higher net present value than the process of the opponent. As explained before, the cost function yields the largest net present value for flowsheets, where C and D leave the process with a high amount and as pure streams.

Figure 2 shows four stages of the agent's development over the progress of the training process specified by the amount of completed training steps. Panel a) in Figure 2 shows an example for a winning flowsheet at the start of the training process. The agent chose to terminate the flowsheet synthesis right at the beginning of the game. This is not

surprising, as the ANN is initialized with random weights and therefore its suggestions do not provide useful solutions to the given problems. After 2000 training steps in panel b), placing a distillation column to separate C and D is the first priority of the agent. This is a useful action since pure C and pure D have a high value. Additionally, the agent already has learnt that placing a reactor at the upper stream is useful, but it is not able to separate C as a pure product yet. Later in panel c) the agent favors to mix the streams first to have a common downstream sequence to save apparatus cost. After 20,000 steps in panel d), the agent has learnt that it is constructive to mix in the CD stream only after the reactor. To do so, it has to prioritize a reactor at the AB stream as first action. Overall, the agent achieved similar results for all of the four feed situations as the agent proposed by Göttl et al. (2021) by employing the described hierarchical action structure.

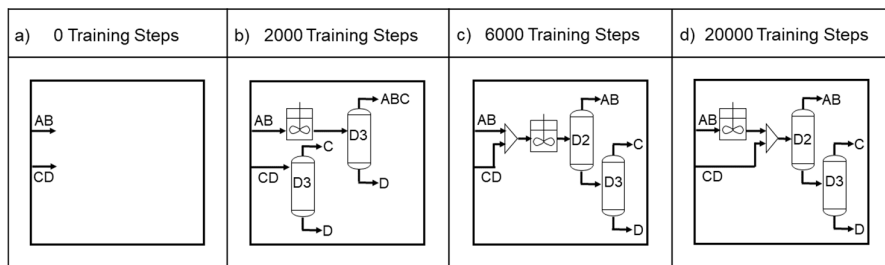


Figure 2: Evolution of the agent over the progress of the training. The shown flowsheets are examples for the winning flowsheet, created by the agent, at different stages during training for situation 1.

4. Conclusions

This contribution gives a further proof-of-concept that reinforcement learning (RL) can be used to train an agent without any prior knowledge to synthesize useful flowsheets for chemical processes in sequential manner. The problem of producing a profitable flowsheet is reformulated into a competitive two-player game that is played by the agent during training. This reformulation improves the efficiency of the training significantly, because it motivates exploration of alternatives. An agent that combines an adjustable artificial neural network (ANN) with an adaptive tree search for forward planning proved efficient and successful. The obtained results in a rather simple process example are promising. The introduction of a hierarchical structure of the actions provides an interesting approach to consider more sophisticated problems in the future (e.g. non-ideal chemical systems, more complex process units). Compared to a flat action space it allows integration of a larger number of process units without the problem of an exploding number of possible actions. Also, additional hierarchy levels could be included, for example to specify continuous parameters (on a discretized range) in the apparatus models.

References

- D. M. Dimiduk, E. A. Holm, S. R. Niezgod, 2018, Perspectives on the Impact of Machine Learning, Deep Learning, and Artificial Intelligence on Materials, Processes, and Structures Engineering, *Integr. Mater. Manuf. Innov.*, Vol. 7, 157–172.

- J. Eason, S. Cremaschi, 2014, Adaptive Sequential Sampling for Surrogate Model Generation with Artificial Neural Networks, *Comput. Chem. Eng.*, Vol. 68, 220–232.
- Q. Göttl, D. G. Grimm, J. Burger, 2021, Automated Synthesis of Steady-State Continuous Processes using Reinforcement Learning, *Front. Chem. Sci. Eng.*, *submitted*, DOI: arXiv:2101.04422.
- A. Khan, A. Lapkin, 2020, Searching for optimal process routes: A reinforcement learning approach, *Comput. Chem. Eng.*, Vol. 141, 107027.
- J. J. Siirola, 1996, Strategic Process Synthesis: Advances in the Hierarchical Approach, *Comput. Chem. Eng.*, Vol. 20, 1637-1643.
- D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. v. d. Driessche, T. Graepel, D. Hassabis, 2017, Mastering the game of Go without human knowledge, *Nature*, Vol. 550, 354-359.
- Z. Zhou, X. Li, R. N. Zare, 2017, Optimizing Chemical Reactions with Deep Reinforcement Learning, *ACS Cent. Sci.*, Vol. 3, 1337–1344.

Operational Analysis of a Proton-Conducting Solid Oxide Electrolysis Cell for Synthetic Fuel Production

Jinjutha Bhichaiphab^a, Dang Saebea^b, Amornchai Arpornwichanop^c, Yaneeporn Patcharavorachot^{a,*}

^a*Department of Chemical Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand*

^b*Research Unit of Developing Technology and Innovation of Alternative Energy for Industries, Department of Chemical Engineering, Faculty of Engineering, Burapha University, Chonburi, 20131, Thailand*

^c*Center of Excellence in Process and Energy Systems Engineering, Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand*

Abstract

Methane and methanol as synthetic fuels can be produced from synthesis gas or syngas. A proton-conducting solid oxide electrolysis cell (H-SOEC) becomes a promising technology that can produce syngas from the co-electrolysis of steam and CO₂. In this work, the synthetic fuel production from syngas produced by a H-SOEC was modelled and simulated through Aspen Plus simulation software. The composition of syngas and synthetic fuel were calculated by using the minimization of Gibbs free energy. Firstly, the steam to CO₂ (S/C) molar ratio in feed was determined to satisfy the suitable stoichiometric number of each fuel. Further, the H-SOEC operating temperature was optimized. The simulation showed that at operation of H-SOEC as 650 °C and 1 atm, the optimal S/C molar ratio for methane and methanol productions is 4.69 and 3.52, respectively. Then, the effect of operation in fuel production was examined. The results indicated that methane flowrate of 0.2 kmol/h can be provided when reactor operates at 250 °C and 3 atm. For methanol production, 0.13 kmol/h of methanol can be generated at reactor operation as 250 °C and 80 atm. In addition, it was found that the methane production does not release CO and CO₂ to nature. Therefore, it can be concluded that the integrated system of H-SOEC and methane production is more attractive feature.

Keywords: Solid oxide electrolysis cell, Co-electrolysis, Proton-conducting, Synthetic fuel, Synthesis gas.

1. Introduction

A solid oxide electrolysis cell (SOEC) is a reverse mode of solid oxide fuel cell that can produce H₂ from separation of steam through electricity. In case of a proton-conducting SOEC (H-SOEC), it can produce more purifier H₂ than the conventional SOEC since no dilution of steam in the product stream (Lei et al., 2020), as seen in Figure 1a. The co-electrolysis of steam and CO₂ through a H-SOEC (Figure 1b) can provide the gas mixture consisting primarily of H₂ and CO which is referred to synthesis gas or syngas. Compared with the conventional method of syngas production (e.g. steam reforming and gasification), the co-electrolysis of H-SOEC is more attractive feature. If the power used

is generated from renewable resource, the syngas production through H-SOEC becomes clean technology. This method not only eliminates the consumption of fossil fuels which have price crisis but also can reduce the CO₂ emission due to the recycling or reusing CO₂ during production.

Although syngas is perceived as energy carriers that tend to replace the fossil fuels in the future, the energy storage in form of other chemicals or fuels is great interest due to ease of use and handling (Léonard et al., 2016). In this work, the productions of methane and methanol have been focused since both fuels are the most promising fuel and feedstock that can be further used in many applications (Bernardi et al., 2019). Interestingly, if this production uses the syngas produced from H-SOEC as an environmentally friendly method, it may be claimed that the obtained synthetic fuels are clean. Moreover, it can call this operation as Power-to-Chemical.

Since, the production of each synthetic fuel requires different stoichiometric number, the operations of H-SOEC to provide syngas for each synthetic fuel may be different. Motivated by this, this works aims to identify the favourable operating conditions of H-SOEC that satisfying each synthetic fuel. This work is performed through the simulation by using Aspen Plus simulator. Furthermore, the influence of reactor operation for synthetic fuel production is studied.

2. Modelling of Synthetic Fuel Production

In this work, the model of the synthetic fuel production from syngas produced by a H-SOEC is developed and designed through Aspen Plus simulation software, as shown in Figure 2. The equation of state as IDEAL is used in the simulation. The simulation of syngas production is firstly performed to identify the appropriate H-SOEC operation for each fuel. Next, the obtained syngas is treated before sending to the reactor for methane or methanol production. For syngas production, steam (STEAM *stream*) and carbon dioxide (CO₂ *stream*) is fed to the anode (ANODE) and the cathode (CATHODE), respectively. The anode is represented by *RStoic* reactor model where the electrochemical reaction of steam is carried out. Therefore, the steam utilization or steam conversion (X), as expressed in Eq. (1), should be specified to determine the current density (i) used for syngas production.

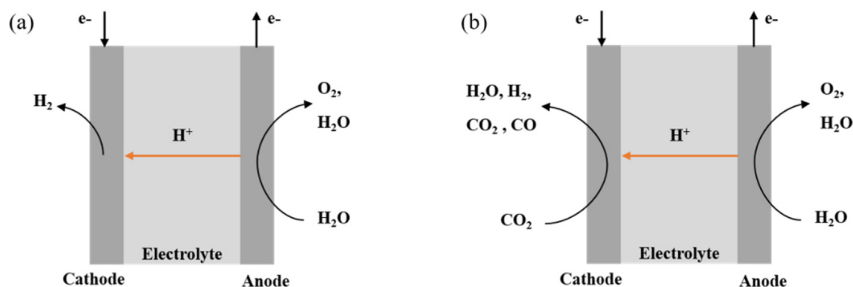


Figure 1 Schematic of H-SOEC for (a) hydrogen production and (b) syngas production.

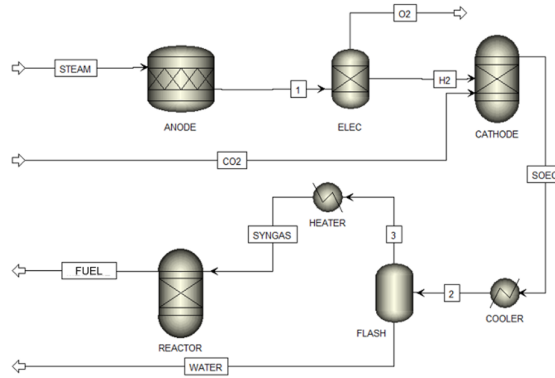


Figure 2 Flowsheet of synthetic fuel production from syngas produced by H-SOEC.

$$X = \frac{iAN_{\text{cell}}}{2F\dot{n}_{\text{H}_2\text{O},\text{in}}} \quad (1)$$

where A is cell area (m^2), N_{cell} is cell number and $\dot{n}_{\text{H}_2\text{O},\text{in}}$ is inlet steam flowrate (kmol/h).

Before the gas product from the cathode is used to produce synthetic fuel, the gas mixture is cooled down in the cooler (COOLER) and sent to the flash drum (FLASH) to remove water. Then, the dry syngas (SYNGAS *stream*) is fed into the reactor (REACTOR) to generate the desired synthetic fuel (FUEL *stream*). The chemical reaction that can be occurred in the reactor depends on type of synthetic fuel. For the methane production, Sabatier and methanation reactions can be carried out in the reactor (Stempien et al, 2015). In case of methanol, the reactions occurred in the reactor may be hydrogenation of CO and CO₂ (Kiss et al., 2016). Since the possible chemical reactions are based on equilibrium, *RGibbs* reactor model is used to calculate the fuel composition.

3. Model Validation

To ensure that the electrochemical reactions can be accurately predicted the performance of H-SOEC, the comparison between the simulation results and experimental data extracted from the literature is performed and the results are summarized in Table 1. The experiment of Ruiz-Trejo and Irvine (2012) used platinum electrodes and BCZY electrolyte that have the thickness as 100 and 200 μm , respectively. 500 cells of H-SOEC is fabricated with specific area of 0.04 m^2 . The anode side is fed with 95% steam and 5% hydrogen. As seen in Table 1, the deviation of results obtained from simulation and experiment is about 10% which is acceptable.

Table 1 The comparison of cell potential obtained from simulation and experimental data (Ruiz-Trejo and Irvine, 2012) at current density of 400 A/m^2

H-SOEC temperature ($^{\circ}\text{C}$)	Cell potential (V)		Deviation (%)
	Experiment	This simulation	
550	1.58	1.39	9.9585
600	1.42	1.30	8.1945
650	1.31	1.22	6.4210

4. Results and Discussion

4.1. H-SOEC operation

In this part, the operation of H-SOEC is determined to provide syngas that is suitable for synthetic fuel production. Firstly, the effect of steam to CO₂ (S/C) molar ratio in feed stream on stoichiometric number (S) as expressed in Eq. (2) is studied.

$$S = \frac{\dot{n}_{\text{H}_2} - \dot{n}_{\text{CO}_2}}{\dot{n}_{\text{CO}} - \dot{n}_{\text{CO}_2}} \quad (2)$$

where \dot{n}_i is the molar flow rate of species i (kmol/h).

For the investigation on the effect of S/C molar ratio, there are two methods performed: (1) steam molar flow rate is increased while CO₂ molar flow rate is constant and (2) CO₂ molar flow rate is decreased with a constant value of steam flow rate. Therefore, the effect of these variations on cell potential is considered as shown in Figure 3a. When the H-SOEC is operated at 650 °C and 1 atm with different S/C molar ratio (0.5, 1, 2, 3, 4 and 5), it can be seen that the variation of steam flow rate has more impact than that of CO₂ flow rate. Since increasing steam flow rate leads to an increase in current density, according to Eq. (1), and this causes significant increases in all voltage losses that include activation, ohmic and concentration losses and thus, the cell potential used for syngas production is higher. Conversely, the variation of CO₂ flow rate does not affect to cell potential. Therefore, to avoid the variation of cell potential or keep current density as constant during the study on the effect of S/C molar ratio, the reduction of CO₂ flow rate is considered.

Figure 3b presents the stoichiometric number as a function of S/C molar ratio. The simulation results indicate that increasing S/C molar ratio or decreasing CO₂ molar flow rate can increase stoichiometric number. However, it can be observed that the stoichiometric number becomes a negative value when S/C molar ratio is below 1. This indicates that the amount of steam fed into H-SOEC is not enough to generate H₂ for reacting with CO₂. From the simulation results, when the suitable stoichiometric number for methane and methanol production as 3 and 2 are specified, the optimal S/C molar ratio can be provided as 4.69 and 3.52 respectively. As expected, since the suitable stoichiometric number of methane production is more than and thus, the required S/C molar ratio is also higher.

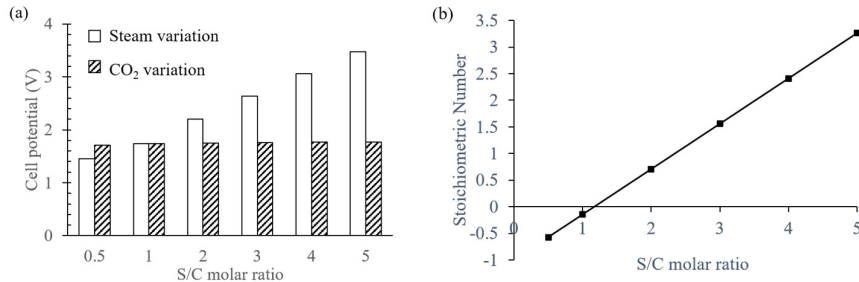


Figure 3 The effect of S/C molar ratio on (a) cell potential and (b) stoichiometric number.

Next, the influence of H-SOEC temperature on stoichiometric number and CO production is investigated, as shown in Figure 4. In this study, the H-SOEC temperature is varied as 550, 600 and 650 K while H-SOEC pressure and S/C molar ratio are constant value as 1 atm and 4.62, respectively. Figure 4a shows that the increment of H-SOEC temperature can reduce the cell potential used for syngas production. This is because higher temperature operation leads to decreases in reversible cell potential, activation loss and ohmic loss. In addition, it is found that increasing H-SOEC temperature can shift the reverse water gas-shift reaction toward and thus, more CO can be produced as seen in Figure 4b. As a result, H-SOEC should be operated at high temperature to provide higher CO while the power consumption is lower.

Under the operation of H-SOEC at 650 °C and 1 atm, for methane production that the S/C molar ratio of 4.69 is specified, it can provide the gas product consisting of 58.3% H₂, 9.2% CO, 7.5% CO₂ and 25% H₂O. For methanol production that setting the S/C molar ratio as 3.52, the product stream is composed of 52.8% H₂, 11% CO, 10.2% CO₂ and 26% H₂O.

4.2. Reactor operation for methane and methanol production

In this part, the operating condition of each reactor on each fuel production is considered to achieve the highest amount of required synthetic fuel. Considering reactor temperature for both fuel productions, it is found that increasing operating temperature cause a reduction of fuel production. This is because the reactions occurred in both reactors are exothermic reaction in which is favourable to low temperature. Therefore, higher temperature operation becomes a negative effect on product composition. The simulation results indicate that the reactor for methane production should be operated at 250 °C and 3 atm in which the methane molar flow rate of 0.2 kmol/h can be produced. While, the operation at 250 °C and 80 atm is suitable for methanol production. Under these operating condition, the methanol of 0.13 kmol/h can be generated. Besides the operation condition of reactor, the effluent of each reactor is considered. The results reveal that the methane production does not release CO and CO₂ to nature. Unlike methane production, there are some amount of CO and CO₂ release during methanol production. Then, comparison between methane and methanol production is considered. It is found that although higher S/C molar ratio is required for methane production, lower operating condition of reactor and no carbon emission are main reasons to make this process is more attractive.

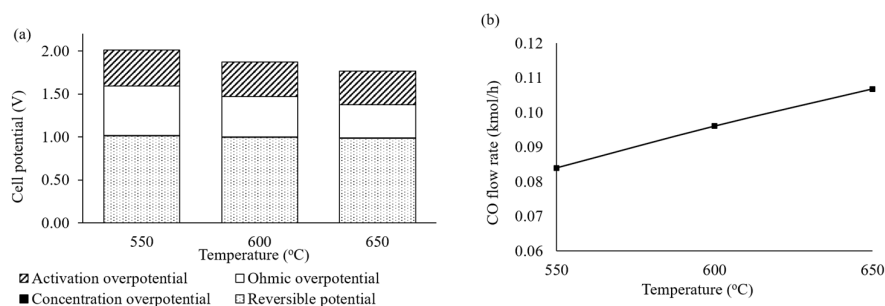


Figure 4 The effect of H-SOEC temperature on (a) cell potential and (b) CO flow rate.

5. Conclusions

This work focused on the methane and methanol production from syngas derived from H-SOEC. The simulation through Aspen Plus simulator was performed to determine the optimal operating condition of H-SOEC and each reactor. Since the methane and methanol production requires different stoichiometric number, the operation of H-SOEC to provide syngas for each fuel is different. One of factors affected to stoichiometric number is S/C molar ratio in feed. The simulation results showed that increasing S/C molar ratio can increase stoichiometric number. For methane production, stoichiometric number of 3 is required which is corresponded to S/C molar ratio of 4.69. In case of methanol production, the H-SOEC must be operated with the S/C molar ratio of 3.52 to satisfy stoichiometric number as 2. Further, the impact of H-SOEC operating temperature on cell potential and gas product is determined. The simulation result revealed that the cell potential used for syngas production is lower when the H-SOEC temperature is higher. In addition, increasing operating temperature leads to an increase in CO amount in gas product. Finally, the operating temperature of reactor used to produce synthetic fuel was examined. For the methane production, the reactor that Sabatier and methanation reactions carried out should be operated at 250 °C and 3 atm. While, the reactor for methanol production must be operated at 250 °C and 80 atm to achieve the highest amount of methanol. Under the operation of each reactor, it was found that the production of methane (0.2 kmol/h) is higher than that of methanol (0.13 kmol/h). Interestingly, it was found that there are no CO and CO₂ emission to environment while some amount of CO and CO₂ can be found in methanol production. Therefore, it can be concluded that the integrated system of H-SOEC and methane production is more attractive feature in terms of production rate and carbon emission.

References

- A. Bernardi, J.E.A. Graciano, B. Chachuat, 2019, Production of chemicals from syngas: an environmental-economic model-based investigation *Computer Aided Chemical Engineering*, 46, 367-372.
- A.A. Kiss, J.J. Pragt, H.J. Vos, G. Bargeman, M.T. de Groot, 2016, Enhanced Process for Methanol Production by CO₂ Hydrogenation, *Computer Aided Chemical Engineering*, 38, 985-990.
- E. Ruiz-Trejo, J.T.S. Irvine, 2012, Ceramic Proton Conducting Membranes for the Electrochemical Production of Syngas, *Solid State Ionics*, 216, 36-40.
- G. Léonard, D. Giulini, D. Villarreal-Singer, 2016, Design and evaluation of a high-density energy storage Route with CO₂ re-use, water electrolysis and methanol synthesis, *Computer Aided Chemical Engineering*, 38, 1797-1802.
- J. Stempien, M. Ni, Q. Sun, S. Chan, 2015, Production of sustainable methane from renewable energy and captured carbon dioxide with the use of solid oxide Electrolyzer: A Thermodynamic assessment, *Energy*, 82, 714-721.
- L. Lei, J. Zhang, R. Guan, J. Liu, F. Chen, Z. Tao, 2020, Energy storage and hydrogen production by proton conducting solid oxide electrolysis cells with a novel heterogeneous design, *Energy Conversion and Management*, 218, 113044.
- L. Namwong, S. Authayanun, D. Seabea, Y. Patcharavorachot, A. Arpornwichanop, 2016, Modeling and optimization of proton-conducting solid oxide electrolysis cell: Conversion of CO₂ into value-added products, *Journal of Power Sources*, 331, 515-526.

Octane Optimization with a Combined Machine Learning and Optimization Approach

Ahmet Can Serfidan,^{a,b} Metin Türkay,^b

^a*TUPRAS, Address, Petrol Cd. No:25 D:No:25, Turkey*

^b*Koc University, Rumelifeneri, Sariyer Rumeli Feneri Yolu*

Abstract

Refinery operations are always sensitive to optimization, and due to the increasingly adverse effects of COVID-19 on energy sectors, its importance had increased significantly. This thesis aims to analyze the reactor temperature that yields a higher RON (octane measurement) value in isomerate product using all available information in the isomerate production network. The main explanatory variables that can affect the RON value can be divided into three categories: feed impurities, isomerization reactor operations, deizohexanizer column operations. Isomerate feed network is quite complex and fed by different crude distillation units and cracker units. Various reactions occur in the isomerization reactors, and depending on the feed content, the reaction mechanism changes. This thesis applies machine learning algorithms to build a model that can capture the relationship between RON and reactor temperature with the other explanatory variables.

We implemented a number of machine learning algorithms to assess their performance on the problem, specifically Linear Regression, Decision Tree, Random Forest, XGBoost, Support Vector Regression, and KNN. Comparing with the linear regression, we achieved 0.82 decreases in the mean absolute error. The mean absolute error of the XGBoost model is 0.08 RON. We find a temperature value with the selected model that yields a higher RON number by trying different temperature values while keeping the same values for the other variables. If we used the suggested temperature by our model, we predict that we could obtain a 0.2 RON increase in the validation zone resulting in an annual profit increase of 528 000 USD Dollar.

Keywords: isomerization unit, octane optimization.

1. Introduction

A refinery produces a wide variety of products. Some of them are high-value light products such as gasoline, jet fuel, and diesel, and others like residual oil can be considered by-products. Gasoline is one of the most common transportation fuels for cars and trucks. The essential quality specification of gasoline is the octane number. It is a measure of the knock resistance and defines gasoline's behavior in the engine during combustion. Refineries adjust the units' operation so that octane in the gasoline meets the specifications. The most common type of octane rating worldwide is Research Octane Number (RON).

One of the most significant contributors to the gasoline pool is the isomerization unit. Isomerization is the principal reaction that takes place in isomerization reactors. It is converting normal paraffin to iso-paraffins to increase the octane number of light straight naphtha (feed). The reaction occurs in a hydrogen atmosphere, over a fixed bed of catalyst, and at operating conditions that promote isomerization and minimize side reactions such as hydrocracking.

In this work, our objective is to find the operating conditions that maximize the octane number in the isomerization unit in Tupras. The primary operating process variable is the reactor inlet temperatures. An upper limit exists for the number of iso-paraffins in the reactor product at any given outlet temperature (equilibrium imposed by thermodynamics). As reactor temperature is raised to increase isomerization, the equilibrium composition will be approached more closely. At very high temperatures, the concentration of iso-paraffins in the product will decrease because of the downward shift in the equilibrium curve, even though high temperatures give a higher reaction rate. Therefore, this work's central question is, "what the reactor inlet temperature should be?" However, of course, the outlet reactor temperature is not the only factor affecting the octane number. There are many other factors such as reactor pressure, deisohexanizer's column operations, undesirable hydrocarbons in the feed inlet, weather conditions, and many more. It is possible to measure some of these variables, such as pressure, continuously. Nevertheless, online analyzers are needed to measure others like feed content. However, online analyzers that measure feed components are absent in many refineries due to high initial and maintenance costs.

This work will select all isomerate product network as our system boundary. Moreover, we will use the most powerful and successful machine learning algorithms to capture linear and nonlinear relationships. For this purpose, we will use Random Forest, Extreme Gradient Boosting, k- Nearest Neighbors, Support Vector Regression, and of course, Linear Regression for the baseline.

2. Method

To maximize the octane number in the isomerization unit, we first have to study which variables affect this number. The most important variable is reactor temperatures. In the normal operation of Isomerization Unit, having once set the pressure, feed rate, and hydrogen flows, the main operating variable is reactor inlet temperatures. There is an upper limit for the amount of iso-paraffins in the reactor product at any given outlet temperature. This is the equilibrium imposed by thermodynamics, and it can be reached only after infinite time. As reactor temperature is raised to increase isomerization, the equilibrium composition will be reached that point faster. However, at excessively high temperatures, the concentration of iso-paraffins in the product will actually decrease because of the downward shift in the equilibrium curve. So, the use of temperatures higher than necessary to reach equilibrium yields nothing other than to increase the amount of hydrocracking. Therefore, the relationship between the temperature and octane number is not linear. As we increase the temperature, considering the other variables fixed, the octane number will rise first, and then it will be decreased. In the modeling part, we take the minus square of temperature to represent this relationship.

Another variable is liquid hourly space velocity (LHSV). It is defined as the volumetric hourly flow of the reactor charge divided by the catalyst volume in the reactors. Increasing LHSV might lead to lower product isomer ratios (low octane number)

The third variable we study is pressure. Higher pressure yields to increase the rate of isomerization reactions.

The fourth variable we pick for this model is C7+ Hydrocarbons. C7 paraffins hydrocrack to C3 and C4, low in octane number. And those which do not hydrocrack will be isomerized to a mixture, again, a lower octane number. Therefore, we do not want a high C7+ content in the isomerization reactors. Unfortunately, unlike the other variables, we cannot measure the C7+ content in the feed, so we have to inferential for this situation.

All of the above variables are related to the reactor or reactor feed. However, there is also Deisohexanizer (DIH) column operation, which can greatly affect the octane number. This column aims to recover product isohexane and pentanes from the stabilized reactor products. The basic aim is to remove all pentanes and 2-2 DMB overhead while minimizing the overhead 3-methylpentane content. Unlike C7+ content in the reactor feed, we have analyzers in here.

Heat input, reflux rate, pressure, all of the column temperatures are good candidates for the inputs for octane prediction.

Two variables set the octane of the main DIH top:

1. The pentane composition of the overhead. The total amount of pentane in this stream is set by the feedstock composition. Namely, the iso-pentane content is set by the C5 isomerization in the reactors. The higher the concentration of pentanes in the feedstock, the lower the product octane.
2. The amount of 2-methylpentane and 3-methylpentane in the overhead. The higher their concentration in the overhead product, the lower the octane.

Especially for the second item, let's imagine that we eliminate all of the methylpentanes. However, this will require more reflux and more recycling to the reactor system. Namely, it is costly. There is an optimization between the product octane vs. utility consumption.

Our target variable is octane number, and it is measured in two different approaches:

- (i) We have an analyzer, and it can sample the data and provide the result roughly in one hour. We have six-year data, which approximately 17500 data (the analyzer's uptime is not 100%). However, generally, the analyzer cannot provide the real result's ground truth.
- (ii) We also have samples taken from the lab, which are approximately around 500 data. It is very low compared to the analyzer, but the reliability is higher than the analyzer.

But before the octane number, the other target variable is feed content in the reactor. In reality, it is quite hard to predict the octane number without predicting the feed content. We do not have any analyzer; however, we know which units directly send to reactor feed, and we know their operating conditions.

Four separate units can feed the isomerization reactors, each with different operating conditions, specifications, and constraints. But all of these four units generate light straight-run naphtha. To infer the feed content, we consider them as possible candidates of those upstream units in the column operations: input flow rates, input temperature, all column temperatures, all column pressures, tray temperature controller, steam feed flows, reflux flow rates, bottom flow rate, distillate flow rate. Also, we generated these features:

reflux overfeed, temperature differences between bottom and top of the columns, distillate over bottom.

I used BarutoSharp to select the most important features. It is a wrapper feature selection method that uses both the Boruta feature selection algorithm with Shapley values [1]. It successfully reduces the column number from 30 to 5-10.

After that, I use other powerful machine learning algorithms, specifically Random Forest, AdaBoost, XGBoost, KNN, SVR, and Linear Regression, to predict the feed content reactor [2].

After obtaining the feed content, I follow a very similar approach to predict the octane number in the isomerate product. The only difference is that since I have not many columns, we do not need to reduce the number of columns, so we skipped the BorutaSharp part. Predicting the octane number is not enough to optimize it. We have to shift the process to optimum conditions. For this aim, the main process variable we need to determine is reactor temperature. Based on the other process variables conditions, we predict the octane number and predict again under the following scenarios: (i) We increased all reactor by λ and kept all the other variables simultaneously. (ii) We increased all reactor temperatures by $\lambda/2$ and kept all the other variables at the same level. (iii) We decreased all reactor temperatures by λ and kept all the other variables at the same level. (iv) We decreased all reactor temperatures by $\lambda/2$ and kept all the other variables at the same level, where λ is the max absolute deviation

3. Results and Discussion

After constructing isomerate product data, we implemented the most powerful machine learning algorithms to predict RON. In the following table, we can see the result of each prediction. We used train data MAE and test data MAE as the evaluate algorithm. Next, we showed the Actual vs. Predicted values only for the test data. Finally, we randomly select validation data to see how that algorithm could predict RON throughout a week.

Table 1. Model Results

Algorithm Name	Hyperparameters	Train MAE	Test MAE	Actual vs. Predicted	Validation Range: 83-88
Decision Tree	Max Depth: 10, best 11 Min Samples Leaf: 2, best 10 Min Samples Leaf: 10, best 10 Max Features: 11 Complexity Parameter: 0.001, best 0.001	0.08	0.11		
Random Forest	Max Depth: 10, best 9 Min Samples Leaf: 2, best 10 Min Samples Leaf: 10, best 10 Max Features: 11 Complexity Parameter: 0.001, best 0.001 Number of Trees: 270, best 270	0.08	0.09		
XGBoost	Max Depth: 10, best 7 Min Samples Leaf: 2, best 2 Min Samples Leaf: 10, best 10 Number of Trees: 10000, best 10000 Learning Rate: 0.1, best 0.1	0.07	0.08		
Linear Regression	All	0.75	0.9		
Support Vector Regression	Grid of Param C: 1000, best 10 Nonlinearity: 0.001, best 0.001 Kernel: RBF, best 0.001	0.05	0.06		
KNN	Nearest Neighbor: 1, best 1 K: 1, best 1	0.06	0.09		

Based on the results, almost all algorithms performed well. The only complex algorithm is linear regression. It has considerable high training and testing MAE. Furthermore, when we check validation data, it has a massive variance in prediction. Other than that, all of the algorithms can be applied in real-life plants. Support Vector Regression provides the minimum error, so let us take a closer look at it.

First, I changed the validation data to a different week, and it looks like it did a great job following the trend again.

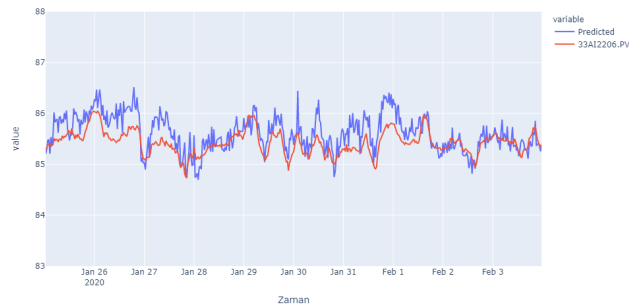


Figure 1: XGBoost RON Model, Validation Data.

Now, let us remember our problem definition. We aim to find the reactor temperature at given conditions to maximize the RON value. So we prepared data composed of the main factors that can affect the RON: the feed conditions, reactor conditions, DIH conditions. To answer our question, we need to keep other conditions (DIH and feed) stable and find the reactor's optimum temperature profile. If the selected algorithm was linear regression, this was a trivial job. By substituting each variable's current values, we could find the optimal temperature except the reactor temperatures. However, we cannot back-calculate the exact temperature value using algorithms like SVR. So, we follow a heuristic approach. Again, we keep the other than reactor conditions simultaneously, but we introduced four different datasets derived from the original (the real one). Before that, we defined the maximum allowable temperature change with a single set change. For example, if the temperature set is 160, we defined the minimum and maximum deviation from 160 to prevent the catalyst from deactivation. Let us call this max absolute deviation λ . We substitute the values into the algorithm for each dataset and predict the RON. Furthermore, for some fixed intervals, or if the RON value is decreasing continuously, we can re-iterate the procedure. This procedure is also suitable for refinery operating conditions since we cannot change the temperatures frequently, and we cannot change it more than the maximum allowable limit.

However, when we implement this on SVR and KNN, we see no change from using these datasets. The main reasons for this situation; for SVR, there are no reactor temperature data in support vectors. So no matter what temperature we choose, we will observe the same RON value. This situation is observed for the KNN algorithm too. One explanation for this situation is that there are explanatory variables with correlation and more critical than that between inputs. For example, there will be more C6+ in the DIH column when there is a low temperature. Thereby, the bottom flow rate and bottom temperature of the DIH column will be increased. The bottom flow rate and bottom temperature support vectors, but the reactor temperature are not. One may suggest removing DIH operation variables. However, in that case, we observe a significant decrease in accuracy, as we will see in the later section.

Besides KNN and SVR reactor temperature insensitivity, all tree algorithms significantly respond to reactor temperature change. Since the Xgboost algorithm leads to minimum test MAE, we selected Xgboost as the best algorithm for our work.

For the same validation dataset, when we implement our heuristic approach, we can find the temperature value, which will give a higher value of RON content.



Figure 2: Temperature Sensivity of the Model.

The blue line is the predicted RON, and the red line is the actual RON value. If we changed the temperature by λ amount, we could see a significant increase during some intervals. On average, we could increase the octane content by around 0.2 just by changing the reactor temperature. However, increasing temperature incur some cost too. So when there is no significant octane change by increasing or decreasing reactor temp, it would be best to minimize the energy loss that is given to the reactor.

4. Conclusions

This study aims to find the reactor temperature that yields a higher RON value in isomerate product. For this purpose, we used all information in the isomerate product network to extract information that can affect RON value. Namely, we prepared data related to the impurities in the feed content, reactor operations, and DIH operations. Then we back-calculated a better temperature value for any give condition by keeping other variables the same and changing only the reactor temperature. After calculating the X-factor, we also gathered data related to the DIH and reactor operations. We employed Linear Regression, Decision Tree, Random Forest, XGBoost, Support Vector Regression, and KNN to predict RON and make a robust model. Although the SVR yielded the least mean absolute error, since it was not sensitive to temperature change, we select the second-best model, XGboost, to model between explanatory variables and the output. The model provided with the 0.08 mean absolute error in the test data, and the model can follow the trend quite well. By feature importance, we see that the most impactful features are created by domain knowledge, and without them, the algorithm performs poor results. With the model obtained, we find a temperature value that yields a higher RON number by substituting different temperature values while keeping the same values for the other values. With such change, we could obtain a 0.2 RON increase in the validation zone. If we expand it to it all year, the overall profit is 528 000 USD Dollar.

5. References

- [1] Kursu, Miron B., and Witold R. Rudnicki. "Feature selection with the Boruta package." *J Stat Softw* 36.11 (2010): 1-13.
- [2] Fortuna, Luigi, Salvatore Graziani, and Maria Gabriella Xibilia. "Comparison of soft-sensor design methods for industrial plants using small data sets." *IEEE Transactions on Instrumentation and Measurement* 58.8 (2009): 2444-2451.

A Platform of Machine Learning-Based Next-Generation Property Estimation Methods for CAMD

Abdulelah S. Alshehri ^{a,b}, Anjan K. Tula ^c, Lei Zhang ^d, Rafiqul Gani ^e, Fengqi You ^a

^a*Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, United States*

^b*Department of Chemical Engineering, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia*

^c*College of Control Science and Engineering, Zhejiang University, Hangzhou-310027, China*

^d*Institute of Chemical Process Systems Engineering, School of Chemical Engineering, Dalian University of Technology, 116024 Dalian, China*

^e*PSE for SPEED, Skyttemosen 6, Allerød, DK-3450, Denmark*

Corresponding authors: F. You: fengqi.you@cornell.edu; R. Gani: rgani2018@gmail.com

Abstract

Physicochemical property estimation methods serve as the basis for the design of molecules that enhance the functionality and efficiency of products and processes. The need to provide reliable pure component properties through quantum chemistry computations and/or experimental measurements is a major bottleneck to the goal of faster and cheaper to market the desired products. Hence, the development of approximate but qualitatively accurate models is vital to progress in the field of Computer-Aided Molecular Design (CAMD), among others. With group contribution (GC) as a dominant molecular representation, semi-empirical methods have been the most popular class in generating approximate property models for CAMD owing to their low computational cost and direct incorporation into optimization models. Recent advances in machine learning have stimulated widespread interest and progress towards closing the gap between semi-empirical and quantum chemistry methods. Herein, we use machine learning and data analysis methods to address the shortcomings of the current GC-based models by synthesizing the next-generation property estimation models and tools for the fast and accurate estimation of 20 physicochemical properties central to CAMD.

Keywords: CAMD, Group-contribution, Machine-learning, Property prediction models.

1. Introduction

In engineering research efforts aiming to improve the functionality and performance of products and processes, the accurate estimation of pure component properties is of paramount importance (Marrero & Gani, 2001). The primary strategy for obtaining

values of these properties and forming theories on their dependency on the structure is experimental observations, involving time-consuming experiments to create a knowledge-based model describing the underlying phenomena. Yet, it rapidly becomes prohibitive to acquire experimental values in literature when searching for ideal sets of properties because of the combinatorial explosion issue of the various potential configurations of atoms (Gani, 2004). Computational techniques, such as Density Functional Theory, have been applied to guide experimental efforts, but, given the high computation time that grows nonlinearly and sharply with molecular size, such methods still form a major bottleneck (Heinen et al., 2020). Consequently, more accurate models connecting molecular representations to properties are needed to build on the approximate underlying functions that describe the needed chemical properties.

Machine learning has recently emerged to play a transformative role in several fields by uncovering complex interactions in high-dimensional domains and systematically searching for a suitable function in the hypothetical space (Jordan & Mitchell, 2015). In the domain of molecular systems, machine learning has been considered for: pure component properties prediction (Yalamanchi et al., 2019), the generation of lead molecules while efficiently exploring the chemical space (Jin et al., 2018), and the creation of competitive novel quantitative structure-property relationship methods (Goh et al., 2017). In the domain of molecular design, the notable method is the GC family of methods (Marrero & Gani, 2001) that have been widely used in CAMD due to their invertibility, easy incorporation within mathematical models, and highly accurate estimates at an affordable computational cost (Alshehri et al., 2020). Properties of chemical structures are described in this class of approaches as functions of the number of molecular fragment occurrences called functional groups (Gani, 2019).

This paper combines the most recent previous works (Hukkerikar et al., 2012) on GC based property estimation approaches with the next-generation property models using machine learning and data analytics techniques for applications in CAMD. Such methods are intended to resolve the shortcomings of current GC models and reduce the accuracy gap between experimentation and estimates. Several models are devised for the following 20 pure component properties given in order of dataset size: octanol-water partition coefficient ($\log P$), normal melting point (T_m), normal boiling point (T_b), oral rat toxicity (LD_{50}), aqueous solubility ($\log Ws$), acid dissociation constant (pK_a), standard enthalpy of formation (H_f), Hildebrandt Solubility Parameter ($HSolp$), liquid molar volume (L_{mv}), critical temperature (T_c), critical pressure (P_c), critical volume (V_c), standard Gibbs energy of formation (G_f), normal enthalpy of fusion (H_{fus}), fathead minnow 96-h LC50 ($LC_{50(FM)}$), photochemical oxidation potential (PCO), bioconcentration factor (BCF), Auto Ignition Temperature (AiT), enthalpy of vaporization at 298K (H_v), and permissible exposure limit ($OSHA-TWA$). The criteria for inclusion of a candidate property for modelling is a minimum of 400 data-points. The dataset has more than 22,000 molecules, but the size of the experimental values of properties ranges from 12,193 to 422 data points for $\log P$ and $OSHA-TWA$, respectively. In the following section, a brief description of the methods involved in synthesizing the GC-simple and GC-ML models and tools is provided. The use of the new models in CAMD is highlighted through two case studies involving crystallization solvent design and surfactant design in the third section, followed by conclusions in the final section.

2. Methods

2.1. Group-Contribution Methods: GC-Simple

The MG (Marrero & Gani, 2001) property prediction model for pure components has the form in Eq. (1) for the vector of the number of occurrences for each group n and the vector of the regressed group contributions, c .

$$f(y) = \sum_{i=1}^{NF} c_i^F n_i^F + \sum_{j=1}^{NS} c_j^S n_j^S + \sum_{k=1}^{NT} c_k^T n_k^T \quad (1)$$

The regression method applied is similar to previous works (Hukkerikar et al., 2012), where the regressed group contributions of different orders are optimized sequentially or simultaneously. These model parameters (ie., regressed group contributions) of the model are obtained at the optimum value of the squared error function with a tuneable parameter w_k Eq. (2), reducing the difference between the true and predicted property values.

$$S(\underline{c}^F, \underline{c}^S, \underline{c}^T) = \min w_k \sum_{j=1}^N (y_j - \hat{y}_j)^2 \quad (2)$$

2.2. Group Contribution Methods: GC-ML

For a vector \underline{n} of group counts and vector \underline{y} of property values for vector \underline{m} data-points (molecules), Machine Learning methods seek to learn distributions over possible functions that fit the given data best according to a specified objective function. The ML-model has the form, given by Eq. (3).

$$f(y) = f_i(\underline{c}_{ML}, \underline{n}, \underline{\theta}) \quad (3)$$

where, \underline{c}_{ML} is a matrix of regressed parameters of size (NK, NP) with NP being the total number of parameters and NK being the size or order of the learning function; \underline{n} is the vector of group counts and $\underline{\theta}$ is a set of additional parameters introduced for the ML model. In principle, Eq. (2) or a probability function is minimized through a regression method. For the uncertainty estimate, the covariance function as well as the Jacobian matrix are needed (section 2.3 for more details).

2.3. Regression of the model parameters

GC-simple models have been obtained by fitting the models defined by Eq. (1) and minimizing the function defined by Eq. (2) using the Support Vector Regression (SVR), method, while, the machine learning-based models have been obtained using the GP regression method, minimizing a probability function. The data-set for each property is divided into 20 parts with 19 of them used for the learning (regression) step and the remaining set used for test (validation) step. The uncertainty estimate of a predicted property value is calculated from the covariance functions and the measured data of the molecules used in the regression step, as in Eq. (4).

$$Uncertainty_{estimate} = \pm [diag J(P^*) COV(P^*) f(P)]^{1/2} .t(v, (\alpha/2)) \quad (4)$$

Where the Jacobian Matrix $J(P^*)$ represents the local sensitivity, of the property model f to variations of the regressed parameter values ($P^* = \underline{c}_{ML}$); $COV(P^*)$ is the covariance matrix of the regressed parameters; is the degrees of freedom (total number of parameters minus the number of not regressed parameters); $t(v, (\alpha/2))$ is the t -distribution corresponding to v and $(\alpha/2)$ percentile. The additional parameters from

the regression step for the GC-ML-model are ℓ , the length scale parameter; and σ the signal variance that regulate the horizontal and vertical variances of the covariance function.

3. Results and Discussion

3.1. Performance analysis of the developed property models

The performances of the regressed models are evaluated using the fraction of experimental data points of properties that lie within 1%, 5%, and 10% of absolute relative error. Under the GC-ML models, the average of the model performance for predicting property values within 1% relative error is boosted to 85% across all properties as visualized in Fig. 1, compared to 32% under simultaneous regression and GC-simple models (not shown in Fig. 1). More details results, including the regressed parameters can be found in Alshehri et al., (2021).

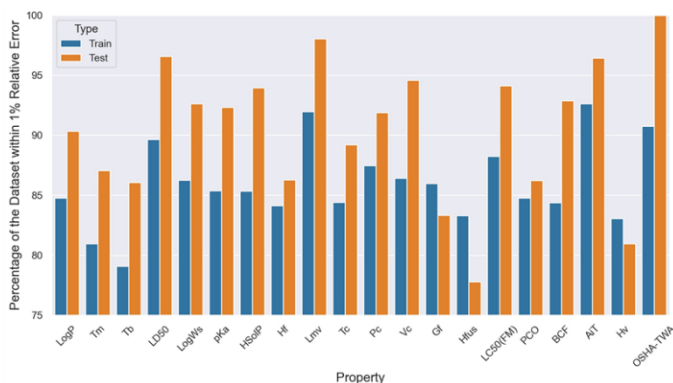


Figure 1: The training and testing performances for the developed property models for molecules predicted under the 1% relative error threshold.

3.2. Data Analysis

The generation of synthetic data can be broadly categorized into two approaches, process-driven and data-driven (Goncalves et al., 2020). In data-driven approaches, generative models based on probability distribution and function approximation methods are used to generate synthetic data. The primary task of generated datasets is to aid in the model development process by complementing the experimental data in tuning the model parameters. Table 1 gives the number of generated data points for each of the properties in the available measured datasets. For these results, the properties having less than 1000 experimental data are selected. The property datasets with the new data points are used to test the developed GC-simple and GC-ML models with acceptable model performance.

Table 1: New datasets generated using data analysis techniques

Chemical property	H_{fus} [kJ/mol]	P_c [bar]	T_c [K]	V_c [cc/mol]	V_c [cc/mol]	G_r [kJ/mol]
Experimental Data	749	774	776	773	425	756
Generated Data	62	100	135	104	55	60

3.3. Application of developed models in hybrid CAMD

Crystallization solvent case study: This case study is taken from (Chai et al., 2020) and the hybrid CAMD method (Liu et al., 2019) has been used, which is a

decomposition-based method to solve mixed-integer nonlinear programming models. Here, in the inner-loop a CAMD problem is solved with the GC-simple models and in the outer-loop the GC-ML models are used to check additional properties. Also, only the uses of pure component property models are highlighted here. 2-Mercaptobenzothiazole (MBT) is an important vulcanization accelerator used in the rubber industry. However, its purification through a solvent-based method replacing the usual acid-base method needs to be investigated. First, a set of building blocks are selected for the inner-loop: CH₃, CH₂, CH, C, OH, CHCOO, aC-CH₃, aCH, aC-Cl. The molecular structure and property constraints used in the inner-loop are given in Table 2. The candidate solvents and their properties from the inner-loop of CAMD are listed in Table 3. For the set of feasible compounds, in the outer-loop additional properties not included in the inner-loop are checked with the GC-ML models and their values are listed in Table 4.

Table 2: List of molecular structure and property constraints.

Name	Lower	Upper
Total Group Number	3	8
Repeat Group Number	0	7
Functional Group Number	1	6
Molecular Weight M_w [g/mol]	80	200
Normal Melting Point T_m [K]	173	310
Normal Boiling Point T_b [K]	373	600
Flash Point F_p [K]	273	393
Hildebrand Solubility Parameter at 298 K δ [MPa ^{1/2}]	18	21
Fathead Minnow 96-hr LC_{50} FM [mol/L]	0	4.8

Table 3: List of feasible solvent candidates for crude MBT crystallization. The values in parentheses are measured experimental values.

Feasible Molecule	Mw [g/mol]	T_m [K]	T_b [K]	F [K]	HSol P [MPa ^{1/2}]	$LC_{50(FM)}$ [mol/L]
1	116.20	231 (239)	451 (449)	340	21.10	3.41(3.51)
2	181.45	290 (290)	486 (486)	379	20.31	4.63(4.80)
3	147.00	255 (256)	449 (453)	341	20.30	4.22(3.41)
4	116.20	244 (243)	437 (432)	334	20.42	3.23
5	88.15	206 (194)	406 (411)	317	21.41	2.83(2.21)
6	102.17	221 (228)	429 (430)	328	21.26	3.12(2.94)

7	102. 17	220 (170)	400 (394)	3 0 9	19.4 7	2.13
8	102. 17	236 (223)	414 (409)	3 2 3	20.5 7	2.94

Table 4: List of predicted properties with GC-ML models of the candidate solvents for crude MBT crystallization.

Feasible Molecule	SMILES	Lo gP	LD50 [mol/kg]	p Ka	B C F
1	OCCCCC C	2.6 2	2.01	9.1 1	2. 23
2	Clc1ccc(Cl) c(Cl)c1	4.0 2	2.00	3.1 5	3. 05
3	Clc1cccc1 Cl	3.4 3	1.97	3.6 5	2. 50
4	CCCCC(O) C	2.2 6	2.11	10. 27	1. 49
5	OCCCCC	1.5 1	1.97	13. 43	1. 30
6	OCCCCC	2.0 3	2.03	10. 35	1. 78
7	CCCC(O)(C)C	1.2 1	2.23	10. 39	0. 49
8	CCCC(O) C	1.7 0	2.11	12. 00	0. 94

Surfactant design: The surfactant design problem (Liu et al., 2019) has also been solved with the new property models (see Table 5). In this problem, T_m , T_b , and $LC_{50(FM)}$ are used in the inner-loop and LogP and OSHA-TWA are used in the outer-loop.

Table 5: List of inner- and outer-loop properties for a candidate surfactant.

SMILES	LC ₅₀ ^(FM) [mol/L]	T _m [K]	T _b [K]	Lo gP	OSHA-TWA
OCCOCCOCCOC(CCCCCC) OCCOCCOCCO	3.16	3 4 9	7 2 3	2.7 5	7.99

The details of the best candidate molecule are given in Table 5 together with the inner and outer-loop properties. More details on this problem can be obtained from the corresponding author.

4. Conclusions

In this work, we introduced the next-generation pure component property models and data analytics tools for the estimation of pure component properties to improve the performance and the application range of CAMD problems. Primarily, the Gaussian Process method is adopted for building nonlinear property models that handle complex molecular structure-property relationships while quantifying uncertainties in the predicted values. Compared to currently-used methods, the new generation of methods yields significantly improved performance. The uses of the simple GC-models and

ML-based models have been highlighted for previously solved CAMD problems in a hybrid scheme. More work is needed to find the optimal computational framework for accommodating a set of models with different complexities and data from different sources in a more practical solution route to CAMD problems with likely applications in drug and other complex molecular design problems. Note that although the uses of the pure property models have been highlighted in the CAMD problems, a lot of other properties (functional and mixture properties) are also needed for their complete solution.

References

- A. S. Alshehri, R. Gani, & F. You, 2020, *Computers & Chemical Engineering*, 141, 107005.
- A. S. Alshehri, A. K. Tula, L. Wang, L. Zhang, R. Gani, F. You, *Next Generation Pure Component Property Estimation Methods*, *AIChE J*, 2021 (under review).
- S. Chai, Q. Liu, X. Liang, Y. Guo, S. Zhang, C. Xu, J. Du, Z. Yuan, L. Zhang, R. Gani, 2020, *Computers & Chemical Engineering*, 135, 106764.
- R. Gani, 2004, *Computers & Chemical Engineering*, 28, 12, 2441-2457.
- R. Gani, 2019, *Current Opinion in Chemical Engineering*, 23, 184-196.
- G. B. Goh, C. Siegel, A. Vishnu, N. O. Hodas, & N. Baker, 2017, arXiv.
- A. Goncalves, P. Ray, B. Soper, J. Stevens, L. Coyle, & A. P. Sales, 2020, *BMC Medical Research Methodology*, 20, 108, 1-40.
- S. Heinen, M. Schwilk, G. F. von Rudorff, & O. A. von Lilienfeld, 2020, *Machine Learning: Science and Technology*, 1, 2, 025002.
- A. S. Hukkerikar, B. Sarup, A. Ten Kate, J. Abildskov, G. Sin, & R. Gani, 2012, *Fluid Phase Equilibria*, 321, 25-43.
- W. Jin, R. Barzilay, & T. Jaakkola, 2018, *International Conference on Machine Learning*, 80, 2323-2332.
- M. I. Jordan, & T. M. Mitchell, 2015, *Science*, 349,6245, 255-260
- Q. Liu, L. Zhang, L. Liu, J. Du, A. K Tula, M. Eden, & R. Gani, 2019, *Computers & Chemical Engineering*, 124, 285-301.
- J. Marrero, & R. Gani, 2001, *Fluid Phase Equilibria*, 183-184, 183-208.
- C. E. Rasmussen, 2003, *Gaussian processes in machine learning*. Summer School on Machine Learning, Springer, 63-71.
- K. K. Yalamanchi, V. C. O. van Oudenhoven, F. Tutino, M. Monge-Palacios, A. Alshehri, X. Gao, & S. M. Sarathy, 2019, *J Phys Chem A*, 123, 8305-8313.

Machine Learning Analysis of the Feed Water Parameters Affecting Reverse Osmosis Membrane Operation

Çağla Odabaşı,^{a,*} Pelin Döloğlu,^a Fatih Gülmez,^b Gizem Kuşoğlu,^a Ömer Çağlar^b

^a TUPRAS Petroleum Refineries Co., R&D Center, Izmit, 41780, Turkey

^b TUPRAS Petroleum Refineries Co., Process Department, Izmit, 41780, Turkey

Abstract

Reverse osmosis (RO) membranes play a key role in wastewater treatment units as they are used to remove salts and other pollutants effectively. RO membrane performance is affected by many different factors such as feed characteristics and operational parameters during operation. The aim of this study is to analyse the influence of feed characteristics of municipal wastewater (conductivity, oxidation reduction potential (ORP), total suspended solids (TSS), turbidity and chemical oxygen demand (COD)) and operational parameters (feed pressure, flow rate and temperature) on RO membrane performance in a municipal wastewater recovery plant using machine learning (ML) techniques. XGBoost, random forest, artificial neural networks (ANNs) and multiple linear regression (MLR) were employed to predict three RO membrane performance indicators (pressure difference across membranes, salt passage and permeate flow rate). The methods that can predict salt passage, permeate flow rate and pressure difference among membranes with the highest accuracy were found as ANNs, random forest and MLR, respectively. Considering the developed models, temperature was found to be the variable affecting all three RO performance parameters. Salt passage was found to be highly affected by feed water conductivity and feed flow rate was determined to be the most influential parameter for the permeate flow rate and pressure difference.

Keywords: machine learning, artificial neural networks, random forest, wastewater treatment, reverse osmosis membranes.

1. Introduction

In wastewater treatment plants, monitoring RO membrane performance is very important. Besides, the factors influencing RO membrane performance are worth investigation to understand the process dynamics for effective operation. Data-driven models are becoming increasingly popular as they can model complex process dynamics without the need for physical models (Del Rio-Chanona et al., 2019). Machine learning techniques are used to develop models which can learn from data. ANNs are common machine learning algorithms used for prediction (Odabaşı et al., 2014; Robert, 2003). They have been employed successfully in various studies in the field of membrane performance (Cabrera et al., 2017; Farahbakhsh et al., 2019; Jawad et al., 2020; Roehl et al., 2018). Random forest and eXtreme Gradient Boosting (XGBoost) methods are other most common tree-based ensemble machine learning algorithms used for supervised machine

learning problems. The random forest method is a popular algorithm because it can handle outliers and noise effectively (Larose, 2005). The XGBoost method (Chen and Guestrin, 2016) is a relatively new algorithm that introduces regularization parameters to prevent overfitting and it was developed for increasing the performance and speed.

In this work, the influence of the feed characteristics of municipal wastewater and operational parameters on RO membrane performance were studied using different machine learning methods (random forest, XGBoost, ANNs and MLR) in a municipal wastewater recovery plant. The best performed models were employed to determine the influence of the input parameters on RO membrane performance indicators (salt passage, pressure difference and permeate flow rate). In this study, we aimed to extract knowledge from a large-scale treatment system and compare the performances of different ML models on different performance indicators. We believe that investigating the influence of the input parameters for RO membrane operation should be given more attention to understand the system dynamics better.

2. Methodology

The two years of historical data of one RO train from the municipal wastewater recovery plant of TUPRAS Izmit Refinery was used. Both operational and laboratory data were used and the data was collected on a daily basis. The database was divided into two sections as training data and test data. Model building, parameter optimisation and validation of the models were performed using training data whereas the accuracy of the model was tested on test data which was not included in model building process. Most widely used ML models such as XGBoost, random forest, ANNs and MLR were used to build models and predict three RO membrane performance indicators (pressure difference between feed and retentate, salt passage and permeate flow rate). The performances of different models were compared using root mean square error (RMSE) as a performance metric. 10 fold cross validation (cv) method, which is the most commonly used validation method, was employed for parameter optimisation of the models. Then, the models which resulted in minimum RMSE on test data prediction were determined as best performed models for predicting RO membrane performance indicators. The influence of input variables were investigated using the best performed models and the details of the input significance analyses are explained in the Results section.

RStudio 1.2.133526 was used for computational analyses. `xgboost` (Chen et al., 2020), `randomForest` (Liaw and Wiener, 2002) and `neuralnet` (Fritsch et al., 2019) packages of RStudio were employed for XGBoost, random forest and ANN methods, respectively. For XGBoost model, the maximum number of rounds (trees to build) was optimised. Learning rate, maximum depth of the trees and early stopping rounds were fixed to 0.3, 6 and 10, respectively. For the random forest model, the number of trees to grow was optimised. For ANN analysis, Z-score normalization was performed before building the model. Then, number of hidden layers and neurons were optimised. Logistic activation function was selected as an activation function and threshold for the partial derivatives of the error function as stopping criteria was set to 0.1.

3. Results

In order to predict RO membrane performance indicators (salt passage, permeate flow rate and pressure difference), different ML methods were implemented and compared. The best performed models that resulted in minimum RMSE on test data prediction are

used to detect the most influential parameters on three RO membrane performance indicators. Feed pressure, flow rate, conductivity, TSS, ORP, COD, turbidity and temperature were used as input parameters. The RMSE values of different models for each RO membrane performance indicator were presented in Table 1.

Table 1. RMSE values of validation and test data prediction using different methods

Methods	Salt passage		Permeate flow rate		ΔP	
	10-fold-cv	Test	10-fold-cv	Test	10-fold-cv	Test
	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
XGBoost	0.91	0.77	3.90	6.64	0.33	0.64
Random forest	0.71	0.65	7.32	10.51	0.35	0.63
ANNs	0.76	0.91	4.36	5.12	0.34	0.58
MLR	0.76	0.85	3.77	4.35	0.39	0.66

For salt passage prediction, a random forest model performed superior to other methods (Table 1). The RMSE of test data prediction was found to be 0.65. The actual and predicted salt passage values of test data are presented in Figure 1a. In random forest method, mean decrease in Gini (IncNode Purity) is used as a measure to analyse the variable importance (Kuhn et al., 2008). This measure represents the performance of each split considering the input parameters using Gini index. If the value of mean decrease in Gini of an input variable is high, this variable has a higher variable importance. Hence, considering the random forest model, it was found that the feed conductivity and temperature had more effect on salt passage than other parameters (Figure 1b).

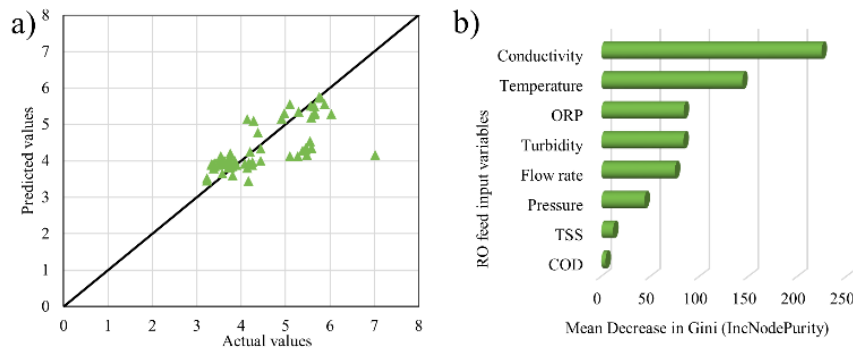


Figure 1. Analysis of salt passage a) actual versus predicted values of test data (number of trees = 100), b) variable importance analysis

Bartels et al. also showed that when the feed water salinity increases, salt passage also increases (Bartels et al., 2005). Hence, the conductivity of the feed water has a significant effect on salt passage and ion removal efficiency. Secondly, a change in the feed water temperature of RO membrane trains causes the RO membrane pore diameters to change. If the feed water temperature increases, pore diameters expand and more ions can pass through the RO membranes. Hence, salt passage increases (Al-Bastaki and Al-Qahtani, 1994; Jin et al., 2009).

For the permeate flow rate prediction, the accuracy of the MLR model was found to be the highest and the RMSE value of test data prediction was found to be 4.35. The actual and predicted permeate flow rate values are shown in Figure 2a. In order to determine the relative importance of the input variables according to MLR model, the significant input variables of which p-values were found to be smaller than 0.05, were considered. In Figure 2b, the absolute values of the coefficients of significant input variables are given. Feed flow rate was found to be the major factor affecting permeate flow rate. Considering the strong linear correlation between feed and permeate flow rate (0.96), the success of the MLR model and the importance of the feed flow rate were expected. Temperature appeared to be the second influential factor for permeate flow rate. Regarding the expansion of the RO membrane pore diameters due to high temperature operation, the feed flow can pass through the pores more easily and the permeate flow rate increase. (Boulahfa et al., 2019).

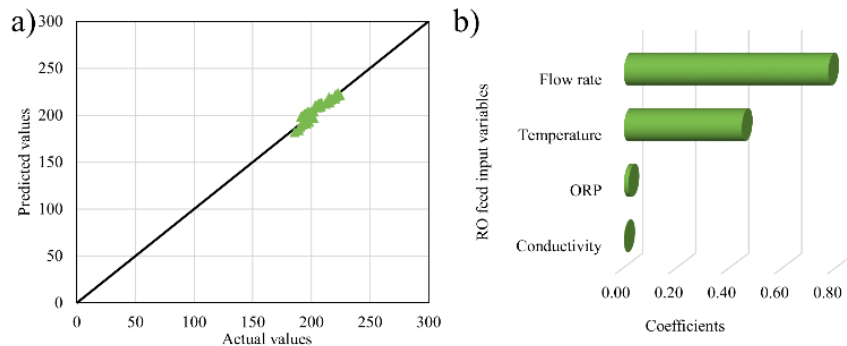


Figure 2. Analysis of permeate flow rate, a) actual versus predicted values of test data, b) variable importance analysis

Lastly, ANN model performed better than other methods for predicting the pressure difference across RO membranes with a minimum RMSE of 0.58 on test data prediction (Table 1). The actual and predicted pressure difference values are given in Figure 3a. In order to determine the variable importance according to ANN model, Olden's method (Olden et al., 2004) was employed. Olden's method is based on connection weight approach that uses raw input-hidden and hidden-output connection weights in the neural network. According to the Olden's method, feed flow rate was found to affect the pressure difference significantly and this situation is expected considering the system hydraulics. Secondly, temperature was found to affect pressure difference. An increase in temperature causes a pore expansion of the RO membranes, hence, the pressure difference between the stages of the train is expected to be lower.

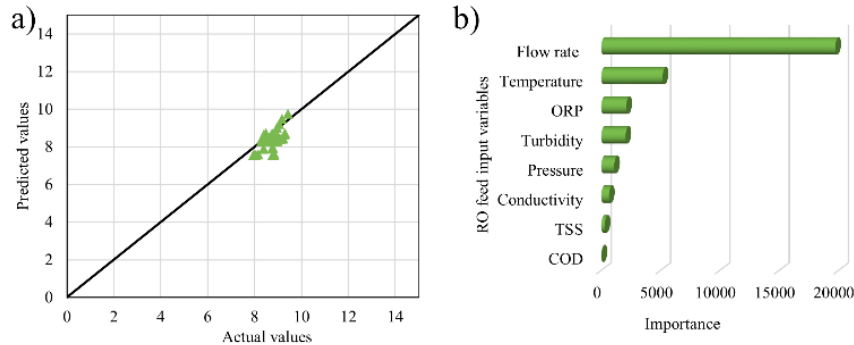


Figure 3. Analysis of pressure difference a) actual versus predicted values of test data (neural network with two hidden layer (2-12 neurons)), b) variable importance analysis.

4. Conclusions

In this study, the factors affecting RO membrane performance were investigated using different ML tools (random forest, XGBoost, ANNs and MLR). Salt passage, permeate flow rate and pressure difference were used as RO membrane performance monitoring indicators whereas the feed characteristics of municipal wastewater (such as conductivity, ORP, TSS, turbidity and COD) and RO membrane feed operational parameters (such as feed pressure, flow rate and temperature) were considered as input parameters. Among four different methods, ANNs and random forest methods performed better for predicting the pressure difference and salt passage, respectively. For permeate flow rate prediction, the MLR method was found to be more successful than other methods due to strong linear correlation between feed and permeate flow rate. According to the variable importance analyses considering the best performed models, temperature was found to be significant for all RO membrane performance indicators whereas conductivity and feed flow rate were found to be major factors affecting salt passage and permeate flow rate, respectively.

References

- Al-Bastaki, N.M., Al-Qahtani, H.I., 1994. Assessment of thermal effects on the reverse osmosis of salt/water solutions by using a spiral wound polyamide membrane. *Desalination* 99, 159–168. [https://doi.org/https://doi.org/10.1016/0011-9164\(94\)00125-1](https://doi.org/https://doi.org/10.1016/0011-9164(94)00125-1)
- Bartels, C., Franks, R., Rybar, S., Schierach, M., Wilf, M., 2005. The effect of feed ionic strength on salt passage through reverse osmosis membranes. *Desalination* 184, 185–195. <https://doi.org/https://doi.org/10.1016/j.desal.2005.04.032>
- Boulaifa, H., Belhamidi, S., Elhannouni, F., Taky, M., El Fadil, A., Elmidaoui, A., 2019. Demineralization of brackish surface water by reverse osmosis: The first experience in Morocco. *J. Environ. Chem. Eng.* 7, 102937. <https://doi.org/https://doi.org/10.1016/j.jece.2019.102937>
- Cabrera, P., Carta, J.A., González, J., Melián, G., 2017. Artificial neural networks applied to manage the variable operation of a simple seawater reverse osmosis plant. *Desalination* 416, 140–156. <https://doi.org/https://doi.org/10.1016/j.desal.2017.04.032>
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*. Association for Computing Machinery, New York, NY, USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>

- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., 2020. xgboost: Extreme Gradient Boosting.
- Del Rio-Chanona, E.A., Ahmed, N.R., Wagner, J., Lu, Y., Zhang, D., Jing, K., 2019. Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses. *Biotechnol. Bioeng.* 116, 2971–2982. <https://doi.org/10.1002/bit.27131>
- Farahbakhsh, J., Delnavaz, M., Vatanpour, V., 2019. Simulation and characterization of novel reverse osmosis membrane prepared by blending polypyrrole coated multiwalled carbon nanotubes for brackish water desalination and antifouling properties using artificial neural networks. *J. Memb. Sci.* 581, 123–138. <https://doi.org/https://doi.org/10.1016/j.memsci.2019.03.050>
- Fritsch, S., Guenther, F., Wright, M.N., 2019. neuralnet: Training of Neural Networks.
- Jawad, J., Hawari, A.H., Zaidi, S., 2020. Modeling of forward osmosis process using artificial neural networks (ANN) to predict the permeate flux. *Desalination* 484, 114427. <https://doi.org/https://doi.org/10.1016/j.desal.2020.114427>
- Jin, X., Jawor, A., Kim, S., Hoek, E.M. V., 2009. Effects of feed water temperature on separation performance and organic fouling of brackish water RO membranes. *Desalination* 239, 346–359. <https://doi.org/https://doi.org/10.1016/j.desal.2008.03.026>
- Kuhn, S., Egert, B., Neumann, S., Steinbeck, C., 2008. Building blocks for automated elucidation of metabolites: machine learning methods for NMR prediction. *BMC Bioinformatics* 9, 400. <https://doi.org/10.1186/1471-2105-9-400>
- Larose, D.T., 2005. *Discovering Knowledge in Data: an Introduction to Data Mining*. John & Wiley & Sons, Inc., Publication.
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2.
- Odabaşı, Ç., Günay, M.E., Yildirim, R., 2014. Knowledge extraction for water gas shift reaction over noble metal catalysts from publications in the literature between 2002 and 2012. *Int. J. Hydrogen Energy* 39. <https://doi.org/10.1016/j.ijhydene.2014.01.160>
- Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecol. Modell.* 178, 389–397. <https://doi.org/https://doi.org/10.1016/j.ecolmodel.2004.03.013>
- Robert, C., 2003. *The Essence of Neural Networks*.
- Roehl, E.A., Ladner, D.A., Daamen, R.C., Cook, J.B., Safarik, J., Phipps, D.W., Xie, P., 2018. Modeling fouling in a large RO system with artificial neural networks. *J. Memb. Sci.* 552, 95–106. <https://doi.org/https://doi.org/10.1016/j.memsci.2018.01.064>

Novel technologies for low-quality natural gas purification

Giorgia De Guido,^{a,*} Matteo Gilardi,^a Laura A. Pellegrini^a

*^aGASP – Group on Advanced Separation Processes & GAS Processing, Dipartimento di Chimica, Materiali e Ingegneria Chimica “G. Natta”, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
giorgia.deguido@polimi.it*

Abstract

Novel low-temperature technologies have been recently studied for CO₂ removal from low-quality natural gas or biogas. The aim of this work is to compare the performances of a low-temperature distillation process with those of other two processes, respectively based on physical absorption and a combination of both, depending on the CO₂ content varied between 10 and 70 mol%. Each process has been simulated, with attention to energy-saving issues, using Aspen HYSYS[®] V9.0 and Aspen Plus[®] V9.0. This study points out that novel technologies based on low-temperature distillation are the best option for CO₂ separation from low-quality natural gas or biogas and that, by combining a bulk CO₂ removal by low-temperature distillation with a finishing step by physical absorption, it is possible to reduce the energy consumptions of the separation achieved by physical absorption only. These outcomes suggest the importance of further investigating novel low-temperature and hybrid CO₂ removal technologies.

Keywords: natural gas purification, physical solvents, low-temperature distillation, hybrid technologies

1. Introduction

The challenge operators in the gas industry have to face in some countries is the economically viable separation of CO₂ from fields where its content is even higher than 70 mol% (Pellegrini et al., 2019). It is well known that, when the acid gases make up an appreciable fraction of the total gas stream, the cost of removing them by heat regenerable solvents may be out of proportion to the value of the treated gas. To overcome this, physical solvents can be used, which are regenerated by pressure reduction. It has been also proved (Langé et al., 2015) that conventional CO₂ separation methods based on chemical absorption are not energetically convenient respect to novel low-temperature (low-T) technologies when the CO₂ content in the feed gas exceeds 8-9 mol%. The recent interest in this type of processes is proved by the intense research on measurement and calculation of CO₂ frost points in natural gas and other mixtures (Pellegrini et al., 2020). To our knowledge, no comparison can be found in the literature between low-T and physical absorption technologies. To fill this gap, this work compares the performances of a process based on low-T distillation and one based on physical absorption using Dimethyl Ether of Polyethylene Glycol (DEPG) for CO₂ removal from natural gas/biogas. In addition to that, a hybrid configuration, consisting of low-T distillation for a bulk CO₂ removal and physical absorption as a finishing step, has been considered taking into account the recent interest in hybrid technologies (mainly for pre-/post-combustion CO₂ capture) due to their lower energy consumption.

2. Methods

The three process schemes taken into account in this work have been simulated with Aspen HYSYS[®] V9.0 (AspenTech, 2016a) and Aspen Plus[®] V9.0 (AspenTech, 2016b). These simulators are powerful aids in the design of processes, provided that the correct property package is selected, as specified in the following, depending on the specific system under investigation. They also offer the possibility of automatically specifying a desired value for some flowsheet variables using tools properly available for this purpose (e.g., the Design-Spec tool in Aspen Plus[®]). In all the simulations, the following specifications for the feed and product streams have been considered.

i) Natural gas (NG) feed stream: it has been assumed as a binary CH₄-CO₂ mixture (CO₂ molar fraction in the range 10-70 mol%), available at 35 °C and 50 bar with a total molar flowrate of 5000 kmol/h.

ii) Treated NG stream: its temperature and pressure have been set, respectively, at 25 °C and 50 bar and a CO₂ content of 2 mol% has been specified to produce a pipeline-quality NG.

iii) CO₂ product stream: it has been subject to further treatment, when needed, to bring it to its bubble point at 50 bar. This pressure value is suitable for its use for Enhanced Oil Recovery (EOR) purposes. For the process schemes involving the use of a solvent, the composition of the CO₂ product recovered from the solvent regeneration unit has been checked to ensure it meets the limit contents acceptable for injection into a depleted oil well and, in particular, that the CH₄ content is below 2 mol%.

The three different CO₂ removal technologies have been compared on the basis of their energy consumptions evaluated by means of the “Net Equivalent Methane” (NEM) method, considering that the following energy requirements are involved: i) heating above ambient temperature, ii) cooling below ambient temperature, iii) gas compression and liquid pumping. The calculation procedure and the parameters needed to convert them into the equivalent amount of methane can be found in the literature (De Guido et al., 2018). In addition to that, also the equivalent CO₂ emissions (Pellegrini et al., 2015) have been determined for each process scheme.

3. Process schemes

3.1. Low-temperature distillation

The process, which performs CO₂ removal by means of extractive distillation (Holmes and Ryan, 1982), has been simulated in Aspen HYSYS[®] V9.0 (AspenTech, 2016a) using the Soave-Redlich-Kwong (SRK) Equation of State. The reader can refer to a previous work (Pellegrini et al., 2018) for the process flowsheet to which some changes have been made in this work since the treated gas is a pipeline-quality gas rather than LNG. The entrainer, i.e. n-butane (nC₄), enters the extractive distillation column on the third tray from the top to avoid its entrainment. Its flow rate has been set at 10 moles/100 moles of feed and its temperature has been adjusted in each simulation to create the minimum discontinuity in column profiles. The feed gas has been subject to cooling and expansion prior entering the tower (25 theoretical trays) to reach its dew point at the operating pressure of 40 bar. The top product stream is, then, compressed to 50 bar and used for feed gas pre-cooling. The bottom product stream, which contains CO₂ and the entrainer, is expanded to 30 bar to remain under the nC₄ critical pressure, and it is fed on the 31st stage (from the top) of the regenerative distillation unit (40 theoretical trays), where CO₂ is separated from nC₄. CO₂ is recovered from the top and is, then, pumped and heated up to reach the final desired conditions. The bottom stream

from this column contains nC_4 (99.99 mol% purity), which is recycled to the extractive distillation tower, after mixing with an appropriate make-up stream, being pumped to 40 bar and cooled down to the desired temperature. This process belongs to the class of low-temperature CO_2 separation processes because of the temperature ($-85.3\text{ }^\circ\text{C}$) reached at the top of the extractive distillation column.

3.2. Physical absorption using DEPG

The physical absorption process using DEPG has been simulated in Aspen Plus[®] V9.0 (AspenTech, 2016b), selecting the “Acid gas: physical solvents” property package, which uses the PC-SAFT Equation of State. The process comprises an absorption section and a solvent regeneration section (Figure 1).

In the absorber, operated at 50 bar and consisting of 10 theoretical trays, the feed gas is contacted counter-currently with the lean solvent that is fed at the top at $20\text{ }^\circ\text{C}$. The treated gas obtained at the top of the absorber is heated up to $25\text{ }^\circ\text{C}$, when needed. The rich-solvent withdrawn at the bottom is sent to the regeneration section, which is performed in three flash drums, with the first two ones operated adiabatically and the last one operated non-adiabatically at atmospheric pressure. This requires some heat to be supplied to reach a final purity for the regenerated solvent set equal to 98.5 mol%. For inlet CO_2 contents up to 20 mol%, the first flash drum has been operated at 18 bar, whereas the pressure of the second one has been adjusted to avoid exceeding a 2 mol% content of CH_4 in the CO_2 product stream. This adjustment has turned out to be no more required at higher inlet CO_2 contents, for which the two flash drums have been operated at 27.6 and 13.8 bar, respectively (Kohl and Nielsen, 1997). The CO_2 stream separated in the regeneration section is compressed and cooled down to reach the final desired conditions. The lean solvent is recycled to the absorber after mixing with a make-up stream, pumping to 50 bar and cooling to $20\text{ }^\circ\text{C}$.

Since the heat required for solvent regeneration in the last flash drum is significant (it accounts for 56-76 % of the overall energy consumption in terms of NEM) as the CO_2 content in the inlet gas decreases from 70 to 10 mol%, heat recovery has been considered for energy-saving purposes. Indeed, the required heat can be provided by the liquid product from the third flash drum, which has to be cooled down before entering the absorber, and by the CO_2 product stream, which is available at temperatures above $500\text{ }^\circ\text{C}$ after compression. However, when the CO_2 content in the feed stream is ≤ 20 mol% additional heating is required, which is supplied by the hot compressed vapours from the intermediate flash drum (not shown in Figure 1).

3.3. Hybrid process: low-temperature distillation and physical absorption using DEPG

The hybrid process (Figure 2) is based on the Dual Refrigerant CO_2 Fractionation + Selexol flow scheme (Ross and Cuellar, 2010) and consists of a bulk removal step that reduces the CO_2 content down to 20 mol% by low-T distillation, and a finishing step performed by physical absorption into DEPG.

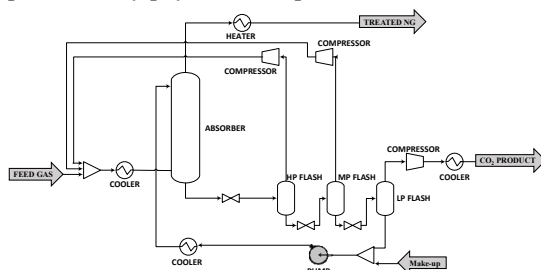


Figure 1. Flowsheet of the physical absorption process using DEPG.

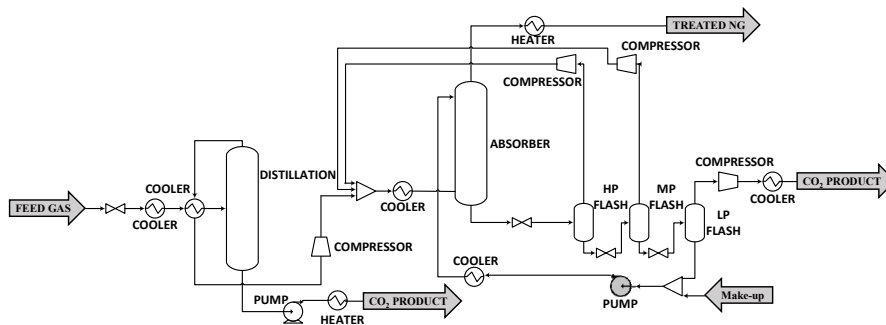


Figure 2. Flowsheet of the hybrid (low-temperature distillation + physical absorption) process.

Simulations for it have been carried out in Aspen Plus[®] V9.0 (AspenTech, 2016b), dividing the flowsheet into two sections, in which the SRK Equation of State and the PC-SAFT Equation of State have been used, respectively. The feed gas is first expanded to 40 bar, the operating pressure of the distillation column, chosen to maintain a certain margin to the critical point, avoiding higher traffic in the column. After expansion, it is cooled down (at least partially) to its dew point using the partially treated NG vapour stream exiting the top of the distillation column, which is available at about -54 °C. From the bottom of this column, pure CO₂ (99.99 mol%) is obtained, which is pumped and heated up to reach the final desired conditions. The top product of the distillation column is compressed to 50 bar and fed to the absorber, after mixing with the recycled streams from the first two flash drums in which the rich-solvent is regenerated. Prior entering the absorber, it is cooled down, assuming cooling water is used as cooling medium. The treated NG is obtained from the top of the absorber close to ambient temperature and it is heated up, when needed, to 25 °C. The rich-solvent is regenerated in three flash drums as in the process based on physical absorption only. The lean solvent is pumped to 50 bar and cooled down to 20 °C before being fed at the top of the absorber. The same energy recoveries as in the physical absorption process described above have been considered in the finishing step for improving the process energy consumptions to avoid supplying heat to the last flash by an external utility.

4. Results and discussion

Figure 3a shows the results of the energy analysis for the three process schemes described above. For CO₂ contents in the feed gas < 60 mol% the process scheme with the lowest energy requirements results to be the one based on low-T distillation. Energy requirements for the physical absorption process are slightly higher up to 25 mol% CO₂, but the difference increases significantly at higher CO₂ molar fractions. Above 60 mol% CO₂ in the feed gas, the hybrid process where the finishing step is accomplished by means of physical absorption into DEPG becomes more convenient. The same observations can be made when considering the equivalent CO₂ emissions, shown in Figure 3b. In Figure 3a, the results for another hybrid process investigated in a previous work (De Guido and Pellegrini, 2019) are also shown. In that case, the finishing step was carried out by chemical absorption into a 40 wt% MethylDiEthanolAmine aqueous solution. This hybrid process turns out to be more convenient than the physical absorption process at CO₂ inlet contents > 50 mol%, but it is more energy-demanding of the other two process configurations in the whole investigated range for CO₂ molar fractions.

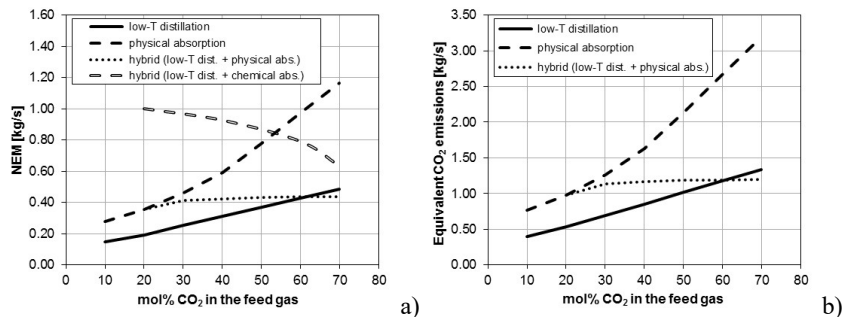


Figure 3. a) NEM consumption [kg/s], and b) equivalent CO₂ emissions [kg/s] for the three process schemes investigated in this work.

It is also important to point out that in the low-T distillation process CH₄ losses with respect to the amount of CH₄ fed to the process (Figure 4a) are practically negligible thanks to the fact that a specification on the CO₂ product purity can be directly imposed. It is also interesting to compare the different processes focusing the attention on the CO₂ product. Figure 4b shows the composition of this stream for the different investigated processes when the CO₂ content in the feed gas is 50 mol%. The bar related to the hybrid configuration refers to the composition of the CO₂ product obtained by mixing the stream withdrawn from the bottom of the distillation column and the stream recovered from the third flash in the finishing step. It is possible to state that the low-T distillation process is the only one that allows getting an almost pure CO₂ product (99.99 mol% purity). When CO₂ removal is performed by physical absorption or the hybrid configuration, the CO₂ product is contaminated with some methane, due to its co-absorption into the solvent, but not with DEPG thanks to its very low volatility.

The results presented in this work can be also interpreted in terms of operating costs. As for investment costs, some preliminary observations can be made. Considering, for example, the number of pieces of equipment, it is possible to observe it is the lowest in the process based on low-T distillation. On the contrary, the hybrid one involves the highest number of pieces of equipment, though they are of lower size with respect to those involved in each standalone CO₂ removal technology the hybrid process combines. A detailed economic analysis is required for a correct assessment of the capital expenses of the process configurations compared in this work.

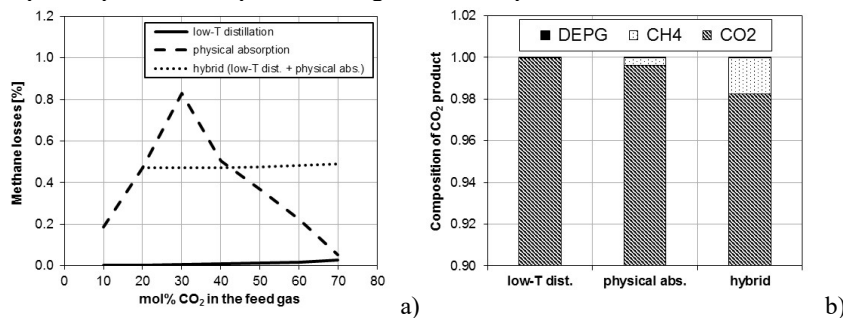


Figure 4. a) Methane losses [%] for the three process schemes investigated in this work; b) composition (molar fractions) of the CO₂-rich product obtained from the three considered process schemes when the CO₂ content in the feed gas is 50 mol%.

5. Conclusions

This work contributes to the current research on novel technologies for CO₂ removal from low-quality natural gas/biogas, which allow overcoming the high energy requirements of conventional technologies. For this purpose, three process schemes respectively based on low-T distillation, physical absorption into DEPG and a combination of the two (hybrid process) have been simulated in Aspen HYSYS® and Aspen Plus®.

The results of the energy analysis suggest that:

- i) low-temperature distillation involves lower energy consumptions and equivalent CO₂ emissions for CO₂ separation from low-quality natural gas or biogas;
- ii) the combination of low-temperature distillation for a bulk removal and physical absorption for a finishing step allows reducing the energy consumption of the process based on physical absorption only;
- iii) if methane losses and the purity of the CO₂ product stream are taken into account, low-temperature distillation has to be preferred to the other CO₂ separation processes.

Future studies will extend the comparison to other physical solvents and hybrid configurations. The latter ones are interesting options for reducing process energy consumptions, but require further research to design competitive processes taking into account factors like the properties of the feed gas and the desired purity of the products.

References

- AspenTech, 2016a, Aspen HYSYS®, Burlington, MA, United States.
- AspenTech, 2016b, Aspen Plus®, Burlington, MA, United States.
- G. De Guido, M.R. Fogli, L.A. Pellegrini, 2018, Effect of heavy hydrocarbons on CO₂ removal from natural gas by low-temperature distillation, *Industrial & Engineering Chemistry Research*, 57(21), 7245-7256.
- G. De Guido, L.A. Pellegrini, 2019, Performances of Novel Technologies for the Exploitation of CO₂-Rich Natural Gas Reserves, *Offshore Mediterranean Conference and Exhibition*, Piacenza, Italy, 27-29 March 2019.
- A.S. Holmes, J.M. Ryan, 1982, Cryogenic Distillative Separation of Acid Gases from Methane, U.S. Patent No. 4,318,723, Washington, DC: U.S. Patent and Trademark Office.
- A.L. Kohl, R. Nielsen, 1997, *Gas purification*, Elsevier.
- S. Langè, L.A. Pellegrini, P. Vergani, M. Lo Savio, 2015, Energy and economic analysis of a new low-temperature distillation process for the upgrading of high-CO₂ content natural gas streams, *Industrial & Engineering Chemistry Research*, 54(40), 9770-9782.
- L.A. Pellegrini, S. Langè, M. Baccanelli, G. De Guido, 2015, Techno-economic analysis of LNG production using cryogenic vs conventional techniques for natural gas purification, *Offshore Mediterranean Conference and Exhibition*, Piacenza, Italy, 25-27 March 2015.
- L.A. Pellegrini, G. De Guido, G. Lodi, M. Saeid, 2018, CO₂ Capture from Natural Gas in LNG Production. Comparison of Low-Temperature Purification Processes and Conventional Amine Scrubbing, *Cutting-Edge Technology for Carbon Capture, Utilization, and Storage*, 283-308.
- L.A. Pellegrini, G. De Guido, V. Valentina, 2019, Energy and exergy analysis of acid gas removal processes in the LNG production chain, *Journal of Natural Gas Science and Engineering*, 61, 303-319.
- L.A. Pellegrini, G. De Guido, S. Ingrosso, 2020, Thermodynamic Framework for Cryogenic Carbon Capture, *Computer Aided Chemical Engineering*, 48, 475-480.
- F.P. Ross, K.T. Cuellar, 2010, Economical Option for CO₂/Methane Separation in Produced Gas Containing a High CO₂ Fraction. 89th Annual GPA Convention, Austin, TX, USA, 21-24 March 2010.

The Influence of Cuttings Polydispersity on Wellbore Cleaning Operations using the Kinetic Theory of Granular Flow and the Dense Discrete Phase Model

Emmanuel I. Epelle, Dimitrios I. Gerogiorgis*

Institute for Materials and Processes (IMP), School of Engineering, University of Edinburgh, The Kings Buildings, Edinburgh, EH9 3FB, United Kingdom

* D.Gerogiorgis@ed.ac.uk

Abstract

In this study, polydispersed spherical particulate systems are examined using the Eulerian-Eulerian (EE) and Lagrangian-Eulerian (LE) modelling approaches in the context of wellbore cleaning operations in the drilling industry. The particle sizes considered here are between 0.5 mm and 1mm, whereas a Power Law rheological model is used for the fluid phase description. The EE approach implemented herein applies the Kinetic Theory of Granular Flow (KTGF) in ANSYS Fluent® and accounts for the particle size differences by representing them as different phases within the flow domain. In addition, the LE approach (employing the Dense Discrete Phase Model – DDPM) represents this difference with the aid of a size distribution model (the Rosin Rammler model). The findings of our computational experiments show that neglecting the size disparity may lead to severe under/overestimation of key variables such as the pressure drop, and particle deposition tendencies; thus, leading to decreased model performance.

Keywords: Polydispersity, cuttings transport, Discrete Phase Model, Rosin-Rammler

1. Introduction

Excessive drillpipe torque, slow drilling rates, and stuck drillpipe are some of the operational challenges faced by drilling engineers during wellbore cleaning (Epelle and Gerogiorgis, 2018a; Zhu et al., 2019). These problems culminate to significantly increased capital and operating expenditure over the drilling and production time horizons. Thus, numerical tools and recent advances in computing developed by the process systems engineering community have been readily applied to understand cuttings transport phenomena with complex non-Newtonian fluids under unfavourable downhole conditions. Particularly, numerous studies that apply Computational Fluid Dynamics (CFD) for the description of annular multiphase (fluid-solid) transport of dense and dilute particulate mixtures have emerged over the past decade (Yilmaz, 2012; Hajipour, 2020; Tang et al., 2020). However, a prevalent assumption in these studies is the existence of a monodispersed system of particles (both in terms of size and shape), which is never the case in real field operations. CFD studies accounting for polydispersed cuttings distribution are scarce, thus motivating the present study. We address this limitation here by applying the Kinetic Theory of Granular Flow (KTGF) and the Dense Discrete Phase Model (DDPM). A comparative assessment of these modelling methods is also presented.

2. Methodology

This study implements the EE model as the first case study for the multiphase flow description in the annulus (Fig. 1a). This model considers the particles of different diameters (Table 1) as continuous separate phases that interact with the non-Newtonian fluid phase via the interphase momentum exchange coefficient. This continuum assumption of the solid phase implies the kinematic and discrete nature of the solid phase are not explicitly accounted for; however, the application of KTGF enables the estimation of the kinematic properties via closure models. Thus, for each particulate phase, we hereby implement closure models for the granular viscosity, granular bulk viscosity, solid pressure, radial distribution and elastic modulus, respectively, using ANSYS Fluent[®].

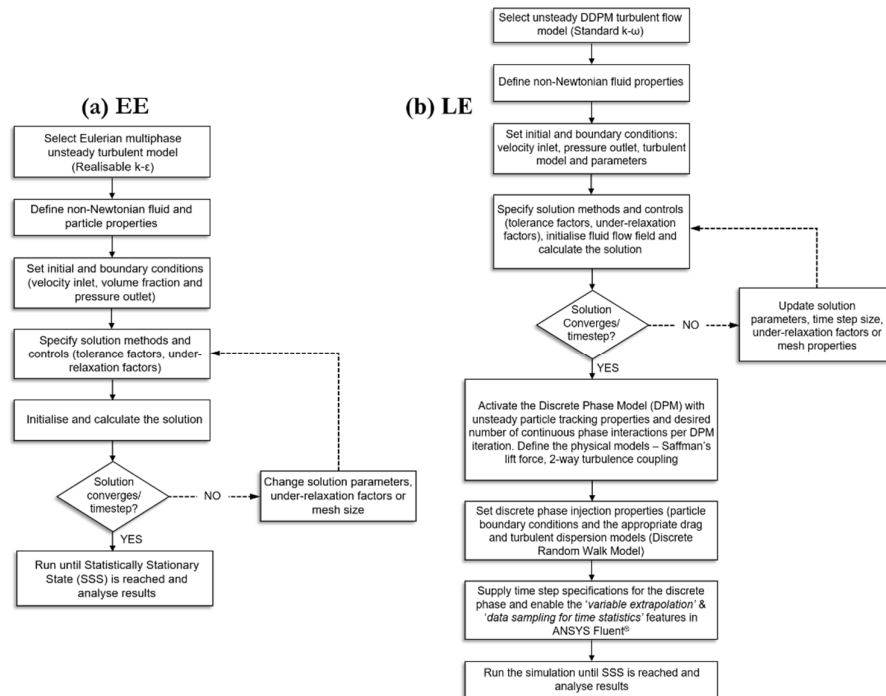


Figure 1: Implemented methodology for the EE and LE model.

Conversely, the LE model which utilises statistically computed particle trajectories coupled with an Eulerian description of the fluid phase is employed. Specifically, the DDPM is applied, given its ability to handle high particle concentrations in the annulus compared to the DPM model, which is limited to dilute particulate suspensions (<12%). For the particulate phase description, the Rosin-Rammler size distribution model is implemented (parameters shown in Table 1). The stepwise procedure for the implementation of this model is shown in Fig. 1b. Although more computationally demanding, this approach better describes the particulate phase motion via the integration of the force balance on the particles in the Lagrangian reference frame. Epelle and Gerogiorgis (2018b) discusses further contrasting differences between these approaches.

A preliminary assessment of the turbulent models in ANSYS Fluent, showed better stability of the *Realisable k-ε* model for the EE simulations and *Standard k-ω* for the LE simulations, respectively. The computational domain was constructed to model wellbore trajectories obtainable in deviated drilling operations, consisting of horizontal, inclined and vertical sections. Discretising the domain with hexahedral elements resulted in a robust mesh with 665,600 elements, and optimal orthogonality, aspect ratio and skewness factors; this yielded grid-independent results. The Quadratic Upstream Interpolation for Convective Kinematics scheme (QUICK) was applied for the discretisation, whereas pressure-velocity coupling was performed using the Semi-Implicit Method for Pressure Linked Equations (SIMPLE algorithm). Supercomputing resources with 16 processing cores (2.4 GHz Intel®-Xeon® CPU) and 32GB of RAM (University of Edinburgh) have been employed to solve the models to full convergence (time step = 0.0005; tolerance factor = 0.001). Table 1 provides further details on the boundary conditions of this study.

Table 1: Simulation input parameters.

	EE model	LE model
Computational geometry		
Casing diameter, d_c (m)	0.113	0.113
Drillpipe diameter, d_{dp} (m)	0.180	0.180
Total length, L (m)	2.340	2.340
Fluid properties		
Composition	0.5% CMC solution	0.5% CMC solution
Fluid density, ρ_l (kg.m ⁻³)	1,000	1,000
Yield stress, τ_0 (Pa)	0	0
Consistency index, K (Pa.s ^{<i>n</i>})	0.5239	0.5239
Flow behaviour index, n	0.60	0.60
Particle properties		
Cuttings density, ρ_p (kg.m ⁻³)	2,800	2,800
Sphericity, ψ	1.00	1.00
Cuttings diameter, d_p (mm) & Rosin-Rammler parameters for the LE model	0.50 (d_{p1}), 0.75 (d_{p2}), 1 (d_{p3})	Min. $d_p = 0.5$; max. $d_p = 1$; mean $d_p = 0.75$; spread parameter = 3.368; number of diameters = 3
Drilling variables		
Cuttings inlet velocity, v_p (m.s ⁻¹)	0.25	0.25
Cuttings inlet volume fraction (-)	0.15 (d_{p1}), 0.2 (d_{p2}), 0.15(d_{p3})	0.5
Drill mud (fluid) circulation velocity, v_l (m.s ⁻¹)	1.5	1.5
Wellbore eccentricity, e	0.6	0.6
Drillpipe rotation (rpm)	100	100

Particles of perfect sphericity ($\psi=1$) are utilised in this study; however, a rather complex but more realistic flow scenario is one in which both size and sphericity distribution models are applied (this is beyond the scope of this study). Computations involving this added complexity will benefit from the availability of experimental data describing such conditions; nonetheless, these are very scarce. The herein described modelling procedure, has been experimentally validated (for monodispersed cuttings) as reported in Epelle and Gerogiorgis (2018a). We present the findings from this study in the subsequent section.

3. Results and Discussion

The cuttings volume fractions for the respective particulate phases of the EE model and the LE model are shown in Fig. 1. Generally, the 0.75 mm particles occupy a significant portion of the annulus, with higher averaged volume fractions observed. This is attributable to its relatively higher inlet volume fraction compared to the other particle sizes (Table 1). However, it can also be observed that the concentration of the smallest particles (0.5 mm) at the annular base is highest, although with significantly unoccupied regions in the upper/wider annular section. Contrary to the observation in a monodispersed scenario (in which smaller particles are more readily transported by the fluid (as shown in Epelle and Gerogiorgis, 2018a), a polydispersed scenario reveals that smaller particles can cause some transport difficulties. It is worth highlighting that the maximum concentration (denoted as the dark red region) has a value of 0.25, and does not reflect a packed bed (volume fraction > 0.63 in ANSYS Fluent[®]). The LE model on the other hand shows a rather unified bulk movement of the particles (although with some faint red regions). This observation can be attributed to the implementation of size distribution model in the flow domain, in contrast to the alternative EE scenario in which the particulate phases are recognised as separate although interacting (Pang et al., 2018).

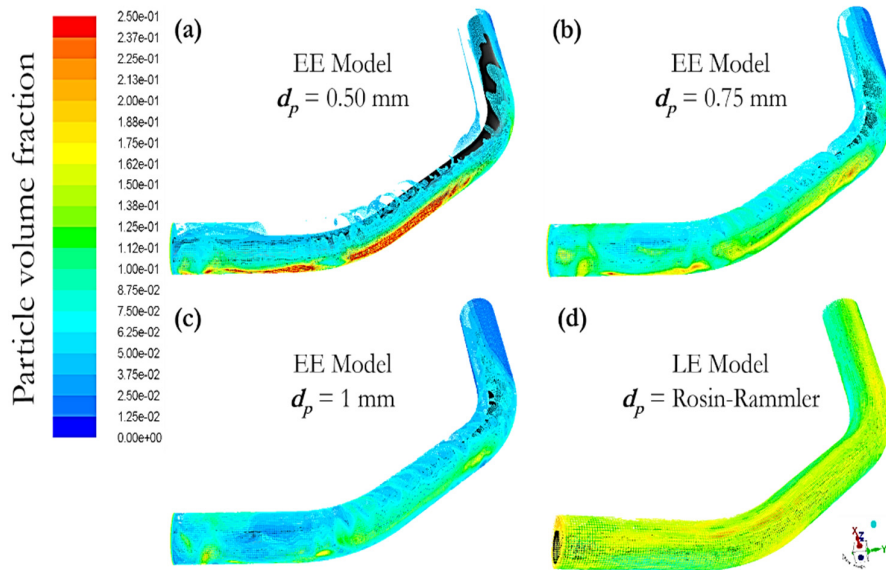


Figure 2: Particle volume fractions as calculated using the EE vs. the LE model.

On analysing the velocity contour plots for the different particle sizes of the EE model, no significant differences in the velocity magnitude were observed; thus, a single particle velocity plot of the EE model is compared with the LE model, as shown in Fig. 3. However, both plots similarly portray the wider annular sections as the regions of high velocity; this observation is attributable to the eccentricity of the domain. Furthermore, the velocities in the horizontal-to-inclined region of the annulus of the EE model shows significantly higher values compared to the LE model. The inherent treatment of the solid phase as a continuous phase and the consequent overestimation of the solid phase velocity is the likely reason for this observation, compared to the LE method, in which the discrete

behaviour of the particles is preserved. It is also worth mentioning that the EE model showed significantly better numerical stability with the specified rotary boundary condition at the drillpipe walls compared to the LE model. Hence, the rotational effect was only gradually applied (and repeatedly increased) to the LE model to ensure stability. This difference in the addition of this extra velocity component (via rotation) has also quite plausibly resulted in the higher velocities which are observed in EE model results.

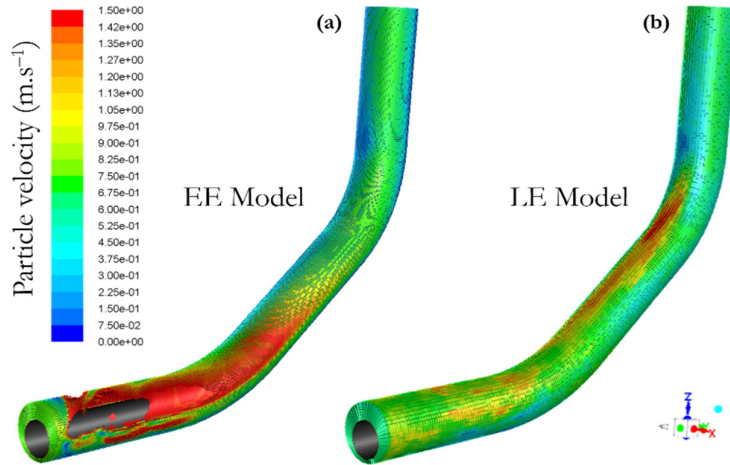


Figure 3: Mean particle velocity contours as calculated using the EE vs. the LE model.

Fig. 4a presents a summary of the volume-averaged properties (over the entire computational domain) for the EE and LE model, respectively. For reasons previously explained, a 27% overestimation of the velocity by the EE model (taking the LE model as the true value) is observed. The radial velocity components for both models are similar; thus indicating that the axial and tangential velocities are the main contributors to the difference observed in the velocity magnitude. Conversely, the EE model under-predicts the volume fraction by 25% as observed in Fig. 4a. This presented volume-averaged value for the EE model was also averaged over the 3 particle diameters. The observed dissimilarity in volume fraction may be associated with the difference in the particle handling methods of both models (separate particulate phases vs. size distribution model).

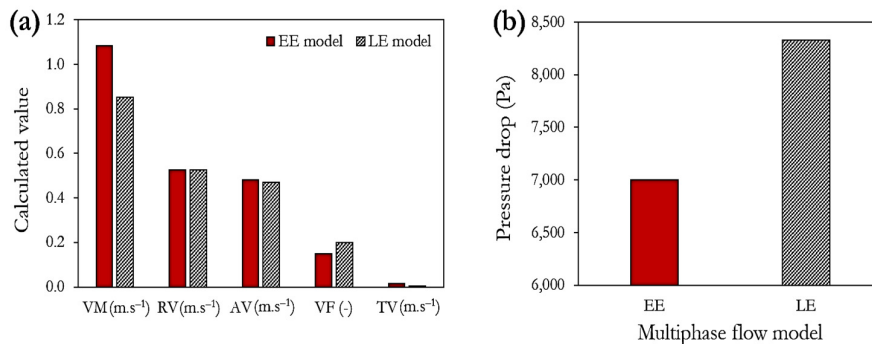


Figure 4: Comparison of velocity components (similar) and pressure drop (considerably different) for EE vs. LE models. (Nomenclature for velocities - VM: velocity magnitude; RV: radial velocity; AV: axial velocity; VF: volume fraction; TV: tangential velocity).

Fig. 4b highlights a clear difference in the calculated pressure drop across the entire computation length for the EE and LE models. A discrepancy of up to 1300 Pa is observed between models. Despite the higher travel velocities observed with the EE model, the pressure drop is lower. Given the more accurate representation of the particle-particle interaction by the LE model (via individually computed particle trajectories), the resulting pressure drop from such momentum exchange is likely to be greater than that for a scenario in which constitutive (closure) relations are used, as is the case for the EE model.

4. Conclusions

This study has evaluated the influence of particle polydispersity on the cuttings transport efficiency in a deviated annulus and comparatively analysed the differences between the EE model (via KTGF) and the LE model (via DDPM). Presented velocity and particle concentration profiles indicate that flow complexity will further increase due to multiple momentum transfer mechanisms (particle-fluid and particle-particle interactions) resulting from polydispersity. The findings of our computational experiments show that neglecting the size disparity in the system may lead to an overestimation of the depositional tendencies of the particles; thus, leading to inaccurate predictions. Thus, incorporating particle polydispersity is a necessary step towards improving the predictive performance of CFD models in wellbore cleaning operations. From a computational cost perspective, the LE model required roughly 2 times the time and effort needed for the EE model. The numerical stability of the LE model proved sensitive to the rotational effect of the drillpipe. Further work may consider the effect of polydispersity in both the particle size and sphericity; such endeavours will certainly require experimental measurements for model validation purposes. Nevertheless, we must note that the CFD approach we have developed and present is based on quantitatively and qualitatively validated sub-models (under monodispersed conditions) that have achieved grid-independent solutions: no significant difference in accuracy is expected for polydispersed transport conditions.

References

- D., Yilmaz, 2012. Discrete Phase Simulations of Drilled Cuttings Transport Process in Highly Deviated Wells. *Master dissertation, Louisiana State University, USA.*
- E.I. Epelle, D.I. Gerogiorgis, 2018a. CFD modelling and simulation of drill cuttings transport efficiency in annular bends: effect of particle sphericity. *J. Pet. Sci. Eng.*, 170, 992-1004.
- E.I. Epelle, D.I. Gerogiorgis, 2018b. Transient and steady-state analysis of drill cuttings transport phenomena under turbulent conditions. *Chem. Eng. Res. Des.*, 131, 520-544.
- M. Hajipour, 2020. CFD simulation of turbulent flow of drill cuttings and parametric studies in a horizontal annulus. *SN App. Sci.*, 2, 1-12.
- M. Tang, G. Yao, S. He, L. He, 2020. Spiral laminar flow of yield-power-law fluids in partially blocked annulus. *J. Dispersion Sci. Technol.*, 1-13.
- X. Zhu, K. Shen, B. Li, Y. Lv, 2019. Cuttings Transport Using Pulsed Drilling Fluid in the Horizontal Section of the Slim-Hole: An Experimental and Numerical Simulation Study. *Energies*, 12, 20, 3939.
- B. Pang, S. Wang, G. Liu, X. Jiang, H. Lu, Z. Li, 2018. Numerical prediction of flow behavior of cuttings carried by Herschel-Bulkley fluids in horizontal well using kinetic theory of granular flow. *Powder Technol.*, 329, 386-398.

A Digital Twin for Process Optimisation in Pharmaceutical Manufacturing

Dimitrios I. Gerogiorgis^{a,*} and Daniel Castro-Rodriguez^b

^a*School of Engineering (IMP), University of Edinburgh, Edinburgh, EH9 3FB, UK*

^b*GlaxoSmithKline (GSK), Montrose, Angus, DD10 8EA, UK*

**D.Gerogiorgis@ed.ac.uk*

Abstract

Organic synthesis is a critical step in producing any Active Pharmaceutical Ingredient (API): it encompasses a series of organic (mostly catalytic) reactions and separations, whose purpose is to arrive at a solution of a target molecule, which must then undergo crystallisation for purification, followed by secondary (downstream) processing towards the final (commercially available) formulation. Consequently, studying and optimising industrial-scale API production requires reliable process unit (especially reactors and separators) and system descriptions, to track key chemical (esp. impurity) concentrations. The present paper describes a Digital Twin (DT) which is currently under development for a particular API target: its goal is to evaluate multicomponent adsorption of Volatile Organic Compound (VOC) emissions from licensed salbutamol synthesis, via industrial FTIR data for mixtures of dichloromethane, chloroform, toluene, methanol and ethanol. A combination of first-principles (e.g. PDE-based adsorption) and data-driven (e.g. PCA) methodologies is essential in order to ensure techno-economically optimal performance.

Keywords: Digital Twin (DT), Hybrid Modelling, Pharmaceutical Manufacturing.

1. Introduction

Pharmaceutical manufacturing plants consist of a wide range of complex unit operations, from upstream processing aimed at synthesis of Active Pharmaceutical Ingredients (APIs) to intermediate purifications and downstream processing for final product formulation. For high plant performance at minimal total cost and solvent waste *under licensed design*, operators continuously assess key units, frequently with only input-output data available. First-principles modelling is desirable, yet parameterisation is often arduous and elusive, especially in case of inherently dynamic operation under sensor accessibility constraints.

Since API production processes are complex (Nagy et al., 2020), with several organic feeds, key intermediates and solvents, many specific interactions are not well understood. Relying on first-principles models (FPM) with hundreds of Partial (PDE) or Ordinary Differential (ODE) and/or Algebraic Equations (AE) is impractical, since such equation systems are costly to formulate and cumbersome to parameterise under time pressure. Statistical (e.g. Latent Variable) methods such as Principal Component Analysis (PCA) are widely used to explore industrial data (if abundant) and discover causality patterns. Employing data-driven models (DDM) to bridge the model fidelity gaps is imperative, especially with industrial production Big Data routinely stored. Operational challenges may however impede DDM use without a FPM basis, especially for licensed processes. For example, an existing (legacy) DDM approach may credibly capture API plant yield and key output quality attributes, but also miss secondary performance characteristics (e.g. VOC emissions) entirely, just because it was neither meant nor built to address them.

Numerous recent publications present elaborate process model implementations aimed at plantwide pharmaceutical process design: frequently high fidelity has been ensured by employing detailed kinetic (e.g. Arrhenius), thermodynamic (e.g. UNIFAC, PC-SAFT) and/or material property submodels based on or developed via previously published data. Techno-economic optimisation studies based on such high-fidelity submodels credibly pinpoint process improvements and quantify savings vs. current API production protocols (Jolliffe & Gerogiorgis, 2016; Diab & Gerogiorgis, 2019), but in all cases require access to (and abundance of) public-domain datasets, invariably unavailable for new ventures.

Hybrid process modelling has emerged as an industrially promising alternative which receives ever-increasing literature attention (Oliveira, 2004; von Stosch et al., 2014; Glassey & von Stosch, 2018, Yang et al., 2020); a recent comprehensive review (Chen et al., 2020) and some implementations, mostly in secondary (downstream) pharmaceutical processing (Ismail et al., 2019; Bascone et al., 2020) have showcased its definite potential. Nevertheless, exploring and securing the validity domain of hybrid modelling approaches is not trivial, especially regarding feasibility (Kahrs & Marquardt, 2007; Bae et al., 2020). The latter study distinguishes two types and three broad categories of hybrid models: in the first (serial) type and category, the black (DDM) only informs the white (FPM) box. In the second (embedded) type, the two are either separate (but fully interacting) entities (with the FPM however being the only handling final outputs), or are entirely fused without executional distinction, featuring full integration at higher development expense. Dynamic optimisation relies on valid domain constraints extracted from the use of DDM.

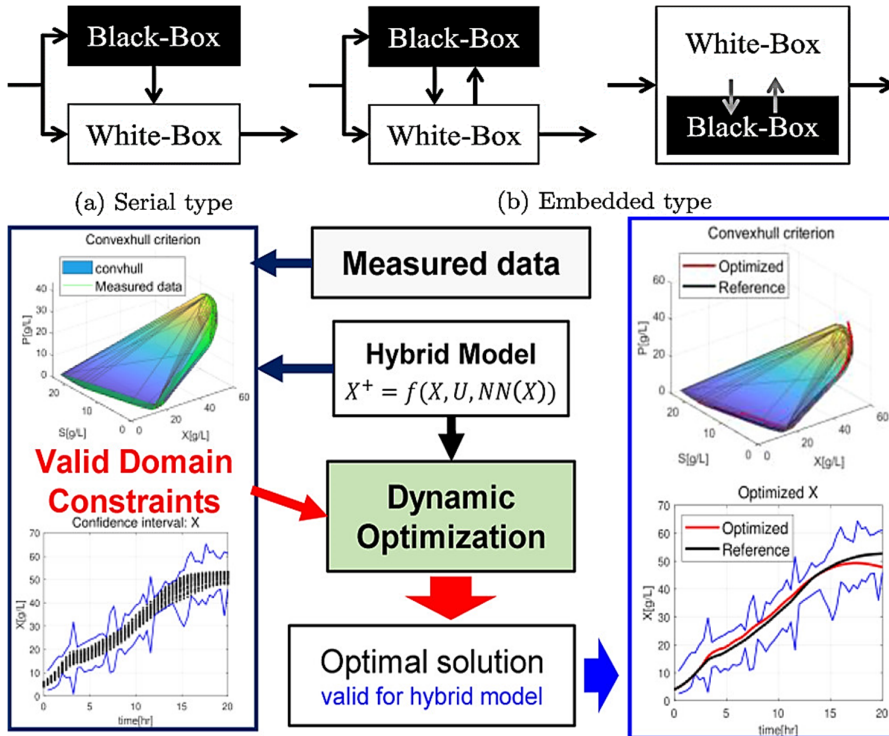


Figure 1: Hybrid model structures and dynamic optimisation strategy (Bae et al., 2020).

2. Salbutamol Manufacturing Process Description

Salbutamol is a proven bronchodilation medication: reliably expanding lung airways to comfort patients with respiratory conditions, it is commonly prescribed against Chronic Obstructive Pulmonary Disease (COPD), bronchitis, emphysema and other ailments. Manufacturing this API requires 4 feedstock compounds (*diacetate*, *glycyl*, *benzyl*, *base*), all produced in the *licensed plant* within a salbutamol batch production schedule (Fig. 2). Prior to this project, the first two (*diacetate* and *glycyl*) subsystems were conclusively identified as the two major VOC sources, thus forming the exclusive focus of our study. Therein, VOC emissions originate from distillation units needed for solvent separation. Mass balance analysis for the majority of VOC emissions has previously indicated that: (1) for *diacetate production*, VOCs come from dichloromethane (DCM) atmospheric and vacuum distillation (occurring in vessel 3/V3), and toluene distillation (in vessel 4/V4). (2) for *glycyl production*, VOCs originate from chloroform and toluene distillation (V3).

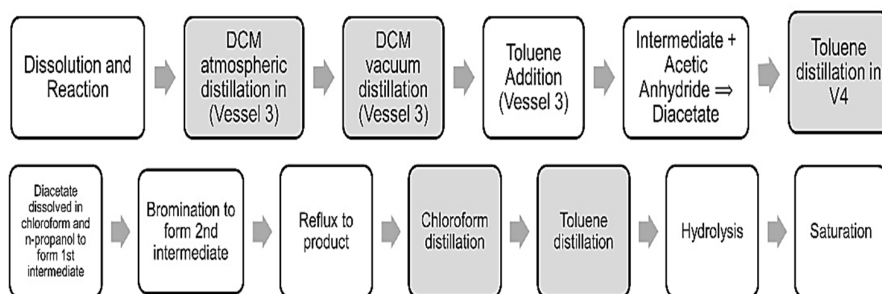


Figure 2: Diacetate (top) and glycyl (bottom) production schedules (grey denotes separation steps).

Fig. 3 shows the licensed VOC abatement system: all solvent emissions from salbutamol production are simultaneously fed to the same treatment (fixed activated carbon bed) unit. Technical (VOC abatement efficiency) but also economic (Operating Expenditure/OpEx) performance strongly depend on this subsystem, whose operation is not fully quantified.

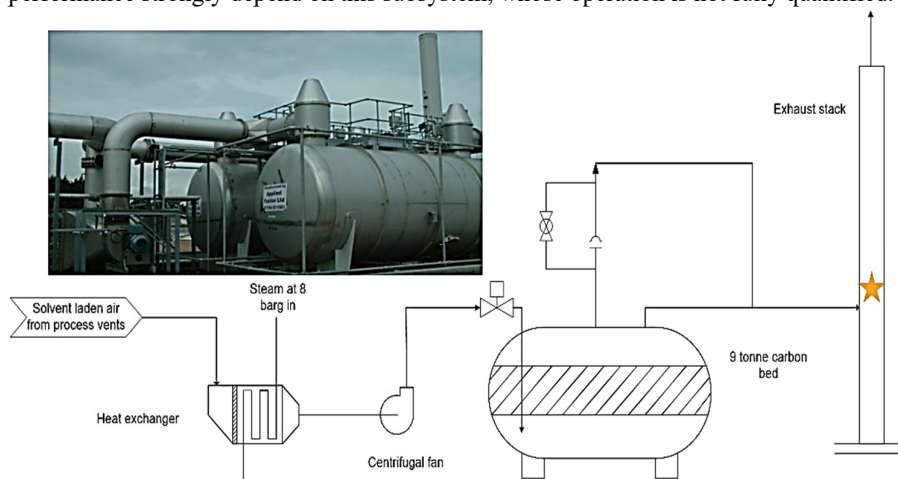


Figure 3: The VOC abatement system (a star denotes the FTIR sensor sampling point of the stack).

3. A Digital Twin (DT) for VOC Capture and its Potential in Pharma 4.0

A *Digital Twin (DT)* is a software entity interfaced with hardware, enabling simulations to probe how design/control changes affect performance, but also e.g. maintenance needs. When interfaced with process (measured) data, DTs can continuously improve and grow (e.g. re-parameterisation via DDM and/or AI), thus ensuring sustainable model relevance. Advances via DT use emerge rapidly in Pharma 4.0 (Casola et al., 2019; Lopes, 2020; Rasheed et al., 2020), at considerably faster pace than other fields (Aversano et al., 2020).

Chemical synthesis of APIs relies on organic solvents (alcohols, ketones, halogenated hydrocarbons) for conducting reactions and separations at adequate yield and safety. Many VOCs evaporate at temperatures much lower than 100 °C: due to high toxicity and catastrophic environmental impact, they must be reliably captured from gas emissions of several units: feed patterns, volumes and conditions govern unit performance (Fig. 4), Consequently, adsorbent material procurement and/or regeneration costs are a challenge. There are *two major technical problems* with the efficiency of modern VOC capture units:

- *First*, the VOC adsorption carbon beds can be quickly and irregularly saturated, due to the widely variant gas feed composition. Consequently, the bed (activated carbon) material must undergo frequent regeneration, a costly task invariably outsourced.
- *Second*, the impure feed stream composition (from the numerous organic synthesis miniplants hosted within a single multi-purpose API manufacturing site) always varies widely in terms of organic solvents, volumes and loads contained in emissions, affecting VOC capture performance, efficiency and above all the total annual cost.

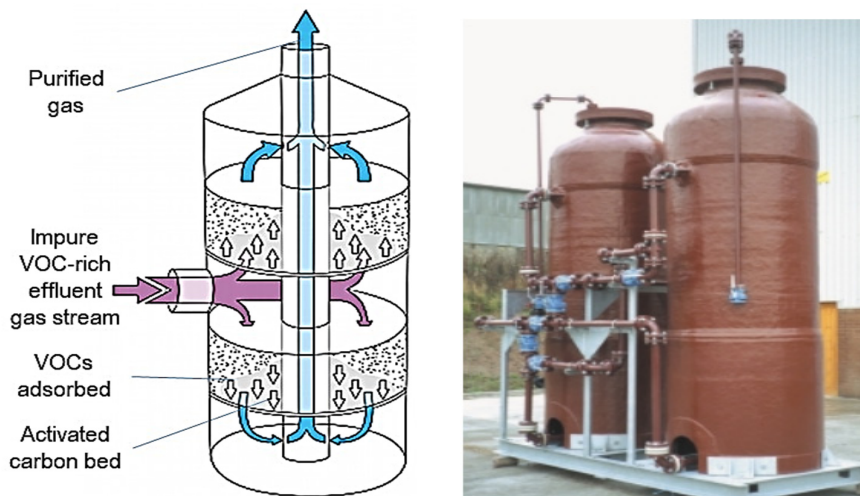


Figure 4: Operating principle and internals of a commercial VOC adsorption column (Forbes, UK).

The present paper addresses the analysis (towards optimisation) of multicomponent VOC adsorption from pharma emission mixtures on industrial activated carbon beds (Fig. 3-4). This project aims to combine first-principles (physics-based) and data-driven process modelling, to create and use a **Digital Twin** (advanced performance evaluation system), to understand, interface datastreams and improve the VOC capture unit at an API plant. The validated Digital Twin can be used to **analyse and predict** carbon bed performance against feed composition variation, and thus comparatively evaluate improvements (e.g. feed gas routing) via **techno-economic (cost) optimisation** (MILP/MINLP) methods.

The key goals of this ongoing project are to: **(1)** Use formal **Data Analytics**, i.e. Principal Component Analysis (**PCA**) to extract process features and elucidate key trends; this has already born fruit, producing key original results towards DDM of VOC capture (Fig. 5), **(2)** Develop **FPM for Multicomponent Adsorption** and study various key VOC mixture combinations, thereby quantifying the effect of feed variability on bed operation (Fig. 5), **(3)** Build and validate the **DT**, evaluating its **predictive potential** vs. VOC bed history, **(4)** Compare **DT simulation predictions** vs. plant efficiency data for given API batches, **(5)** Perform **DT-based dynamic optimisation** to reduce Operating Expenditure (OpEx) via optimal gas routing and mixing, under API plant throughput and emission constraints. Our DT combines best-in-class components: physics-based (FPM) descriptions rely on detailed equation systems capturing macro- (flow) and micro- (adsorption) transport phenomena, emission thermodynamics/phase equilibria, and VOC capture vessel design. Industrial data on VOC emissions vs. production conditions have been compiled from a dedicated gas monitoring station for the salbutamol API plant (Montrose, Scotland, UK), with special attention to VOC-critical production stages (namely distillation separations). Statistical (DDM) descriptions successfully use Principal Component Analysis (PCA), an established methodology we employ to correlate feed compositions (and changes) to bed output patterns, since breakthrough times (esp. due to selective adsorption) vary greatly.

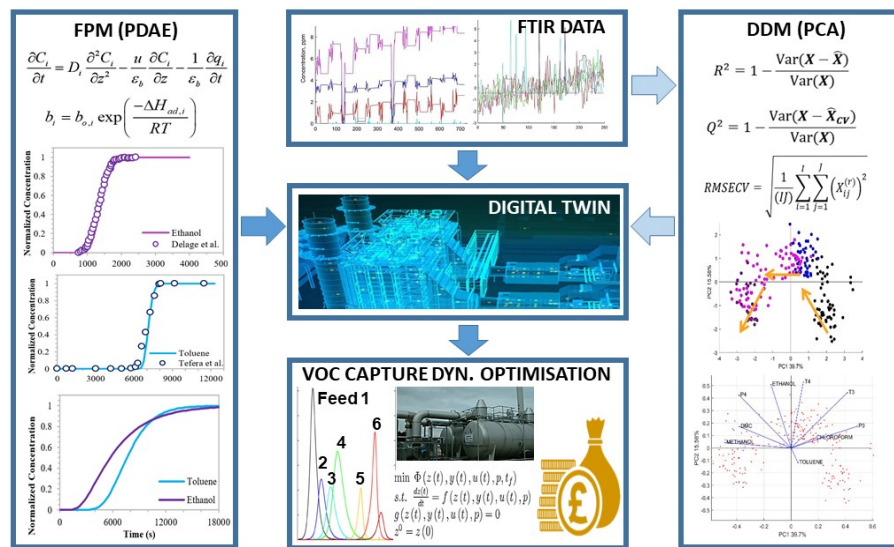


Figure 5. DT architecture: FPM and DDM frameworks and results towards dynamic optimisation.

A DT attains real economic value when *interfaced with process measurements*, hence datastream procurement from both inlet and outlet points of the VOC capture bed is key. The combination and complementarity of FPM and DDM offers unprecedented potential. Design (column flow topology) and operational (feed gas stream mixing and/or routing) modifications can be tested (and benefits computed) *in silico*, prior to hopeful adoption. Once the VOC capture DT is conclusively validated vs. industrial data, it will be used to:

- Quantify VOC capture (carbon bed) performance in detail
- Map VOC capture efficiency vs. feed/condition variations
- Evaluate and identify superior gas emission handling protocols
- Perform a comparative techno-economic analysis of all promising modifications
- Propose respective VOC capture trial runs for the API manufacturing plant.

4. Conclusions

Elaborate plantwide process models describing complex upstream (API) manufacturing need accurately capture thermodynamics (phase equilibria), separations and economics for many operating scenarios and conditions (pressure, temperature, feeding, mixing), but parameterising them is a laborious affair, given the multiphase and multicomponent mixtures and the resulting complications ubiquitous in API syntheses (Nagy et al., 2020). Licensed pharmaceutical manufacturing inevitably relies on significant solvent volumes for many (especially reaction/separation) uses, with environmental repercussion risks. The Digital Twin (DT) addressing VOC emissions abatement uses formal Data Analytics, has identified causality patterns via industrial production data, and develops predictive potential for dynamic cost optimisation of a VOC capture unit (feed patterns, conditions).

Acknowledgements

The authors gratefully acknowledge a SRPe-NMIS PhD Scholarship provided to the DG group, as well as a Royal Society Short Industrial Fellowship (2020-21) awarded to D.I.G.

References

- G. Aversano, M. Ferrarotti, A. Parente, 2020. Digital twin of a combustion furnace operating in flameless conditions: reduced-order model development from CFD simulations, *Proceedings of the Combustion Institute* (in press).
- J. Bae, H.J. Lee, D.H. Jeong, J.M. Lee, 2020. Construction of a valid domain for a hybrid model and its application to dynamic optimization with controlled exploration, *Industrial & Engineering Chemistry Research*, **59**(37), 16380–16395.
- D. Bascone, F. Galvanin, N. Shah, S. Garcia-Munoz, 2020. A hybrid mechanistic-empirical approach to the modelling of twin screw feeders for continuous tablet manufacturing, *Industrial & Engineering Chemistry Research*, **59**(14), 6650–6661.
- G. Casola, C. Siegmund, M. Mattern, H. Sugiyama, 2019. Data mining algorithm for pre-processing biopharma drug product manufacturing records, *Computers & Chemical Engineering* **124**, 253–269.
- Y.J. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, M. Ierapetritou, 2020. Digital twins in pharmaceutical and biopharmaceutical manufacturing: A review, *Processes* **8**(9), 1088.
- S. Diab, D.I. Gerogiorgis, 2019. Technoeconomic MINLP optimization for design of LLE cascades in CPM of atropine, *AIChE Journal* **65**, e16738.
- J. Glassey, M. von Stosch, 2018. *Hybrid Modeling in Process Industries*, CRC Press.
- H.Y. Ismail et al., 2019. Developing ANN-Kriging hybrid model based on process parameters for prediction of mean RTD in twin-screw wet granulation, *Powder Technology*, **343**, 568–577.
- H.G. Jolliffe, D.I. Gerogiorgis, 2016. Plantwide design and economic evaluation of two CPM cases: Ibuprofen and artemisinin, *Computers & Chemical Engineering*, **91**, 269–288.
- O. Kahrs, W. Marquardt, 2007. The validity domain of hybrid models and its application in process optimization, *Chemical Engineering & Processing: Process Intensification* **46**, 1054–1066.
- M.R. Lopes, A. Costigliola, R. Pinto, S. Vieira, J.M.C. Sousa, 2020. Pharmaceutical quality control laboratory digital twin – A novel governance model for resource planning and scheduling, *International Journal of Production Research*, **58**(21), 6553–6567.
- Z.K. Nagy, A. El Hagrasy, J. Litster, 2020. *Continuous Pharmaceutical Processing*, Springer.
- R. Oliveira, 2004. Combining first principles modelling and artificial neural networks (ANN): a general framework, *Computers & Chemical Engineering*, **28**(5), 755–766.
- A. Rasheed, O. San, T. Kvamsdal, 2020. Digital twin: values, challenges and enablers from a modeling perspective, *IEEE Access*, **8**, 21980–22012.
- M. von Stosch, R. Oliveira, J. Peres, S.F. de Azevedo, 2014. Hybrid semi-parametric modeling in process systems engineering, *Computers & Chemical Engineering*, **60**, 86–101.
- V. Steinwandter, D. Borchert, C. Herwig, 2019. Data science tools and applications on the way to Pharma 4.0, *Drug Discovery Today*, **24**(9), 1795–1805.
- S. Yang, P. Navarathna, S. Ghosh, B.W. Bequette, 2020. Hybrid modeling in the era of smart manufacturing, *Computers & Chemical Engineering*, **140**, 106874.

Development of Hybrid RTO approaches in the absence of a rigorous dynamic model by the use of Hammerstein model structures

Pedro de A. Delou^{a,*}, Rodrigo Curvelo^a, Maurício B. de Souza Jr.^{a,b} and Argimiro R. Secchi^a

^a*Chemical Engineering Program/COPPE, Federal University of Rio de Janeiro, Rio de Janeiro - Brazil*

^b*Chemical Engineering Department/School of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro – Brazil*
pdelou@peq.coppe.ufrj.br

Abstract

One of the main drawbacks of the so-called two-step approach in Real-time Optimization (RTO) is the long waits for stationary operation. To overcome this issue, a hybrid RTO (HRTO) approach has been proposed in the literature in which a dynamic estimation is carried out, followed by economic optimization. Despite generally presenting a good performance, the original proposition is highly dependent on the availability of a dynamic process model, which could hinder or even make the design of a plant-wide HRTO infeasible. In this work, an HRTO strategy able to perform in the absence of a rigorous dynamic model is developed. A Hammerstein model structure is proposed to replace the dynamic model, and three HRTO architectures are introduced. These architectures are compared to the original HRTO and to a two-step static RTO without the steady-state detection stage. The approaches with less internal model variation present better performance over the approaches that are more sensitive to changes due to the arising of undesirable oscillatory behavior. This issue must be further investigated and overcome.

Keywords: Real-time Optimization, Parameter Estimation, Kalman Filter, Surrogate models, Williams-Otto Reactor.

1. Introduction

Real-time optimization (RTO) is a term commonly used in reference to a class of optimization strategies that uses real-time measurements in an optimization feedback loop in order to overcome uncertainties and achieve process optimality under a certain set of constraints (Darby et al., 2011). One of the most accepted classifications of RTO methods divide them on how the measurement is used to adapt the optimization problem. In the implicit methods, the adaption is carried out directly on the input variables such as in Self-Optimizing Control (Skogestad, 2003), Necessary Condition of Optimality (NCO) tracking (Srinivasan et al., 2003), and Extremum-Seeking Control (Morosanov, 1957). Within the class of explicit methods, an additional subdivision considers whether the adaption is carried out in the process model or in the optimization problem itself. In the first subclass, it is assumed that the main source of uncertainty is parametric such as in the two-step approach (Jang et al., 1987), in which a static

parameter estimation problem is solved before an economic optimization. In the second one, it is assumed that the source of uncertainty is due to structural mismatch between the process model and plant. Hence, modifiers are added to the optimization problem in order to match the Karush-Kuhn-Tucker (KKT) conditions of the model-based problem to the real plant problem, such as in ISOPE (Gao and Engell, 2005), Constraint Adaption (Marchetti et al., 2008) and Modifier Adaption (Chachuat et al., 2008).

The so-called two-step approach is the most widespread RTO method in industry (Câmara et al., 2016). However, the long waits for steady condition, especially in frequently disturbed or slow dynamic processes, is one of the main drawbacks of the strategy (Darby et al., 2011). In this context, two different research groups proposed to replace the static parameter estimation with a dynamic parameter estimation, such as the Extended Kalman Filter, giving origin to the Hybrid Real-Time Optimization (HRTO) approach (Matias and Le Roux, 2018; Krishnamoorthy et al., 2018). Despite generally presenting satisfactory performance results, the HRTO depends on the availability of a rigorous dynamic process model, structurally similar to the static process model, which is costly, time-consuming, and could even make the design of a plant-wide HRTO infeasible. Herein, the term “rigorous dynamic process model” is used in reference to a dynamic phenomenological model based on first principles of mass, energy, and momentum balances. In order to overcome this drawback, in this work, three HRTO architectures that are able to perform in the absence of a rigorous dynamic model are developed and compared.

2. Hammerstein model structure

It is assumed that an adequate static model of the plant is available:

$$\begin{aligned} 0 &= F(x^s, u, \theta) \\ y^s &= G(x^s, u, \theta) \end{aligned} \quad (1)$$

in which, $x^s \in R^{n_x}$ are the state variables in stationary state, $y^s \in R^{n_y}$ are the measured variables in stationary state, $u \in R^{n_u}$ are the input variables and $\theta \in R^{n_\theta}$ are the vector of parameters and unmeasured disturbances on which uncertainty occurs. In addition, the functions $F: R^{n_x} \times R^{n_u} \times R^{n_\theta} \rightarrow R^{n_x}$ and $G: R^{n_x} \times R^{n_u} \times R^{n_\theta} \rightarrow R^{n_y}$ are the static mapping of the states and the outputs related to the inputs and parameters, respectively. For the sake of simplifying notation, this model will be represented as $y^s = y^s(u, \theta)$.

A general Hammerstein structure consists of a static mapping of the inputs and parameters into the output variables and a linear dynamic model to make the transition from the current state to the mapped steady state. Considering a discrete linear dynamic model of the type of autoregressive with exogenous input (ARX) to approximate the plant dynamics, the Hammerstein model is given by:

$$y_k^h = A_k y_{k-1}^h + B_k y_{k-1}^s(u_{k-1}, \theta_{k-1}) \quad (2)$$

in which, $y_k^h \in R^{n_y}$ is the vector of output variables provided by the Hammerstein model at instant k . In addition, the squared matrices A_k and B_k gather the dynamic parameters to be identified from process data during operation, but can also be kept constant. Here, two models are proposed which exploit this structure, they are:

- i. model 1: $B_k = (I - A_k)$;
- ii. model 2: A_k and B_k are independent.

It is noteworthy that model 1 presents the property to match the stationary conditions of the static model, and it has half of the number of dynamic parameters to be identified compared to model 2.

3. HRTO architectures

The proposed HRTO architectures based on the Hammerstein model presented in Section 2 are described below:

- **HRTO-HEKF-fixed**: the adjustable parameters are estimated by an Extended Kalman Filter (EKF) layer that considers the Hammerstein model as its internal dynamic model, named Hammerstein EKF (HEKF), wherein the dynamic matrices of the model are kept constant. The estimated parameters are then kept fixed in a subsequent static optimization layer;
- **HRTO-HEKF-adaptive**: the adjustable parameters and the dynamic matrices of the Hammerstein model are simultaneously estimated by the HEKF layer;
- **HRTO-HEKF-RELS**: the adjustable parameters are estimated by the HEKF layer, but the dynamic matrices of the Hammerstein model are estimated by a Recursive Extended Least Squares (RELS) estimator with a forgetting factor.

These approaches are compared with the following architectures:

- **HRTO-EKF**: the original HRTO, in which the true dynamic model of the process is known and the adjustable parameters are estimated by an EKF layer;
- **RTO-LSE**: the original two-step RTO, in which the adjustable parameters are estimated by a nonlinear Least Squares Estimator (LSE) subjected to the static model. In this approach, no steady-state detection is performed, so the LSE runs regardless of the dynamic nature of the data.

3.1. Hammerstein EKF algorithm

The algorithm for the parameter estimation in the EKF framework is very well described in the Appendix of Krishnamoorthy et al. (2018). Using their nomenclature, the application for Hammerstein models gives the following augmented *a priori* state estimation:

$$\begin{bmatrix} y_{k|k}^h & \theta_{k|k} \end{bmatrix} = \begin{bmatrix} A_k y_{k-1}^h + B_k y_{k-1}^s(u_{k-1}, \theta_{k-1}) & \theta_{k|k-1} \end{bmatrix} + \begin{bmatrix} w_{y,k} & w_{\theta,k} \end{bmatrix} \quad (2)$$

in which, $w_{y,k}$ and $w_{\theta,k}$ are artificial zero-mean noise, so that $w_{y,k} \sim N(0, Q_y)$ and $w_{\theta,k} \sim N(0, Q_\theta)$. Therefore, the process noise covariance matrix is augmented as

$\tilde{Q} = \text{diag}([Q_y, Q_\theta])$. In the HRTO-HEKF-adaptive the elements of the dynamic matrices are estimated simultaneously to the adjustable parameters, so the columns of these matrices are stacked up and concatenated with vector θ .

3.2. RELS algorithm

The algorithm of the RELS with forgetting factor α is very well presented in Rodríguez-Blanco et al. (2017). Using their nomenclature, the application for the Hammerstein models are $\hat{\phi}_k = y_k^h - y_k^s$, $\varphi_{k-1}^T = y_{k-1}^h - y_{k-1}^s$ and $\hat{\beta}_k = (A_k)^T$ for model 1 and $\phi_k = (y_k^h)^T$, $\varphi_{k-1}^T = \left[(y_{k-1}^h)^T, (y_{k-1}^s)^T \right]$ and $\hat{\beta}_k = \left[(A_k)^T, (B_k)^T \right]^T$ for model 2 and the initial covariance matrix of the prediction error is given by $\Sigma_0 = I/\alpha$.

4. Case Study: The Williams-Otto Reactor

The Williams-Otto Reactor with three reactions was exploited as a case study. It is a classical benchmark for real-time optimization studies, where several phenomena can be observed, such as inverse response and change of the gain sign. Its equations, parameter values, and notations can be found in Forbes and Marlin (1996).

In this study, we consider X_C as an unmeasured variable, so the set of measured variables are $y = [X_A, X_B, X_E, X_P, X_G]^T$, the degrees of freedom of the optimizer are $u = [F_B, T_R]^T$ and the set of adjustable parameters were in fact two unmeasured disturbances $\theta = [F_A, W]^T$. The kinetic parameters were considered fixed and perfectly known. To initialize the Hammerstein models, an offline identification problem was run around a nominal point, $u_n = [2, 70]^T$ in $[kg/s, ^\circ C]$ and $\theta_n = [1.8275, 2105]^T$ in $[kg/s, kg]$, distant from the nominal optimum point, $u^* = [4.2, 85.9]^T$ in $[kg/s, ^\circ C]$. Finally, a scenario of parameter variation was designed to test the approaches. For the approach HRTO-HEKF-RELS, two values of the forgetting factor α were tested, 0.5 and 0.99. A sampling time of 60s was considered in a simulation window of $N = 450$ sampling times in which a decreasing ramp was subjected to the reactor holdup, W , and a pulse disturbance was subjected to the feed flow rate of A , F_A , as shown in Figure 1.

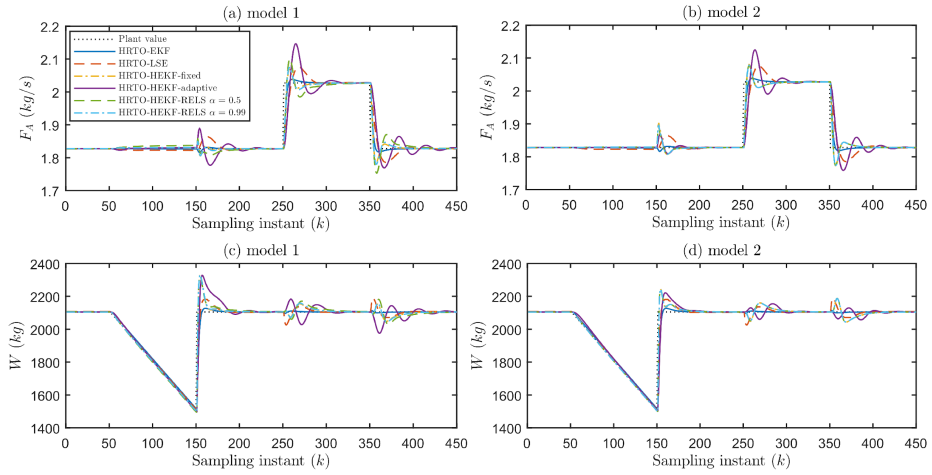


Figure 1: Performance of the different HRTO architecture over a scenario of parametric variation: (a) estimation of F_A with HEKF and RELS using model 1; (b) estimation of F_A with HEKF and RELS using model 2; (c) estimation of W with HEKF and RELS using model 1; (d) estimation of W with HEKF and RELS using model 2.

In order to compare all approaches in this scenario, a normalized Mean Squared Error (nMSE) for each variable is shown in Table 1.

Table 1: nMSE comparison between HRTO architectures

Variables	HRTO-EKF	RTO-LSE	model 1				model 2			
			HRTO-HEKF		HRTO-HEKF-RELS		HRTO-HEKF		HRTO-HEKF-RELS	
			fixed	adaptive	$\alpha = 0.5$	$\alpha = 0.99$	fixed	adaptive	$\alpha = 0.5$	$\alpha = 0.99$
X_A	$4.6 \cdot 10^{-6}$	$3.4 \cdot 10^{-6}$	$5.8 \cdot 10^{-6}$	$8.4 \cdot 10^{-5}$	$6.0 \cdot 10^{-6}$	$5.8 \cdot 10^{-6}$	$5.9 \cdot 10^{-6}$	$6.6 \cdot 10^{-5}$	$5.9 \cdot 10^{-6}$	$5.9 \cdot 10^{-6}$
X_B	$3.8 \cdot 10^{-6}$	$4.1 \cdot 10^{-6}$	$4.8 \cdot 10^{-6}$	$3.2 \cdot 10^{-5}$	$4.7 \cdot 10^{-6}$	$4.8 \cdot 10^{-6}$	$4.7 \cdot 10^{-6}$	$3.0 \cdot 10^{-5}$	$4.7 \cdot 10^{-6}$	$4.7 \cdot 10^{-6}$
X_E	$2.4 \cdot 10^{-6}$	$3.4 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$	$1.7 \cdot 10^{-5}$	$3.2 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$	$1.6 \cdot 10^{-5}$	$3.2 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$
X_p	$6.9 \cdot 10^{-7}$	$2.5 \cdot 10^{-6}$	$8.5 \cdot 10^{-7}$	$4.6 \cdot 10^{-6}$	$8.5 \cdot 10^{-7}$	$8.5 \cdot 10^{-7}$	$9.1 \cdot 10^{-7}$	$5.1 \cdot 10^{-6}$	$9.1 \cdot 10^{-7}$	$9.1 \cdot 10^{-7}$
X_G	$2.3 \cdot 10^{-6}$	$2.4 \cdot 10^{-6}$	$3.3 \cdot 10^{-6}$	$1.8 \cdot 10^{-5}$	$3.2 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$	$2.4 \cdot 10^{-5}$	$3.2 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$
J	$4.1 \cdot 10^{-2}$	$1.8 \cdot 10^{-1}$	$4.8 \cdot 10^{-2}$	$2.8 \cdot 10^{-1}$	$5.4 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$	$5.5 \cdot 10^{-2}$	$3.0 \cdot 10^{-1}$	$5.4 \cdot 10^{-2}$	$5.4 \cdot 10^{-2}$
F_A	$1.7 \cdot 10^{-4}$	$3.4 \cdot 10^{-4}$	$1.8 \cdot 10^{-4}$	$6.0 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$	$2.1 \cdot 10^{-4}$	$4.9 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$	$2.0 \cdot 10^{-4}$
W	$5.9 \cdot 10^{-1}$	$7.0 \cdot 10^{-1}$	$7.5 \cdot 10^{-1}$	1.4	$8.5 \cdot 10^{-1}$	$7.6 \cdot 10^{-1}$	$6.3 \cdot 10^{-1}$	$9.8 \cdot 10^{-1}$	$6.3 \cdot 10^{-1}$	$6.2 \cdot 10^{-1}$

The nMSE for each variable i , using the true plant value as reference, is given by:

$$nMSE_i = \sum_{k=1}^N \frac{\hat{y}_{i,k} - y_{i,k}^p}{y_{i,k}^p} \quad (2)$$

in which, the hat emphasis $\hat{(\cdot)}$ and the superscript $(\cdot)^p$ represent the estimated value and the true plant value, respectively.

It is straightforward from Figure 1 that all approaches were able to estimate the true value of the parameters after a stabilization period, which confirms the potential of the Hammerstein approaches to substitute the rigorous dynamic model in the architecture of the HRTO, opening up a vast range of possibilities of applications. Surprisingly, the RTO-LSE presented a good performance in parameter estimation during transient regions, suggesting that HRTO could be applied directly to classic RTO structures, just by removing the steady-state detection stage and increasing the running frequency of the optimization loop. However, this approach is the most costly one among the other structures analyzed, as shown in Table 2, it is almost 11.5 times more costly than EKF and 23 times more costly than the proposed HEKF. This preliminary conclusion about the RTO-LSE should be verified by testing with many other case studies.

Table 2: Average time consumption for each stage in loop (ms)

Stage	HRTO-EK F	RTO-LSE	model 1				model 2			
			HRTO-HEKF		HRTO-HEKF-RELS		HRTO-HEKF		HRTO-HEKF-RELS	
			fixed	adaptive	$\alpha = 0.5$	$\alpha = 0.95$	fixed	adaptive	$\alpha = 0.5$	$\alpha = 0.95$
Plant Optimization	2.04	2.08	2.08	2.08	2.25	2.05	2.07	2.08	2.40	2.15
Estimation	1.94	22.94	0.74	0.98	0.95	0.86	0.69	1.03	0.98	0.82
Model Optimization	11.92	12.81	11.95	12.03	12.95	12.03	11.99	12.09	13.69	12.23
Total simulation loop	16.95	39.31	15.89	16.35	17.35	16.07	15.84	16.49	18.35	16.31

In general, approaches using model 1 presented a greater adaption capability over approaches using model 2, due to the lesser number of parameters to estimate. However, this capability does not necessarily reflect better parameter estimation; in fact approaches using model 2 presented a greater overall accuracy on parameter estimation. In addition, the approach HRTO-HEKF-RELS with $\alpha = 0.99$ was very similar to the HRTO-HEKF-fixed for both models, which is a reflection of the low adaptability capacity of the RELS strategy with high value of the forgetting factor. For the approaches HRTO-HEKF-RELS with $\alpha = 0.5$ and HRTO-HEKF-adaptive, the higher adaptability capacity was able to reduce the nMSE for the output variables, but presented a worst parameter estimation accuracy than the HRTO-HEKF-fixed, even though the models used in this approach were identified far from the operating point and in a fixed parameter scenario.

It is noteworthy that the approach HRTO-HEKF-adaptive presented an undesirable oscillatory behavior, which appears due to the arising of oscillatory modes in the Hammerstein models during operation. This effect should be avoided and further investigated.

5. Conclusion

In this work, different HRTO architectures were proposed based on the use of a Hammerstein model structure that combines the available static process model with a linear ARX identified from past data to provide approximate dynamics. Therefore, the requirement of availability of a rigorous dynamic process is removed from the original proposition of the HRTO, enabling it to be used in a large range of applications. The proposed methodologies showed satisfactory performances in parameter estimation and adequate computational costs. However, they presented no resources to prevent the arising of unstable and undesirable oscillatory modes on the Hammerstein model during the adaptive process, this should be further investigated.

References

- M. M. Câmara, A. D. Quelhas, J. C. Pinto, 2016. Performance evaluation of real industrial RTO systems. *Processes* 4 (4), 1–20.
- B. Chachuat, B. Srinivasan, D. Bonvin, 2008. Model parameterization tailored to real-time optimization. In: B. Braunschweig, X. Joulia (Eds.), 18th European Symposium on Computer Aided Process Engineering. Vol. 25 of *Computer Aided Chemical Engineering*. Elsevier, pp. 1 – 13.
- M. L. Darby, M. Nikolaou, J. Jones, D. Nicholson, 2011. Rto: An overview and assessment of current practice. *Journal of Process Control* 21 (6), 874 – 884.
- J. Forbes, T. Marlin, 1996. Design cost: a systematic approach to technology selection for model-based realtime optimization systems. *Computers & Chemical Engineering* 20 (6), 717 – 734.
- W. Gao, S. Engell, 2005. Iterative set-point optimization of batch chromatography. *Computers & Chemical Engineering* 29 (6), 1401 – 1409, selected Papers Presented at the 14th European Symposium on Computer Aided Process Engineering.
- S. Jang, B. Joseph, H. Mukai, 1987. On-line optimization of constrained multivariable chemical processes. *AIChE Journal* 33 (1), 26–35.
- D. Krishnamoorthy, B. Foss, S. Skogestad, 2018. Steady-state real-time optimization using transient measurements. *Computers & Chemical Engineering* 115, 34 – 45.
- A. Marchetti, B. Chachuat, D. Bonvin, 2008. Real-time optimization via adaptation and control of the constraints. Vol. 25 of *Computer Aided Chemical Engineering*. Elsevier, pp. 393 – 398.
- J. O. Matias, G. A. Le Roux, 2018. Real-time optimization with persistent parameter adaptation using online parameter estimation. *Journal of Process Control* 68, 195 – 204.
- I. Morosanov, 1957. Method of extremum control. *Automatic and Remote Control* 18, 1077–1092.
- T. Rodríguez-Blanco, D. Sarabia, J. Pitarch, C. de Prada, 2017. Modifier adaptation methodology based on transient and static measurements for rto to cope with structural uncertainty. *Computers & Chemical Engineering* 106, 480 – 500.
- S. Skogestad, 2003. Self-optimizing control: From key performance indicators to control of biological systems. In: B. Chen, A. W. Westerberg (Eds.), *Process Systems Engineering 2003*. Vol. 15 of *Computer Aided Chemical Engineering*. Elsevier, pp. 340 – 345.
- B. Srinivasan, D. Bonvin, E. Visser, S. Palanki, 2003. Dynamic optimization of batch processes: Ii. role of measurements in handling uncertainty. *Computers & Chemical Engineering* 27 (1), 27 – 44.

Study of alternative configurations for methanol purification

Stefania Moioli*, Laura A. Pellegrini

GASP, Group of Advanced Separation Processes and GAS Processing, Dipartimento di Chimica, Materiali e Ingegneria Chimica "Giulio Natta", Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
stefania.moioli@polimi.it

Abstract

Methanol is a clean fuel and is considered an important chemical feedstock. It is usually produced by catalytic conversion and can be produced from carbon dioxide and thus could significantly contribute to the reduction of direct CO₂ emissions to the atmosphere. Despite a number of technologies have been developed for its synthesis, post-production treatment is fundamental for obtaining a high purity methanol stream. Purification is generally carried out by distillation, with significant energy consumption.

The raw product exiting from the methanol production section is mainly composed of methanol and water, with traces of other components and is fed to the purification section. This section is composed of two separation units. The first unit aims at removing the low boiling impurities which exit from the top and the second one, the most energy demanding, treats the heavier stabilized methanol-rich stream to increase its concentration in methanol for obtaining a product with high purity.

This work focuses on the optimization of the process, with the study of possible modifications for obtaining energy saving and the selection of the best one on the basis of a techno-economic analysis.

Simulations have been carried out by using the commercial software ASPEN Plus[®]. The thermodynamic method based on the Non-Random-Two-Liquid (NRTL) model, suitable for this system and for the low pressure conditions, has been employed and a rate-based approach has been considered for simulating the columns.

The alternative configurations favour a reduction of the influence of the methanol purification section on the total costs of the production plant.

Keywords: methanol, distillation, energy saving, configurations, economic analysis.

1. Introduction

Methanol can be produced from fossil or renewable sources and is employed for energy production in the marine, automotive and electricity sectors (Ptasinski et al., 2002), in addition to be a fundamental commodity in the chemical industry, used for plastics, paints and cosmetics (Methanol Institute, 2020). The chemist George Andrew Olah was a great supporter of the "methanol economy" for the capability of this compound to provide an efficient mean to store energy and of being used as a convenient fuel as well as a raw material for synthetic hydrocarbons and their products, thinking of it as a possible substitute to natural oil and gas (Olah, 2004, 2005, 2013; Olah et al., 2018).

Methanol is mainly produced from syngas or by other routes starting from methane (Lange, 1997, 2001). Generally, most of the production comes from natural gas, though about 10% originates from distillation of heavy cuts of oil or from coal in locations where

natural gas is not available or cannot be easily supplied. Its production is usually obtained by catalytic conversion also starting from the CO₂ removed from other streams, therefore it could significantly contribute to the reduction of direct greenhouse gas emissions to the atmosphere (Van-Dal and Bouallou, 2013). The methanol price is characterized by fluctuations depending on the market conditions and on the market locations, so that values ranged from 500 \$/ton to 190 \$/ton in the last three years (MMSA, 2020). For this reason, a low-cost production is fundamental. In particular, post-production treatment is needed for obtaining a high purity methanol stream and is generally carried out by distillation, with relevant energy consumptions (Zhang et al., 2010).

In the literature, many works focused on the production path of methanol. Some researchers proposed different kinetic models for the synthesis of this substance, considering different catalysts, while other authors studied the overall scheme of production (Zhang et al., 2016) starting from different raw materials (Rivarolo et al., 2016), including sugarcane bagasse (Renó et al., 2011), CO₂ removed from other industrial processes (Abdelaziz et al., 2017; Atsonios et al., 2016; Kourkoumpas et al., 2016; Pérez-Fortes et al., 2016; Szima and Cormos, 2018) and also renewable hydrogen (Galindo Cifre and Badr, 2007) or using different types of energy as the solar one (Kim et al., 2011). Only few papers focused on a reduction of the economics of the process through the analysis and the comparison of different schemes (Shahandeh et al., 2014; Shahandeh et al., 2015). Doulas and Hoadley (Douglas and Hoadley, 2006) performed a thermal and economic comparison of schemes with two and three columns, with enhanced heat integration. Sun et al. (Sun et al., 2012) considered a 5-column heat integrated distillation scheme and compared the obtained savings with the 4-column scheme though with a higher capital cost and complexity of operating many distillation columns. Kiss et al. (Kiss et al., 2016) proposed the addition of a stripping unit with wet hydrogen to remove CO and CO₂ from the water-methanol mixture. Many works (Abdelaziz et al., 2017; Atsonios et al., 2016; Pérez-Fortes et al., 2016; Szima and Cormos, 2018; Van-Dal and Bouallou, 2013), then, considered a scheme based on the presence of a flash and a distillation column, though without focusing on the optimization of the purification section. This type of scheme has been considered in the present work.

The aim of this paper is the study of different configurations for the section of purification of the raw methanol stream in order to determine the one which reduces the investment costs and the operating costs, mainly due to the energy consumptions on the basis of a techno-economic analysis.

2. Methodology

The mixture considered in this paper contains mainly methanol and water (the ratio methanol / water is equal to 1.96), with traces of ethanol and other light gases residual from the previous steps of reaction. The purification section must produce a methanol-rich stream of AA grade, according to the International Methanol Producers and Consumers Association (IMPCA), for which a minimum weight fraction of methanol in the product stream of 0.9985 is required, with a content of water lower than 0.1 % wt. As additional specification, in this work a recovery of the methanol entering the purification section of at least 90% has been considered.

The simulations for the design and the evaluation of the performances of the different schemes have been carried out in ASPEN Plus[®], selecting for the thermodynamic description of the system a γ/ϕ method based on the Non-Random Two Liquids (NRTL) theory (Renon and Prausnitz, 1968), which has been found accurate for the representation of the phase equilibria of the components of the feed stream (Allocca, 2020). The tool

Sensitivity provided within the process simulator has been employed for analyzing the influence of the main parameters of the column.

All the obtained results have been used as input for the economic analysis, which includes the evaluation of both investment and operating costs, performed on the basis of the *Preliminary Design Estimate* method and the average operating costs reported in Turton et al. (Turton et al., 2012) and the optimal configuration has then been identified.

3. The considered schemes

In this section the three considered configurations are described. The first one (*Scheme A*), taken as reference, is the simplest one while the other ones, *Scheme B* and *Scheme C*, are characterized by the addition of a heat exchanger and also of a splitter in order to obtain economic advantages.

Indeed, since the reboiler duty represents the main operating cost of this plant and energy management is recognized as a key point in the design of chemical processes (Luyben, 2020), *Scheme B* and *Scheme C* have been analysed and optimized in order to minimize both investment costs and operating costs.

3.1. Scheme A

One single distillation column cannot perform the separation for obtaining a product stream with all the above-mentioned specifications, mainly because of the presence of the residual gases from reaction in the feed stream. Therefore, a flash or another distillation column are usually employed to remove this incondensable species upstream of the column used for the separation of methanol and water.

The raw methanol stream (stream 1) is fed to the flash unit, where most of the lightest components are removed and exit from the top (stream 2). The liquid stream (stream 3), rich in methanol and water, is fed to the distillation column, provided with a partial condenser and a partial reboiler (Figure 1).

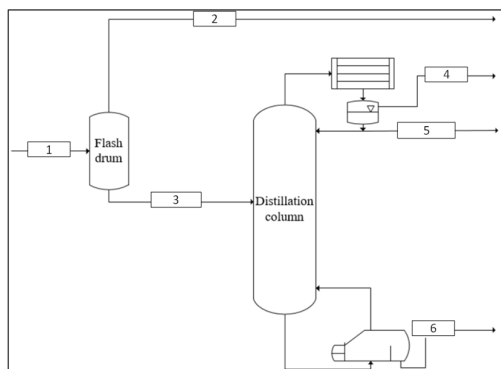


Figure 1. Base scheme (*Scheme A*) for the purification section of the methanol production plant.

The methanol-rich stream (stream 2), with the required specifications, is obtained at the top of the column as liquid distillate, while a water-rich stream with a small amount of ethanol exits from the bottom of the unit. The vapor distillate is composed of the incondensable gases which have not been removed in the previous flash. The pressure of the feed, equal to 2 bar, is reduced to 1.25 bar to separate most of the gases from the liquid in the flash, before entering the distillation column, operated at atmospheric pressure.

The characteristics of the column (height, feed stage, bottom to feed ratio) have been optimized with the aim of fulfilling the required specifications while maintaining low investment and operating costs.

3.2. Scheme B

A common method for energy saving consists in pre-heating the stream fed to the distillation column with the bottom product stream exiting the reboiler in a process-process heat exchanger. The water-rich stream is fed to the hot side of the unit for heat transfer, where it is cooled, transferring its heat content to the cold fluid, the liquid exiting from the flash unit to be warmed.

The variation in the feed temperature influences the profiles inside the distillation column, so that for the same column height of *Scheme A*, the optimal bottom to feed ratio and feed stage have been determined.

3.3. Scheme C

A splitter for dividing the flash liquid product stream before the heat exchanger can be added to achieve a reduction in the reboiler duty (Deshmukh et al., 2005) because of the higher temperature obtained for the preheated stream. In particular, the liquid exiting from the flash (stream 3 in Figure 1) is split into two streams, with the same composition and with the flowrate depending on the splitting fraction. One part of the stream is directly fed to the distillation column without variations in its temperature, while the other part is fed to a process-process heat exchanger similar to the one of *Scheme B* for pre-heating and is then fed to the distillation column at a different height. When the splitting fraction, defined as the fraction of stream directly fed to the distillation tower, is equal to 0, the process layout of *Scheme B* applies, with all the stream being pre-heated before being separated in the distillation column, while when the splitting fraction is equal to 1 the base scheme (*Scheme A*) is considered, with no pre-heating.

The splitting fraction and the feed stages of the two streams entering the column have been optimized for *Scheme C*. To determine the values of these variables an iterative procedure has been applied, since the result of each one depends on the value of the other ones.

4. Results

Table 1 reports the results of the economic evaluation of the plant, as for the investment costs and for the operating costs. When compared to *Scheme A*, *Scheme C* results in the largest savings, with an 11.23% reduction in capital costs and a 13.32% reduction in operating expenses.

Table 1. % variation in investment and operating costs obtained with *Scheme B* and with *Scheme C* if compared to *Scheme A*.

	% variation in investment costs	% variation in operating costs
<i>Scheme B</i>	+1.93%	-1.24%
<i>Scheme C</i>	-11.23%	-13.32%

For *Scheme B*, the exchange of heat between the hot stream, available at a pressure slightly above the atmospheric one and at about 100°C, and the cold stream at 40°C makes the overall stream fed to the column at a temperature higher than the one in *Scheme A*. However, because the flowrate of the water-rich stream exiting the reboiler is much lower

than the one of the cold stream, this latter reaches only the temperature of 55°C if a $\Delta T_{\text{approach}}$ of 5°C is considered.

As for *Scheme C*, the optimal result is for a splitting fraction equal to 0.9, which allows for obtaining the lowest reboiler duty. The resulting profile in the process-process heat exchanger is reported in Figure 2. Due to the lower flowrate of stream to be heated, the temperature increases much more than the one of the hot stream in *Scheme B*. The hot stream enters the column with a vapour fraction of about 20%, reducing the energy requirements for the reboiler.

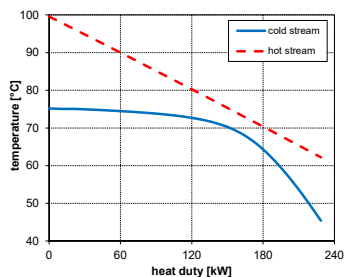


Figure 2. Temperature profiles for the hot and cold streams in the process-process heat exchanger for *Scheme C*.

5. Conclusions

This work focused on the study of a methanol purification plant, being methanol a relevant clean fuel and a chemical feedstock. The analysis of different configurations for the section of purification of the raw methanol stream has been carried out in order to obtain a methanol product with the specifications required by IMPCA while minimizing the costs. Three different schemes have been considered, a base case with a flash unit and a distillation column and two more complex cases, characterized by the pre-heating of all the stream fed to distillation (*Scheme B*) or by the splitting of this stream to have only a partial pre-heating before entering the column (*Scheme C*). On the basis of the techno-economic analysis carried out, *Scheme C* resulted the most performing one, allowing for both a reduction of the energy consumption and of the investment costs.

References

- O.Y. Abdelaziz, W.M. Hosny, M.A. Gadalla, F.H. Ashour, I.A. Ashour, C.P. Hulteberg, 2017, Novel process technologies for conversion of carbon dioxide from industrial flue gas streams into methanol, *Journal of CO₂ Utilization*, 21, 52-63.
- D. Allocca, 2020, Biomethanol production plant: design and optimization of the purification section. Politecnico di Milano.
- K. Atsonios, K.D. Panopoulos, E. Kakaras, 2016, Thermocatalytic CO₂ hydrogenation for methanol and ethanol production: Process improvements, *International Journal of Hydrogen Energy*, 41, 792-806.
- B.F. Deshmukh, R.K. Malik, S. Bandyopadhyay, 2005, Efficient feed preheat targeting for distillation by feed splitting, in: Puigjaner, L., Espuña, A. (Eds.), *Computer Aided Chemical Engineering*. Elsevier, pp. 751-756.
- A.P. Douglas, A.F.A. Hoadley, 2006, A process integration approach to the design of the two- and three-column methanol distillation schemes, *Applied Thermal Engineering*, 26, 338-349.
- P. Galindo Cifre, O. Badr, 2007, Renewable hydrogen utilisation for the production of methanol, *Energy Conversion and Management*, 48, 519-527.

- J. Kim, C.A. Henao, T.A. Johnson, D.E. Dedrick, J.E. Miller, E.B. Stechel, C.T. Maravelias, 2011, Methanol production from CO₂ using solar-thermal energy: process development and techno-economic analysis, *Energy & Environmental Science*, 4, 3122-3132.
- A.A. Kiss, J.J. Pragt, H.J. Vos, G. Bargeman, M.T. de Groot, 2016, Novel efficient process for methanol synthesis by CO₂ hydrogenation, *Chemical Engineering Journal*, 284, 260-269.
- D.S. Kourkoumpas, E. Papadimou, K. Atsonios, S. Karellas, P. Grammelis, E. Kakaras, 2016, Implementation of the Power to Methanol concept by using CO₂ from lignite power plants: Techno-economic investigation, *International Journal of Hydrogen Energy*, 41, 16674-16687.
- J.-P. Lange, 1997, Perspectives for Manufacturing Methanol at Fuel Value, *Industrial & Engineering Chemistry Research*, 36, 4282-4290.
- J.-P. Lange, 2001, Methanol synthesis: a short review of technology improvements, *Catalysis Today*, 64, 3-8.
- W.L. Luyben, 2020, Energy management in distillation preheat systems, *Chemical Engineering and Processing - Process Intensification*, 156, 108074.
- Methanol Institute, 2020, <https://www.methanol.org/>
- MMSA, 2020, <https://www.methanolmsa.com/methanol/>
- G.A. Olah, 2004, After Oil and Gas: Methanol Economy, *Catalysis Letters*, 93, 1-2.
- G.A. Olah, 2005, Beyond Oil and Gas: The Methanol Economy, *Angewandte Chemie International Edition*, 44, 2636-2639.
- G.A. Olah, 2013, Towards Oil Independence Through Renewable Methanol Chemistry, *Angewandte Chemie International Edition*, 52, 104-107.
- G.A. Olah, T. Mathew, A. Goeppert, G.K. Surya Prakash, 2018, Difference and Significance of Regenerative Versus Renewable Carbon Fuels and Products, *Topics in Catalysis*, 61, 522-529.
- M. Pérez-Fortes, J.C. Schöneberger, A. Boulamanti, E. Tzimas, 2016, Methanol synthesis using captured CO₂ as raw material: Techno-economic and environmental assessment, *Applied Energy*, 161, 718-732.
- K.J. Ptasinski, C. Hamelinck, P.J.A.M. Kerckhof, 2002, Exergy analysis of methanol from the sewage sludge process, *Energy Conversion and Management*, 43, 1445-1457.
- M.L.G. Renó, E.E.S. Lora, J.C.E. Palacio, O.J. Venturini, J. Buchgeister, O. Almazan, 2011, A LCA (life cycle assessment) of the methanol production from sugarcane bagasse, *Energy*, 36, 3716-3726.
- H. Renon, J.M. Prausnitz, 1968, Local compositions in thermodynamic excess functions for liquid mixtures, *AIChE Journal*, 14, 135-144.
- M. Rivarolo, D. Bellotti, L. Magistri, A.F. Massardo, 2016, Feasibility study of methanol production from different renewable sources and thermo-economic analysis, *International Journal of Hydrogen Energy*, 41, 2105-2116.
- H. Shahandeh, J. Ivakpour, N. Kasiri, 2014, Feasibility study of heat-integrated distillation columns using rigorous optimization, *Energy*, 74, 662-674.
- H. Shahandeh, M. Jafari, N. Kasiri, J. Ivakpour, 2015, Economic optimization of heat pump-assisted distillation columns in methanol-water separation, *Energy*, 80, 496-508.
- J. Sun, F. Wang, T. Ma, H. Gao, P. Wu, L. Liu, 2012, Energy and exergy analysis of a five-column methanol distillation scheme, *Energy*, 45, 696-703.
- S. Szima, C.-C. Cormos, 2018, Improving methanol synthesis from carbon-free H₂ and captured CO₂: A techno-economic and environmental evaluation, *Journal of CO₂ Utilization*, 24, 555-563.
- R. Turton, R.C. Bailie, W.B. Whiting, J.A. Shaeiwitz, D. Bhattacharyya, 2012, *Analysis, Synthesis & Design of Chemical Processes*. 4th ed. Prentice Hall International Series in the Physical and Chemical Engineering Sciences, Upper Saddle River, New Jersey, USA.
- É.S. Van-Dal, C. Bouallou, 2013, Design and simulation of a methanol production plant from CO₂ hydrogenation, *Journal of Cleaner Production*, 57, 38-45.
- C. Zhang, K.-W. Jun, G. Kwak, Y.-J. Lee, H.-G. Park, 2016, Efficient utilization of carbon dioxide in a gas-to-methanol process composed of CO₂/steam-mixed reforming and methanol synthesis, *Journal of CO₂ Utilization*, 16, 1-7.
- J. Zhang, S. Liang, X. Feng, 2010, A novel multi-effect methanol distillation process, *Chemical Engineering and Processing: Process Intensification*, 49, 1031-1037.

Power-to-OME₁ via Direct Oxidation of Methanol: Process Design and Global Flowsheet Optimization

Dominik Bongartz^{a,*}, Jannik Burre^a, Anita L. Ziegler^a and Alexander Mitsos^{b,a,c}

^aProcess Systems Engineering (AVT.SVT), RWTH Aachen University, Aachen, Germany

^bJARA-ENERGY, 52056 Aachen, Germany

^cEnergy Systems Engineering (IEK-10), Forschungszentrum Jülich, Jülich, Germany

*dominikbongartz@alum.mit.edu

Abstract

Dimethoxymethane (OME₁) is a potential blend component for diesel fuel that enables dramatic reductions in pollutant formation. With current technology, however, the production of OME₁ from renewable electricity would be less efficient than other power-to-fuel processes. We therefore present an alternative process based on a different synthesis route: direct oxidation of methanol. We use deterministic global optimization to maximize the exergy efficiency of the process. The problem can be solved globally despite the complexity of the flowsheet and the detailed phase equilibrium and enthalpy models used for most units. In contrast, only simple models are used for distillation. The effects of this simplification and that of heat integration are quantified. The optimized process has a higher exergy efficiency than the benchmark and eliminates all external heat demand.

Keywords: Design, global optimization, power-to-fuel, oxymethylene ether, methylal

1. Introduction

Alternative fuels are urgently needed for reducing the greenhouse gas and pollutant emissions of the transport sector. A promising candidate is dimethoxymethane (OME₁), which can be used as a compression ignition fuel itself with suitable additives or blended in diesel fuel. In either case, the formation of soot and nitrogen oxides can be drastically reduced compared to diesel fuel (Omari et al., 2017). OME₁ is also an intermediate in pathways for producing longer-chain oxymethylene ethers (Burger et al., 2010), which can directly replace diesel fuel with only minor engine modifications.

The current production of OME₁ is based on methanol and aqueous formaldehyde solution (Weidert et al., 2017). By combining such an OME₁ production process with formaldehyde production from methanol and methanol production from renewable hydrogen (H₂) and carbon dioxide (CO₂), OME₁ can be produced in a power-to-fuel process chain, which we call the *benchmark process chain* in the following. However, this benchmark process chain for converting H₂ and CO₂ to OME₁ has an exergy efficiency of only 73 %, in contrast to around 90 % for methane, methanol, or dimethyl ether (Bongartz et al., 2019). This lower efficiency can be attributed to the accumulation of losses in the three process steps. Among these, formaldehyde production has the lowest efficiency.

A number of alternative synthesis pathways for OME₁ that avoid a separate formaldehyde production process have been explored (Sun et al., 2019). Among these, the direct oxida-

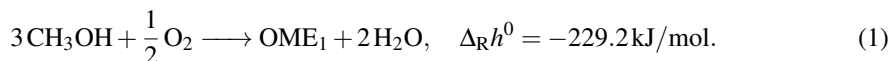
tion of methanol to OME₁ (Tatibouët, 1997) is promising for short-term implementation because heterogeneous catalysts with high conversion and selectivity have been developed. In this direct oxidation, formaldehyde is converted in-situ to OME₁ immediately after its formation over a bifunctional catalyst. Despite the considerable work on catalyst development, no optimized process has been presented and analyzed yet.

In this contribution, we thus design a power-to-fuel process for producing OME₁ from H₂ and CO₂ via direct oxidation of methanol. To achieve maximum efficiency, we aim at leveraging recent developments in global flowsheet optimization using our open-source optimizer MAiNGO (Bongartz et al., 2018). In the following, we first introduce the proposed process concept. Next, we describe the flowsheet optimization problem and its solution with MAiNGO. To ensure tractability of the global flowsheet optimization, only simple models are used for distillation columns, and no heat integration is considered during optimization. Finally, we quantify the impact of the simple distillation models and conduct pinch analysis to enable comparison with the benchmark process chain.

2. Process Concept

The proposed process concept consists of two sub-processes: conversion of H₂ and CO₂ to methanol, and conversion of methanol to OME₁ (see Fig. 1). For the first sub-process, multiple concepts have already been proposed. Herein, we slightly adapt the concept for methanol production of Van-Dal and Bouallou (2013): We omit the low-pressure flash drums for separating remaining traces of CO₂ from the produced methanol, because we do not expect CO₂ to impact the downstream reaction of methanol to OME₁.

In the second sub-process, the main reaction is



Reaction (1) occurs in the gas phase in fixed-bed reactors at ambient pressure and moderate temperatures (Sun et al., 2019). For the present analysis, we select the catalyst by Liu et al. (2008), which enables peak values of >90% for methanol conversion and selectivity to OME₁ at temperatures of 393 K to 423 K. Conversion increases with temperature whereas selectivity decreases. The main side products are methyl formate (MF) and formaldehyde (FA). Although the experiments were conducted with an O₂-N₂ mixture with 17 mol% of O₂, we assume that the performance is the same with air as oxidant.

After the reactor, gas-liquid separation is achieved through low-temperature cooling. The gas stream consists of N₂ and unreacted O₂. The liquid stream contains water, methanol, OME₁ as well as some MF and FA. Mixtures of water, methanol, OME₁, and FA exhibit several azeotropes (Kuhnert et al., 2006). The addition of MF, in contrast, does not introduce additional azeotropes according to model predictions (Deutz et al., 2018). Since MF is also the most volatile component, it is withdrawn as top product of column D2. The remaining mixture is similar to the one in OME₁ production from methanol and aqueous FA. In that context, Weidert et al. (2017) proposed a pressure-swing distillation to overcome the methanol-OME₁ azeotrope. The first column of their setup has a reactive section to convert leftover FA to OME₁, and a side product draw to remove methanol. Herein, we remove the side draw and instead obtain a methanol-water mixture as bottom product of D3 that is recycled to D1, since D1 already separates methanol from water.

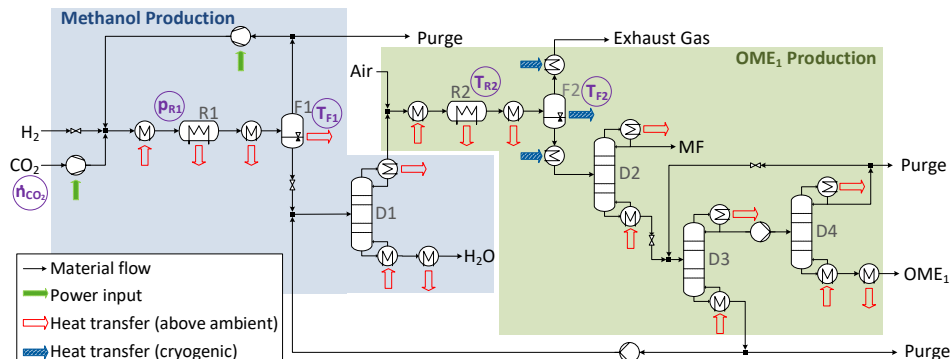


Figure 1: Process flow diagram for the production of OME₁ from H₂ and CO₂ via direct oxidation of methanol. Figure adopted from Bongartz (2020). Combustion of purge streams and the MF stream for energy recovery is not shown for simplicity.

3. Flowsheet Optimization

Given the flowsheet in Fig. 1, we aim at maximizing its exergy efficiency, which is a thermodynamically consistent measure for the quality of energy conversion processes. Such flowsheet optimization problems are nonconvex and should thus be solved with global optimization methods. However, these are computationally expensive. To ensure tractability, we use a three-stage procedure: First, we optimize the entire flowsheet globally using simple models for the most challenging units in the flowsheet, i.e., the distillation columns. Second, we use the results as initial values to optimize the columns locally using rigorous models. Third, we conduct pinch-based heat integration for the fixed optimized process.

3.1. Global Optimization with Simple Distillation Models

The global optimization of the flowsheet is described in detail by Bongartz (2020) and only key points are summarized here. The design variables are shown in circles in Fig. 1. Their allowable ranges are given in Table 1. The objective is the exergy efficiency

$$\eta_{\text{ex}} = \frac{\dot{n}_{\text{OME}_1} e_{\text{OME}_1} + \sum_{k \in \mathcal{Q}_{\text{out}}} \dot{E}_{Q,k}}{\dot{n}_{\text{H}_2} e_{\text{H}_2} + \dot{n}_{\text{CO}_2} e_{\text{CO}_2} + \sum_m \dot{W}_m + \sum_{k \in \mathcal{Q}_{\text{in}}} \dot{E}_{Q,k}}, \quad (2)$$

with \dot{n}_i and e_i denoting the molar flow rate and molar exergy of species i , respectively, $\dot{E}_{Q,k}$ the exergy of heat flow k , and \dot{W}_m the power consumption of compressor m . No internal heat integration is considered for simplicity. Instead, all heat flows that correspond to exergy leaving and entering units are assigned to the sets \mathcal{Q}_{out} and \mathcal{Q}_{in} , respectively.

The process is modeled with species molar flow rates and overall enthalpy flow rates. The balances are limited to species that occur in appreciable quantities in the corresponding parts of the flowsheet. Enthalpies are computed with the DIPPR-107 equation for heat capacity, the DIPPR-106 equation for enthalpy of vaporization, and the AspenPlus correlation for enthalpy of solution of Henry components. For the compressors, we use a fixed isentropic efficiency and an entropy model also based on the DIPPR-107 equation. In reactor R1, we enforce equilibrium of methanol formation and reverse water gas shift. In reactor R2, we correlate the yields of Liu et al. (2008) as a function of temperature

Table 1: Design variables for flowsheet optimization (adopted from Bongartz (2020)).

Symbol	Description	Unit	Range	Solution Point
\dot{n}_{CO_2}	CO ₂ inlet flow rate	kmol/s	[0, 1]	0.251
p_{R1}	Inlet pressure reactor R1	bar	[60, 70]	70.0
T_{F1}	Temperature flash drum F1	°C	[30, 110]	100
T_{R2}	Temperature reactor R2	°C	[120, 140]	124
T_{F2}	Temperature flash drum F2	°C	[-100, 25]	-94.6

using second degree polynomials. FA formation is neglected (i.e., counted as unreacted methanol), which is conservative because FA would be further converted to OME₁ in the reactive section of D1. Phase equilibrium is modeled with the extended Antoine equation and Henry’s constant correlation and solvent mixing rules from AspenPlus. In the second sub-process, we additionally use NRTL for the nonideal liquid mixture. For the columns D1–D4, we use the Underwood shortcut for the minimum reflux ratio. The actual reflux ratio is chosen 30 % higher for finite column height. The azeotropes at the top of columns D3 and D4 are introduced as pseudo-components as proposed by Vogelpohl (2002).

The problem is written in a reduced-space formulation (Bongartz and Mitsos, 2017), which is a hybrid between equation-oriented and sequential-modular formulations: The optimizer does not see all model variables, but rather only the degrees of freedom and few variables for decoupling the equation systems, e.g., those in tear streams. The other variables are computed sequentially from the model equations. Flash calculations are handled as described by Bongartz and Mitsos (2019). In total, the problem has 58 optimization variables, 53 equalities, and 19 inequalities. An equation-oriented formulation would have several hundred variables and equalities. The problem is implemented via the C++ API of MAiNGO v0.2.0.4 (Bongartz et al., 2018)¹ and solved to a relative optimality tolerance of 10^{-2} on 12 cores of an Intel Platinum 8160 with 2.1 GHz in 12 h of wall clock time using MPI parallelization. The optimal values of the design variables are given in Table 1. The exergy efficiency computed via Eq. (2) is 83 %. Local instead of global optimization results in anything from the global to substantially suboptimal solutions (e.g., 11% efficiency) or no feasible points, depending on the initial guess. For details, see Bongartz (2020). Note that this optimal efficiency cannot be compared to the benchmark process chain because of the simplified treatment of distillation and heat integration.

3.2. Local Optimization of Distillation Columns with Rigorous Models

To obtain reliable values for the energy demand for distillation, we optimize the columns locally with equilibrium tray-to-tray models, using the results from the previous sections as initial points. To achieve reasonable column designs, we minimize the total cost comprising heating, cooling, and capital cost estimated with the method of Guthrie (1969). Minimizing exergy destruction would result in excessively large columns. Design variables are the reboiler and condenser duties, the number of stages, and the feed stage. Purity constraints are imposed for OME₁ and wastewater (≥ 99.5 mol %) and the MF sent to combustion (≥ 80 mol %). The problems are formulated and solved as described by Kraemer et al. (2009): The integer variables for the number of trays and the feed tray are

¹The model is available in the C++ and text input formats of MAiNGO and in GAMS format under the name *DirectOxidationOME* in the GloPSE library at <https://git.rwth-aachen.de/avt.svt/public/glopse>.

Table 2: Reboiler duties of the distillation columns per unit of OME₁ produced (MJ/kg).

	D1	D2	D3	D4
Simple model	5.13	1.75	1.87	6.67
Rigorous model	2.61	1.37	1.04	0.57

Table 3: Utility demands per unit of OME₁ after heat integration of the optimized power-to-OME₁ process using rigorous (vs. simple) distillation models.

	Energy / MJ/kg	Exergy / MJ/kg
Heat at ≥ 136 °C	0 (1.98)	0 (0.54)
Cooling at -25 °C	-1.18	0.24
Cooling at -95 °C	-2.05	1.38
Heat at 128 °C	-5.47 (-4.90)	-1.41 (-1.26)

relaxed, and integrality is enforced through Fischer-Burmeister constraints in a series of successive nonlinear programs. The problems are solved with SNOPT (Gill et al., 2005) through GAMS, using external functions for the thermodynamics. The reboiler duties computed with the rigorous column models differ significantly from those with the simple models (see Table 2), in particular for D4, which has an azeotrope as top product. The overall heat demand of the process is 50 % less than predicted with the simple models. Still, the exergy efficiency according to Eq. (2) only rises to 84 % compared to 83 % with the simple models, since the exergy of the reboiler duties is small compared to that of the H₂. The rigorous column models are thus not crucial for determining the overall exergy efficiency. This may however be different for economic or environmental objectives.

3.3. Heat Integration

In the flowsheet optimization, all heat flows were considered directly in Eq. (2). This corresponds to utilities with matching temperatures being available for all heat flows, which is unrealistic. Therefore, we apply the same procedure as for the benchmark process chain: We conduct pinch analysis for full heat integration and assume a small number of utilities for the remaining heating and cooling demand. The process needs cryogenic cooling at -25 °C and -95 °C and no external heating, while it can export some heat at 128 °C (see Table 3). The resulting exergy efficiency of the process is 77 %, which is significantly lower than the 84 % predicted with the simplified treatment of heat flows. Nevertheless, the efficiency is higher than that of the benchmark process chain (73 %). Furthermore, the fact that no heat input is required is a clear advantage over the benchmark process chain. If we relied on the simple distillation models, we would erroneously conclude that the process did have a net heat demand at 136 °C (cf. Table 3).

4. Conclusion

A process for producing OME₁ from H₂ and CO₂ via direct oxidation of methanol was presented. The exergy efficiency of the process was maximized with the deterministic global optimization solver MAiNGO using detailed models for most of the flowsheet, but treating distillation and heat integration in a simplified way. According to an analysis with rigorous distillation models and a realistic setting for heat integration, the exergy

efficiency is 77 % as compared to 73 % for the literature benchmark. Still, equivalent comparison in a simulator as well as cost optimization would be desirable. The results show that global optimization of relatively complex flowsheets with moderately many degrees of freedom is possible with MAiNGO. However, more work is needed to allow for detailed distillation models. The results also underline the importance of realistic scenarios for heat integration and utilities, which should be considered in the optimization.

Acknowledgments: The authors gratefully acknowledge funding by the Federal Ministry of Education and Research within the Kopernikus Project P2X: Flexible use of renewable resources – exploration, validation and implementation of Power-to-X concepts.

References

- D. Bongartz, 2020. Deterministic global flowsheet optimization for the design of energy conversion processes. Ph.D. thesis, RWTH Aachen University, <https://doi.org/10.18154/RWTH-2020-06052>.
- D. Bongartz, J. Burre, A. Mitsos, 2019. Production of oxymethylene dimethyl ether from hydrogen and carbon dioxide – Part I: Modeling and analysis for OME₁. *Industrial & Engineering Chemistry Research* 58, 4881–4889.
- D. Bongartz, A. Mitsos, 2017. Deterministic global optimization of process flowsheets in a reduced space using McCormick relaxations. *Journal of Global Optimization* 69 (4), 761–796.
- D. Bongartz, A. Mitsos, 2019. Deterministic global flowsheet optimization: Between equation-oriented and sequential-modular methods. *AIChE Journal* 65 (3), 1022–1034.
- D. Bongartz, J. Najman, S. Sass, A. Mitsos, 2018. MAiNGO – McCormick-based Algorithm for mixed-integer Nonlinear Global Optimization. Tech. rep., Process Systems Engineering (AVT.SVT), RWTH Aachen University, <http://permalink.avt.rwth-aachen.de/?id=729717>.
- J. Burger, M. Siegert, E. Ströfer, H. Hasse, 2010. Poly(oxymethylene) dimethyl ethers as components of tailored diesel fuel: Properties, synthesis and purification concepts. *Fuel* 89, 3315–3319.
- S. Deutz, D. Bongartz, B. Heuser, A. Kätelhörn, L. Schulze Langenhorst, A. Omari, M. Walters, J. Klankermayer, W. Leitner, A. Mitsos, S. Pischinger, A. Bardow, 2018. Cleaner production of cleaner fuels: wind-to-wheel – environmental assessment of CO₂-based oxymethylene ether as a drop-in fuel. *Energy & Environmental Science* 11, 331–343.
- P. E. Gill, W. Murray, M. A. Saunders, 2005. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review* 47 (1), 99–131.
- K. M. Guthrie, 1969. Capital cost estimation. *Chemical Engineering* 24, 114–142.
- K. Kraemer, S. Kossack, W. Marquardt, 2009. Efficient optimization-based design of distillation processes for homogeneous azeotropic mixtures. *Industrial & Engineering Chemistry Research* 48 (14), 6749–6764.
- C. Kuhnert, M. Albert, S. Breyer, I. Hahnenstein, H. Hasse, G. Maurer, 2006. Phase equilibrium in formaldehyde containing multicomponent mixtures: Experimental results for fluid phase equilibria of (formaldehyde + (water or methanol) + methylal) and (formaldehyde + water + methanol + methylal) and comparison with predictions. *Industrial & Engineering Chemistry Research* 45 (14), 5155–5164.
- J. Liu, Y. Fu, Q. Sun, J. Shen, 2008. TiO₂ nanotubes supported V₂O₅ for the selective oxidation of methanol to dimethoxymethane. *Microporous and Mesoporous Materials* 116 (1-3), 614–621.
- A. Omari, B. Heuser, S. Pischinger, 2017. Potential of oxymethylenether-diesel blends for ultra-low emission engines. *Fuel* 209, 232–237.
- R. Sun, I. Delidovich, R. Palkovits, 2019. Dimethoxymethane as a cleaner synthetic fuel: Synthetic methods, catalysts, and reaction mechanism. *ACS Catalysis* 9 (2), 1298–1318.
- J. M. Tatibouët, 1997. Methanol oxidation as a catalytic surface probe. *Applied Catalysis A: General* 148 (2), 213–252.
- É. S. Van-Dal, C. Bouallou, 2013. Design and simulation of a methanol production plant from CO₂ hydrogenation. *Journal of Cleaner Production* 57, 38–45.
- A. Vogelpohl, 2002. On the relation between ideal and real mixtures in multicomponent distillation. *Chemical Engineering & Technology* 25 (9), 869–872.
- J.-O. Weidert, J. Burger, M. Renner, S. Blagov, H. Hasse, 2017. Development of an integrated reaction-distillation process for the production of methylal. *Industrial & Engineering Chemistry Research* 56, 575–582.

A Machine Learning Approach for Device Design from Materials and Operation Data

Luis A. Briceno-Mena, Gokul Venugopalan, Christopher C. Arges, Jose A.
Romagnoli^{*,*}

*Department of Chemical Engineering, Louisiana State University, Baton Rouge,
Louisiana 70803, United States*

**jose@lsu.edu*

Abstract

Machine Learning allows for the modelling and analysis of complex systems for which little mechanistic knowledge is available and is therefore envisioned as a powerful tool for the development of new designs with applications in engineering problems. In this work, we propose a framework based on dimension reduction, clustering, and self-organizing maps for the modelling and analysis of devices from materials and operation data, from which useful information can be drawn to inform future designs and developments. We demonstrate the applicability of this approach by analysing a high-temperature polymer electrolyte membrane fuel cell (HT-PEMFC). It was found that out of the 12 input variables studied, temperature, oxygen stoichiometric ratio, and ionomer binder ion exchange capacity are the most influential for achieving high power HT-PEMFC. This framework could be extended as new data becomes available about the different device components.

Keywords: Machine learning, materials science, energy conversion, device design.

1. Introduction

In recent years, there has been an increasing interest in the benefits of Machine Learning (ML) for materials discovery, i.e. the prediction of materials properties from molecular or structural information (Ong, 2019). However, the integration of materials and device level modelling could accelerate the deployment of newly discovered materials in real-life applications. One such application suitable for this approach is that of the development of low cost polymer electrolyte membrane fuel cells for automobile applications. High-temperature polymer electrolyte membrane fuel cells (HT-PEMFCs) pose a promising option for lowering the cost of hydrogen-powered vehicles. These fuel cells operate in the temperature range of 120 to 250 °C and offer simpler heat and water management in comparison to the conventional low-temperature fuel cells (LT-PEMFCs) (Xiao et al., 2005). Because of the high temperature, the cell stack radiator is minimized and the humidification of the feed gas is no longer needed. This is particularly important for stacks used in medium and heavy duty vehicles. Additionally, HT-PEMFCs tolerate CO in the hydrogen fuel stream, which enables the use of cheaper hydrogen (Venugopalan et al., 2020). These aspects significantly lower the cost of

hydrogen-powered setups for vehicles, which can in turn can lead to a more widespread adoption of cleaner transportation technologies. Despite its benefits, however, current designs of HT-PEMFCs show lower power density, high platinum metal loadings, and limited temperature range and water tolerance (Lee et al., 2019; Quartarone et al., 2017). Therefore, new developments are needed for HT-PEMFCs to be suitable for transportation applications. In this work, we implemented an ensemble of machine learning techniques, integrated with a semi-empirical model to bridge materials and device level modelling, and to explore the design search space for a high-temperature polymer electrolyte membrane fuel cell.

2. Related Methods

Figure 1 shows a schematic of the data analysis process employed. We used Uniform Manifold Approximation and Projection (UMAP) for dimension reduction and Hierarchical Density Based Spatial Clustering of Applications with Noise (HDBSCAN) for clustering to classify a variety of design configurations according to their peak power density as a measure of performance. We then fed the clustered data set into a self-organizing map (SOM) implementation so that the SOM algorithm would be induced to find three clusters. After this, component maps were generated to gauge the influence of each variable over the design performance. The overall result is a tool for the exploration of a high dimensional design space with the capability to related the lower dimensional representations to physically meaningful variables. Following is a brief description of the methods used.

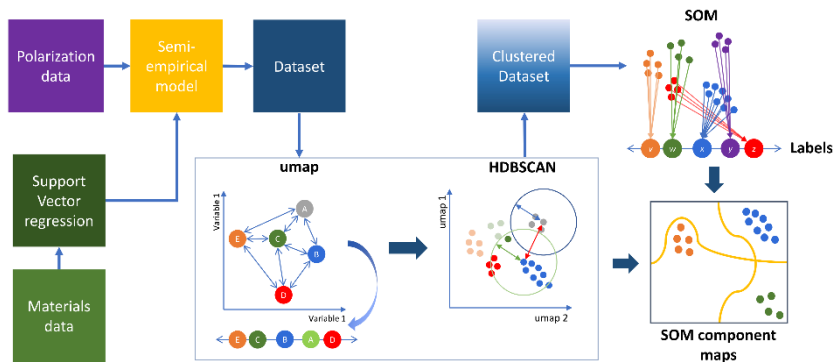


Figure 1. Schematic representation of the proposed framework for the sensitivity analysis.

2.1. Synthetic data

The data for the machine learning-based analysis was obtained from a 0-D polarization model substantiated using experimental data. A detailed description of the experimental procedures can be found elsewhere (Venugopalan et al., 2020). The 0-D, semi-empirical model was based upon existing HT-PEMFC models that are physics-informed with distributed parameters (Kregar, 2020; Oh et al., 2014; Yin et al., 2020) and features some semi-empirical expressions (Cheddie, 2007; Scott & Mamlouk, 2009; Sousa et al., 2010) to describe overpotential losses. This approach was selected because it accurately predicts data with few adjustable parameters, and it has relevant descriptors that factor how H_3PO_4 content in the membrane and CO in the feed gas stream impact HT-PEMFC polarization. The key assumptions for the model were: continuum level, steady state

operation, ideal gas behaviour, no gas crossover, no water transport considerations, and uniform distribution of catalyst and ionomer binder and gas transport properties across the electrodes and gas diffusion layers. The model uses two Support Vector Regression (SVR) with a radial basis kernel function (Drucker et al., 1997) to predict materials conductivity from their composition at a given temperature. In particular, the SVR implementation proposed by Chang and Lin (Chang & Lin, 2011) was used. The model serves as a bridge between the materials data and the device performance prediction, allowing the simulation of new configurations that will be later added to the training dataset for the ML procedures. To deal with the variability of the polarization model, we introduced gaussian noise following the expected distribution of this variability. This noise will later be removed using a denoising capable clustering technique.

The synthetic dataset was obtained using a stochastic generator. To produce different configurations 12 input variables were allowed to take different values around a base case in a uniform distribution within a given interval (e.g., \pm base value \times 40%). For example, a given configuration would include a lower catalyst loading, and higher temperature and pressure while another configuration would have the same catalyst loading but with a smaller membrane thickness.

2.2. Dimension reduction and Clustering

A density-based clustering (DBC) approach was adopted. DBC finds clusters of any shape, as opposed to centroid based clustering which assumes a shape (i.e. a sphere of equal variance) for the clusters. Also, in DBC, data points in sparse regions are not required to be assigned to a cluster but rather identified as noise. Additionally, introducing hierarchical clustering (HC) finds nested relationships amongst data instead of flat partitions in the data. This characteristic enables a more complex clustering with higher resolution. In this research we used the implementation proposed by McInnes and Healy (McInnes & Healy, 2017). Prior to the clustering we used UMAP for dimension reduction. UMAP is a neighbour graphs-based dimension reduction technique, meaning that a graph is first built for the high dimensional data and then it is embedded in a low dimensional space using a force directed layout using cross entropy to measure the distance between the high dimensional graph and the low dimensional graph. The resulting low dimensional embedding is denser, providing a better starting point for HDBSCAN (McInnes & Healy, 2018).

2.3. Self-Organizing Maps

SOMs are a type of artificial neural networks that use competitive and collaborative learning to generate a low dimensional representation of a dataset based on a measure of similarity amongst the data. SOMs are based on the notion of vector quantization, in which the input data is partitioned in number contiguous regions, each one represented by a model vector (Kohonen, 2013). Every input data is then associated with the model vector that best matches with it (competitive learning) and then its spatial neighbours are modified for better matching as well (collaborative learning). The result is a graphical representation of the high-dimensional dataset in a low-dimensional (typically two-dimensional) plot in which regions of similar entities are close together and far apart from dissimilar ones. The contribution of the different features of the data to the distribution on the map can be traced by examining the model vectors. Here, we employed the implementation designed by Moosavi et al (Moosavi et al., 2014).

3. Results and Discussion

Figure 2a and 2b present the SVR models for the prediction of conductivity, κ^{H^+} , as a function of ion exchange capacity (IEC) and temperature T for both the membrane and the ionomer binder. Figure 2c shows the accuracy for the prediction of the polarization data. In general, the three models show a good fit with the experimental, thus providing a reliable source for synthetic data.

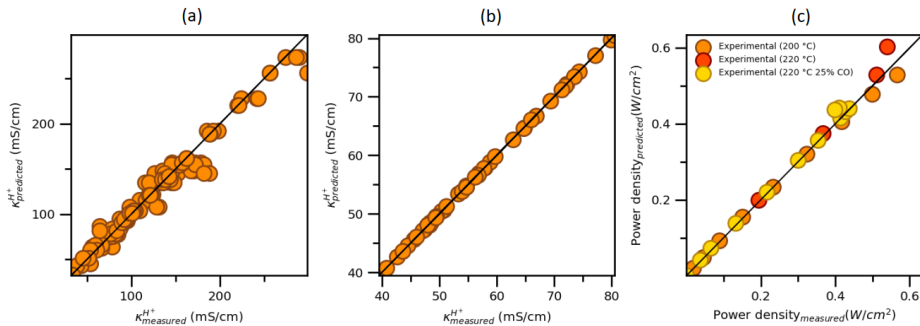


Figure 2. Model prediction against experimental data for (left) SVR for membrane conductivity, (centre) SVR for ionomer binder conductivity, (right) semi-empirical model for polarization. Measured data from (Venugopalan et al., 2020).

The implementation of UMAP for dimension reduction and HDBSCAN for clustering using the generated dataset resulted in 5 clusters. These clusters were further grouped into 3 and identified as low, medium, and high-performance, according to the peak power density.

Figure 3 shows a representation of the clusters projected onto the SOM. The training of the SOM was successful in the sense that it distinctly segregates the three regions of low (Figure 3a), medium (Figure 3b), and high (Figure 3c) performance (as measured by peak power density). With this mapping, the component maps can now be obtained to identify which of the variables under study has more influence on the high-performance cluster.

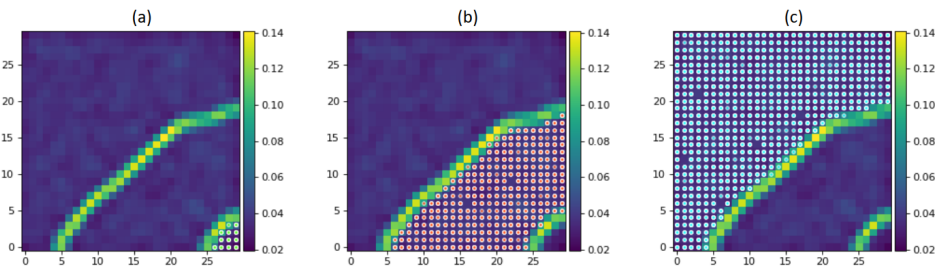


Figure 3. UMAP-HDBSCAN clusters projected onto the self-organizing map. The third one from the left shows the cluster for higher performance as measured by peak power density. Colours represent model weights difference (blue being small differences and green being large differences). Dots represent the clusters.

The component maps for the 3 variables that are most influential over the high-performance cluster are shown in Figure 4. Out of the 12 variables explored,

oxygen stoichiometric ratio, SO_2 (Figure 4a); ion exchange capacity of the ionomer binder, IEC_{io} (Figure 4b); and operating temperature, T (Figure 4c); were found to be the most influential over the high-performance cluster. These observations were cross validated using a Space Greedy Search (SGS) analysis (Figure 5) as described by Zhu *et al.* (Zhu *et al.*, 2018). For the SGS analysis the original clusters from the UMAP-HDBSCAN stage were used. Although the influence of these variables over the other clusters is not so regular, a clear tendency to be highly relevant for the high-performance cluster is observed. Therefore, optimized values for T , IEC_{io} , SO_2 are needed to achieve high performance. Furthermore, the component contributions show the relative importance of temperature and oxygen availability with respect to the other operating conditions. On one side, this suggests that mass transfer phenomena such as oxygen permeability (associated with oxygen availability) plays a major role on the fuel cell performance. On the other hand, operating temperature could provide a greater boost in performance than, for example, pressure, which should be considered for new designs.

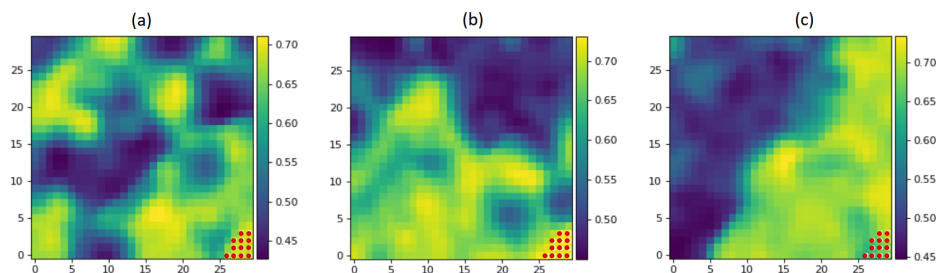


Figure 4. Component maps for the 3 most relevant variables to the high-performance cluster: (left) oxygen stoichiometric ratio, SO_2 ; (centre) ion exchange capacity of the ionomer binder, IEC_{io} ; and (right) operating temperature, T . Colours represent degree of influence (blue being small contributions and green being large contributions). Dots represent the clusters.

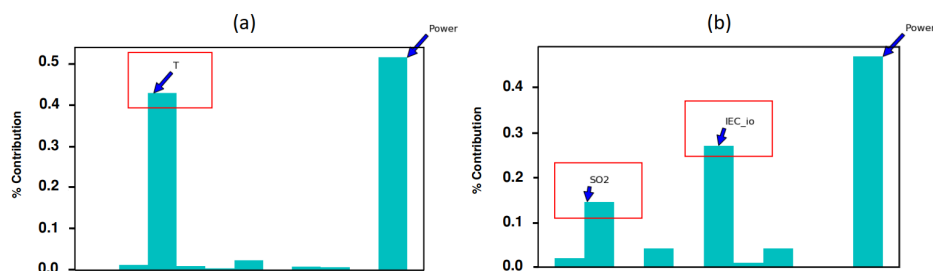


Figure 5. Contribution plot from the Space Greedy Search analysis between clusters (a) 5-2, (b) 5-3. Red boxes indicate the highest contributions.

4. Conclusions

Our results show that density-based clustering, combined with neighbor graphs dimension reduction, and SOM analysis, helps identify physically meaningful patterns in the operation data. These patterns can in turn be related to the materials properties

predicted by support vector regression via the semi-empirical model, providing information for both the design of the materials and the appropriate device conditions needed for a cost-effective operation. The modular nature of this approach provides an easy way to incorporate it into the manufacturing process of such materials and devices.

5. Acknowledgments

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office (AMO) Award Number DE-EE0009101. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Luis A. Briceno-Mena thanks the support received from Universidad de Costa Rica.

6. References

- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), Article 27.
- Cheddie, D. F., & Munroe, N. D. (2007). A two-phase model of a in intermediate temperature PEM fuel cell. *International Journal of Hydrogen Energy*, 32(7), 832-841.
- Drucker, H., Burges, C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support Vector Regression Machines. *Adv. Neural. Inf. Process. Syst.*, 9, 155-161.
- Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37, 52-65.
- Kregar, A., Tavcar, G., Kravos, A., & Katrasnik, T. (2020). Predictive system-level modeling framework for transient operation and cathode platinum degradation of high temperature proton exchange membrane fuel cells. *Applied Energy*, 263.
- Lee, A., Choe, Y.-K., Matanovic, I., & Kim, Y. (2019). The energetics of phosphoric acid interactions reveals a new acid loss mechanism. *J. Mater. Chem. A*, 7, 9867-9976.
- McInnes, L., & Healy, J. (2017). *Accelerated Hierarchical Density Clustering*. Paper presented at the 2017 IEEE International Conference on Data Mining Workshops, New Orleans.
- McInnes, L., & Healy, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. *ArXiv*, abs/1802.03426.
- Moosavi, V., Packmann, S., & Valles, I. (2014). SOMPY: A Python Library for Self Organizing Map (SOM). GitHub.[Online]. Available: <https://github.com/sevamoo/SOMPY>.
- Oh, K., Jeong, G., Cho, E., Kim, W., & Ju, H. (2014). A CO poisoning model for high-temperature proton exchange membrane fuel cells comprising phosphoric acid-doped polybenzimidazole membranes. *International Journal of Hydrogen Energy*, 39(36), 21915-21926.
- Ong, S. P. (2019). Accelerating materials science with high-throughput computations and machine learning. *Computational Materials Science*, 161, 143-150.
- Quartarone, E., Angioni, S., & Mustarelli, P. (2017). Polymer and composite membranes of proton-conducting, high-temperature fuel cells: a critical review. *Materials*, 10.
- Scott, K., & Mamlouk, M. (2009). A cell voltage equation for an intermediate temperature proton exchange membrane fuel cell. *International Journal of Hydrogen Energy*, 34(22), 9195-9202.
- Sousa, T., Mamlouk, M., & Scott, K. (2010). An isothermal model of a laboratory intermediate temperature fuel cell using PBI phosphoric acid membranes. *Chemical Engineering Science*, 65(8), 2513-2253-.
- Venugopalan, G., Chang, K., Nijoka, J., Livingston, S., Geise, G. M., & Arges, C. G. (2020). Stable and Highly Conductive Polycation–Polybenzimidazole Membrane Blends for Intermediate Temperature Polymer Electrolyte Membrane Fuel Cells. *ACS Applied Energy Materials*, 3(1), 573-585. doi:10.1021/acsaeam.9b01802

- Xiao, L., Scanlon, E., Ramanathan, L., Choe, E., Rogers, D., Apple, T., & Benicewicz, B. (2005). High-Temperature Polybenzimidazole Fuel Cell Membranes via a Sol-Gel Process. *Chemistry of Materials*, 17(21).
- Yin, C., Gao, Y., Li, T., Xie, G., Li, K., & Tang, H. (2020). Study of internal multi-parameter distributions of proton exchange membrane fuel cell with segmented cell device and coupled three-dimensional model. *Renewable Energy*, 147, 650-662.
- Zhu, W., Sun, W., & Romagnoli, J. (2018). Adaptive k-Nearest-Neighbor Method for Process Monitoring. *Industrial & Engineering Chemistry Research*, 57(7), 2574-2586. doi:10.1021/acs.iecr.7b03771

A Computer-Aided Platform for Simultaneous Process Synthesis and Intensification

Mohammed Sadaf Monjur, Salih Emre Demirel, Jianping Li, M. M. Faruque Hasan*

*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843-3122, USA.
hasan@tamu.edu*

Abstract

We develop SPICE (Synthesis and Process Intensification of Chemical Enterprises), which is a software tool for chemical process design and discovery. Primarily based on the building block representation of various physicochemical phenomena, SPICE offers a comprehensive set of toolkits to perform (i) benchmarking to determine the theoretical bounds on process design opportunities and maximum intensification targets, (ii) conceptual process design, (iii) process synthesis and optimization, (iv) process simulation and analysis of both intensified and non-intensified unit operations, (v) identification of intensification hotspots for existing processes and (vi) property-performance mapping towards elucidating new materials for intensification purposes. The benefits and the applicability of SPICE as a process design platform are demonstrated through multiple case studies.

Keywords: Process Synthesis, Intensification, Building Blocks, Software Prototype

1. Introduction

Increased competition, environmental regulation, rise of unconventional feedstocks, and volatile market conditions have contributed to renewed interest in process design and intensification methods for the chemical process industry. Several new design and synthesis platforms/prototypes (e.g., ProCAFD, MIPSYN, Pyosyn, and Synopsis, among others) have emerged. Computer-aided process intensification is now considered as an integral part of conceptual design. However, discovering novel processes without relying on pre-postulated conceptual design alternatives remains a challenge. One can potentially address this by adopting “bottom-up” representations that use generic but finite number of design elements encapsulating all the fundamental physicochemical phenomena and functions that constitute a chemical process. Recently, Hasan and co-workers (Demirel et al., 2017) proposed such a representation based on design building blocks that allow us to generate numerous designs of unit operations, process configurations and flowsheets without explicit consideration or knowledge of all plausible alternatives beforehand. This departure from unit operations-based representation to design elements provide a systematic new way for computer-aided optimal process design and intensification.

In this work, we describe the development of a design software tool based on the building block representation. We call this as SPICE (Synthesis and Process Intensification of Chemical Enterprises). The overall SPICE framework is depicted in Figure 1. SPICE offers a comprehensive toolkit with the following functionalities:

- (i) fast benchmarking and targeting for determining the theoretical bounds on process design opportunities and maximum intensification targets,
- (ii) conceptual process design,
- (iii) process synthesis and optimization,
- (iv) simulation, and analysis of intensified and non-intensified unit operations,
- (v) identification of intensification hotspots for existing processes, and
- (vi) property-performance mapping to elucidate materials for intensification purposes.

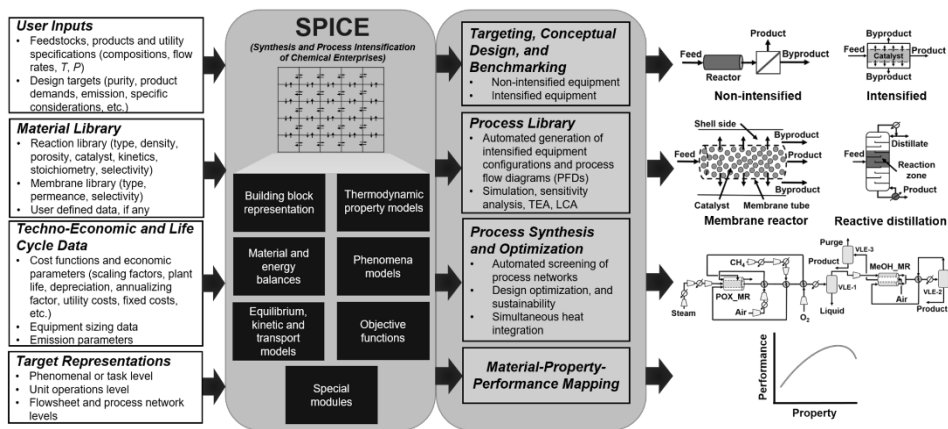


Figure 1: SPICE framework.

For specified/user-selected inputs of feedstocks, product specifications, material properties, and techno-econo-environmental parameters, SPICE generates a rank-ordered list of design configurations with optimal sizing and a comprehensive analysis of various target metrics. Combining the thermodynamic minimum energy calculation with the Feinberg Decomposition theory (Feinberg and Ellison, 2001), SPICE can determine the theoretical attainable limits for any reaction-separation system. Users can further perform detailed process simulation, sensitivity analysis, and techno-economic analysis (TEA). Furthermore, SPICE provides options for multiobjective optimization at the conceptual design stage to automatically generate Pareto solutions considering the techno-economic and environmental trade-offs. Finally, under the material property-performance mapping section, users can perform parametric studies to find the effect of material properties on the process performance in terms of product purity, quality, conversion, energy, and cost.

2. SPICE Framework

2.1. User Interface

Users specify the feedstock availability and compositions, minimum product purity, productivity and demands, and the operating ranges for temperatures, pressures and

flow rates. They can also select the set of potential phenomena to be considered. For a system involving reactions, the material library provides information to the users on reaction stoichiometry, equilibrium and kinetics. For separations, users can indicate their choices for VLE, membranes, solvents and absorbents. Additionally, users can define new materials with their own data. SPICE also accepts user-defined cost functions and economic parameters such as scaling factor, plant life, depreciation rate, annualized factor, utility costs, raw material costs, fixed costs, etc. Finally, users can select the level of design details (short-cut or rigorous) by selecting the appropriate models.

SPICE also allows specialized submodules focusing on specific set of phenomena and intensification options (Figure 2). For new designs, users can choose SPICE_GEN. Alternatively, for a known process, SPICE_RF finds retrofitting hotspots and SPICE_OPT optimizes the process. Users can also restrict to only SPICE_HI (heat integration), SPICE_MARS (membrane-assisted reactive separation), SPICE_RD (reactive distillation), SPICE_MS (membrane separation), SPICE_MD (membrane distillation), SPICE_ED (extractive distillation), SPICE_DWC (dividing wall column), SPICE_MRD (membrane-assisted reactive distillation), and SPICE_MED (membrane-assisted extractive distillation).

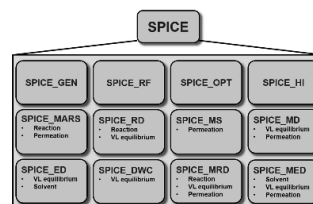


Figure 2: SPICE sub-modules for specific phenomena.

2.2. Building Block-Based Representation

The building block-based superstructure (Demirel et al., 2017; Li et al., 2018; Demirel et al., 2019) is used to represent the overall design problem, which is formulated as a single mixed-integer nonlinear program (MINLP). Each block has two design elements, namely the block interior and the block boundaries. A block interior is characterized by a temperature, a pressure, a phase, and a composition of chemical species. The interactions between two adjacent blocks are controlled by a common boundary. A boundary can be either unrestricted, semi-restricted or completely restricted (Figure 3a). Streams go through an unrestricted boundary without restrictions. A semi-restricted boundary indicates that there exists a mass-transfer interceptor which represents separation phenomena (e.g., membrane, VLE interface). A completely restricted boundary implies zero-flow across. External feed streams are introduced and products can be withdrawn from each block. Additionally, jump flows allow interaction between non-adjacent blocks.

A combination of the block interior and the boundaries are used to represent different phenomena (Figure 3b). For instance, a reactor block interior is filled with catalyst. Two blocks separated by a common semi-restricted boundary represents separation phenomena (phase contact, permeation, pervaporation, etc.). When several blocks are arranged in a two-dimensional grid, we call it a superstructure (Figure 3c). This arrangement embeds different equipment (an example is shown in Figure 3d). By

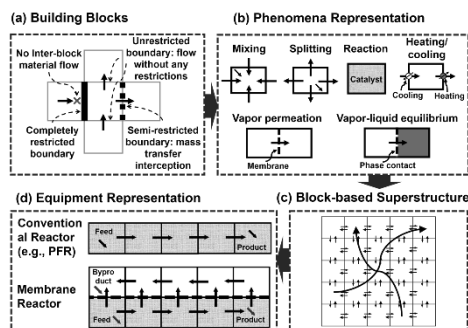


Figure 3: Building block representation.

arranging many such building blocks in different orders, numerous flowsheets can be generated and screened.

2.3. Material and Energy Balances

The MINLP model includes fundamental material and energy balance constraints for each block. For block $B_{i,j}$, where i and j denote the coordinate of the block in the 2-D grid, the material balance (Eq. 1) includes species flowrates through each of the four boundaries. $F_{i,j,k}$ and $R_{i,j,k}$ denote horizontal and vertical flows, respectively; $G_{i,j,k}$ captures the consumption/generation due to chemical reaction, $M_{i,j,k,f}$ and $N_{i,j,k,p}$ represent feedstock f and product p flows, respectively; and $J_{i,j,i',j',k}$ allows jump flows from block $B_{i,j}$ to block $B_{i',j'}$ for species k . The energy balance (Eq. 2) ensures that the enthalpies and the heat and work exchanges via internal transformation and external utility are preserved.

$$F_{i,j-1,k} + R_{i-1,j,k} - F_{i,j,k} - R_{i,j,k} + G_{i,j,k} + \sum_{f \in FS} M_{i,j,k,f} - \sum_{p \in PS} N_{i,j,k,p} + \sum_{(i',j') \in Link} J_{i',j',i,j,k} - \sum_{(i,j) \in Link} J_{i,j,i',j',k} = 0, \quad \forall i, j, k \quad (1)$$

$$EF_{i,j-1} + ER_{i-1,j} - EF_{i,j} - ER_{i,j} + EG_{i,j} + EM_{i,j} - EN_{i,j} + EJ_{i,j}^f - EJ_{i,j}^p + W_{i,j}^{comp} - W_{i,j}^{exp} + Q_{i,j}^h - Q_{i,j}^c = 0, \quad \forall i, j \quad (2)$$

$$f_{i,j,k,ph}^{ph} - M(1 - z_{i,j,ph}) \leq \Phi_{i,j,k} \leq f_{i,j,k,ph}^{ph} + M(1 - z_{i,j,ph}), \quad \forall i, j, k, ph \quad (3)$$

$$f_{i,j,k,ph}^{ph} = f^{rxn} (T_{i,j}, P_{i,j}, y_{i,j,k'}, V_{i,j,ph}, z_{i,j,ph}), \quad \forall i, j, k, ph \in Rxn \quad (4)$$

$$f_{i,j,k,ph}^{ph} = f^{equil} (K_{i,j,k,ph}^{eq}, T_{i,j}, P_{i,j}, y_{i,j+1,k'}, z_{i,j,ph}), \quad \forall i, j, k, ph \in Equil \quad (5a)$$

$$f_{i,j,k,ph}^{ph} = f^{equil} (K_{i,j,k,ph}^{eq}, T_{i,j}, P_{i,j}, y_{i+1,j,k'}, z_{i,j,ph}), \quad \forall i, j, k, ph \in Equil \quad (5b)$$

$$f_{i,j,k,ph}^{ph} = f^{rate} (\lambda_{k,ph}, DF_{i,j,k,ph}, d_{i,j,ph}, z_{i,j,ph}), \quad \forall i, j, k, ph \in Rate \quad (6)$$

2.4. Thermodynamic Property Model

Thermodynamic property models are an important and integral part of any process design and synthesis platform to accurately estimate the properties of each associated species. However, these models are highly nonlinear, nonconvex and involve large number of equations, which increase the complexity. To overcome this, SPICE exploits data-driven surrogate thermodynamic models for several properties like specific enthalpy, specific heat capacity, density, and saturation pressure. These fitted models often offer high accuracy ($R^2 > 0.99$) over a specified range while reducing the computational demand.

2.5. Phenomena Models

Eq. 3 is the phenomena assignment constraint, where, ph denotes the phenomena set. Here, $\Phi_{i,j,k} := G_{i,j,k}$ for $ph \in Rxn$, $\Phi_{i,j,k} := y_{i,j,k}$ for $ph \in Equil$, and $\Phi_{i,j,k} := \{F_{i,j,k}, R_{i,j,k}\}$

for $ph \in Rate$. Based on the value of $z_{i,j,ph}$, the corresponding phenomena model $f^{ph}(T_{i,j}, P_{i,j}, y_{i,j,k}, V_{i,j,ph}, \lambda_{k,ph}, DF_{i,j,k,ph}, d_{i,j,ph})$ is activated. Basic models representing various chemical phenomena, including reaction and separation, are included within the framework (Eqs. 4-6). Separation phenomena are classified in two groups, namely the equilibrium-based phenomena (Eqs. 5a-b) and the rate-based phenomena (Eq. 6). Here, $T_{i,j}$, $P_{i,j}$, $y_{i,j,k}$, $V_{i,j,ph}$ and $DF_{i,j,k,ph}$ are block interior variables representing temperature, pressure, composition, catalyst amount, and driving force across the boundary, respectively, while $\lambda_{k,ph}$ and $d_{i,j,ph}$ are block boundary variables denoting mass transfer coefficient and mass transfer area, respectively. Equilibrium-based phenomena includes vapor-liquid, liquid-liquid, and vapor-vapor phase contact, and are observed in distillation, flash, solvent extraction, absorption operations. Rate-based phenomena includes gas permeation, liquid permeation and pervaporation that are

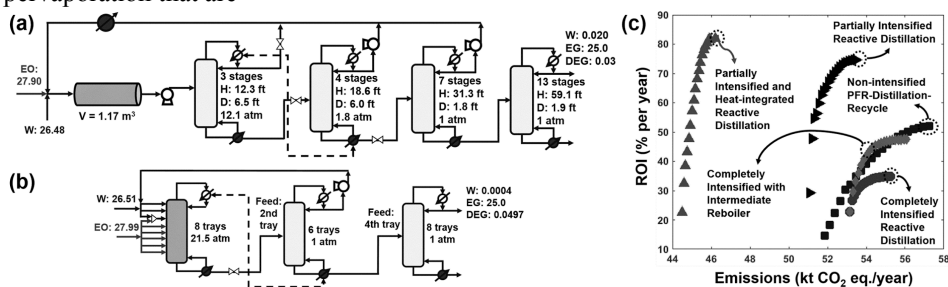


Figure 4: SPICE generated flowsheets (a-b) and multiobjective optimization results (c).

common to membrane-based separation. The assignment of phenomena to the block interiors and the boundaries is dictated by the binary variable $z_{i,j,ph}$.

2.6. Design Objectives

SPICE allows to use several built-in objective functions, such as, maximize {total annual profit (TAP), return on investment (ROI), yield, and conversion}, or minimize {total annual cost (TAC), energy, emission, size, and maximum allowable reactor temperature}, among others. Additionally, users can modify these functions based on their needs.

3. Case Studies

3.1. Ethylene Glycol Production

Here, we demonstrate the SPICE framework in designing an intensified process for ethylene glycol (EG) production from ethylene oxide (EO) and water (W) with minimum TAC while producing 25 kmol/h EG with 99.8 % (mol) purity. As this process involves reaction and vapor liquid equilibrium separation (distillation), we select SPICE_RD module. First, a unit operation based process synthesis and integration is performed, which yields a process with TAC of 12.42 MMS\$/y. The resultant flowsheet (Figure 4a) contains an adiabatic reactor and four distillation columns to achieve the target purity. Additionally, heat integration is considered between the condenser of the first distillation column and the re-boiler of the second distillation column. When phenomena-based synthesis approach is considered, a novel intensified flowsheet (Figure 4b) is obtained with 12.33 MMS\$/y TAC, utilizing only

three major equipment (Demirel et al., 2019). We have also considered sustainable process intensification using multiobjective optimization for EG production considering both economic and environmental objectives. The resultant pareto solutions from SPICE_RD for several intensified, partially intensified and non-intensified designs are shown in Figure 4c (Demirel et al., 2020).

3.2. Shale Gas Conversion to Methanol

Methanol processes are equilibrium-limited and energy intensive, and have lower methanol yield. We have applied the SPICE framework to search for better process configurations with maximum TAP. First, we optimize a base-case flowsheet (Ehlinger et al., 2014) using the SPICE_OPT module. The resultant flowsheet has a TAP of 12.7 MM\$/y and a ROI of 22 %. We then switch to SPICE_RF and use the optimized base-case as a warm-start when searching for flowsheets with higher TAPs. This generates a new intensified and heat-integrated flowsheet as presented in Figure 5. This flowsheet has a TAP of 38 MM\$/y (200 % improvement) and a ROI of 52 % (136 % improvement). The improvements are due to several changes. First, it considers only two reactors and excludes the WGS reactor and the CO₂ separator. Second, the reactors are now replaced by membrane reactors. The membrane reactor for partial oxidation (POX_MR) separates

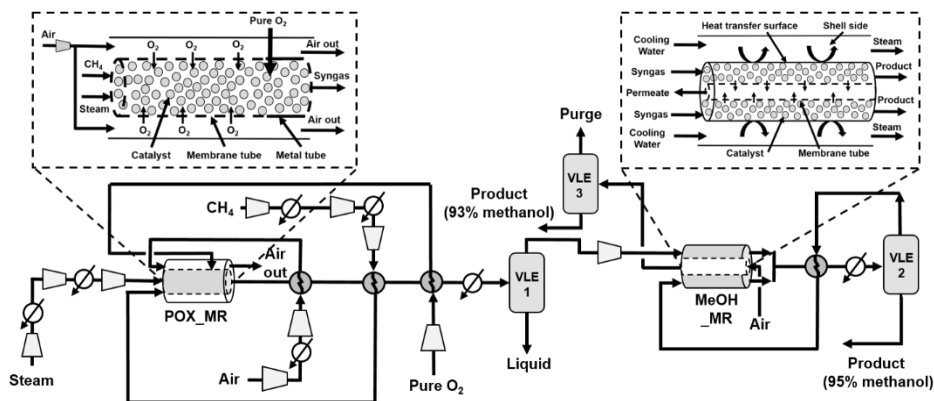


Figure 5: Intensified and heat-integrated process flowsheet for methanol synthesis.

the required O₂ from air and performs the reaction simultaneously, which reduces the consumption cost of expensive pure O₂ by 33 %. The intensified MeOH_MR, on the other hand, increases the methanol yield by 17 % through in-situ removal of the products from the reaction mixtures. Finally, heat integration between the inlet and outlet of POX_MR reduces the utility cost by 58 %.

4. Conclusions

SPICE is a comprehensive process design software prototype. It is a toolkit with the capabilities to benchmark, design, synthesize and intensify chemical process systems. Tests suggest that new processes can be systematically designed with significant reduction in energy and cost for important design problems. The new pathways may not always use fully-intensified systems, specially in the presence of multiple conflicting objectives, which again highlights the need for a comprehensive design package such as SPICE.

5. Acknowledgements

The authors gratefully acknowledge the support from NSF and AIChE/RAPID Institute.

References

- S. E. Demirel, J. Li, M. M. El-Halwagi, M. F. Hasan, 2020. Sustainable process intensification using building blocks. *ACS Sustainable Chemistry & Engineering*.
- S. E. Demirel, J. Li, M. F. Hasan, 2017. Systematic process intensification using building blocks. *Computers & Chemical Engineering* 105, 2–38.
- S. E. Demirel, J. Li, M. F. Hasan, 2019. A general framework for process synthesis, integration, and intensification. *Industrial & Engineering Chemistry Research* 58 (15), 5950–5967.
- V. M. Ehlinger, K. J. Gabriel, M. M. Noureldin, M. M. El-Halwagi, 2014. Process design and integration of shale gas to methanol. *ACS Sustainable Chemistry & Engineering* 2 (1), 30–37.
- M. Feinberg, P. Ellison, 2001. General kinetic bounds on productivity and selectivity in reactor-separator systems of arbitrary design: Principles. *Industrial & engineering chemistry research* 40 (14), 3181–3194.
- J. Li, S. E. Demirel, M. F. Hasan, 2018. Process synthesis using block superstructure with automated flowsheet generation and optimization. *AIChE Journal* 64 (8), 3082–3100.

Studying the effect of solvent properties in treatment design within CO₂ Integration Networks

Dhabia Al-Mohannadi^a, Sabla Y. Alnouri^{b,c}

^aChemical Engineering Program, *Texas A&M University at Qatar*

^bGas Processing Centre, *College of Engineering, Qatar University*

^cThe Baha and Walid Bassatne Department of Chemical and Petroleum Engineering, *American University of Beirut, P.O. Box 11-0236, Riyad El-Solh, Beirut, Lebanon*

dhabia.al-mohannadi@tamu.qatar.edu

Abstract

Carbon Dioxide Integration mainly targets the recovery of CO₂ streams, and assesses the allocation of those recovered streams into CO₂-using sinks, with an overall aim of attaining a minimum cost of a CO₂ network that could meet a prescribed emission target. Meeting emission targets often introduces numerous challenges, especially when energy intensive processes are involved. Hence, CO₂ Integration aims to identify appropriate CO₂ capture, allocation, and utilization options. There exists numerous processes (or CO₂-using sinks) that convert CO₂ streams into value added products, while generating revenue in some cases. However, many CO₂ sink options may be costly, especially if a high purity CO₂ stream is required in the process. Thus, the incorporation of treatment units is imperative, to separate CO₂ gas from the remaining gaseous emission material, which is in fact the most expensive design step in CO₂ integration systems. This study focuses on assessing the effect of varying the temperature of the solvent to be used in the treatment stage of CO₂ on the overall performance of the entire integration network attained. A MINLP formulation has been established based on key design parameters whilst taking into consideration important technical performance constraints that effectively describe the system to be assessed. A more detailed cost model for the treatment units was developed, to establish the effect of varying the emission reduction targets onto the overall performance of CO₂ integration systems.

Keywords: Carbon Integration, GHG emissions, Treatment, Carbon dioxide, Cost models

1. Introduction

Carbon capture utilization and sequestration (CCUS) has emerged as one of the most promising solutions to reduce large amounts of CO₂ especially as the global community adopts climate targets. CCUS involves many steps starting from capture of CO₂ from emission sources, transportation of CO₂ and finally either biological or geological sequestration or in utilization in CO₂ converting technologies. Each step creates a challenge especially due to the large energy requirement. However, the major influence on CCUS cost depends on energy intensive CO₂ treatment especially when dealing with CO₂ dilute streams (IPCC, 2005). There has been a plethora of work in the area of carbon capture using adsorption, absorption, and membrane processes (Ochedi et al. 2020) each varying in efficiency and cost. The most widely applied CO₂ separation technology in industry and that work well with dilute streams is CO₂ adsorption process (Hussin and

Aroua, 2020). Thus, this work aims to understand the effect of varying temperature and pressure in CO₂ adsorption units on the cost of overall CCUS network. Many of the cost items are lumped into single estimation parameters. This study aims to more accurately capture the design of such systems based on a detailed cost model for the absorption-stripping stage within the network.

2. Methodology

To be able to study the various effects of solvent properties within the treatment stage on the overall design of CCUS networks, it is imperative to find appropriate models that can correlate those effects. In doing so, such models can then be integrated onto the original CO₂ Integration model (Al-Mohannadi and Linke, 2016, Al-Mohannadi et al., 2015), with appropriate adjustments and modifications. In this paper, the primary focus was on the following three different solvent properties: density, viscosity, and Henry's constant. It should be noted that all those properties are a function of temperature. Since most common solvents that are used in the absorption stage for CO₂ removal from flue gas effluents are amines, the appropriate correlations that describe amine property behaviour as a function of temperature were identified as follows, and have been used in this study:

The liquid (solvent) density ρ_L as a function of temperature can be found using Eq. (1) below:

$$\rho_L = a_1 + a_2T + a_3T^2 \quad (1)$$

Where a_1 , a_2 and a_3 are equation parameters obtained from DiGuilio et al. (1992), and T is the temperature of the solvent.

The liquid (solvent) viscosity μ_L as a function of temperature can be found using Eq. (2) below:

$$\ln \mu = b_1 + \frac{b_2}{T - b_3} \quad (2)$$

Where b_1 , b_2 and b_3 are equation parameters obtained from DiGuilio et al. (1992). To find Henry's constant, Eq. (3) below can be used from Li (2017)

$$\ln H = c_1 + c_2T + c_3T^2 \quad (3)$$

Where c_1 , c_2 and c_3 are equation parameters. Values for those 3 equation parameters are based on the correlations provided by H.B. Liu et al. (1999).

3. Mathematical Formulation

The mathematical formulation of the main optimization problem for CO₂ integration falls under the general form of a Non-Linear problem, in which a central objective function ($f(x)$) is subjected to network equality and inequality constraints $h(x)$ and $g(x)$.

$$\min f(x) \quad (4)$$

$$h_i(x) = 0, i \in I = \{1, \dots, a\} \quad (5)$$

$$g_j(x) \leq 0, j \in J = \{1, \dots, b\} \quad (6)$$

Eq. (4) represents the objective function, $f(x)$, which aims to minimize the overall cost of the overall CCUS network in the form of piping, compression, pumping, treatment and sink processing. The overall cost covers both capital and operating costs for which the network is designed to operate. Cost and revenue calculations were based on the model by Al-Mohannadi and Linke, 2016, except for the treatment unit cost, which was developed based on solvent property data, and the respective column sizing calculations (both diameter and height). Moreover, the material of construction was also included in the treatment cost evaluation. Hence, the treatment unit cost in this work is a function of the dimensions of the treatment units (diameter and height), which in turn are functions of the gas feed flow, gas stream properties, the solvent feed flow and the respective solvent feed properties (Brunazzi et al 2002). The cost of the system generated by the model is determined by summing the operating and annualized capital costs (assuming 20 years plant lifetime). The central variable in the optimization problem (x) is the carbon-dioxide flowrate (which can either be treated or untreated). Even though temperature can also be a decision variable, studying the effect of varying temperature on the system would not be possible, since only 1 optimal solution (at the optimal temperature) would be reported. Hence, temperature was kept as a parameter, for the purpose of this study.

This objective function combined to two different sets of constraints: equality constraints $h(x)$ (Eq. (5)), and inequality constraints $g(x)$ (Eq. (6)). Equality constraints primarily cover the mass and component balances of captured and uncaptured (treated and untreated) CO₂ flows throughout the CCUS network. Inequality constraints $g(x)$, on the other hand, involve production and performance requirements, which involve imposing a minimum net CO₂ emission requirement onto the overall CCUS network, in addition to quality constraint limits on each CO₂ within the network to ensure acceptable CO₂ quality and effective sink performance. The exact mathematical expressions associated with all equality and inequality constraints are outlined in Al-Mohannadi and Linke, 2016. In order to be able to solve the optimization problem, What'sBest9.0.5.0 LINDO Global Solver for Microsoft Excel 2010 (Lindo, 2010), was used to implement the model, on a laptop with Intel® Core™ i7-2620M, 2.7 GHz, 8.00 GB RAM, 64-bit Operating System.

4. Results and Discussion

The effect of varying the solvent temperature within the treatment unit was investigated in this work for Methyl Diethanolamine, also known as MDEA. This solvent is well-known for having quite effective CO₂ removal efficiencies. A case study that involves four different CO₂ emission sources (involving an Ammonia plant, a Power-plant, a Refinery and a Steel plant) and six different CO₂ utilization sinks (involving Enhanced Oil Recovery, a Greenhouse option, a Storage option, a Methanol plant, a Urea plant, and an Algae space), i.e, 4 source options that could potentially be matched with 6 different sinks, to meet a prescribed CO₂ target. The setup resembles the case that has been previously presented by Al-Mohannadi and Linke (2016), apart from the treatment handling. A treatment model that enables the assessment of solvent property variations, whilst providing sizing calculations for all treatment units has been used. The operating lifetime assumed was 20 years. In order to explore the effect of varying the temperature on the solvent, and hence onto the overall CCUS network design, six different optimization problems have been attempted, where each is assigned a different overall net capture target as follows: 5%, 10%, 20%, 25%, 35% and 45%, which are in line with

the capture targets reported by Al-Mohannadi and Linke (2016). The temperature could vary between 20 and 27 °C for each of the six different cases involving different net capture targets. Higher temperature ranges could be investigated in a similar manner, subject to data availability. A total of 42 runs have been made, and the results obtained are summarized in Figures 1-3. First off, Figure 1 presents the treatment cost variation as a function of solvent temperature, for the different net capture targets that have been considered.

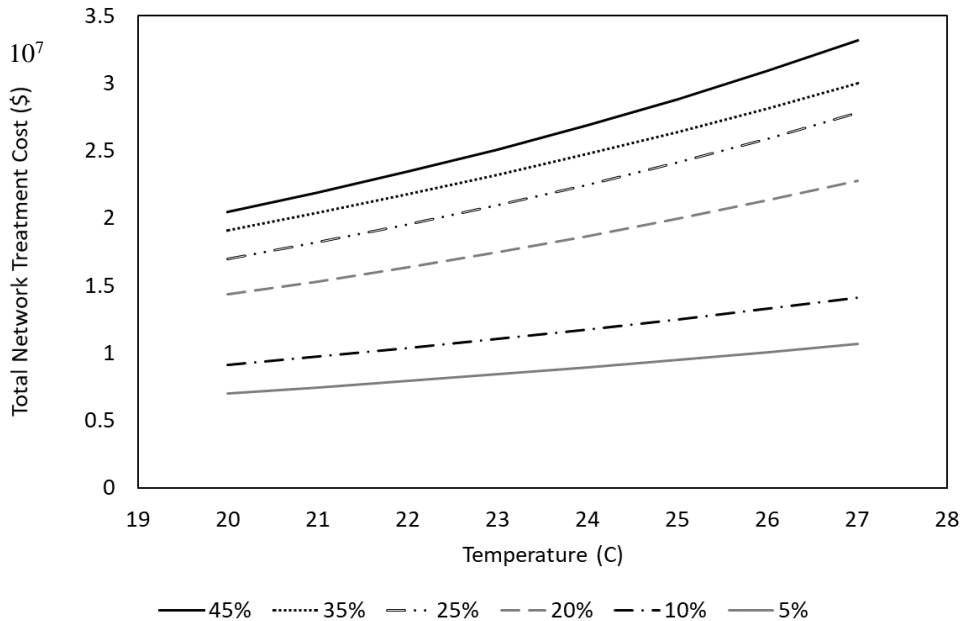


Figure 1. Treatment cost variation as a function of solvent temperature for different net capture targets

A consistent increase in the overall treatment cost for all cases was observed, when comparing the various capture target cases, due to associated changes in solvent properties that require larger treatment units (in terms of diameter and height). Nevertheless, it should be noted that the highest treatment cost jump was observed between the 10% and 20% capture cases. Moreover, the 5% case was the least sensitive to temperature variations since this case involved the least amount of treatment to meet the capture target. When the total network cost comes into perspective, the trend observed slightly differs since the total network cost involves the costing of other entities that are independent of treatment solvent properties. Figure 2 shows the total network cost variation as a function of solvent temperature for the different net capture targets. It is evident that both the 5% and 10% capture cases become the least sensitive to those temperature variations, while the 45% capture case being the most expensive case explored. The big cost gap that was observed between the 45% capture case and the rest of the cases was attributed to the sink processing costs, which were found to be the least in the 45% capture case. Some of the cases have reported a negative total cost for the run, indicating a revenue-generating case in which the total cost of the network was less than the revenue generated.

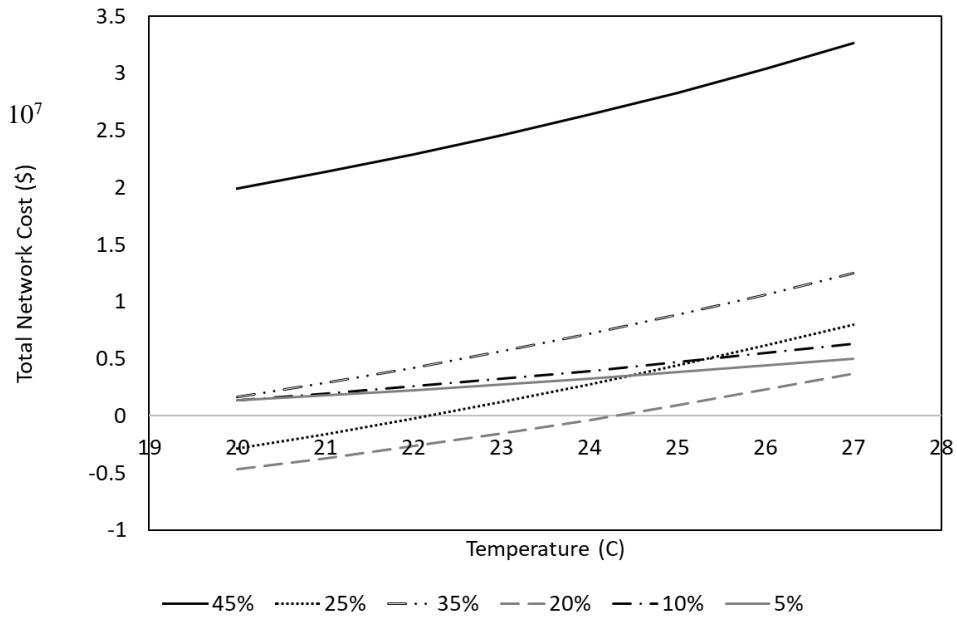


Figure 2. Total Network cost variation as a function of solvent temperature for different net capture targets

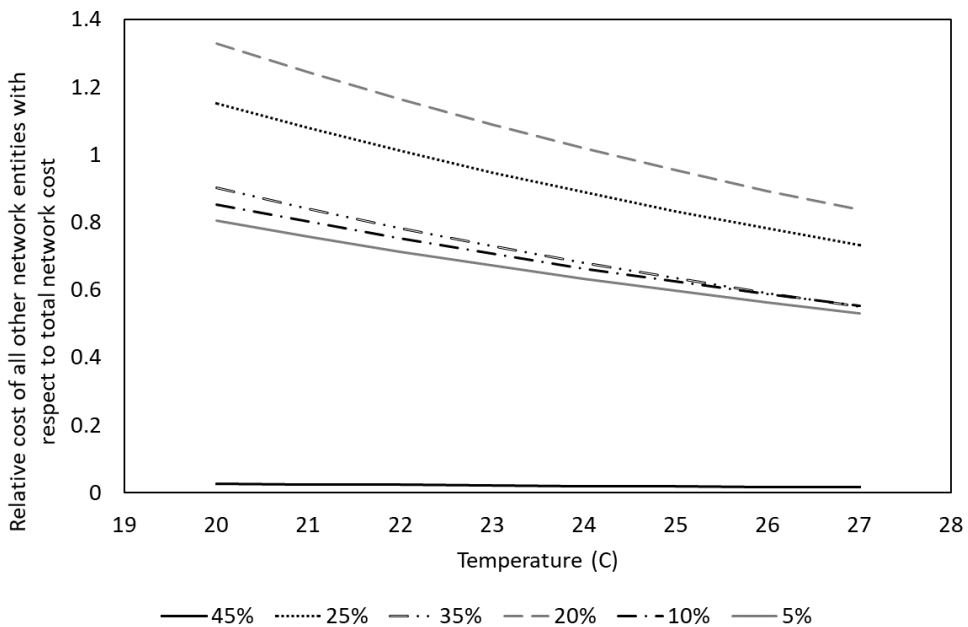


Figure 3. Relative cost variation of all remaining entities within CCUS network (compression, pumping, piping, sink processing) as a function of solvent temperature, for different net capture targets

Furthermore, the 45% capture case required the most treatment, when compared to the rest of the cases. Finally, the relative cost variation of all remaining entities within CCUS network (compression, pumping, piping, sink processing), as a function of solvent temperature for different net capture targets are illustrated in Figure 3. It is evident that when the relative cost of all those entities combined (with respect to the total cost) is observed, and a decrease in the overall cost of those entities combined has been observed across the temperature range which has been studied. The least sensitive case was the 45% capture case, as it required the most treatment, to meet the capture target.

5. Conclusions

The solvent temperature effect was investigated for different net capture CO₂ reduction targets. It was observed that the total network cost increases with solvent temperature. It was also found that at low CO₂ targets, the temperature did not affect the overall cost, while the 45% capture case was the most expensive. The increased cost was not solely due to the temperature of the solvent but also due to the deployment of non-revenue sinks such as geological storage.

References

- D. M. Al-Mohannadi, P. Linke, 2016, On the systematic carbon integration of industrial parks for climate footprint reduction, *J. Clean. Prod.*, 112 (5) 4053-4064.
- D. M. Al-Mohannadi, P. Linke, S. K. Bishnu, S.Y Alnouri, 2015, Interplant Carbon Integration Towards Phased Footprint Reduction Target, *Comput. Aided Chem. Eng.*, 37, 2057-2062.
- E. Brunazzi, G. Nardini, and A. Paglianti, 2002, An Economical Criterion for Packed Absorption Column Design, *Chem. Biochem. Eng. Q.* 15 (4) 199-206
- F. Hussin, M. K. Aroua, 2020, Recent trends in the development of adsorption technologies for carbon dioxide capture: A brief literature and patent reviews (2014-2018), *J. Clean. Prod.*, 253, 119707 DOI: 10.1016/j.jclepro.2019.119707
- F. O. Ochedi, J. Yu, H. Yu, Y. Liu, A. Hussain, 2020, Carbon dioxide capture using liquid absorption methods: a review, *Env. Chem. Lett.* DOI: 10.1007/s10311-020-01093-8
- H. B. Liu, C. F. Zhang, and G. W. Xu, 1999, A Study on Equilibrium Solubility for Carbon Dioxide in Methyl-diethanolamine-Piperazine-Water Solution *Ind. Eng. Chem. Res.* 38, 4032-4036
- IPCC, 2005, IPCC Special Report on Carbon Dioxide Capture and Storage. Prepared by Working Group III of the Intergovernmental Panel on Climate Change [Metz, B., O. Davidson, H. C. de Coninck, M. Loos, and L. A. Meyer (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 442
- Lindo Systems, 2010, What'sBest! 9.0.5.0 -Excel Add-In for Linear, Nonlinear, and Integer Modeling and Optimization. <http://www.lindo.com>.
- L. Li, M. Maeder, R. Burns, G. Puxty, S. Clifford, H. Yu, 2017, The Henry Coefficient of CO₂ in the MEA-CO₂-H₂O System, *Energy Procedia*, 114, 1841-1847
- R. Klaimi, S. Alnouri, D. Al-Mohannadi, J. Zeaiter, P. Linke, 2018, Synthesis of Carbon Integration Networks Coupled with Hydrate Suppression and Dehydration Options, *Chem. Prod. Process.*, 13(4), 20180019.
- R. M. DiGuilio, R. J. Lee, S. T. Schaeffer, L. L. Brasher, and A. S. Teja, 1992, Densities and viscosities of the ethanolamines, *J. Chem. Eng. Data.* 37(2) 239-242

A Multi-generation system in Sabkha-Tah operating with nanofluids and a novel RO unit.

Yasser M. Abdellatif^a, Eric C. Okonkwo^a, Yusuf Bicer^a, Tareq AlAnsari^{a,b*}

^a *Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.*

^b *Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.*

talansari@hbku.edu.qa

Abstract

The extremely high temperatures in the Western Sahara region has to lead to difficulties in the supply of water and food. Sabkha-Tah in Morocco is one of the regions in the Western Sahara, which possesses conditions that prevent the farming of certain crops. Nevertheless, the region has the advantage of being in a zone 60 m below sea level. This geographical advantage has the potential to reduce energy consumption in the reverse osmosis unit (RO). An energy, water, and food nexus study that takes advantage of the region's geographical characteristics are performed. A multi-generation system that produces cooling, power, and freshwater for a greenhouse situated in the Sabkha-Tah is proposed and thermodynamically analyzed. The freshwater is supplied through a RO unit that uses the hydrostatic pressure of falling water from the Atlantic Ocean. Cooling is supplied through an absorption cooling cycle that uses the waste heat of the Rankine cycle condenser, while a solar-powered Rankine cycle is used to feed the system with the required power. The system uses a parabolic trough collector (PTC), which utilises the high solar irradiance in Sabkha-Tah to generate the energy needed by the Rankine cycle. The PTC uses Cu/Therminol VP1 nanofluid as the working fluid to enhance the efficiency of the system. The study results show a reduction in the RO's power consumption from 5.561 to 2.51 kWh/m³ by using an energy recovery turbine (ERT) and the hydrostatic pressure of the falling water. The system shows that 46.18 kW of pump work can be saved when using the hydrostatic water pressure. The overall energy and exergy efficiencies for the multi-generation system are calculated to be 60.81% and 29.76%, respectively, using TherminolVP1 in PTC, while the energy efficiency increased by 0.014% using nanofluid (Cu/TherminolVP1) with a volume fraction of 0.02.

Keywords: Reverse osmosis; nanofluids; gravitational energy; nexus; multi-generation.

1. Introduction

A multigeneration system that supplies a greenhouse in Sabkha-Tah with cooling and freshwater is proposed in the study. Figure 1 shows the diagram for the studied system with all subsystems included. The sun supplies the energy used to power the system which produces cooling, freshwater, and electricity. The system consists of several subsystems such as a greenhouse, RO unit, absorption cooling cycle, PTC, energy storage, water storage, and Rankine cycle. The annual values of solar irradiance and temperatures of the region were used in this study (Climate, 2019; Solargis, 2019). Sabkha-Tah region has a large area of 250 km² which allows for building a greenhouse with a proposed dimension of 20 m by 20 m and is to be used to grow tomatoes. The system is powered by a parabolic

trough collector (PTC) that uses Therminol VP1 and Cu/ Therminol VP1 nanofluids as the working fluids. The electricity produced by the Rankine cycle is used to meet the electrical needs of the greenhouse and also feeds the community. The desalination technology used is the reverse osmosis unit as it needs high pressure which can be met using the natural advantage of the region. The hydrostatic pressure obtained from the 60 m column of water between the Atlantic Ocean level and the depleted zone level is used along with the energy recovery turbine to reduce the energy required in the RO pump. An ammonia-water absorption cooling cycle is used to provide cooling for the greenhouse and surrounding buildings. At night-time, the greenhouse will require no cooling as the maximum temperature at night in this region is 20 °C. Moreover, an overhead water storage unit uses its potential energy to supply water to the greenhouse at night instead of the pumps.

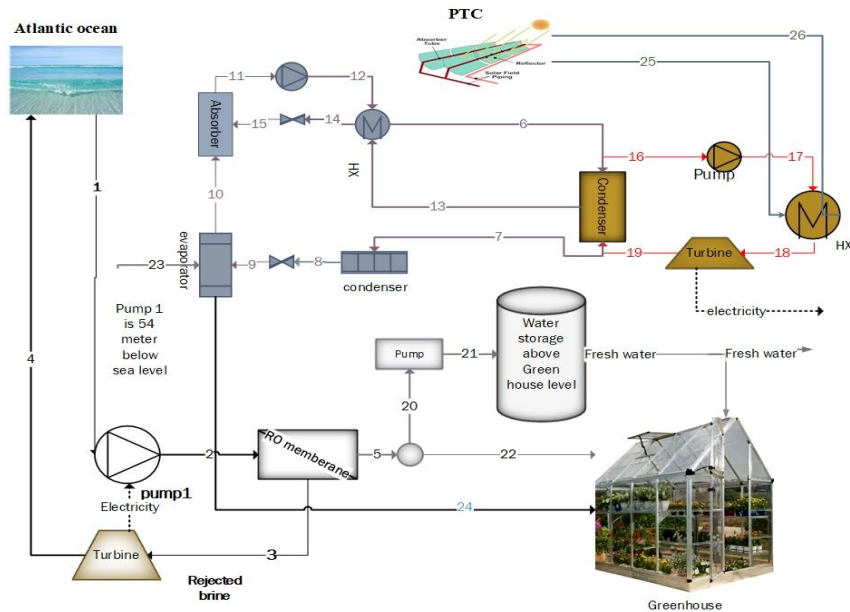


Figure 1: Schematic diagram of the proposed system for the depleted zone in Sabkha-Tah, Western Sahara, Morocco.

2. Methodology/Analysis

The proposed system is analysed using the first and second laws of thermodynamics, properties such as enthalpy, entropy, and specific exergy are determined for each component. The mass, energy and exergy balance equations were used to calculate the thermodynamics properties for each state point (Sorgulu & Dincer, 2019). The area required for one tomato plant and the water consumption per plant is calculated based on published studies (Snyder, 1914). The solar collector is modelled based on the mathematical approach presented in studies related to parabolic trough collectors (PTC) (Okonkwo, Abid, & Ratlamwala, 2019; Okonkwo, Abid, Essien, et al., 2019). Cu/TherminolVP1 is the heat transfer fluid used in the PTC and its properties were calculated using published studies (Azizi et al., 2015; Bruggeman, 1935). The mass, energy, and exergy balance equation for the components of the absorption cooling cycle are shown (Molani & Ziapour, 2019). The design specifications of the RO unit used in this work are similar to the design presented in (Kianfard et al., 2018) and the RO

mathematical model was validated against an actual RO plant that uses spiral wound type membrane. The mathematical model used for the current work is based on (Nafey & Sharaf, 2010). Finally, the overall energy and exergy efficiencies of the integrated system are defined as follows (Sorgulu & Dincer, 2019):

$$\eta_{en,ov} = \frac{\dot{W}_{net} + (\dot{m}_{fw} \times h_{fw})_{RO} + Q_{ACS,ev}}{\dot{Q}_{solar} + (\dot{m}_{sw} \times h_{sw})_{RO}} \eta_{en,ov} = \frac{\dot{W}_{net} + (\dot{m}_{fw} \times h_{fw})_{RO} + Q_{ACS,ev}}{\dot{Q}_{solar} + (\dot{m}_{sw} \times h_{sw})_{RO}} \quad (1)$$

$$\eta_{ex,ov} = \frac{\dot{W}_{net} + (\dot{m}_{fw} \times ex_{fw})_{RO} + \dot{E}x^{Q_{ACS,ev}}}{\dot{E}x^{Q_{solar}} + (\dot{m}_{sw} \times ex_{sw})_{RO}} \quad (2)$$

$$\eta_{ex,ov} = \frac{\dot{W}_{net} + (\dot{m}_{fw} \times ex_{fw})_{RO} + \dot{E}x^{Q_{ACS,ev}}}{\dot{E}x^{Q_{solar}} + (\dot{m}_{sw} \times ex_{sw})_{RO}}$$

where, \dot{m}_{rb} , \dot{m}_{fw} and \dot{m}_{sw} , \dot{m}_{rb} , \dot{m}_{fw} and \dot{m}_{sw} represents the mass flowrates of the rejected brine, freshwater, and seawater, respectively, while \dot{W}_{net} , \dot{Q}_{solar} and $Q_{ACS,ev}$ are the net power, input solar radiation and cooling from absorption cooling cycle, respectively. $\dot{E}x$ represent the exergy term.

3. Results and Discussion

Each state point in the system was thermodynamically analysed. Temperature (T), pressure (P), enthalpy (h), entropy (s), and exergy (ex), the concentration of ammonia (C_{AM}) and the mass flow rate (m) at each state point are presented in Table 1. The working fluid used is different in each cycle. The Cu/Therminol VP1 nanofluid is the working fluid used in the analysis of the PTC. Table 1 presents the physical properties at each state point.

Table 1: Thermodynamics proprieties for all state points in case of the nanofluid.

NO	T (K)	P (kPa)	h (kJ/kg)	ex (kJ/kg. K)	X (g/kg)	C_{Am} (%)	m (kg/s)
1	298	101.3	98.56	4889	43	-	83.3
2	298.6	5666	105.8	4904	43	-	83.3
3	298.6	5099	102.40	4906	61.33	-	58.33
4	298.6	656.6	98.510	4901	61.33	-	58.33
5	298.3	101.3	106.10	4898	0.240	-	25
6	346.2	1167	1609	351.1	-	100	4.35
7	303.1	1167	341.6	326.8	-	100	4.35
8	288.2	728.8	341.6	324.2	-	100	4.35
9	288.2	728.8	1476	280.9	-	100	4.35
10	303.1	728.8	-103.4	-184.5	-	53.4	36.63
11	303.2	1167	-102.6	-184	-	53.4	36.63
12	340.3	1167	112.1	-166.5	-	53.4	36.63
13	327.5	1167	4.695	-181.3	-	47.12	32.28
14	273.1	1167	-238.9	-181.5	-	47.12	32.28
15	273.2	728.8	-238.9	-182.1	-	47.12	32.28
16	353.2	47.37	334.9	18.18	-	-	1.36
17	353.5	3000	338.7	21.33	-	-	1.36
18	667.1	3000	3217	1163	-	-	1.36

19	353.2	47.37	2557	358.8	-	-	1.36
20	298.4	101.3	106100	4898	0.240	-	25
21	300.9	297.3	106500	4891	0.2408	-	12.5
22	298.4	101.3	106100	4898	0.2408	-	12.5
23	303.2	101.3	92.47	0	-	-	165.4
24	293.2	101.3	49.03	0.4493	-	-	165.4
25	670.7	141.26	-	-	-	-	7.08
26	356.5	141.26	-	-	-	-	-

Figure 2 shows how the region's depth affects the specific power consumption in the RO unit. The RO unit utilized in this study has a recovery ratio of 0.3, and it is integrated with an energy recovery turbine (ERT), which achieves lower specific energy consumption than the conventional RO unit. The power required for both units decreases as the region's depth below sea level increases. A depth of 534 m below sea level is required to generate enough pressure for the proposed RO unit to operate with no energy input from the pumps. The energy efficiency for the overall system and that of the PTC with varying solar irradiance change is presented in Figure 3. A little improvement in the PTC's energy efficiency is noticed as the solar irradiance increases. The working fluid's volumetric flow rate is optimized at solar radiation of 600 W/m^2 , beyond this, the energy efficiency is constant and begins to drop as radiation gets higher. Using Cu/Thermnoil VP1 nanofluid increases the PTC energy efficiency by 0.08% as compared to the base fluid, as can be seen in Figure 3. The overall energy efficiency is simply a ratio of the useful output from the system to the energy input. The absorption cooling cycle coefficient of performance sharply decreases as the solar radiation increases. This increase in solar radiation causes the heat available in the Rankine cycle condenser to also increase, all the excess heat is consumed in the absorption cycle's condenser. As a result, cooling remains the same and finally leads to lower overall energy efficiency. The enhancement in the overall energy efficiency was as low as 0.014% because the nanofluid usage can reduce only the input power to the PTC circulating pump. The electricity consumption in that pump is deficient compared to the overall input power to the system, hence resulting in a small enhancement.

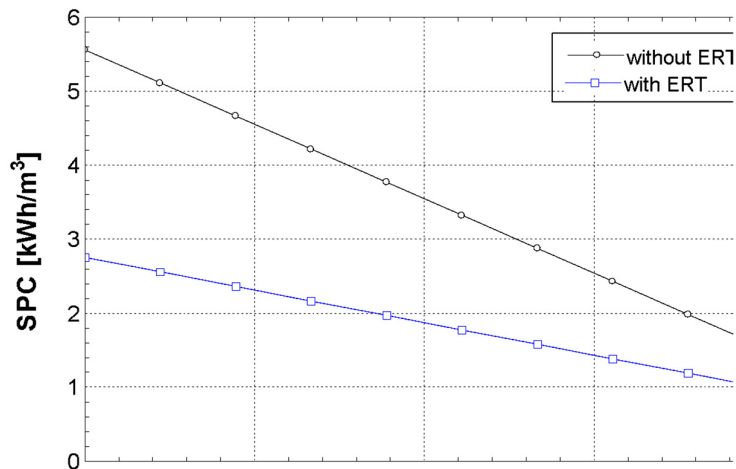


Figure 2: Effect of seawater salinity on the RO's specific power consumption (SPC).

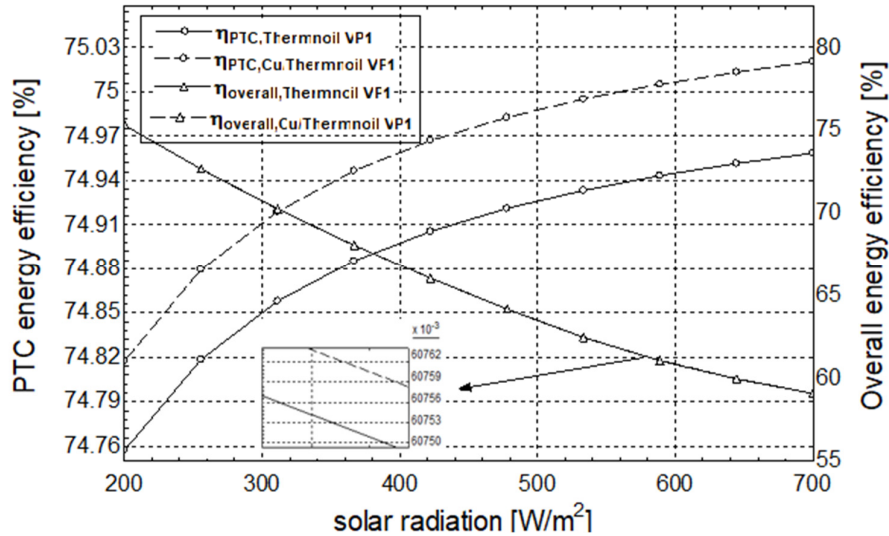


Figure 3: overall and PTC energy efficiencies versus solar radiation.

Table 2 presents the results of some key calculated outputs at an irradiance of 600 W/m 2 . It also shows a comparison between the outputs when using the nanofluid and only the base fluid in the PTC.

Table 2: The main calculated values at solar irradiance of 600 W/m 2

Parameter	Therminol VP1	Cu/Therminol VP1 nanofluid
Electricity consumption in RO	226 kW	226 kW
The volume flowrate of freshwater	90 m 3 /h	90 m 3 /h
Absorption cycle cooling output	4936 kW	4936 kW
Absorption cycle heat input	3042 kW	3044
Heat input to Rankine cycle	3940 kW	3943
Electricity production of the Rankine cycle	897.9 kW	898.9 kW
Area of greenhouse	9389 m 2	9389 m 2
Number of tomato plants	25239 plants	25239 plants
Electricity production of the whole system	636.3 kW	637.5 kW

4. Conclusions

In this study, a multi-generation system is proposed for use in Sabkha-Tah to produce energy, cooling, and freshwater for a tomato greenhouse. The system uses the geographical feature of the region, which is situated 60 m below sea level. In the proposed system, water from the Atlantic Ocean is pumped into the RO membrane to produce fresh water. The electricity and cooling needs are supplied using the absorption cooling cycle and Rankine cycle, respectively. The study presents a novel reduction in the power consumption of the reverse osmosis unit. The RO system's required power is reduced when using an energy recovery turbine and by taking advantage of the region's depleted

level, which is 60 m below the sea level. The RO-specific power consumption decreases from 3.065 to 2.511 kWh/m³. The system's overall energy and exergy efficiencies are calculated to be 60.8% and 29.76%, respectively, when base fluid (TherminolVP1) is used in the PTC. However, when using Cu/TherminolVP1 nanofluid in the PTC, the overall energy and exergy efficiency is improved by 0.08% and 0.014%, respectively.

References

- Azizi, Z., Alamdari, A., & Malayeri, M. R. (2015). Convective heat transfer of Cu – water nanofluid in a cylindrical microchannel heat sink. *ENERGY CONVERSION AND MANAGEMENT*, 101, 515–524. <https://doi.org/10.1016/j.enconman.2015.05.073>
- Bruggeman, D. A. G. (1935). Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen. *Ann. Phys.*, 5(24), 636–664.
- “Climate and Average Monthly Weather in Tarfaya, Morocco.” *World Weather & Climate Information*, 2019, weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine,Tarfaya,Morocco.
- Kianfard, H., Khalilarya, S., & Jafarmadar, S. (2018). Exergy and exergoeconomic evaluation of hydrogen and distilled water production via combination of PEM electrolyzer, RO desalination unit and geothermal driven dual fluid ORC. *Energy Conversion and Management*, 177(September), 339–349. <https://doi.org/10.1016/j.enconman.2018.09.057>
- Molani, F., & Ziapour, B. M. (2019). Performance Analyses of Ammonia–Water Absorption Cooling Cycle Combined with Enhanced Passive PVT System. *Journal of Energy Engineering*, 145(3), 04019008. [https://doi.org/10.1061/\(asce\)ey.1943-7897.0000603](https://doi.org/10.1061/(asce)ey.1943-7897.0000603)
- Nafey, A. S., & Sharaf, M. A. (2010). Combined solar organic Rankine cycle with reverse osmosis desalination process: Energy, exergy, and cost evaluations. *Renewable Energy*, 35(11), 2571–2580. <https://doi.org/10.1016/j.renene.2010.03.034>
- Okonkwo, E. C., Abid, M., Essien, E. A., Kavaz, D., & Ratlamwala, T. A. H. (2019). Olive Leaf-Synthesized Nanofluids for Solar Parabolic Trough Collector — Thermal Performance Evaluation. *Journal of Thermal Science and Engineering Applications*, 11(4), 041009 (1-13). <https://doi.org/10.1115/1.4043820>
- Okonkwo, E. C., Abid, M., & Ratlamwala, T. A. H. (2019). Comparative study of heat transfer enhancement in parabolic trough collector based on modified absorber geometry. *Journal of Energy Engineering- ASCE*, 145(3), 1–16. [https://doi.org/10.1061/\(ASCE\)EY.1943-7897.0000602](https://doi.org/10.1061/(ASCE)EY.1943-7897.0000602).
- Snyder, R. (1914). *How To Calculate Element Concentration in a Fertilizer Greenhouse Tomato Handbook*.
- Sorgulu, F., & Dincer, I. (2019). Design and analysis of a solar tower power plant integrated with thermal energy storage system for cogeneration. *International Journal of Energy Research*, 43(12), 6151–6160. <https://doi.org/10.1002/er.4233>
- SolarGIS. *Global Solar Atlas*, 23 Oct. 2019, globalsolaratlas.info/download/morocco.

A POD-ROM Methodology for Optimal Structuring of a Non-isothermal Fixed-bed Reactor for Process Integration

Katarzyna Bizon,^a Gaetano Continillo,^{b,*} Enrico A. Cutillo,^b Alfonso D'Onofrio^b

^a*Faculty of Chemical Engineering and Technology, Cracow University of Technology, ul. Warszawska 24, 31-155 Kraków, Poland*

^b*Dipartimento in Ingegneria, Università degli Studi del Sannio, Piazza Roma 21, 82100 Benevento, Italy*

gaetano.continillo@unisannio.it

Abstract

The structure of a catalytic reactor integrating two reactions is yield-optimized in terms of distribution of active sites of two catalysts in a hybrid pellet, using a two-scale (pellet and reactor) approach. POD-Galërkin model reduction is adopted to reduce the burden of the optimization algorithm. Results show that optimal structuring can have a key role in catalyst process intensification, in terms of yield, chosen as the performance indicator and thus also as the objective function for the optimization problem.

Keywords: hybrid catalyst pellet, process integration, optimal reactor structuring.

1. Introduction

A catalytic fixed-bed reactor integrating two or more reactions poses a problem of performance optimization in terms of best selection of its structural parameters. Beyond temperature and residence time, the distribution of the catalytic active sites in the solid as well as the size of the pellets play a key role. In an earlier work, the yield-optimal distribution of catalytic active sites was determined for a single spherical hybrid pellet where two consecutive isothermal chemical reactions with one reversible step under arbitrary kinetics occur (Bizon and Continillo, 2019). A Proper Orthogonal Decomposition (POD) model reduction approach was employed to limit the computational effort related to the resolution of the optimization problem. Later on, the non-isothermal case of a hybrid pellet was tackled (Bizon and Continillo, 2020), and a Discrete Empirical Interpolation Method (DEIM) approach was combined with POD-Galërkin to best deal with the strong non-linearities due to the presence of the Arrhenius terms. In this work, the study is extended to the whole reactor, and the optimal distribution of catalytic active sites within a single pellet is determined with reference to the process enhancement on the apparatus level. This is a multiscale problem, in that the state variables are distributed along the reactor length and within the pellets, thus, to limit the computational expenses, the discretized model of the pellet is reduced using POD.

2. Mathematical model and solution methodology

Two reversible chemical reactions of the first order in a catalytic fixed-bed reactor made of spherical hybrid pellets take place under non-isothermal and steady-state conditions:



The mass and energy balance equations for a pellet, written in dimensionless form, are:

$$\frac{d^2 \beta_A}{d\xi^2} + \frac{2}{\xi} \frac{d\beta_A}{d\xi} - \Phi_1^2 f_1 \frac{\tilde{r}_1}{\tilde{r}_{1,ref}} = 0 \quad (2a)$$

$$\frac{d^2 \beta_B}{d\xi^2} + \frac{2}{\xi} \frac{d\beta_B}{d\xi} + \Phi_1^2 f_1 \frac{\tilde{r}_1}{\tilde{r}_{1,ref}} - \Phi_2^2 f_2 \frac{\tilde{r}_2}{\tilde{r}_{2,ref}} = 0 \quad (2b)$$

$$\frac{d^2 \beta_C}{d\xi^2} + \frac{2}{\xi} \frac{d\beta_C}{d\xi} + \Phi_2^2 f_2 \frac{\tilde{r}_2}{\tilde{r}_{2,ref}} = 0 \quad (2c)$$

$$\frac{d^2 \theta_p}{d\xi^2} + \frac{2}{\xi} \frac{d\theta_p}{d\xi} + \delta_1 \Phi_1^2 f_1 \frac{\tilde{r}_1}{\tilde{r}_{1,ref}} + \delta_2 \Phi_2^2 f_2 \frac{\tilde{r}_2}{\tilde{r}_{2,ref}} = 0 \quad (2d)$$

where $\beta_i = C_{p,i}/C_{ref}$ denotes reactant concentration within the pellet, $\theta_p = T_p/T_{ref}$ is the temperature of the pellet and $\xi = r/R_p$ is the radial coordinate with R_p being the pellet radius. Symbols Φ_j and δ_j are, respectively, Thiele modulus and dimensionless enthalpy, and f_1 and f_2 in Eq. (2) are the local volume fractions of the pellet occupied by active sites catalysing, respectively the first and the second step of the process described by Eq. (1). The boundary conditions associated with the above balances are:

$$\left. \frac{d\beta_i}{d\xi} \right|_{\xi=0} = 0, \quad i = A, B, C; \quad \left. \frac{d\theta_p}{d\xi} \right|_{\xi=0} = 0 \quad (3a)$$

$$\left. \frac{d\beta_i}{d\xi} \right|_{\xi=1} = \text{Bi}_m (\alpha x_i - \beta_i(1)); \quad \left. \frac{d\theta_p}{d\xi} \right|_{\xi=1} = \text{Bi}_q (\theta_g - \theta_p(1)) \quad (3b)$$

Under the assumptions that the gas, flowing in a plug flow through the catalytic bed, obeys the ideal gas law and that the pressure drop can be neglected (this last assumption clearly limits the applicability of this particular model to the cases in which the assumption is verified to a significant extent), the component mass balances, the continuity equation and the energy balance for the gas are:

$$\frac{d}{d\zeta} (\alpha x_i v) + \Gamma_m [\alpha x_i - \beta_i(1)] = 0, \quad i = A, B, C \quad (4a)$$

$$\frac{d}{d\zeta} (\alpha v) + \Gamma_m \sum_i [\alpha x_i - \beta_i(1)] = 0 \quad (4b)$$

$$v \frac{d\theta_g}{d\zeta} + \Gamma_{q1} [\theta_g - \theta_p(1)] + \Gamma_{q2} (\theta_g - \theta_q) = 0 \quad (4c)$$

with x_i being component molar fraction, $v = u/u_{ref}$ – dimensionless interstitial gas velocity and $\zeta = z/L_r$ – dimensionless reactor length. The conditions at the reactor inlet are:

$$x_i|_{\zeta=0} = x_{m,i}; \quad \alpha|_{\zeta=0} = \alpha_m; \quad \vartheta_g|_{\zeta=0} = \vartheta_m; \quad v|_{\zeta=0} = v_m \quad (5)$$

More details on the model formulation and definitions of the dimensionless parameters are reported in a previous work (Bizon and Continillo, 2021). The optimization problem is to find f_1 maximizing the yield of C with respect to A, defined as:

$$\max_{f_1} Y_{CA,r} = \frac{\dot{n}_{C,out}}{\dot{n}_{A,in}} \quad \text{subject to } 0 \leq f_1 \leq 1 \text{ and } f_2 = 1 - f_1. \quad (6)$$

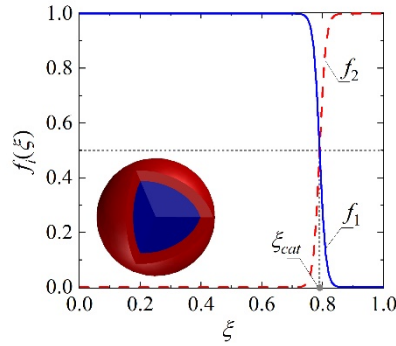


Figure 1. Distribution of f_1 and f_2 within the pellet for $\zeta_{cat} = 0.7937$.

Considering that the core-shell arrangement of the active sites within bifunctional pellets is a promising approach to process integration and intensification (Sánchez-Contador et al., 2019; Bizon et al., 2020), the optimization procedure seeks the optimal radius of the pellet core ζ_{cat} . Functions f_1 and f_2 are defined by sigmoid functions as:

$$f_1(\xi) = \frac{1}{1 + \exp[-A \cdot (\xi - \xi_{cat})]} \quad \text{and} \quad f_2(\xi) = 1 - f_1(\xi) \quad (7)$$

Figure 1 shows the distribution of f_1 and f_2 determined according to Eq. (7) with $A = 100$ and $\zeta_{cat} = 0.7937$, that gives equal pellet volume occupied by each of the two catalysts. Considering that the catalyst arrangement with the active sites enhancing the first step of the process located in the pellet core proved to overperform the reverse arrangement, only the arrangement with catalyst one located in the pellet core and catalyst two in the shell is considered in this study. The solution of the optimization problem given by Eq. (6) requires the resolution of the two-scale reactor model. The most straightforward approach consists in the transformation of the infinite-dimensional problem (Eq. (2)-(5)) into a finite-dimensional form using the finite difference method. Given the core-shell nature of the pellet, as many as $N = 101$ discrete nodes are necessary to obtain satisfactory solution accuracy for the pellet sub-model. On the other hand, to account for possible hot spots, $M = 501$ discrete nodes along the reactor are needed to discretize the gas phase balance equations. Thus, to reduce the computational effort related to resolution of the optimization problem, a model-reduction technique is adopted. Following the approach described in Reference (Bizon and Continillo, 2021), in this study Proper Orthogonal

Decomposition (POD) combined with Gal rkin projection (Holmes et al., 1996) is employed to reduce the dimensionality of the smaller scale model, i.e. the pellet sub-model. The POD-Gal rkin procedure consists in the orthogonal projection of the full-order finite-dimensional system onto a low number of POD modes generated from the full-order model (FOM) snapshots. Consider the following parameter-dependent system of N algebraic equations resulting from the discretization of the boundary value problem (Eq. (2)-(3)):

$$\mathbf{A}\mathbf{y}(\mu) + \mathbf{F}(\mathbf{y}(\mu)) = 0 \quad (8)$$

For the case here considered, the model μ represents the location of pellet within the bed and the value of ζ_{cat} for which the FOM was resolved. The POD basis, $\Psi = [\psi_1, \psi_2, \dots, \psi_N] \in \mathbb{R}^{N \times N}$, is then determined by solving the eigenvalue problem:

$$\mathbf{C}\Psi = \Lambda\Psi \quad \text{where} \quad \mathbf{C} = \frac{1}{M_s} \mathbf{Y}\mathbf{Y}^T \quad (9)$$

where \mathbf{Y} is the matrix of snapshots obtained from FOM simulation and M_s is the number of snapshots used for the determination of the POD basis. Since in this study POD-Gal rkin is used to reduce the pellet model coupled with the gas model, both embedded in the optimization algorithm, it is necessary to construct a POD basis that is able to approximate accurately the system behavior as it develops with the variation of the decision variable. Over the years, different sampling strategies were proposed to construct the so-called global POD basis (Kuan et al., 2019), however, until now there is no reliable sampling policy. In this study, the procedure based on k -means clustering (Bizon and Continillo, 2021) is employed to extract the most informative snapshots from a large number of solution profiles within the pellet. Considering that the optimal value of ζ_{cat} is expected to be close to 0.7937 (radius at which core and shell have equal volume in a sphere), the FOM was simulated for three values of ζ_{cat} , i.e. 0.75, 0.8, 0.85, 0.9 and 0.95, and then the k -means algorithm was run on the ensemble of intrapellet solution profiles collected along the reactor for all values of ζ_{cat} . For each state variable, a separate POD basis was determined using the $M_s = 100$ most informative snapshots.

3. Results and discussion

The values of the main parameters used in the simulations are summarized in Table 1.

Table 1. Main model parameters used in the numerical simulations.

Parameter	Value	Parameter	Value
D_{eff}	$10^{-6} \text{ m}^2 \cdot \text{s}^{-1}$	R_p	$2.5 \cdot 10^{-3} \text{ m}$
$k_1(T_{in}) = k_2(T_{in})$	2.95 s^{-1}	T_{in}	400 K
$k_{.1}(T_{in}) = k_{.2}(T_{in})$	0.15 s^{-1}	u_{in}	$0.7 \text{ m} \cdot \text{s}^{-1}$
L_r	8 m	λ_{eff}	$10^{-4} \text{ kW} \cdot \text{m}^{-1} \cdot \text{K}^{-1}$

Representative centroid of clusters employed to construct the POD basis for the concentration of intermediate product B within the pellet are shown in Fig. 2a, whereas Fig. 2b. shows the leading POD modes determined for β_B . The intrapellet solution profiles vary with the location of the pellet in the reactor, due to ongoing chemical process and

variation of the composition of the flowing gas, and also with the variation of the catalyst arrangement represented by the variable ζ_{cat} . As shown in Fig. 2a, the k -means procedure permits to select the most informative profiles while skipping those that are practically the same, and thus do not contribute, in terms of information content, to the POD basis.

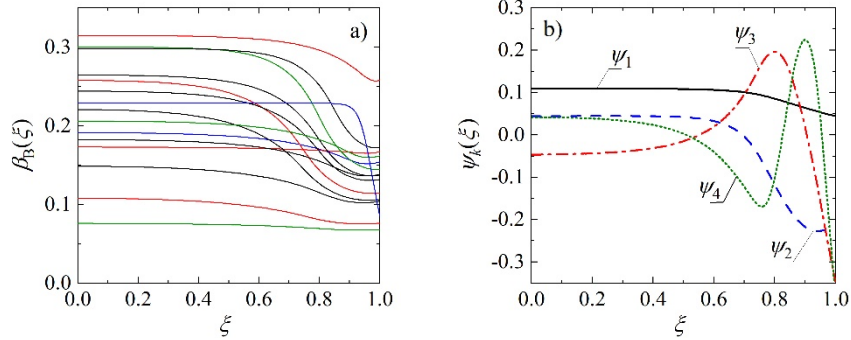


Figure 2. Representative solution profiles of $\beta_B(\xi)$ within the pellet selected using k -means clustering (a) and leading POD modes for the concentration of component B (b).

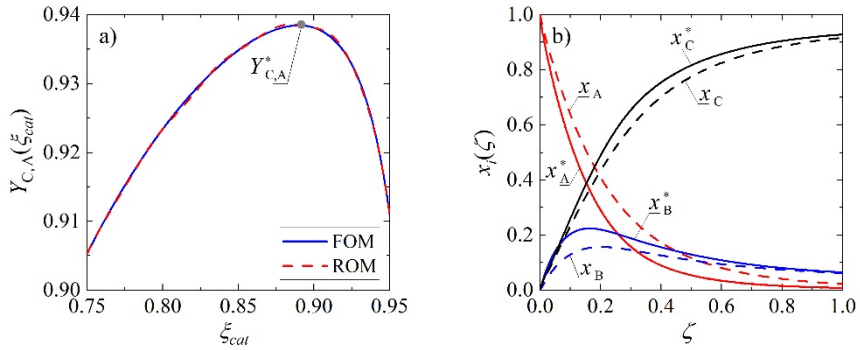


Figure 3. Values of the objective function $Y_{C,A}$ determined using FOM and ROM (a), and molar fractions of the reactants in the gas phase along the reactor determined using ROM for optimal value of ζ_{cat}^* (solid lines) and for reference value $\zeta_{cat} = 0.7937$ (dashed lines).

After analysing the convergence of the ROM constructed using different truncation order, the value of $K = 19$, for each intrapellet state variable, was selected to run the optimization algorithm. Figure 3a shows the comparison of the objective function determined using FOM and ROM, together with the location of the optimum, determined by the MATLAB *fmincon* function with the submerged *fsolve* function for the resolution of the full- and reduced-order reactor model. It can be observed that the ROM approximates very accurately both the objective function and the location of the optimum. The optimal value of ζ_{cat} results to be higher than the one corresponding to the 1:1 volume ratio, $\zeta_{cat} = 0.7937$, of the pellet occupied by the first and the second type of the catalytic active centres. Increasing the value of ζ_{cat} from the reference value equal to 0.7937 to optimal value, i.e. $\zeta_{cat}^* = 0.8901$, permits in fact (Fig. 3b) to enhance the conversion of A near the reactor inlet. Values of the optimum location and corresponding values of the objective function are in Table 2 together with the computational time expenses. Despite the relatively high number of the POD modes used to build the ROM, mainly as a

consequence of the core-shell structure of the pellet, the ROM still allows considerable time saving, both when performing a single simulation and computing the optimization of the pellet structure.

Table 2. Comparison of optimal values and computational time of FOM and ROM ($K = 19$).

Model	ξ_{cat}^*	Y_{CA}^*	Single sim. time, s	Optimization time, s
FOM	0.8901	0.9384	88.064	1573
ROM	0.8846	0.9386	18.904	385

4. Conclusions

Adopting the optimal structuring found in this study will positively impact the production and is foreseen as a practical gain in many processes of industrial interest. Although more detailed than most literature models, this approach has still some limitations in that it does not address reactor non-idealities: the hybrid plug-flow fixed-bed reactor model is in fact adopted, and the pressure drop is assumed to be negligible. The model reduction procedure complicates the preparation of the numerical code, but the computational gain is expected to be well worth the effort as the dimension of the problem is increased. Even in this low dimensional case, POD-Galërkin helps even though, with this particular problem, accuracy requires a relatively large number of modes to be included in the ROM.

Acknowledgements

G.C. is grateful to Università del Sannio for granting a sabbatical leave that made this and other research works possible. The research was partly financed by the Polish National Science Centre, project number 2017/26/D/ST8/00509.

References

- K K. Bizon, G. Continillo, 2019, Determination of the optimal distribution of active centers in a multifunctional catalyst pellet using global searching combined with reduced-order modeling approach, *Computer Aided Chemical Engineering*, 46, 1015-1020.
- K. Bizon, G. Continillo, 2020, Optimal design of a non-isothermal hybrid catalyst pellet based on POD-DEIM reduced-order methodology, *Computer Aided Chemical Engineering*, 48, 271-276.
- K. Bizon, K. Skrzypek-Markiewicz, G. Continillo, 2020, Enhancement of the direct synthesis of dimethyl ether (DME) from synthesis gas by macro- and microstructuring of the catalytic bed, *Catalysts*, 10, 852.
- K. Bizon, G. Continillo, 2021, Efficient optimization of a multifunctional catalytic fixed-bed reactor via reduced-order modeling approach, *Chemical Engineering Research and Design*, 165, 214-229.
- P. Holmes, J.L. Lumley, G. Berkooz, 1996. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, UK.
- L. Kuan, J. Yulin, C. Yushu, Y. Yongfeng, H. Lei, Z. Zhiyong, L. Zhonggang, G. Choa, 2019. Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems, *Mechanical Systems and Signal Processing*, 123, 264-297.
- M. Sánchez-Contador, A. Ateka, M. Ibáñez, J. Bilbao, A.T. Aguayo, 2019, Influence of the operating conditions on the behavior and deactivation of a CuO-ZnO-ZrO₂@SAPO-11 core-shell-like catalyst in the direct synthesis of DME, *Renewable Energy*, 138, 585-59

Effect of Pseudo-Homogeneous Concentration Based Kinetics on Hybrid Reactive Distillation Columns for Selectivity Engineering

Deepshikha Singh, Rajdeep Mukherjee, Harinder Singh, Ankur Gaur, Shabih Ul Hasan*

*Department of Chemical Engineering, Motilal Nehru National Institute of Technology
Allahabad, Prayagraj-211004, India*

E-mail: shasan@mnnit.ac.in

Abstract

Knowledge of appropriate kinetic models for the design of the reactive distillation column is critical and directly affects the performance parameters. In many industrially important reactions, a pseudo-homogeneous concentration-based kinetic model is more suitable for process design. The present work offers a design procedure of hybrid reactive distillation columns (HRD) to obtain the desired selectivity of intermediate product in reaction schemes for which a pseudo-homogeneous concentration-based kinetic model is best suited. This work is in continuation with our previous work on selectivity engineering with HRD columns [Hasan et al., 2013a, b; Hasan et al., 2014a, b; Hasan et al., 2015; Singh et al. 2018; Singh et al. 2019] in which we developed generalized algorithms for the design of HRD columns to obtain the desired selectivity of intermediate products in different types of multi reaction schemes. However, all the previous investigations are restricted to mole fraction-based kinetics. As a result, here, we examined the potential of HRD for pseudo-homogeneous concentration-based kinetic models and developed an algorithm to obtain HRD column designs of the desired selectivity.

Keywords: Hybrid Reactive Distillation, Pseudo-homogeneous, Concentration-based kinetics, Selectivity, Design

1. Introduction

For designing HRD columns or any large reactor, one should know well-determined kinetics, which includes the effects of all parameters on elementary and side reactions that can predict the actual mechanism of reactions. Although the kinetics of industrially relevant reactions having rigorous thermodynamic approach with activity coefficients are more accurate for the study purpose, however pseudo-homogeneous concentration-based kinetics equally holds good results and best suited for process design [Ronnback et al., 1997]. In this work, we developed a conceptual design algorithm to design HRD columns to obtain desired selectivity for the multi reaction systems with pseudo-homogeneous concentration-based kinetics. To illustrate the above point, a model series type reaction scheme has been studied as an example and a combined graphical-simulation algorithm is developed using the concepts of attainable region approach.

2. Design Methodology

It is well known that in a CSTR, the reaction occurs at the outlet conditions. Therefore, the reaction and mixing vectors are collinear given by material balance [Glasser et al. (1987)]. In any other CSTR type arbitrary reactor, the reaction will not occur at the outlet conditions, and hence the collinearity condition is no more valid. The composition in this arbitrary reactor can be controlled by similarly introducing separation attributes as in the case of reactive distillation.

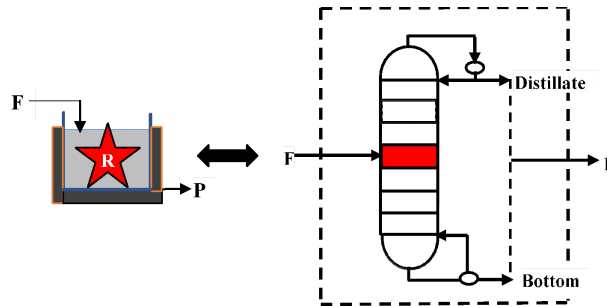


Fig. 1. The analogy between Arbitrary reactor (R) and HRD Column

Consider HRD column with only one reactive stage as shown in Fig. 1. The reaction does not occur at product stream composition (P), which is the end composition found by substantially mixing the top (distillate), bottom streams. The composition at which reactions occur in it is governed by separation attributes (vapor to feed ratio, distillate to feed ratio, number of stages, etc.). Therefore, potential reaction stage compositions that provide the desired selectivity in composition space can be obtained as discussed in the next section.

2.1 Locus of Reactive Concentrations (LRCs)

Consider a series reaction scheme (Eq. 1) with kinetic rate constants $[k_1 \ k_2] = [1 \ 1]$. The reaction occurs in liquid phase only with feed as pure reactant A.



$$[r_A \ r_B \ r_C] = [-k_1 C_A^*, k_1 C_A^* - k_2 C_B^*, k_2 C_B^*]$$

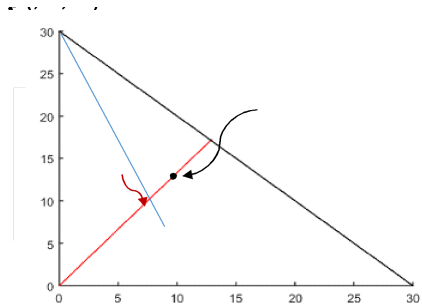


Fig. 2a. Selected reactive concentration, C^* on locus of reactive concentrations (LRCs).

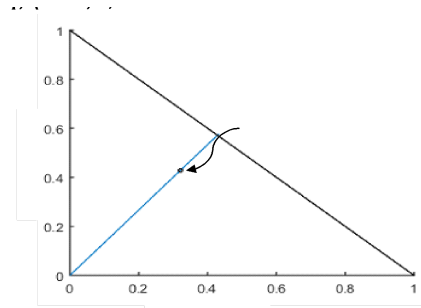


Fig. 2b. Transformed, X^* on locus of reactive stage compositions (LRSC).

Assuming constant density and volume system. The bounds in concentration space can be taken as equivalent to the reactant's initial concentration in case of single reactant schemes. Based on desired selectivity choose a point, P as shown in Fig. 2a. Now by taking the component mass balances on arbitrary reactor one can easily write Eqn. (2). The LHS of this equation is nothing but the slope of line AP. Using Eqn. (2) plot the locus of reactive concentrations, as shown in Fig. 2a.

$$\frac{C_{BP} - C_{BF}}{C_{AP} - C_{AF}} = \frac{k_1 C_A^* - k_2 C_B^*}{-k_1 C_A^*} \quad (2)$$

where,

C_{AF}, C_{BF} = Concentration of component A and B in inlet stream

C_{AP}, C_{BP} = Concentration of component A and B in outlet stream

It should be noted here that by choosing the point, P in the concentration space the conversion and selectivity gets fixed as given by Eqn. (3). For constant density and volume system ($v_0=v$).

$$\text{Selectivity} = (C_{BP}v - C_{BF}v_0)/(C_{AF}v_0 - C_{AP}v) \quad (3)$$

$$\text{Conversion} = (C_{AF}v_0 - C_{AP}v)/(C_{AF}v_0)$$

Now, one can transform the reactive concentrations into reactive mole fractions by using Eqn. (4) and thereby plot the locus of reactive stage compositions (LRSCs) in the mole fraction composition space as shown in Fig. 2b.

$$X_i^* = \sum_{i=1}^{N_c} C_i^* V_r / n, \quad X^* = [X_A^* X_B^* X_C^*] \quad (4)$$

In the RD column, if the reaction occurs at a composition belonging to any of the LRSC, our desired selectivity is ensured.

2.2 Algorithm Steps

The design methodology is based on the fact that if we separate reaction and distillation features of HRD configuration as shown in Fig. 3a, then the resulting configuration (Fig. 3b) consist of arbitrary reactor and separator system that gives the performance same as HRD configuration (Fig. 3a). This will enable us to find the desired operating parameters required to obtain the design of the desired selectivity.

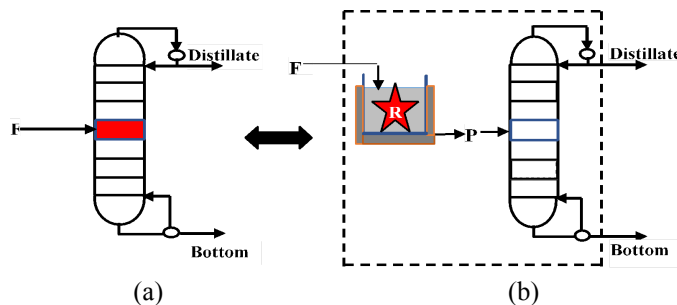


Fig. 3. (a) Hybrid (RD) column. (b) Arbitrary Reactor Separator System (Non-RD).

- i. Select any point P (Fig. 2a) in the concentration space according to the intermediate product's desired selectivity, B.

- ii. Join points, A (30, 0) and P (9, 5.25) to find the slope of line AP. Using Eqn. (2) plot the locus of reactive concentrations (LRCs). One can calculate the selectivity and conversion using Eqn. (3). For the present example, selectivity = 0.25 and conversion = 0.70.
- iii. Select any C^* on the locus of reactive concentrations as shown in Fig. 2a. For the ongoing example we have selected $C^* = [C_A^* \ C_B^* \ C_C^*] = [12, 9.6, 2.3]$. The third concentration is calculated by the summation equation (applicable for constant density and volume system only).
- iv. Specify the volumetric feed flow rate ($v_0 = 0.5$ m³/sec) to the arbitrary reactor and calculate the volume of reactor, V_R using any of the Eqn. (5) or (6).

$$C_{AF} - C_{AP} = (V_R / v_0) * r_A(C^*) \quad (5)$$

$$C_{BF} - C_{BP} = (V_R / v_0) * r_B(C^*) \quad (6)$$

- v. Calculate the total number of moles, n using Eqn. (6).

$$n = \sum_{i=1}^{N_c} n_i; \text{ where, } n_i = C_i^* V_r \quad (7)$$

- vi. Transform the selected reactive stage concentration, $C^* = [C_A^* \ C_B^* \ C_C^*] = [12, 9.6, 2.3]$ into reactive stage mole fraction, $X^* = [X_A^* \ X_B^* \ X_C^*] = [0.4267, 0.32, 0.2533]$ using Eqn. (4).
- vii. Calculate damkohler number, Da by employing the material balance of either component A or B on arbitrary reactor using Eqn. (8) or (9). Also find the concentration of component C in the product stream using Eqn. (10).

$$C_{A,0} - C_{A,P} + Da \left(\frac{-k_1 C_A^*}{k_{ref}} \right) = 0 \quad (8)$$

$$C_{B,0} - C_{B,P} + Da \left(\frac{k_1 C_A^* - k_2 C_B^*}{k_{ref}} \right) = 0 \quad (9)$$

$$C_{C,0} - C_{C,P} + Da \left(\frac{k_2 C_B^*}{k_{ref}} \right) = 0 \quad (10)$$

Where, H' = volumetric catalyst hold-up & Damkohler number,

$$Da = (H' \times k_{ref}) / v_0, \quad k_{ref} = 1$$

- viii. Calculate the flow rate of product stream P, of arbitrary reactor from Eqn. (11) that is sent to non-RD column as shown in Fig. 3b.

$$F_P = v_0 (C_{AP} + C_{BP} + C_{CP}) \quad (11)$$

- ix. Specify the hypothetical feed to Non-RD column using Eqn. (12).

$$X_{feed_NonRD} = [X_{afeed} \ X_{bfeed} \ X_{cfeed}] \quad (12)$$

where, $X_{ifeed} = (C_{iP} * v_0) / F_P$, for $i = A, B, C$

- x. Next, define the total number of stages ($N=12$) and location of feed ($N_{fl}=6$).
- xi. Find out operating parameters so that the feed stage composition of the Non-RD column coincides with chosen reactive stage composition, X^* . To achieve this draw family of curves as shown in Fig. 5, for different $V:F$ (boil-up to feed ratio) and for a range of $D:F$ (Distillate to feed ratio). Adjust the operating parameters to meet the required condition. After finding suitable parameters with the help of family of curves, plot Non-RD column profile in composition profile.

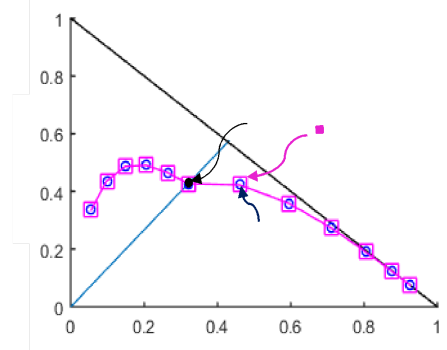
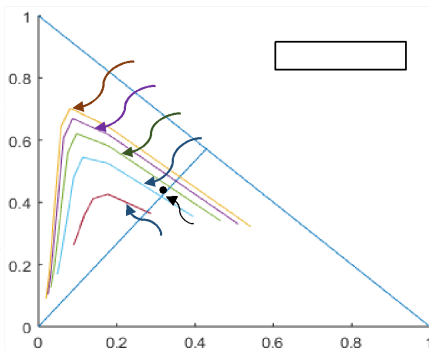


Fig. 5. Family of feed stage compositions at different values of $V:F$ (range of $D:F$)

Fig. 6. Identical non-RD and RD column profiles

- xii. Simulate the RD column using the same operating parameters and plot the column profile of RD column. The close match of column profiles shown in Fig. 6, indicates feasible design obtained with the help of a simple HRD column. Table 1 shows the feasible design specifications meeting our requirements of desired selectivity for the reaction scheme given by Eq. (1).

Table1: Design specifications of hybrid RD Column for reaction scheme having pseudo-homogeneous concentration-based kinetic model

Column configuration	Non-RD Column	RD Column
Number of components, NC	3	3
Rate Constant [$k1, k2$]	[1,1]	[1,1]
Volatility order [$A B C$]	[$B > A > C$]	[$B > A > C$]
Reaction Volume	-	0.8203
Number of stages, N	12	12
Feed location, N_{fl}	6 th Stage	6 th Stage
Feed flow rate, F	15	15
Feed composition, X_f	(0.3, 0.1750, 0.5250)	(1, 0, 0)
Feed stage composition, X^*	(0.4267, 0.3200, 0.2533)	(0.4267, 0.3200, 0.2533)
Number of reactive stages, N_r	-	1
Location of Reactive stage, N_{rl}	-	6 th Stage
Damkohler Number	-	1.6406
Distillate to feed ratio, $D:F$	0.1377	0.1377

Vapor to feed ratio, $V:F$	2.381	2.381
End Compositions after mixing the the distillate, side draw and bottom streams	(0.3, 0.1750, 0.5250)	(0.3, 0.1750, 0.5250)

3. Conclusions

A conceptual design algorithm of the hybrid RD column is developed to obtain the desired selectivity design in case of a series reaction scheme with a pseudo-homogeneous concentration-based kinetic model. Presented methodology generates multiple feasible designs which is reliant on feed location, number of moles involved in the reaction, volumetric flow rate, number of stages etc., and can be accessed in terms of operating and capital cost to determine a good/optimal design for the given set of design goals. It is applicable to single feed configurations with reactant(s)/intermediate boiling in residue curve map. The developed method provides good initialization for rigorous simulation and optimization for industrially relevant multireaction systems.

References

- Glasser, D., Hildebrandt, D., Crowe, C., 1987. A geometric approach to steady flow reactors: the attainable region and optimization in concentration space. *Ind. Eng. Chem. Res.*, 26 (9), 1803–1810.
- Hasan, S.U., Malik, R.K., Mahajani, S., 2013, Selectivity engineering with single-feed hybrid reactive distillation columns, *Chemical Engineering Science*, 99, 324–334.
- Hasan, S.U., Mahajani, S., Malik, R.K., 2013, A conceptual design algorithm for single-feed hybrid reactive distillation column involving azeotropic systems, *Computer Aided Chemical Engineering*, 32, 313–318.
- Hasan, S.U., Mahajani, S., Malik, R.K., 2014, Selectivity Engineering with Simple and Complex Hybrid Reactive Distillation Columns, *Industrial and Engineering Chemistry Research*, 53, 18526–18538.
- Hasan, S.U., Mahajani, S., Malik, R.K., 2015, Selectivity Engineering with Single Feed Multi-Side Draw Hybrid Reactive Distillation Columns, *Chemical Engineering Journal*, 278, 479–491. 5.
- Singh, D., Dutta, A., Gaur, A., Hasan, S.U., 2018, Selectivity Engineering with Hybrid Reactive Distillation Column: Mixtures containing Inerts and Multi-Azeotropes, *Chemical Engineering Transactions*, 69, 619–624.
- Singh, D., Dutta, A., Gaur, A., Hasan, S.U., 2019, Effect of non-ideal kinetics on hybrid reactive distillation models for selectivity engineering, *Computer Aided Chemical Engineering*, 46, 409–414.
- Ronnback, R., Salmi, T., Vuori, A., Haario, H., Lehtonen, J., Sundqvist, A., Tirronen, E., 1997, Development of a kinetic model for the esterification of acetic acid with methanol in the presence of a homogeneous acid catalyst, *Chemical Engineering Science*, 52 (19), 3369–3381.

An MILP model of post-combustion carbon capture based on detailed process simulation

Lukas Weimann^a, Guus Dubbink^{a,b}, Louis van der Ham^b, Gert Jan Kramer^a, and Matteo Gazzani^{a,*}

^a*Copernicus Institute of Sustainable Development, Utrecht University, 3584 CS Utrecht, The Netherlands*

^b*University of Twente, 7522 NB Enschede, The Netherlands*

Abstract

This contribution presents an integer-linear model of a post-combustion carbon capture plant featuring discrete sizing, part load operation and dynamic behavior. In particular, the model covers a design space from 200 t_{CO₂}/h (size of Petra Nova) down to 7.5 t_{CO₂}/h, a part load operation range from 50% to 100%, and a CO₂ concentration range from 7.5% to 12%. Starting with detailed, partly rate-based, models in Aspen Plus and deriving linear performance planes thereof allows to bring process model information into a linear system modeling domain for a reasonable range of design specifications.

By applying this model to a low-emission energy system design optimization, its practicability and added value could be demonstrated. The simulations show that especially for systems with high non-dispatchable energy generation, the information about the carbon capture plant's dynamic behavior is essential. To fit the scope of this paper, the mathematical formulation of the model is reported in a condensed manner. However, all information required to formulate the model is provided.

Keywords: CCS, MILP, energy system design, low emission

1. Introduction

The importance of carbon capture and storage (CCS) for decarbonizing our society is widely recognized in the scientific community and among policy makers. Although the topic of CCS experienced extensive research in the last decades, its role in integrated, low-emission energy systems has started to be addressed only recently. A reason for this is the high level of complexity of detailed energy system models, which further increases by applying detailed CCS models. Nevertheless, with the shift of conventional electricity generators like gas turbines from baseload operation to backup operation for non-dispatchable renewable energy sources, the ability of integrating CCS into energy system modeling is crucial. We aim at filling this gap by developing a linear post-combustion capture model which is compliant with the multi-energy system (MES) modeling framework developed in earlier works (Gabrielli et al., 2018a, 2018b). The model is based on simulations of a monoethanolamine (MEA) carbon capture process in Aspen Plus and features discrete sizing, part load operation and dynamic behavior. To showcase the model, we apply it to design a low-emission energy system which supplies the electricity demand of the Dutch province of Zeeland. In a multi-objective optimization for total system cost and annual CO₂ emissions, the trade-offs between renewable energy technologies (solar and wind energy) and conventional ones (gas

turbine) with CCS are investigated. Section 2 will describe the model and its verification, Section 3 shows the results of the model's application to the energy system design, and Section 4 will conclude this contribution.

2. Modeling

2.1. Process model

The model in Aspen Plus is a complete description of the MEA-based carbon capture process including absorber, stripper, heat exchangers, pumps, and a compressor for the lean vapor compression. Note, however, that compression of the CO₂ stream for any downstream activities like storage is not included. A water wash section to recover liquid from the CO₂ product stream was omitted to limit the number recycles. The equilibrium based process model was refined with a rate-based absorber model according to Amirkhosrow et al. (2020). To couple the two domains, the Murphree efficiencies obtained from the standalone rate-based absorber model were applied to the equilibrium absorber column. To account for the change of the Murphree efficiencies with changes in operating conditions, the equation reported by Zhang et al. (2016), describing the effect of flue gas flow rate, CO₂ concentration and solvent flow rate, was used. Since the solvent-to-flue gas ratio is fixed in this work, the original equation was adapted accordingly. Key constraints of the process are a capture rate of 90% and a CO₂-purity of at least 95%.

2.2. Linear performance planes

The linear model is supposed to describe the electricity and heat demands as a function of flue gas flow rate and CO₂ concentration. The flue gas flow rate is treated as part load operation for a plant of fixed size. To simplify the model, three discrete sizes are considered, i.e. design point flue gas flow rates of 15 kg/s, 80 kg/s, and 350 kg/s. For each of those design sizes, the flue gas flow rate was varied between 50% and 100% of the design flow rate (part load), and the CO₂ concentration was varied between 7.5% and 12%. Starting with a minimum number of 12 data points, enough simulations were conducted to achieve an R² value of 95% for the linear performance plane described as

$$\chi_e = \alpha_e \cdot F_e + \beta_e \cdot [CO_2] + \gamma_e \quad (1)$$

where χ_e is the demand of energy carrier e , α , β , and γ are the fitting parameters, F is the flue gas flow rate, and $[CO_2]$ is the CO₂ concentration. The parameters are shown in Table XY. The performance planes were implemented according to the following equations

$$Q_{e,t} = \sum_{i \in I} \alpha_{e,i} \cdot F_{i,t} + \left(\beta_{e,i} \cdot [CO_2]_t + \gamma_{e,i} \right) \cdot u_{i,t} \quad \forall t \in T, \quad e \in E \quad (2)$$

$$M_t^{CO_2} = \sum_{i \in I} [CO_2]_t \cdot F_{i,t} \cdot \theta \quad \forall t \in T \quad (3)$$

$$J = \sum_{i \in I} \left(k_i^c + k_i^m \right) \cdot x_i \quad (4)$$

where Q is the energy consumed, M^{CO_2} is the mass of CO_2 captured, and C is the cost; k^c and k^m are the cost coefficients for investment cost and operation and maintenance costs, respectively; u and x are booleans for operation and installation, respectively; i and t indicate the discrete plant size and time instance, respectively. Furthermore, the model is subject to the following constraints: non-negativity for all variables; only one plant size can be installed and operated; upper and lower limit for flue gas flow rate according to plant size and operation state (flow rate is zero when plant is not operating); upper and lower limit for flue gas CO_2 concentration. The dynamic behavior is constrained as described in the following section.

Table 1: Overview of fitting parameters for linear performance planes

Design Flue Gas Flow Rate kg/s	Electricity			Heat		
	α kg/s	β wt-%	γ -	α kg/s	β wt-%	γ -
15	0.004	0.05	-155.20	0.60	40.9	-7.07
80	0.004	0.30	-0.05	0.68	239.2	-44.0
350	0.004	1.39	-0.22	0.58	897.8	-150

2.3. Dynamic behavior

The dynamic response to changes in the flue gas flow rate was assessed using literature values. The dynamics for changes in CO_2 concentration were neglected since this value is not expected to fluctuate significantly. To model dynamic behavior in linear models, three key parameters are required, namely the time until equilibration, the maximum rate of change, and the time constant. The latter is obtained by fitting a first-order dynamic function such that its integral fits the integral of the real dynamic data. For more details about this approach, the reader is referred to Gabrielli et al. (2018b).

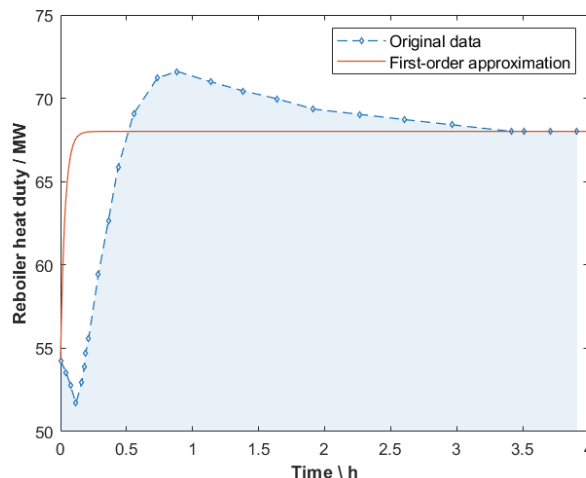


Figure 1: Original response curve and first-order approximation for the reboiler heat duty

For ramping behavior, the time until equilibration was found to be 3 h and the maximum rate of change is 20% (Bui et al., 2020; Luyben, 2013). Using the dynamic data reported in (Luyben, 2013) to derive the time constant showed that over the course of the re-equilibration process (3 h) the delayed response is compensated by a subsequent overshoot. This leads to an almost stepwise ramping behavior for the first-order dynamics. Figure 1 shows this on the example of the reboiler duty, but similar outcomes were obtained for all data of interest. Hence, the time constant was neglected and the dynamic behavior described as an instant increase in the response variable to a change of maximum 20% of flue gas flow rate, followed by 3 hours of steady-state operation, i.e. no additional change in flue gas flow rate can occur in this time period.

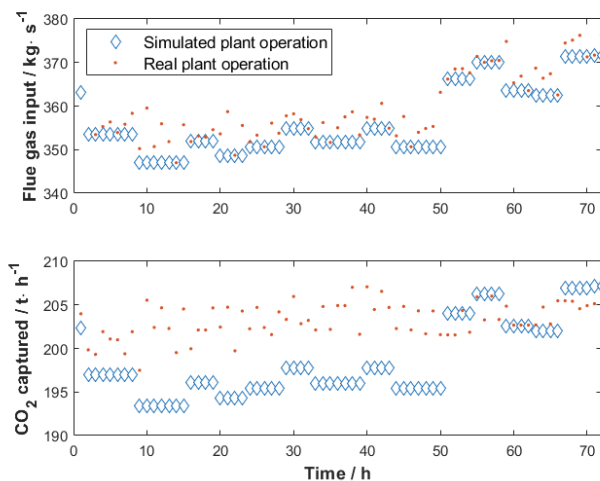


Figure 2: Comparison of real data from the Petra Nova plant with the model performance.

2.4. Model validation

The model was validated using the publicly available operation data of Petra Nova over a 72 h time period (Mitsubishi Heavy Industries, 2017). Assuming a constant CO_2 concentration, the flue gas flow rate was calculated from the published data on captured CO_2 and capture rate. Figure 2 shows the result of feeding the model with the derived flue gas data and maximizing the CO_2 captured. The dynamic limitation of the model is clearly visible, and the total CO_2 captured in the analyzed 72 h period only deviates 2% from the real data. A major contributor to this deviation is that the real capture rate is not exactly at 90% but slightly decreases from 93.6% to 87.6%.

3. Application

3.1. System description

To showcase the model developed in this work, an energy system consisting of photovoltaic panels, wind turbines, batteries, a single-cycle natural gas turbine, and the post-combustion capture unit was designed to satisfy the electricity demand of the Dutch province of Zeeland (3.24 TWh/y). The multi-objective optimization for total system cost and CO_2 emission was carried out based on hourly resolved weather and demand data from the year 2019. The natural gas price is assumed constant at 30

EUR/MWh (average 2008-2019) (Statista, 2020) and a CO₂ tax of 25 EUR/t, inspired by current European Emission Allowance prices, was applied.

3.2. Post combustion capture in low emission energy systems

To investigate the effect of carbon capture on the energy system, a set of Pareto-optimal designs was generated for the same system with and without the option for carbon capture. The resulting Pareto-fronts can be seen in Figure 3. Due to large uncertainties in the cost estimates, the findings for cost should be used for comparative rather than absolute interpretation. Furthermore, the cost of carbon sequestration or utilization is not included. It can be observed that the addition of PCC to the system improves the Pareto front. This results from the fact that the gas turbine with PCC replaces most of the wind turbine and PV installations for a wide range of designs. Only for deep decarbonization (below 0.1 t_{CO2}/MWh) renewable energy sources are used at significant extent. This can be explained by the carbon capture rate, which limits the decarbonization to 90% if all electricity was to be supplied by gas turbines. Hence, the capacity of installed gas turbines below this emission limit decreases until it reaches zero for a carbon-free design. Once passing the threshold of 0.1 t_{CO2}/MWh, the operation mode of the gas turbine switches from baseload to backup for renewables abruptly. Figure 4 shows the operation of the gas turbine (using CO₂ produced as a proxy) and the PCC units for the design allowing for 100 kt_{CO2}/y, zoomed in on a 72 h time period. The flexible operation of the gas turbine is obvious. However, less obvious but more important, the operation is constrained by the dynamics of the PCC unit which shows in the periods of steady state operation which never violate the 3-hour constraint. While the dynamics of the gas turbine and the PCC unit are not strictly coupled, the limit on CO₂ emissions does not allow the gas turbine to operate out of sync with the PCC unit.

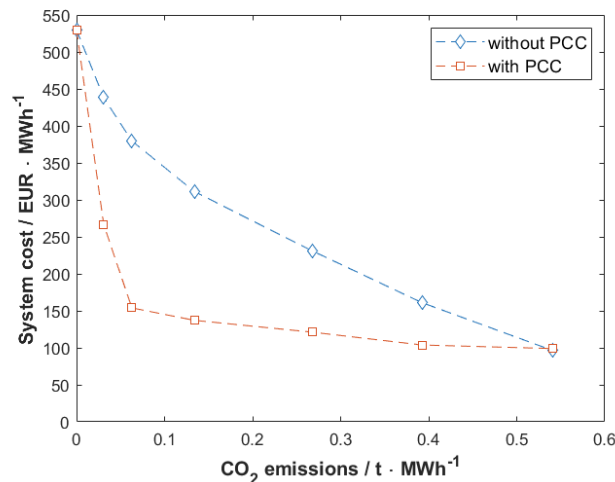


Figure 3: Pareto front for the energy system design for the province of Zeeland.

4. Conclusion

In this contribution, we derived linear performance planes to describe the electricity and heat demands as (a) a continuous function of flue gas flow rate and CO₂ concentration and (b) a discrete function of size. These performance planes are valid for a parameter range wide enough to cover most fossil energy conversion technologies found in practice. Together with the information provided on the dynamic behavior, a complete MILP can be formulated. The application of the derived model to a simple case study showed the importance of having such a model available for the design of low-emission energy systems. In particular, the dynamic behavior was found to play a significant role for deep decarbonization. The key novelty of this work is the achieved flexibility in the design of energy systems with CCS. Thus far, CCS has been applied in these kinds of optimizations in a coupled, non-flexible manner, e.g. a coal-fired power plant with 90% of its CO₂ captured. The model presented herein decouples those units and allows the optimization algorithm rather than the user/operator to decide whether CCS should be used at all, to which extent it should be used, and, if properly implemented, with which technologies it should be combined. Furthermore, the link from system level to process simulation level has proven significant added value.

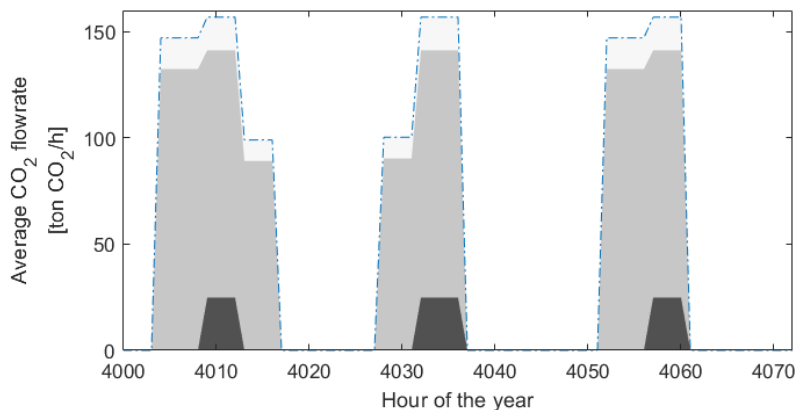


Figure 4: Operation of the gas turbine and PCC units for the design allowing for 0.03 $t_{\text{CO}_2}/\text{MWh}$, zoomed in on a 72-h time period. The dashed-blue line shows the CO₂ produced from the gas turbine and therefore represents the gas turbine operation

References

- Amirkhosrow, M., Pérez-Calvo, J.F., Gazzani, M., Mazzotti, M., Nemati Lay, E., 2020. Rigorous rate-based model for CO₂ capture via monoethanolamine-based solutions: effect of kinetic models, mass transfer, and holdup correlations on prediction accuracy. *Sep. Sci. Technol.* <https://doi.org/10.1080/01496395.2020.1784943>
- Bui, M., Flø, N.E., de Cazenove, T., Mac Dowell, N., 2020. Demonstrating flexible operation of the Technology Centre Mongstad (TCM) CO₂ capture plant. *Int. J. Greenh. Gas Control.* <https://doi.org/10.1016/j.ijggc.2019.102879>
- Gabrielli, P., Gazzani, M., Martelli, E., Mazzotti, M., 2018a. Optimal design of multi-energy systems with seasonal storage. *Appl. Energy* 219, 408–424. <https://doi.org/10.1016/j.apenergy.2017.07.142>
- Gabrielli, P., Gazzani, M., Mazzotti, M., 2018b. Electrochemical conversion technologies for optimal design of decentralized multi-energy systems: Modeling framework and

- technology assessment. *Appl. Energy* 221, 557–575.
<https://doi.org/10.1016/j.apenergy.2018.03.149>
- Luyben, W., 2013. *Distillation Design and Control Using Aspen™ Simulation, Distillation Design and Control Using Aspen™ Simulation*. John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9781118510193>
- Mitsubishi Heavy Industries, 2017. MHI's carbon capture technology.
<https://www.co2conference.net/wp-content/uploads/2017/12/4-MHI-Slides-on-the-PetroNova-Project.pdf> (accessed 10.28.20).
- Statista, 2020. Industry prices of natural gas in the Netherlands 2008-2019.
<https://www.statista.com/statistics/595650/natural-gas-price-netherlands/> (accessed 10.28.20).
- Zhang, Q., Turton, R., Bhattacharyya, D., 2016. Development of Model and Model-Predictive Control of an MEA-Based Postcombustion CO₂ Capture Process. *Ind. Eng. Chem. Res.*
<https://doi.org/10.1021/acs.iecr.5b02243>

Numerical study of fertilizer granules dynamics within rotary drum granulator

Safae Elmisaoui,^{a,b,*} Lhachmi Khamar,^{a,c} Saad Benjelloun,^a Mohamed Khamar,^b Jean-Michel Ghidaglia^{a,d}

^aMSDA, Mohammed VI Polytechnic University (UM6P), Ben Guerir, Morocco

^bLGCE, Mohammed V University in Rabat, Morocco

^cLIPIM, ENSA-Khouribga, Sultan Moulay Slimane University, Morocco

^dENS Paris-Saclay, Université Paris-Saclay, CNRS, Centre Borelli, F-91190, Gif-sur-Yvette, France

Safae.elmisaoui@um6p.ma

Abstract

The granulation process using rotary drums technology, is an essential unit operation to form good quality granules in many industries such as food processing, pharmaceutical, petrochemical and mineral processing.

The main objective of this work is to study and analyze, by CFD approach, the dynamics of solid fertilizer granules and the fluid phase, within an industrial inclined rotary drum granulator. Particularly, we studied the impact of the rotational speed, and angle of inclination of the equipment on the particles size distribution (PSD) within the solid phase. The three-dimensional CFD model that we are adapting is based on an Euler-Euler multiphase flow approach coupled with the kinetic theory of granular flow (KTGF), the appropriate model and numerical schemes were selected based on previous studies (Zheng et al. (2019)). As could be proved by experiments and simulation results, the particles size distribution depends on the rotational speed and show regime transition along the rotary drum. Our study demonstrates the capability of multiphase CFD model to predict the solid and fluid flow behaviors, and analyze the PSD of fertilizer granules inside the inclined rotary drum granulator.

Keywords: Fertilizer, CFD, PSD (Particles Size Distribution), Kinetic Theory of Granular Flow (KTGF), rotary drum, granulator

1. Introduction

Rotating drums are present in many industrial processes; pharmaceutical, fertilizers, and food industries. Their areas of applications include granulation, cooling, coating, milling, reactions, drying, and mixing (Zheng et al. (2019)). As a process, it may penalize these industries in terms of profitability, quality of finished product and respect of commercial and environmental constraints. Indeed, this operation presents a serious challenge for manufacturers because its performance depends heavily on the operating parameters and the geometrical characteristics (Epelle and Gerogiorgis 2018).

CFD is used extensively to solve numerous industrial challenges due to the maturity of computational methods and advanced computing speeds and its ability to predict the

results of real experiments. Different numerical approaches have been proposed to predict the transverse and axial particle motion and particles size distribution in rotary equipment. Physical models for the particles flow analysis have been developed to relate the flow behavior to the key operating parameters as rotational speed, filling level, and inclination angle (Santos et al. (2013)).

2. Mathematical Model development

Computational Fluid Dynamics (CFD) has proven its effectiveness as a predictive tool for complex flows modeling. Regarding multiphase CFD modeling of granular flows in rotary drums, both the Eulerian approach and the Lagrangian approach are used. Several studies (Arruda et al. (2009); Nascimento et al. (2015)) used Discrete Element Method (DEM) as lagrangian simulation approach to study different rotary drum configurations: with and without flights. The numerical results agreed well with experimental observations, but for all this cases the simulations were done either with a reduced dimension of the drum; or the particles diameter with a specific ratio that define the number of particles in the axial direction. The mean advantage of DEM simulations is their ability to consider the particles shape characteristics and allowing to the track the particles through the computational domain. However, it is limited to reduced scale equipment due to the computational time.

Alternatively, the Eulerian-Eulerian (E-E) approach is commonly adopted for CFD simulations involving several (N) phases, and these phases are assumed to form interpenetrating continua. The dynamic behavior of a multiphase system with air and fertilizer granules can be described through Multiphase Navier-Stokes (NS) equations composed by continuity and momentum equations to be solved for each phase in the Eulerian framework. Also, the E-E approach is less computationally expensive, and is the most widely applied approach in the study of hydrodynamics of multiphase flows including granular phases. Due to the continuum assumption of the particle phase, the E-E approach requires additional closure models. The kinetic theory of granular flow (KTGF) is a tool developed from the kinetic theory of gases to describe the kinematics of a granular media. It is used to describe the particle motion in fluid-solid flows due to the inherently random particle movement caused by the fluid mechanics. It (KTGF) proves its effectiveness in rotary drum systems hydrodynamic studies, when coupled to NS equations in CFD model.

2.1. Model description

To model a multiphase flow, it is essential to take into account the conservation laws and the interactions between the present phases. In this work, the Eulerian multiphase model coupled with a KTGF model is used to model the air and fertilizer granules flow inside the inclined rotary drum granulator. The model solves momentum and continuity equations for each phase. The continuity equations for gas-solid mass conservation are given by the following equations:

$$\frac{\partial(\alpha_g \rho_g)}{\partial t} + \nabla \cdot (\alpha_g \vec{v}_g \rho_g) = 0 \quad (1)$$

$$\frac{\partial(\alpha_s \rho_s)}{\partial t} + \nabla \cdot (\alpha_s \vec{v}_s \rho_s) = 0 \quad (2)$$

Where, \vec{v} and α and ρ are respectively velocity, volume fraction and density of air (gas) and fertilizer particles (solid) phases. The air volume V_g is defined by :

$$V_g = \int_V \alpha_g dV \quad (3)$$

Details on the air and solid momentum, particles phase stress description, collision and drag forces, and the particles characteristics from the KTGF are given in Santos et al. (2013).

The conservation of the gas and solid momentum is given by the following equations:

$$\frac{\partial}{\partial t} (\alpha_g \rho_g \vec{v}_g) + \nabla \cdot (\alpha_g \rho_g \vec{v}_g \vec{v}_g) = -\alpha_g \nabla p + \nabla \cdot \bar{\tau}_g + \alpha_g \rho_g \vec{g} + \alpha_g \rho_g (\vec{F}_{q,g} + \vec{F}_{lift,g} + \vec{F}_{vm,g}) \pm (K_{gs}(\vec{v}_g - \vec{v}_s)) \quad (4)$$

$$\frac{\partial}{\partial t} (\alpha_s \rho_s \vec{v}_s) + \nabla \cdot (\alpha_s \rho_s \vec{v}_s \vec{v}_s) = -\alpha_s \nabla p - \nabla p_s + \nabla \cdot \bar{\tau}_s + \alpha_s \rho_s \vec{g} + \alpha_s \rho_s (\vec{F}_q + \vec{F}_{lift,s} + \vec{F}_{vm,s}) + (K_{gs}(\vec{v}_g - \vec{v}_s)) \quad (5)$$

\vec{F}_q , \vec{F}_{lift} , \vec{F}_{vm} are the external body force, lift force and virtual mass force, respectively. Only drag and gravity are considered in the present investigation. Details on the air and solid momentum exchange coefficient, the particles phase stress description, the collision and drag forces, also the particles characteristics integrated by the KTGF are given following the model of Santos et al. (2013).

2.2. Numerical simulation conditions

The CFD code used for this study is the commercial ANSYS Fluent 2020 R2, and calculations have been developed to predict the flow behavior of a multiphase bed composed by granules and air within the flow domain defined as a rotating cylinder representing the rotary drum granulator. The rotating drum is inclined by 3.57° , and with real industrial dimensions as defined in Table 1. At the initial state, the granulator (Gr) contains a bed of granules with a height of 1250 mm and air. The granules are supposed spherical and the initial volume fraction is set to 0.65 to define the granular bed as a portion of the total domain, and the remaining spaces are filled with fluid phase (air).

The PSD is defined in the solid region by patching two superposed zones, each zone contains a granules bed with the unified diameter (Table1). We simulated the system with three different rotational speeds: 8 rpm (rotation per minute), 16 rpm and 24 rpm. The physical characteristics of the studied system and the simulated phases are presented in the following table: The Multiple Reference Frame (MRF) approach was used to define the rotational motion of the flow domain. For the boundary conditions at the inlet, a constant flow rate is specified and at the outflow of the rotary drum, pressure outlet boundary condition is assumed. No velocity slip exists at solid walls and the standard wall functions are used for near-wall treatment. A velocity magnitude is set at the inlet and zero-gauge pressure is set at the outlet.

2.3. Computational domain

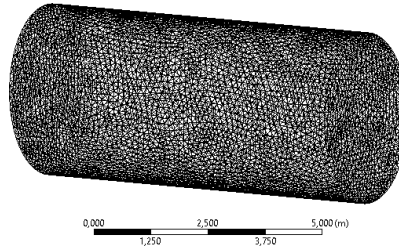


Figure 1: Mesh of the computational domain

One important step in CFD modeling is the mesh construction that represents the computational grid. A grid convergence test is done, in which the computational domain is discretized using three different grids containing a coarse, medium and fine mesh tetrahedral cells. The solid fraction distribution was determined using the different grid sizes with 8372, 66 330, and 508 440 mesh elements, representing the coarser, medium and refined mesh respectively as showed in Figure 2. Since the difference between numerical predictions obtained using the fine and the medium grids is negligible, the medium grid (Figure1) was selected to maintain a balance between computation load and numerical accuracy.

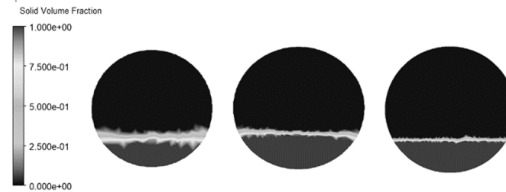


Figure 2: Cross sections from grid independence test

3. Results and analysis

3.1. Flow behavior and regimes transition for air-granules simulation

In order to study and to analyse the different regimes of the granular bed motion with an homogeneous particles distribution, the rotational speed of the granulator was varied between 8 and 24 rpm. Results in Figure 3 show the transition between the known typical transverse bed motions: rolling, cascading, and cataracting (Mellmann (2001)) during rotation. When the speed increases the arch of the kidney-shaped bed increases.

for model validation, CFD results are compared with published experiments and simulated data (Santos et al. (2013)). The hydrodynamic of the solid phase presents a transition inside the vessel at different levels, depending on the inclination angle degree. This shows the im- portance of the rotational speed in optimizing the granulation process. Each efficiency in monitoring the granulator key parameters will control the hydrodynamic flow that govern the particles size enlargements in the granulation process.

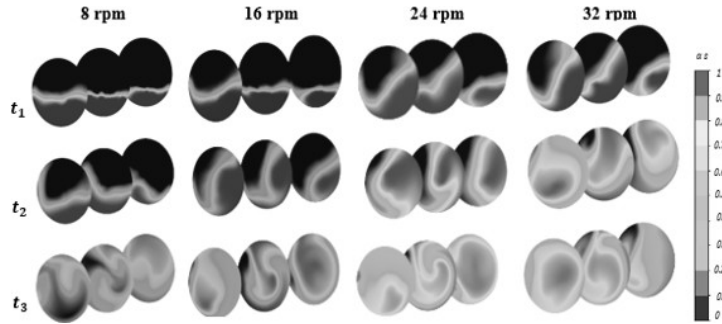


Figure 3: Regime transition for different levels of rotational speeds

3.2. Transverse flow behavior with particles size distribution (PSD)

The fertilizer granules bed was set by defining two zones with a unified diameter for each one. The diameters of particles are respectively: $1.00 \cdot 10^{-3}$ mm, and $1.00 \cdot 10^{-2}$ mm.

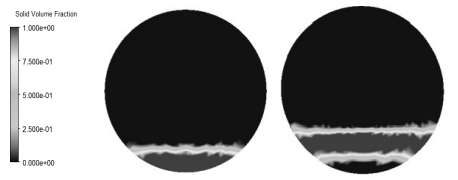


Figure 4 : Initial states of two zones with solid fraction

When the PSD is considered, the inclination angle effect is clearly distinct as the key parameter controlling the regimes transitions within the rotary drum. Figure 5 exposes the results for the three simulations with different rotational speeds. At the inlet part the dominant regime is the "slipping" one.

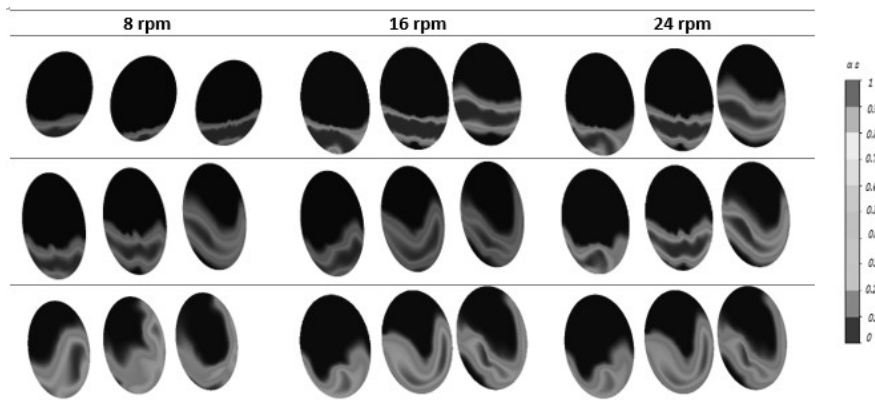


Figure 5: Particles distribution along the rotary drum (particles with d_2)

By increasing the rotational speed, the transition from the "slipping" to the "rolling regime" at the middle of the fluid domain, also the gravitational forces impact more the

particles flow behavior due to the slope at that level, and in the zone until the outlet of the drum a "crushing regime" characterizes the flow.

The computational time of the dynamic particles motion also increases exponentially with the increase of the approximately in double times the computational cost exists between the less and high values degree of rotational speeds.

4. Conclusion

Solid–gas two phase particle dynamics in a rotary drum granulator was modeled and simulated using the Eulerian Granular Multiphase Model using the Euler–Euler approach along with the Kinetic theory of granular flow. Regimes transitions is studied within the flow behavior in an industrial-size granulator and have been investigated using a three-dimensional CFD model. The simulation results for an inclination angle of 3.57 degree and rotational speed between 8 and 24 rpm, allow a prediction of the multiphase flow behavior with identification of the inclination angle and rotational speed as key parameters that control the mixing and the size enlargement of particles. The simulation results were compared to experimental studies already done using glass bills (Rong et al. (2020)). The numerical results agreed well with experiment studies measurements.

References

- E. Arruda, F. Lobato, A. Assis, M. Barrozo, 2009. Modeling of fertilizer drying in roto-aerated and conventional rotary dryers. *Drying Technology* 27 (11), 1192–1198.
- EPELLE, Emmanuel I. et GEROGIORGIS, Dimitrios I, 2018. A CFD investigation of the effect of particle sphericity on wellbore cleaning efficiency during oil and gas drilling. In : *Computer Aided Chemical Engineering*. Elsevier, p. 127-132.
- J. Mellmann, 2001. The transverse motion of solids in rotating cylinders—forms of motion and transition behavior. *Powder technology* 118 (3), 251–270.
- S. Nascimento, D. Santos, M. Barrozo, C. Duarte, 2015. Solids holdup in flighted rotating drums: An experimental and simulation study. *Powder Technology* 280, 18–25.
- W. Rong, Y. Feng, P. Schwarz, P. Witt, B. Li, T. Song, J. Zhou, 2020. Numerical study of the solid flow behavior in a rotating drum based on a multiphase cfd model accounting for solid frictional viscosity and wall friction. *Powder Technology* 361, 87–98.
- D. Santos, I. Petri, C. Duarte, M. Barrozo, 2013. Experimental and cfd study of the hydrodynamic behavior in a rotating drum. *Powder technology* 250, 52–62.
- X. Zheng, B. Jin, Y. Zhang, Y. Zhang, C. Zhou, 2019. Numerical simulation of flow characteristics in an inclining rotating kiln with continuous feeding. *International Journal of Chemical Reactor Engineering* 17 (10).

Application of PharmaPy in the digital design of the manufacturing process of an active pharmaceutical ingredient

Daniel Casas-Orozco^a, Daniel Laky^a, Vivian Wang^b, Mesfin Abdi^b, Xin Feng^b, Erin Wood^b, Gintaras V. Reklaitis^a, Carl Laird^{a,c}, Zoltan K. Nagy^{a,*}

a. Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN 47907 USA.

b. Food and Drug Administration, Silver Spring, MD 20993 USA

c. Sandia National Laboratories, Albuquerque, NM 87123 USA

** Corresponding author: zknagy@purdue.edu*

Abstract

Flowsheet design and optimization constitute one of the key challenges in the chemical engineering and process optimization communities. Software tools for digital design and flowsheet simulation are readily available for traditional chemical processing problems such as distillation and hydrocarbon processing, however tools for pharmaceutical manufacturing are much less widely developed. This paper introduces, PharmaPy, a Python-based modelling platform for pharmaceutical facility design and optimization. The versatility of the platform is demonstrated in simulating continuous, batch, and hybrid process flowsheets.

1. Introduction

Pharmaceutical products have been traditionally manufactured in batch facilities, permitting the flexible use and operation of equipment for the production of a variety of low volume – high value products. However, the transition to continuous operation is desirable given the advantages with regard to controllability, safety, consistency of product quality, reduced manual labor, and reduced equipment sizes, often leading to significant decreases in capital and operating expenditures.

Although end-to-end continuous operation is a desirable goal, the physics and chemistry of the phenomena involved can prevent the processing tasks required by the process from being operated in a truly continuous mode, making the best operating mode a combination of batch/continuous/semibatch processing steps. This hybrid operation poses challenges for process modeling and design as a consequence of the dynamic and often discontinuous nature of the resulting flowsheet model. Moreover, the combinatorics of the potentially large number of possible flowsheet configurations resulting from the possible choices of unit operation types and operating modes can present significant computational challenges.

To address these challenges, it is desirable to develop a tool that allows simulation of a representative set of dynamic drug substance (DS) and/or drug product (DP) unit operations and that offers a consistent framework for material/energy interchange between these operations. Such a modeling tool would enable rigorous comparison between unique flowsheet configurations in terms of technical viability and economic

performance, and would allow process design to be aligned with the Quality-by-Design and Quality-by-Control paradigms (Food and Drug Administration (FDA) 2004).

Commercial flowsheet simulation tools, such as Aspen Plus (2013) and gPROMS (Process Systems Enterprise 2020), are widely available and certainly are sufficient for conventional chemical process simulation and design. However, the focus of these software packages is not on pharmaceutical manufacturing. PSE does offer gPROMS FormulatedProducts as a platform for digital twin analysis of active pharmaceutical ingredient (API) synthesis and drug product processing, but the scripting capabilities, such as those seen within the IDAES framework (Lee et al. 2021), are lacking. We want to address the gap in available software by providing an open-source tool that enables flexibility in user control while still offering adequate modeling fidelity and numerical accuracy. Note that the goal is not to replace tools like gPROMS FormulatedProducts, rather it is to offer a complementary platform to fill the gap in end-to-end hybrid process design and operation that these established tools do not address.

2. Description of PharmaPy

The software tool PharmaPy is implemented in Python to incorporate powerful, pre-existing computational packages and utilize flexible scripting capabilities. It has two major themes as a platform for the design and optimization of pharmaceutical facilities (Casas-Orozco et al. 2020). The first theme addresses process modeling and simulation. PharmaPy utilizes state-of-the-art numerical tools, such as Sundials (Hindmarsh et al. 2005), which are required for the efficient solution and inter-communication of dynamic DS/DP pharmaceutical unit operation models. The package employs the sequential modular equation-solving strategy (Hillestad and Hertzberg 1986) and an object-oriented software architecture and organization of the model library.

The second theme addresses superstructure and flowsheet optimization. Here, PharmaPy utilizes a semi-heuristic algorithmic structure. First, we coordinate the generation of flowsheet alternatives using full-space combinatorial analysis. Subsequently, the full space of flowsheet designs is reduced using design feasibility constraints along with an experimentally conscious rule-based system. Finally, the reduced set of flowsheet alternatives may be compared in a simultaneous optimization scheme in Pyomo.GDP (Chen et al. 2018). In addition to these two computational themes, PharmaPy provides an open-source collection of common pharmaceutical unit operation models, which will be accessible to users through the PharmaSystems Hub (<https://pharmahub.org/groups/pharmasystems>).

The process of utilizing PharmaPy for flowsheet simulation and/or optimization begins with the creation by the user of a data file containing the physical and thermodynamic properties of participating species. Next, the pharmaceutical unit operations and their connections are selected and/or defined using text-based, object-oriented programming in Python. The flowsheet may then be executed in a manner appropriate to the function being performed, i.e. parameter estimation, process simulation or flowsheet optimization. The text-based, open-source representation of PharmaPy allows the process of running flowsheets, especially when analysing various operating conditions and/or equipment specifications, to be systematized with a high degree of user control.

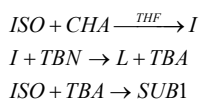
Ultimately, users may then employ the second theme, mentioned above, to compare and contrast alternative flowsheet representations in an automated, direct optimization framework. This second theme is still in development stages, thus flowsheet alternatives in this study are generated and compared manually. In the next section, analysis focuses

on the first theme of PharmaPy and is illustrated with simulations of three concrete flowsheet alternatives, comparing and contrasting API production under continuous, batch, and hybrid operating modes.

3. Case studies

3.1. Description of the flowsheets

Three flowsheets describing the synthesis and separation of lomustine were used as proof-of-concept case studies to demonstrate the capabilities of PharmaPy. Lomustine is an API for the treatment of brain cancer, for which a continuous synthesis framework has been recently reported (Jaman et al. 2019). Lomustine (L) is produced in a sequence of two reactions: in the first reaction, an intermediate (I) is synthesized from 2-chloroethyl isocyanate (ISO) and cyclohexylamine (CHA); in the second reaction, species I reacts with tert-butyl nitrite (TBN) to produce lomustine and tert-butyl alcohol (TBA). A third, undesired reaction between ISO and TBA yields a condensation subproduct (SUB1).



The first flowsheet alternative for lomustine synthesis is the continuous process (except the last recovery step, done by batch filtration) shown in Figure 1. The flowsheet model consists of two plug flow reactors (PFRs) in series, followed by a solvent switch operation (VAP01) in which THF is replaced by Heptane (C7), facilitating improved lomustine crystallization. Afterwards, a slurry containing lomustine crystals is formed in a mixed suspension, mixed product removal (MSMPR) cooling crystallizer (CR01), modelled by a one-dimensional population balance including nucleation, size-independent growth, and dissolution. The output stream from CR01 is collected in a dynamic receiver HOLD01 (which operates in semibatch mode), and then, the collected material is sent to a batch filter (F01), where lomustine crystals are recovered. Reactors R01 and R02 are initially filled with THF, whereas VAP01 and CR01 have their holdups initially filled with C7. All units are initialized at standard temperature and pressure, except for VAP01, which is required to operate at vapor-liquid equilibrium under vacuum.

The reactions are assumed to be first order with respect to participant species. Lomustine solubility (S) in C7 as a function of temperature (T) was fitted to a polynomial form of the type $S = A + BT + CT^2$, using experimental data, and further modified to the form $S' = (1 + \alpha x_{THF}) \cdot S$ (x_{THF} : liquid mole fraction of THF, α : empirical factor) to account for the significant increase of lomustine solubility caused by the presence of THF in the solution (Mackey et al. 2020). The inhibition effect of species other than lomustine on its crystallization growth kinetics was represented by a simplified multi-impurity model (Borsos, Majumder, Nagy 2016):

$$p_{imp} = 1 - \alpha \sum_i \left(\frac{K_i C_i}{1 + \sum_j K_j C_j} \right)$$

where α is an effectiveness factor, C_i are mass concentrations (kg m^{-3}) and K_i are adsorption constants ($\text{m}^3 \text{kg}^{-1}$). The second flowsheet is an end-to-end batch version of Figure 1, with R01 and R02 modeled as batch, stirred tank reactors. The solvent switch operation in VAP01 is carried out by mixing the contents from R02 with C7 after the

reaction is completed, and then boiling off the solution until a desired THF composition is attained. Batch crystallization is performed with a linear cooling temperature profile.

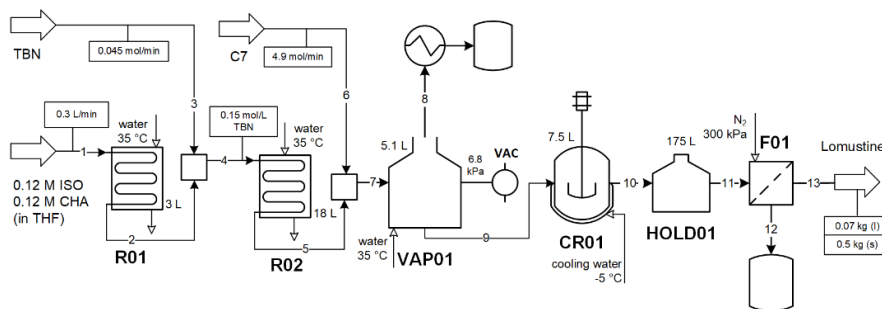


Figure 1. Schematic of the continuous flowsheet

The third flowsheet is a hybrid operation, where lomustine is synthesized continuously in R01 and R02 (Figure 1), and the resulting effluent (collected in HOLD01) is transferred to VAP01, where semibatch solvent switch is executed by vaporizing the solution while continuously feeding C7. Finally, batch crystallization and batch filtration are carried out under the same conditions as the end-to-end batch process.

Regardless of the operation mode, two key factors are to be considered for process design purposes: i) degradation temperature of lomustine (308 K, Mackey et al. 2020), and lomustine solubility in the liquid phase. These factors must be continuously monitored during the solvent switch operation in VAP01, where significant temperature and solvent composition changes are expected. The dynamic models used to describe the operation of the evaluated operating modes consist of systems of ODE/DAEs, and are drawn from the literature (Rawlings, Miller, and Witkowski 1993; Sahlodin, Watson, and Barton 2016). The balance equations which are PDEs (i.e. the PFRs and crystallizers) are discretized either by upwind or high-resolution schemes (LeVeque 2002). The continuous and hybrid operation explicitly considers the start-up of each piece of equipment from the initial conditions described above. Lomustine production for each flowsheet is 0.5 kg of dry lomustine in the exiting filtration cake. Residence times, and cycle times, in R01 and R02 were kept equivalent between the two flowsheets respectively.

3.2. Results

As shown in Table 1, more operation time must be allocated to produce 0.5 kg lomustine in the continuous flowsheet compared to the batch and hybrid flowsheets. It should be noted, however, that the initial transient state that must be completed before attaining steady-state operation contributes to this deficiency. This effect can be observed in Figure 2, where there is an initial period with no lomustine in CR01 and HOLD01.

This start-up period results in no lomustine crystal production during the first 1.6 hours of operation.

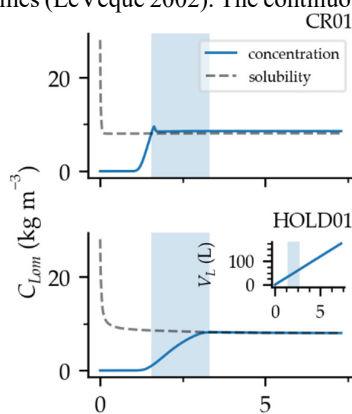


Figure 2. Concentration profiles for continuous crystallization-receiver set.

Table 1. Processing times

Unit	Time (min)				
	Batch	Residence		Operation	
		Continuou s	Hybri d	Continuou s	Hybri d
R01	10	10	10		
R02	60	60	60		
VAP01	102	5	--	438	
CR01	60	18.8	--		60
F01	9	--	---	90	2
Total	241	--	---	528	370

As seen in Figure 2, by the time crystals are first formed in CR01 (ca. 1.6 h), 37.3 L of liquid without any lomustine (a mixture of mostly C7 and THF) is already accumulated in HOLD01. This means that for the time period covered by the shaded region in Figure 2, crystals in the stream entering to HOLD01 get diluted to some extent, since this slurry gets mixed with a volume of sub-saturated liquid. This ultimately contributes to low lomustine productivity observed for the continuous process.

Another important factor to consider is the attainable degree of separation in VAP01 for the continuous case. Figure 3 shows a steady-state VAP01 composition of 22 and 76 mol% THF and C7, respectively, operating between 6.3 kPa and 6.9 kPa and a heat transfer media temperature of 308 K. This final remaining composition of THF in the outlet stream of VAP01 increases the solubility of lomustine in C7 by a factor of 2.4, inhibiting lomustine crystallization in the continuous case.

Moreover, the equilibrium temperature in VAP01 for the continuous case must be relatively low, here at 291 K, to ensure solubility constraints are met by the THF/C7 solvent mixture. In contrast, vaporization in the batch case realizes much greater separation of THF in VAP01. These conditions result in a final solution with 12.7 kg/m³ of lomustine in VAP01 (lower than the continuous case, Figure 3) but with much lower THF content (1.4 mol%) and higher temperature (301 K). The differences in vaporization lead to differing final crystallization temperatures, 268 K in continuous and 283 K in batch, to achieve similar recovery of lomustine. These results suggest that diverting material from CR01 during the startup period could increase crystal yield. Moreover, an improved solvent switch operation such as a staged distillation column may be necessary to circumvent the high solubility caused by the remaining THF, especially in the continuous operation, after the vaporization step proposed in this work.

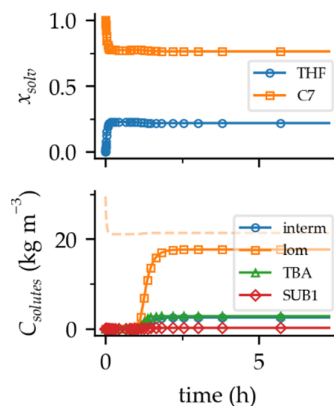


Figure 3. Solvent molar fractions (top) and mass concentration of solutes (bottom) in VAP01 for the continuous flowsheet. Dotted line: Lomustine solubility.

Under the hybrid mode, the semibatch solvent switch allows operation at lower temperatures compared to the batch case (slow addition of C7) which results in slower evaporation of THF and subsequently a final liquid with higher THF composition. Given that THF increases lomustine solubility, batch crystallization in the hybrid case operates at higher initial concentration, which leads to increased yield and less reaction time needed to manufacture 0.5 kg of lomustine crystals.

4. Conclusions

This work demonstrates the features of PharmaPy to analyze a diverse set of pharmaceutical flowsheets, through the use of dynamic, first-principle models. In the case study presented, start-up effects and thermodynamic limitations mark the key differences and challenges in translating an end-to-end batch process to a fully continuous operating.

5. Acknowledgments

Funding for this work was made possible, in part, by the Food and Drug Administration through grant (U01FD006738). Views expressed in written materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. Disclaimer: This paper describes objective technical results and analysis. Any subjective views or opinions that may be expressed in this paper do not necessarily represent the views of the U.S. Department of Energy of the United States Government.

References

- Aspen Plus, 2013, Aspen Technology, Inc. - USA
- Borsos, Majumder, and Nagy. 2016. "Multi-Impurity Adsorption Model for Modeling Crystal Purity and Shape Evolution during Crystallization Processes in Impure Media." *Cryst. Growth Des.* 16 (2): 555–68.
- Casas-Orozco, Laky, Reklaitis, Laird, Lee, Abdi, Feng, Wood, and Nagy. 2020. "PharmaPy: A Process Synthesis Framework for Hybrid Pharmaceutical Manufacturing Systems." In *AICHE Meet.*, 2. Paper 235e.
- Food and Drug Administration (FDA). 2004. "Pharmaceutical CGMPs for the 21st Century - A Risk-Based Approach." Rockville, MD. <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/Manufacturing/QuestionsandAnswers/CurrentGoodManufacturingPracticescGMPforDrugs/UCM071836>.
- Hillestad, and Hertzberg. 1986. "Dynamic Simulation of Chemical Engineering Systems by the Sequential Modular Approach." *Comput. Chem. Eng.* 10 (4): 377–88.
- Hindmarsh, Brown, Grant, Lee, Serban, Shumaker, and Woodward. 2005. "Sundials." *ACM Trans. Math. Softw.* 31 (3): 363–96.
- Jaman, Sobreira, Mufti, Ferreira, Cooks, and Thompson. 2019. "Rapid On-Demand Synthesis of Lomustine under Continuous Flow Conditions." *Org. Process Res. Dev.* 23 (3): 334–41.
- Lee, Ghouse, Eslick, Laird, Siirola, Zamarripa, Gunter, et al. 2021. "The IDAES Process Modeling Framework and Model Library - Flexibility for Process Simulation, Optimization and Control." *J. Adv. Manuf. Process.* <https://idaes.org>.
- LeVeque. 2002. *Finite Volume Methods for Hyperbolic Problems*. New York: Cambridge University Press.
- Mackey, Mufti, Lee, Abdi, Feng, Wood, and Thompson. 2020. "Process Design and Development of a Small Scale Hybrid Manufacturing System for the Cancer Drug Lomustine." In *AICHE Annu. Meet.* Paper 642e.

Process Systems Enterprise. 2020. "Process Systems Enterprise, GPROMS." 2020.
www.psenderprise.com/products/gproms.

Rawlings, Miller, and Witkowski. 1993. "Model Identification and Control of Solution
Crystallization Processes: A Review." *Ind. Eng. Chem. Res.* 32 (7): 1275–96.

Sahlodin, Watson, and Barton. 2016. "Nonsmooth Model for Dynamic Simulation of Phase
Changes." *AIChE J.* 62 (9): 3334–51.

Computer-Aided Screening of Deep Eutectic Solvent Systems for the Associative Extraction of α -Tocopherol from Deodorizer Distillate

Zhen Song,^{a,b,*} Teng Zhou,^{a,b} Zhiwen Qi,^c Kai Sundmacher^{a,b}

^a*Process Systems Engineering, Otto-von-Guericke University Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany*

^b*Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany*

^c*School of Chemical Engineering, East China University of Science and Technology, 130 Meilong Road, 200237, Shanghai, China*
songz@mpi-magdeburg.mpg.de

Abstract

This work presents a computer-aided framework for the screening of deep eutectic solvent (DES) systems by using the α -tocopherol extraction from methylated oil deodorizer distillates (MODD) as an example of practical relevance. Taking advantage of the differences in the hydrogen bond donating abilities of α -tocopherol and methyl linoleate (model compounds in MODD), the DES screening task is to select suitable components that can achieve the associative extraction of α -tocopherol by *in situ* DES formation at high selectivity against methyl linoleate. The COSMO-RS model is employed in the DES screening for two purposes: (1) The solid-liquid equilibria between α -tocopherol and each candidate component are calculated to check the potential for DES formation and identify the existence of a proper liquid window. (2) The infinite dilution capacity and selectivity of different components for the α -tocopherol/methyl linoleate extraction are predicted to estimate their potential for the separation task. The components preselected by COSMO-RS evaluation are further examined regarding other physical as well as environmental, health, and safety (EHS) properties.

Keywords: Eutectic solvent, component screening, COSMO-RS, α -tocopherol extraction

1. Introduction

Eutectic solvent systems, in most cases also coined as deep eutectic solvents (DES), not only have physicochemical properties similar to ionic liquids (ILs), such as negligible vapour pressure and a wide liquid phase existence range, but also feature green and sustainable characteristics such as non-toxicity, biodegradability and low cost (Smith et al., 2014). For these reasons, DES are widely expected as promising alternative to conventional organic solvents in many chemical processes, especially for the extraction of natural products from bio-resources, for instance, the extraction of α -tocopherol from methylated oil deodorizer distillates (MODD) (Qin et al., 2017). However, despite the great interests in DES, most current studies still rely on empirical knowledge to experimentally test different combinations of components (Abbott et al., 2017; Qin et al., 2019) while computer-aided methods guiding DES selection are still scarce. So far, only few works (Bezold et al., 2017; Salleh et al., 2017) have attempted to theoretically design or screen DESs, which are limited to a small set of already reported DESs or hypothetical component combinations assuming eutectic system formation. Moreover, while mainly

concentrating on the application performance of DES, their environmental, health, and safety (EHS) properties are usually taken for granted, even without an estimation of such properties of the pure components. Very recently, our group developed a quantitative correlation model for the prediction of freezing point depression to facilitate the search of novel DES and subsequently proposed a systematic DES screening method integrating the estimation of EHS properties of components (Song *et al.*, 2020a). Nevertheless, due to the limited availability of experimental data, this method only applies to DES with choline chloride (ChCl) as the hydrogen bond acceptor (HBA) and cannot directly estimate the eutectic temperature without knowing the eutectic composition. In the present contribution, we propose a general framework for screening DES systems that could virtually cover any component of interest. The extraction of natural α -tocopherol from MODD, an abundant by-product in vegetable oil refining, is taken as an industrially relevant case study.

2. Method

As α -tocopherol is a weak hydrogen bond donor (HBD) while methylinoleate (model of the undesirable MODD components) is a weak hydrogen bond acceptor (HBA), the DES screening task in the case study is formulated as to select suitable HBA components that can *in situ* form DES with α -tocopherol at high selectivity against methylinoleate. In this context, a computer-aided DES screening framework is proposed as shown in Figure 1. It consists of five steps: (1) collection of candidate components, (2) generation of COSMO files, (3) estimation of DES formation, (4) evaluation of extraction potential, and (5) assessment of physical and EHS properties. In the following, each of these steps is briefly described.

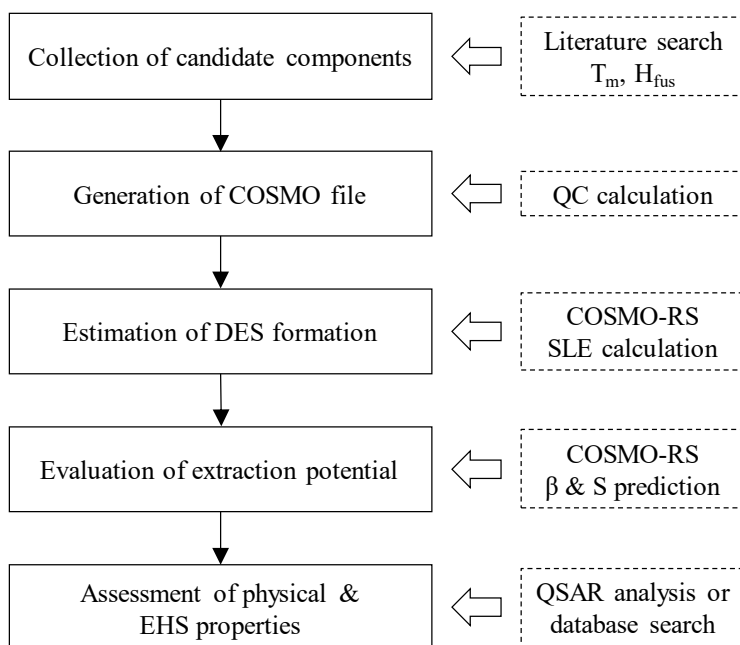


Figure 1. Framework for the computer-aided screening of DES.

2.1. Collection of candidate components

To maximize the probability of finally identifying a suitable component, the first step is to collect a significant number of candidates. However, estimating the DES formation between two components essentially requires not only the evaluation of the non-ideality of mixtures, but also the pure component fusion properties (i.e., melting temperature and fusion enthalpy or entropy). Therefore, we only collect the candidate components with known fusion properties, which include already reported components of DES systems as well as other potential ionic liquids or salts.

2.2. Generation of COSMO file

As the next two steps are based on the COSMO-RS model (Eckert and Klamt, 2002), the COSMO files of all the collected candidate components are required. In this work, all the COSMO files of the collected conventional candidates, and α -tocopherol and methylinoleate in the target mixture are already covered in the current COSMObase, and can be directly taken for use. For salt components, their COSMO files need to be calculated as ion pairs (cation-anion) since Abranches et al. (2019) have demonstrated that this manner is more reliable than the one as electroneutral mixture (cation, anion) for SLE prediction of DES systems by COSMO-RS. The details for the COSMO file calculation of salt components can refer to our earlier work (Song et al., 2020b).

2.3. Estimation of DES formation

The COSMO-RS solid-liquid equilibria (SLE) prediction is performed to check whether component combinations can form DES or not. With the required COSMO files and fusion properties of each component, the SLE of binary systems can be calculated from Eq. (1) by neglecting the secondary term related to heat capacity upon melting:

$$\ln x_i^L \gamma_i^L = -\frac{\Delta H_{fus,i}}{RT} \left(1 - \frac{T}{T_{m,i}} \right) \quad (1)$$

where ΔH_{fus} and T_m represent the fusion enthalpy and melting temperature; x and γ stand for the mole fraction and activity coefficient in the liquid phase, respectively. The liquidus curves (solubility curves) of the two component in the systems are calculated separately, and their intersection point, if it exists, is identified as the eutectic point (T_e). By comparing the predicted T_e with the T_m of two individual components, the potential formation of DES systems can be estimated.

2.4. Evaluation of extraction potential

After checking the possibility of forming DES systems, different candidate components should also be evaluated with regard to their extraction potential. In this step, the infinite dilution capacity (C) and selectivity (S) on the mass basis of different candidates for the extraction task are calculated as defined in Eqs. (2) and (3).

$$C = \frac{1}{\gamma_i^\infty} \times \frac{MW_i}{MW_{com}} \quad (2)$$

$$S = \frac{\gamma_j^\infty}{\gamma_i^\infty} \times \frac{MW_i}{MW_j} \quad (3)$$

Here, γ_i^∞ and γ_j^∞ are the infinite dilution activity coefficients of the target solute (i) and the dilute (j) in each candidate component to be evaluated; MW_i , MW_j , and MW_{com} denote the molecular weights of i , j , and candidate component, respectively. Similar criteria have

been demonstrated to be effective for screening ionic liquids as solvent with high mass-based extraction performance (Song et al., 2017) and thus are also used here for DES component selection.

2.5. Assessment of physical and EHS properties

As there is currently no general method for predicting the physical and EHS properties of DES systems, such properties of the individual components are analysed instead as empirical reference. In this step, the candidates retained after the above steps are first searched through chemistry databases (e.g., PubChem) to check whether there are experimental data on the related properties; if no experimental data are available, quantitative structure-activity relationships (QSAR) models from literature are then sought for preliminary analysis, as introduced by Song et al. (2020a). The candidates with desirable physical properties and potential EHS compatibility are finally retained as promising DES components for further studies.

3. Application

In the case study for the extraction of α -tocopherol from methylolinoleate, overall 133 candidates with known fusion properties are collected from literature. This candidate database includes 21 salts and 62 conventional components that have been involved in already reported DES systems as well as 50 salt compounds (in some cases can be referred as ionic liquids) newly introduced here.

The COSMO files of all the collected salt components are calculated by Gaussian 09 (Frisch et al., 2009) following the same procedure as introduced earlier (Song et al., 2020b). As it has been demonstrated that the conformers of salt component have a notable effect on the SLE prediction by COSMO-RS, different anion locations around cation are taken into account as initial configuration for the geometric optimization in the procedure. For each salt component, two to six stable conformers are correspondingly obtained and are selected simultaneously in the SLE calculation.

To screen components that can form a DES if mixed with α -tocopherol, the binary SLE of {candidate component + α -tocopherol} are calculated by COSMO-RS and compared with those of {candidate component + methylolinoleate}. As both α -tocopherol and methylolinoleate are of low polarity (as implied by their structures of long alkyl chain, Figure 2) and high fusion enthalpy (56.28 and 48.16 kJ/mol, respectively, as estimated by the Joback method (Joback and Reid, 1987)), COSMO-RS can be considered as a reliable approach for predicting the T_c of such systems (Song et al., 2020b).

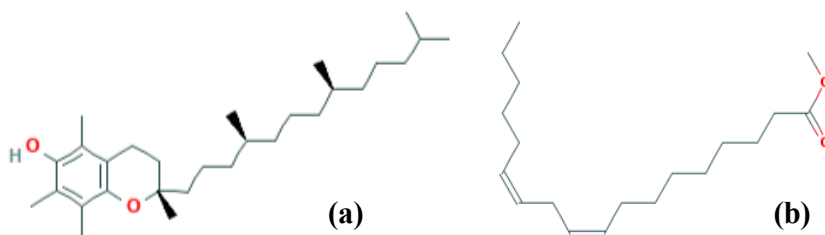


Figure 2. Molecular structures of (a) α -tocopherol and (b) methylolinoleate.

After the SLE calculation, the DES formation of different binary systems is judged by:

$$\min(T_{m,A}, T_{m,B}) - T_{c,cal} > 5 \text{ K} \quad (4)$$

where $T_{m,A}$ and $T_{m,B}$ represent the melting temperatures of two individual components, and $T_{e,cal}$ is the eutectic temperature calculated by COSMO-RS. By applying Eq. (4), only the combinations with a predicted T_e that is at least 5 K lower than the melting temperatures of both two components are selected for potential DES formation. It is worth mentioning that the judgement of “true” DES may be more physically complex than simply based on the difference between the T_e of mixture and melting temperature of pure components, as discussed by Martins et al (2019). However, as we mainly focus on the application potential here, Eq. (4) can be used for the purpose of fast screening. Based on this constraint, 36 of the 133 candidates are retained. Besides, 9 candidates with melting points below 293.15 K that are directly applicable as extraction solvents are also kept for comparison of the separation potential in the next step.

The C and S of the 45 candidates retained above for the α -tocopherol and methylinoleate extraction at 293.15 K are calculated based on Eqs. (2) and (3). As compared in Figure 3, the 9 additional candidates (with T_m below 293.15 K) have much lower C and S than the 36 candidates selected based on the judgement of DES formation. This comparison well demonstrates the advantage of the strategy of associative extraction of α -tocopherol from methylinoleate over the direct extraction approach. The best candidate with the highest C and S is $[P_{4,4,4,4}][MeSO_3]$ and the second one is $[N_{4,4,4,4}][Cl]$. Following that, $[N_{2,2,2,2}][Cl]$, $[N_{6,6,6,6}][Br]$, R-camphor, and $[N_{5,5,5,5}][Br]$ are better than the remaining ones. These six candidates can be regarded as the top six candidates from the point of view of their extractive power. When further investigating these candidates in PubChem, R-camphor is found to be flammable and hazardous to health while the other 5 salt candidates have currently not been reported to violate critical EHS criteria. It is interesting to note that $[N_{4,4,4,4}][Cl]$ has already been experimentally demonstrated to be able to form DES with α -tocopherol. It exhibits an excellent extraction performance for this separation task (Qin et al., 2017). This finding confirms the suitability of the proposed DES component design method, and indicates that the other selected candidates, especially $[P_{4,4,4,4}][MeSO_3]$, are very interesting to be assessed in future experimental studies.

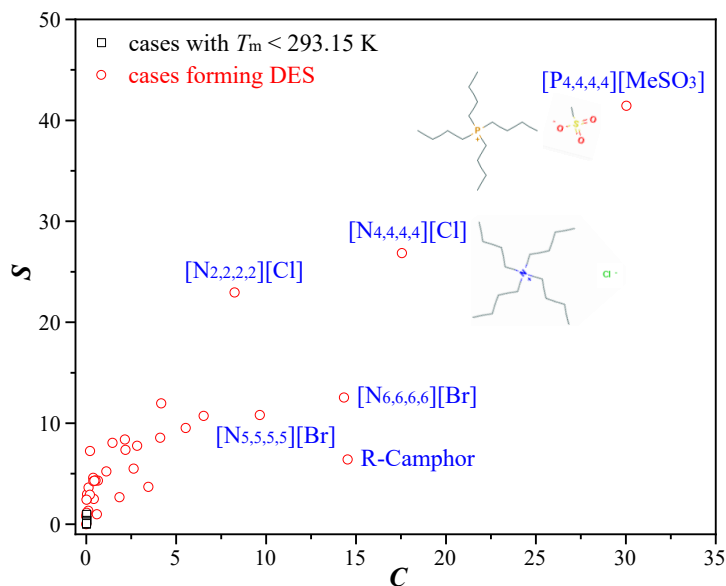


Figure 3. Comparison of the extraction potential of the 45 retained candidate components.

4. Conclusions

In this work, a computer-aided framework for the screening of DES systems based on the SLE prediction by COSMO-RS is presented and applied to the case study of searching suitable HBA components for the associative extraction of α -tocopherol from methylolinoleate. This framework could be readily extended or adjusted to other similar processes to rationally guide the DES or DES component selection prior to costly experiment studies.

Acknowledgements

The authors gratefully acknowledge the support of the Sino-German joint research project led by Deutsche Forschungsgemeinschaft (DFG) and National Natural Science Foundation of China (NSFC) under the grants SU 189/9-1 and 21861132019.

References

- E.L. Smith, A.P. Abbott, K.S. Ryder, 2014. Deep eutectic solvents (DESs) and their applications. *Chem. Rev.* 114, 21, 11060-11082.
- L. Qin, J. Li, H. Cheng, L. Chen, Z. Qi, W. Yuan, 2017. Association extraction for vitamin E recovery from deodorizer distillate by in situ formation of deep eutectic solvent. *AIChE J.* 63, 6, 2212-2220.
- A.P. Abbott, E.I. Ahmed, K. Prasad, I.B. Qader, K.S. Ryder, 2017. Liquid pharmaceuticals formulation by eutectic formation. *Fluid Phase Equilibr.* 448, 2-8.
- H. Qin, Z. Song, Q. Zeng, H.Y. Cheng, L.F. Chen, Z.W. Qi, 2019. Bifunctional imidazole-PTSA deep eutectic solvent for synthesizing long-chain ester IBIBE in reactive extraction. *AIChE J.* 65, 675-683.
- F. Bezold, M.E. Weinberger, M. Minceva, 2017. Computational solvent system screening for the separation of tocopherols with centrifugal partition chromatography using deep eutectic solvent-based biphasic systems. *J. Chromatogr. A*, 1491, 153-158.
- Z. Salleh, I. Wazeer, S. Mulyono, L. El-Blidi, M.A. Hashim, M.K. Hadj-Kali, 2017. Efficient removal of benzene from cyclohexane-benzene mixtures using deep eutectic solvents - COSMO-RS screening and experimental validation. *J. Chem. Thermodyn.* 104, 33-44.
- Z. Song, X.T. Hu, H.Y. Wu, M.C. Mei, S. Linke, T. Zhou, Z.W. Qi, K. Sundmacher, 2020a. Systematic Screening of Deep Eutectic Solvents as Sustainable Separation Media Exemplified by the CO₂ Capture Process. *ACS Sustain. Chem. Eng.* 8, 8741-8751.
- F. Eckert, A. Klamt, 2002. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J.* 48(2), 369-385.
- D.O. Abranches, M. Larriba, L.P. Silva, M. Melle-Franco, J.F. Palomar, S.P. Pinho, J.a.P. Coutinho, 2019. Using COSMO-RS to design choline chloride pharmaceutical eutectic solvents. *Fluid Phase Equilibr.* 497, 71-78.
- Z. Song, J.W. Wang, K. Sundmacher, 2020b. Evaluation of COSMO-RS for the solid-liquid equilibria prediction of binary eutectic solvent systems. *Green Energy Environ.* Doi: 10.1016/j.gee.2020.11.020.
- Z. Song, T. Zhou, Z.W. Qi, K. Sundmacher, 2017. Systematic method for screening ionic liquids as extraction solvents exemplified by an extractive desulfurization process. *ACS Sustain. Chem. Eng.* 5, 3382-3389.
- M. Frisch, G. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, 2009. Gaussian09. Inc., Wallingford CT 201.
- K.G. Joback, R.C. Reid, 1987. Estimation of pure-component properties from group-contributions. *Chem. Eng. Commun.* 57(1-6), 233-243.
- M.A.R. Martins, S.P. Pinho, J.a.P. Coutinho, 2019. Insights into the Nature of Eutectic and Deep Eutectic Mixtures. *J. Solution Chem.* 48, 962-982.

Investigating critical design criteria for photo-electrochemical reactors

Mohammed Al-Breiki*, Yusuf Bicer

Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

E-mails: malbreiki@hbku.edu.qa

Abstract

A photo-electrochemical reactor has been gaining interest and has been seeking improvements in design, production efficiency, and cost reduction. As these improvements can make a photo-electrochemical reactor compatible with renewable fuel production devices, photo-electrochemical reactors' design aspects are critical to reaching optimal performance through distinct and optimal reactor architectures. Thus, this study demonstrates designing factors related to single and double reactors following prior design decisions and performance indicators as these decisions play a significant role in achieving optimal design. Therefore, the results demonstrate that simplicity, functional convenience, and operational costs are the major decisions that need to be taken before the designing phase. Overall operation yield, current efficiency, energy efficiency, overall selectivity, and specific energy consumptions are performance indicators used to compare designed reactors. These decisions and indicators can provide insights into practical reactors' designs.

Keywords: PEC, single reactor, double reactor, design, design indicator

1. Introduction

The most abundant energy sources is solar energy, but it is scattered and intermittent, this requiring storage through a fuel conversion (e.g., ammonia, hydrogen, etc.). Photosynthesis is a route that is a viable low temperature for solar fuel generation. Artificial photosynthesis applies the same concept as photosynthesis using semiconductors that capture lights and attached to electrodes covered by catalysts. Then, catalysts generate hydrogen or other products and generate oxygen through electrochemical reactions. Extensive research has been devoted to developing solar-driven electrolysis materials; however, the design aspects of the electrochemical system have gained little attention. The design aspects of these systems are critical because the combination of the materials that offer optimum performance in such a system is significantly dependent on the system's architecture and operating conditions.

Significant efforts have been devoted to the development of photo-electrochemical (PEC) reactor performance for solar-driven fuels. One of the first works conducted by Fujishima and Honda in 1972 proposed the use of flat plates of single-junction light-absorbers attached in an electrolyte as PEC (Fujishima and Honda,

1972). After their work, multi-junction PEC reactors have been suggested. Multi-junction PECs increase photocatalysis efficiency due to improvements in the sunlight absorption (Khaselev and Turner, 1998). Additionally, two electrode reactors: a photoactive semiconducting and a metallic counter (e.g., SrTiO₃ and KTaO₃), have been suggested for solar fuel production (Domen et al., 1980). PEC systems can be used for various fuels such as hydrogen, ammonia, and methanol. Recent studies have investigated designs based on nano-structured and micro-structured components where the embedded nanoparticles in the electrolyte serve as electrocatalytic reaction sites and as a light absorber. Consequently, such a structure allows a larger specific surface for the reactions to occur and enhances the sunlight (James et al., 2009).

One of the significant components in double or compact PEC reactors is separators. Separators facilitate product extraction, increase safety, increase performance, and limit products and electrolytes to crossover, thus limit unwanted reactions to happen. Separators can be polymer membranes, thin capillaries, and porous media. Prior research proposed using multi-functional membranes that provide multiple processes, including product separation, ionic conduction, reactivity, and structure support (Yang, 2010). Such membranes can be obtained using nano-composite materials.

The design aspects of photo-electrochemical reactors are critical to reaching optimal performance through distinct and optimal reactor architectures. Some of the significant design criteria for photo-electrochemical reactors focuses on; (i) maximization of solar photons' absorption by photo-electrodes for higher production yield, (ii) minimization of product gases reaction/recombination by proper separation, and (iii) minimization of possible losses occurring between electrodes. The second and the third design criteria are typical for solar-driven reactors and conventional (electricity-driven) reactors. In contrast, the first design criteria are only for solar-driven reactors as it targets maximization of solar photons absorption. Previous studies emphasize that PEC without solar concentrating optics can only operate at a peak current density of 10-30 mA cm⁻². On the other hand, if an external electricity source is applied, the PEC is used at the current density of 1 A cm⁻². Such differences impacted the optimal PEC design (Haussener et al., 2012).

There exists a considerable body of literature on PEC reactor materials. However, the design aspects and criteria of PEC reactors are still faded. Reactor performance indicators for designers to compare and evaluate the manufactured PEC are not collected in solid form, and general rules to follow before designing the PEC are also missing in the literature. Thus, additional studies to understand the fundamental tenets of design criteria are required. In this regard, the objectives of this study are as follows:

- Presenting advantages and disadvantages of implementing electrochemical reactions in single and double reactors.
- Revealing design criteria for selection of reactor type.
- Demonstrating general rules before designing PEC reactors.
- Providing reactor performance indicators to evaluate and compare PEC reactors.

2. Advantages and Disadvantages of PEC Reactor Types

Differences between two types: single and double reactors are presented in the proceeding sections, following a brief description of each reactor.

2.1. Single PEC Reactor

Open beaker cell is a single and the most basic PEC reactor. The cell is filled with electrolyte in which the reactor body is submerged. Anode and cathode electrodes are positioned in the same electrolyte. There is no membrane in such a reactor. So, for ammonia production via a single PEC reactor, hydrogen passes through the anode and is converted into protons, which then transports through the electrolyte to the cathode over which nitrogen passes. Moreover, one of the issues facing open baker cells is that cathode and anode reaction in the same reactor body makes product quantification less convenient. Another problem of a single PEC reactor is the electrolyte leaking and air diffusion into the reactor.

Several studies suggested using proper sealing to prevent air diffusion into the reactor and electrolyte leaking (Bosserez et al., 2015). The air distribution into the reactor can cause several issues, such as overestimating oxygen production in the cell as the oxygen can act as an electron acceptor at the anode. Similarly, underestimation of hydrogen production in the cell can occur as the oxygen can recombine with hydrogen. Regarding PEC single reactor body materials, the body of these reactors can be made of glass such as Quartz and Pyrex since these material types are visible and UV-light illumination. However, these materials are expensive, so they are not preferable. Hence, polymer materials (e.g., Teflon and Poly Ether Ketone) are used in building such reactors (Bosserez et al., 2015). A single PEC reactor's significant drawback is recombination reactions due to a mixture of products (hydrogen and oxygen gases). There is also a safety concern if the hydrogen concentration reaches the explosion limit (Schröder et al., 2004). Thus, proper PEC design can collect produced hydrogen separately.

2.2. Double PEC Reactor

One of a double PEC reactor's main features is separating cathode and anode electrodes via a membrane. Various conditions can be maintained in each compartment in the double reactor. Higher Coulombic efficiency can be obtained because of the diffusion of oxygen into an anode compared to a single reactor. Most of the double PEC reactor designs are based on a proton exchange membrane (PEM) with the occurrence of water formation reaction in the opposite electrochemical direction. The membrane is placed in the middle of the two compartments where anodic and cathodic reactions occur, and it is placed with appropriate sealing to prevent leaking. In such a reactor, the critical parameters for improving reactor performance are the suitable choice of electrolyte, electrolyte volume, and ionic path length (Hernández-Pagán et al., 2012). One of the main disadvantages of double PEC is the membrane occurrence as the membrane increases the set-up cost and increases complexity. As reported in the literature, the membrane's price is 24% of the total fuel cell (Perry et al., 2020).

3. Design Criteria – Single or Double Reactor

One of the first decisions researchers/designers can take in PEC reactor design is using a single or double reactor. The single reactor operates where the cathode and anode are in a single electrolyte compartment, whereas the double reactor operates with separated cathodic and anodic electrolyte compartments by a membrane. The absence of a membrane is one of the main advantages of a single reactor. The reactor's design becomes more straightforward. The cost becomes lower as the materials associated with

the membrane, such as gaskets and fittings, are not needed in a single chamber, and membrane degradation is not considered. As recorded in the literature, the membrane's cost in a PEM fuel cell accounts for about 24% of its total cost (Perry et al., 2020). Furthermore, a single type photo-electrochemical reactor without a membrane has more negligible ohmic resistance than a double type as the membrane impedance is absent.

On the other hand, a double photo-electrochemical reactor avoids producing unwanted side reactions due to the separation of catholyte and anolyte electrodes. One of the complex decisions in the double reactor is selecting membrane-type, whether cationic or anionic. The membrane-type needs to maintain the material balance of both reactions in cathode and anode to maintain neutrality and avoid pH changes. Cationic membranes repel neutral molecules and anions, and they are stable in the alkaline environments as they produce a chemical imbalance when OH^- is consumed. Whereas anionic membranes repel neutral molecules and cations and are unstable in alkaline environments, they keep the chemical balance by replacing OH^- from the electrolyte. Besides, bipolar membranes have been introduced to overcome the challenges limitation of cationic or anionic membranes. Bipolar membranes have two layers structure of cationic and anionic, which prevent unwanted products from crossover. However, these membranes still need care as some undesired ions can cross over (Blommaert et al., 2020). The disadvantages of bipolar membranes are delamination and dehydration (Shen et al., 2017).

4. Initial Decisions During Reactor Design

Despite the diversity of PEC reactors, some general rules are required in reactor design. Noting that these rules sometimes provide conflicting requirements of the reactor design. Hence, a critical thinking approach needs to be applied to achieve a suitable design.

- *Simplicity*: The reactor design needs to meet the process requirements while keeping the design structure as simple as possible. This can be achieved by lowering the capital and operational costs and rendering the used technology.
- *Reliability and functional convenience*: The reactor must handle routine process operations such as product extraction and maintenance. For example, the reactor should facilitate inspection of electrodes and membranes followed by replacing parts or cleaning. The reactor also needs to operate in a high degree of safety, reliability, and security.
- *Integration*: The reactor should integrate into the overall process or environment directly and conveniently. Hence, the shape, space, and reactor ancillaries must be considered. It may be advantageous to have all operations such as gas separation and solvent extraction within the reactor body in some cases.
- *Reaction engineering parameters*: This category covers optimization of selectivity, mass transport rate, current distribution, and production rate. To achieve the optimal speed and selectivity of production, current distribution, and control of the potential are required. For example, high production rates involve a uniformly high mass transport over the electrode.
- *Operational costs*: These costs can be minimized by several factors, including (a) low and reliable reactor components, (b) small reactor voltage, which results in a

suitable choice of the material and shape of the electrodes, and (c) absence of high power mechanical devices for pumping purposes or electrode movement.

5. Reactor Performance Indicators

Since there is a variety of reactor designs, the need for reactor performance indicators is essential. Table 1 shows the reactor performance indicators and the expressions of the indicators and their observations. These indicators help designers to evaluate the performance of the manufactured reactor.

Table 1: Reactor performance indicators

Reactor Performance Indicator	Definition	Expression	Observations
Fractional conversion (x)	The fraction of reactant, that is consumed by the electrochemical reaction	$x = \frac{m_i - m_t}{m_i}$	m_i , initial molar amount of reactant m_t , molar amount at time t
Overall operational yield (θ_p)	The maximum molar amount of desired product obtained from 1 mole of reactant	$\theta_p = \frac{m_p}{n_p m_i}$	m_p , moles of reactant converted to product n_p , number of moles reactant consumed
Current Efficiency (\emptyset)	The yield based on electrical charge passed during electrolysis	$\emptyset = \frac{WnF/M}{q}$	W , weight of active material n , number of electrons transferred per mole reaction F , Faraday's constant M , molecular weight of the material q , total charge
Overall Selectivity (S_p)	The ratio of desired product to total products	$S_p = \frac{m_p/n_p}{\sum m_{p,i}/n_{p,i}}$	m_p , moles desired product of P n_p , number of moles of wanted product of P $m_{p,i}$, moles of all product $n_{p,i}$, number of moles of all product
Specific Energy Consumption (SEC)	The amount of substance on a mass	$SEC = \frac{-nFE_{cell}}{\emptyset M}$	E_{cell} , cell potential
Energy Efficiency (E)	It can be indicated with respect to energy yield	$E = \frac{\Delta G \emptyset}{nFE_{cell}} = -\frac{(E_e^C - E_e^A)\emptyset}{E_{cell}}$	E_e^C , equilibrium potential for cathode E_e^A , equilibrium potential for anode

6. Conclusions

This study summarizes the main design criteria for PEC reactors. As there are various PEC reactor designs, this study lists some reactor performance indicators to compare and evaluate PEC reactors, including energy efficiency, fractional conversion, current

efficiency, etc. Besides, we listed few general rules to consider when designing PEC, such as simplicity, reliability and functional convenience, integration, reaction engineering parameters, and operational costs. Design criteria for selecting the type of reactor (single or double) have been presented, and the results have shown that a double reactor has more advantages compared to a single reactor as different conditions can be maintained in each compartment of the double reactor, higher Coulombic efficiency can be obtained because of diffusion of oxygen into the anode. However, the existence of a membrane causes more complexity in the reactor.

References

- Blommaert, M.A., Verdonk, J.A.H., Blommaert, H.C.B., Smith, W.A., Vermaas, D.A., 2020. Reduced Ion Crossover in Bipolar Membrane Electrolysis via Increased Current Density, Molecular Size, and Valence. *ACS Appl. Energy Mater.* 3, 5804–5812. <https://doi.org/10.1021/acsaem.0c00687>
- Bosserez, T., Rongé, J., van Humbeeck, J., Haussener, S., Martens, J., 2015. Design of Compact Photoelectrochemical Cells for Water Splitting. *Oil Gas Sci. Technol. – Rev. d'IFP Energies Nouv.* 70, 877–889. <https://doi.org/10.2516/ogst/2015015>
- Domen, K., Naito, S., Soma, M., Onishi, T., Tamaru, K., 1980. Photocatalytic decomposition of water vapour on an NiO-SrTiO₃ catalyst. *J. Chem. Soc. Chem. Commun.* 543–544. <https://doi.org/10.1039/C39800000543>
- Fujishima, A., Honda, K., 1972. Electrochemical Photolysis of Water at a Semiconductor Electrode. *Nature* 238, 37–38. <https://doi.org/10.1038/238037a0>
- Haussener, S., Xiang, C., Spurgeon, J.M., Ardo, S., Lewis, N.S., Weber, A.Z., 2012. Modeling, simulation, and design criteria for photo-electrochemical water-splitting systems. *Energy Environ. Sci.* 5, 9922–9935. <https://doi.org/10.1039/c2ee23187e>
- Hernández-Pagán, E.A., Vargas-Barbosa, N.M., Wang, T., Zhao, Y., Smotkin, E.S., Mallouk, T.E., 2012. Resistance and polarization losses in aqueous buffer-membrane electrolytes for water-splitting photo-electrochemical cells. *Energy Environ. Sci.* 5, 7582–7589. <https://doi.org/10.1039/c2ee03422k>
- James, B.D., Baum, G.N., Perez, J., Baum, K.N., 2009. Technoeconomic Analysis of Photoelectrochemical (PEC) Hydrogen Production.
- Khaselev, O., Turner, J.A., 1998. A monolithic photovoltaic-photoelectrochemical device for hydrogen production via water splitting. *Science* (80-.). 280, 425–427. <https://doi.org/10.1126/science.280.5362.425>
- Perry, S.C., Ponce de León, C., Walsh, F.C., 2020. Review—The Design, Performance and Continuing Development of Electrochemical Reactors for Clean Electrosynthesis. *J. Electrochem. Soc.* 167, 155525. <https://doi.org/10.1149/1945-7111/abc58e>
- Schröder, V., Emonts, B., Janßen, H., Schulze, H.P., 2004. Explosion limits of hydrogen/oxygen mixtures at initial pressures up to 200 bar. *Chem. Eng. Technol.* 27, 847–851. <https://doi.org/10.1002/ceat.200403174>
- Shen, C., Wycisk, R., Pintauro, P.N., 2017. High performance electrospun bipolar membrane with a 3D junction. *Energy Environ. Sci.* 10, 1435–1442. <https://doi.org/10.1039/c7ee00345e>
- Yang, P., 2010. Semiconductor nanowires for energy conversion. *ACS Natl. Meet. B. Abstr.* <https://doi.org/10.1109/inec.2010.5424441>

Advanced exergoeconomic analysis of a double effect evaporation process in an instant coffee plant

Tinoco-Caicedo D.L.,^{a,b,*} Feijoó-Villa E.,^a Calle-Murillo J.,^a Lozano-Medina A.,^b and Blanco-Marigorta A.M.^b

^a *Facultad de Ciencias Naturales y Matemáticas, Escuela Superior Politécnica del Litoral, 090903 Guayaquil, Ecuador*

^b *Department of Process Engineering, Universidad de las Palmas de Gran Canaria, 35017 Gran Canaria, Spain*

**dtinoco@espol.edu.ec*

Abstract

The advanced exergoeconomic analysis performed on different industrial processes is used to determine the avoidable exergy destruction and inversion cost rates in order to increase the rentability and sustainability of a factory. This study focuses on the performance of an advanced exergoeconomic analysis of a double effect evaporation process (DEEP) of coffee extract in a factory located in Ecuador. The avoidable and unavoidable exergy destruction cost rate and the avoidable investment cost rate were determined for each component by exergoeconomic balances using the results from the process simulation with the best operational conditions and with the worst operational conditions. The avoidable exergy destruction cost rate represents 18.3% of the overall exergy destruction cost rate. It was estimated that around 70.3 \$/h of the overall operational costs could be saved if the exergetic efficiency of the first double effect evaporator (D-101) and the steam condenser (E-103) increased from 46.3% to 60.7% and 52.3% to 61.6%, respectively. Additionally, an increment of the initial concentration of soluble solids in the extract can reduce the avoidable operational costs by 15%.

Keywords: Advanced exergoeconomic analysis, Process simulation, Double effect evaporation, Avoidable cost rate

1. Introduction

Instant coffee is a widely consumed product worldwide, its market is projected to grow in the next seven years (Coherent Market Insights, 2019). One of the steps in industrial-scale instant coffee production is the evaporation of water from the extract in order to increase the concentration of soluble solids in preparation for the drying process. This evaporation process is energy-intensive, and consumes large amounts of fossil fuels since water has a high latent heat. Additionally, previous studies show that this process has low energetic and exergetic efficiencies (Mojarab Soufiyan et al., 2016), which lead to a high level of energy waste, and consequently raise production costs. In order to address these inefficiencies, it is not only necessary to identify and quantify the losses, but also to determine the fraction of avoidable losses. The advanced exergoeconomic analysis developed with computational tools allows for the determination of the exergy destruction rate and the avoidable and unavoidable costs of each component of the system. This analysis therefore becomes an important decision-making tool in production industries to reduce operational costs and increase the sustainability of the processes.

Some conventional exergy analyses have been done in food industries that have evaporation as part of their process, such as the production of tomato paste (Mojarab Soufiyan et al., 2016), powdered milk (Bühler et al., 2018), and yogurt (Mojarab Soufiyan and Aghbashlo, 2017). However there are no advanced exergoeconomic analyses performed in food industries; these studies would allow for the determination of the real potential for improvement of each component of the process (Liu et al., 2020).

In this context, this work presents an advanced exergoeconomic analysis of the double effect evaporation process of coffee extract, by using real operational data from a factory located in Ecuador. The aims of this study are to identify the main sources of exergy destruction that significantly affect the operational cost, and to quantify the avoidable cost that could be reduced. Additionally, a parametric study is performed to analyze the effect the initial concentration of soluble solids has on the operational cost indicators.

2. Methodology

2.1. Process Description

Figure 1 describes the double effect evaporation process of coffee extract in a factory located in Ecuador. Coffee extract (stream 1) is an aqueous solution with an initial concentration of soluble solids of 18 w/w%, from *Robusta* and *Arabica* beans. This extract is pumped to a heat exchanger (E-102) for pre-heating it up to 50°C by using steam. The steam is generated in the boiler (B-201) by using fuel oil N°6. Meanwhile, an already concentrated extract (stream 7) leaves the second effect (D-102); part of it is mixed with the heated extract and recirculated to D-102. The other part is sent to the first effect (D-101). The evaporated water (stream 11) in D-102 enters the condenser (E-101), where the temperature is reduced from 50°C to 32°C. The condensate water (stream 15) is mixed with condensate from de D-102 (stream 16) and then it is discarded. The concentrated coffee (stream 8) that leaves the D-101 reaches a concentration of 50 w/w% and then it is cooled from 66°C to 11°C in a heat exchanger of multiple flow (E-103), where cooling-tower water (C1) and chilled water (W3) are used.

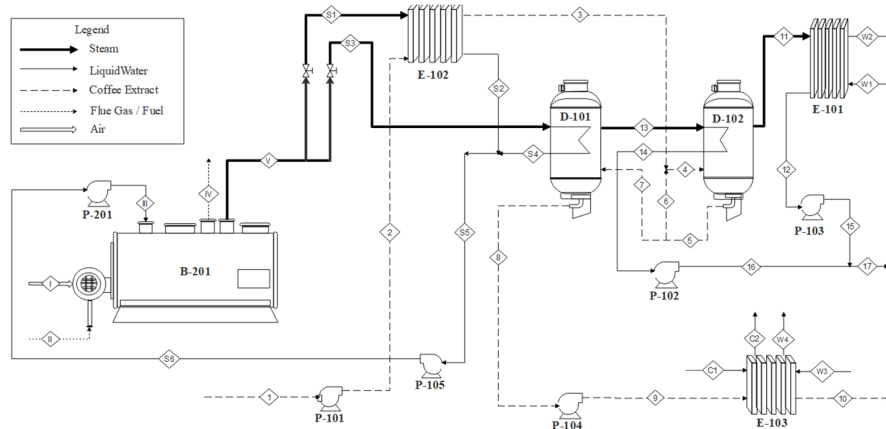


Figure 1. Process flow diagram of the double effect evaporation of coffee extract and the steam generation unit.

The process simulation was developed taking into account the following assumptions:

- The system was at steady state and the elevation in the coffee extract boiling point, due to concentration of solution, was assumed to be negligible.
- The heat loss rate and the pressure lost in all the components were negligible.
- A complete combustion was assumed in the combustion chamber.
- The steam and gases were considered as ideal gases when the pressure was below 1000 kPa. For higher pressures, the SRK-Equation was used as the equation of state.
- The coffee extract and liquid water were considered as ideal solutions.

2.2. Advanced Exergoeconomic Analysis

The advanced exergoeconomic analysis of the double effect evaporation process of coffee extract was performed with the aim to determine the avoidable (AV) and unavoidable (UN) exergy destruction cost rates and investment cost rates at a component level. This analysis was done with the real operational data (ROD) from the plant. The dead state in the system was considered at 298 K and 1 atm. For each component the best operational conditions (BOC) and the worst operational conditions (WOC) were established based on a previous study (Morosuk and Tsatsaronis, 2019) as shown in Table 1. The BOC were used for the process simulation in Pro/II[®]. The values of each state and the results from the k th component were used to perform the exergetic analysis (to calculate $\dot{E}_{D,k}^{UN}$), the economic analysis (to calculate $c_{F,k}$) and the exergoeconomic analysis at a component level (to calculate $\dot{C}_{D,k}^{UN}$). The thermodynamic model formulation was performed in Engineering Equation Solver (ESS), software.

Table 1. Values of Parameters for Different Operation Conditions.

Component	Parameter	ROD	BOC	WOC
D-101	ΔT_{min} [°C]	35.0	9.5	50.2
D-102	ΔT_{min} [°C]	15.1	5.0	20.0
P-101	η_{is} [-]	0.80	0.95	0.70
P-102	η_{is} [-]	0.80	0.95	0.70
P-103	η_{is} [-]	0.80	0.95	0.70
P-104	η_{is} [-]	0.80	0.95	0.70
P-105	η_{is} [-]	0.80	0.95	0.70
E-101	ΔT_{min} [°C]	4.0	2.0	23.0
E-102	ΔT_{min} [°C]	77.0	10.0	85.0
E-103	ΔT_{min-cw} [°C]	10.0	2.0	20.0
E-103	ΔT_{min-rw} [°C]	6.0	2.0	20.0
B-201	λ [-]	1.24	1.00	1.93
P-201	η_{is} [-]	0.80	0.95	0.70

The equations used for the exergetic analysis, the economic analysis (TRR methodology) (Bejan et al., 1996) and the advanced exergoeconomic analysis (Petraokopoulou et al., 2012) are presented in Table 2. The auxiliary equations and specific costs for were obtained from a previous study (Tinoco-Caicedo et al., 2020).

Table 2. Equations used for Exergy, Economic and Advanced Exergoeconomic Analyses.

Exergetic/Economic Analyses		Advanced Exergy/Exergoeconomic Analyses	
Parameters	Equations	Parameters	Equations
Exergy Balance	$E = E^{PH} + E^{CH}$	Exergoeconomic Balance	$\dot{C}_{p,k} = \dot{C}_{F,k} + \dot{Z}_k - \dot{C}_{D,k}$
Exergy destruction rate	$\dot{E}_{D,k} = \dot{E}_{F,k} - \dot{E}_{P,k}$	Exergy destruction cost rate ($\dot{C}_{D,k}$)	$\dot{C}_{D,k} = c_{F,k} \dot{E}_{D,k}$
Levelized Total Revenue	$TRR_L = CRF \sum_{j=1}^n \frac{TRR_j}{(1 + i_{eff})^j}$	Relative cost difference (τ_k)	$\tau_k = \frac{c_{p,k} - c_{F,k}}{c_{F,k}}$

Exergetic/Economic Analyses		Advanced Exergy/Exergoeconomic Analyses	
Parameters	Equations	Parameters	Equations
Requirement levelized carrying charges	$CC_L = TRR_L - FC_L - OMC_L$	Exergoeconomic Factor (f_k)	$f_k = \frac{\dot{Z}_k}{\dot{Z}_k + \dot{C}_{D,k}}$
Annual carrying charges (\dot{Z}_k^{CI})	$\dot{Z}_k^{CI} = \frac{CC_L}{\tau} * \frac{PEC_k}{\sum_k PEC_k}$	Unavoidable Exergy destruction rate ($\dot{E}_{D,k}^{UN}$)	$\dot{E}_{D,k}^{UN} = \dot{E}_{X_{P,k}} \left(\frac{\dot{E}_{X_{D,k}}}{\dot{E}_{X_{P,k}}} \right)^{UN}$
Annual operating and maintenance costs (\dot{Z}_k^{OM})	$\dot{Z}_k^{OM} = \frac{OMC_L}{\tau} * \frac{PEC_k}{\sum_k PEC_k}$	Unavoidable Exergy destruction cost rate ($\dot{C}_{D,k}^{UN}$)	$\dot{C}_{D,k}^{UN} = c_{F,k} \dot{E}_{D,k}^{UN}$ $\dot{C}_{D,k}^{AV} = \dot{C}_{D,k} - \dot{C}_{D,k}^{UN}$
Annual total costs (\dot{Z}_k)	$\dot{Z}_k = \dot{Z}_k^{CI} + \dot{Z}_k^{OM}$	Unavoidable investment cost rate (\dot{Z}_k^{UN})	$\dot{Z}_k^{UN} = \dot{E}_{X_{P,k}} \left(\frac{\dot{Z}_k}{\dot{E}_{X_{P,k}}} \right)^{UN}$ $\dot{Z}_k^{AV} = \dot{Z}_k - \dot{Z}_k^{UN}$

Furthermore, the unavoidable investment cost for the kth component was determined following the same methodology (simulation, exergetic analysis, economic analysis and exergoeconomic analysis), using the worst operational conditions (WOC) which implies the lowest possible component purchase costs (Gungor et al., 2015).

3. Results and Discussion

The results from the advanced exergoeconomic analysis are shown in Table 3. The components that have the highest avoidable exergy destruction rate ($\dot{E}_{D,k}^{AV}$) are (in descending order): B-201, D-101 and D-102. However, the components that have the highest avoidable exergy destruction cost rate are (in descending order): D-102, E-103 and D-101. These results are interesting because although B-201 is responsible for 71.7% of the overall avoidable exergy destruction rate, it only represents the 4.6% of the overall avoidable operating costs ($\dot{C}_{D,k}^{AV} + \dot{Z}_k^{AV}$). On the other hand, despite having a low $\dot{E}_{D,k}^{AV}$, the E-103 is responsible for 15.9% of the overall avoidable exergy destruction cost rate. This is due to this component having the highest specific cost of the system (2×10^{-3} \$/kJ) because it uses tower water and chilled water for cooling the concentrated extract.

Table 3. Results of the Advanced Exergoeconomic Analysis

Comp.	f_k (%)	r_k	$\dot{E}_{D,k}^{AV}$ (kW)	$\dot{E}_{D,k}^{UN}$ (kW)	$\dot{C}_{D,k}^{AV}$ ($\frac{\$}{h}$)	$\dot{C}_{D,k}^{UN}$ ($\frac{\$}{h}$)	\dot{Z}_k^{AV} ($\frac{\$}{h}$)	\dot{Z}_k^{UN} ($\frac{\$}{h}$)
E-101	2.0	4.5	0.4	51.0	2.4	300.1	0.7	5.4
D-102	55.0	1.5	27.9	12.7	57.2	26.2	7.8	93.5
D-101	86	6.4	27.7	60.6	4.3	9.3	4.1	78.1
E-103	8	8×10^{-4}	1.6	2.0	13.1	16.5	8×10^{-3}	2.7
B-201	8	2.5	168.5	596.9	4.2	15.0	0.2	1.3
E-102	23	7.4	6.8	13.3	1.1	2.2	4×10^{-2}	0.9
P-101	89	2.8	0.6	0.3	3×10^{-2}	0.1	1×10^{-2}	0.7
P-102	90	2.7	0.6	0.2	0.1	2×10^{-2}	4×10^{-2}	0.7
P-104	91	2.9	0.2	0.2	2×10^{-2}	2×10^{-2}	2×10^{-2}	0.4
P-103	91	2.6	0.6	0.2	0.1	2×10^{-2}	3×10^{-2}	0.7
P-105	93	3.0	0.1	0.1	1×10^{-2}	1×10^{-2}	3×10^{-3}	0.3
P-201	93	3.8	1×10^{-2}	0.1	1×10^{-3}	1×10^{-2}	1×10^{-3}	0.2

Additionally, component D-102 was found to have the greatest improvement potential, given that 68.6% of its exergy destruction costs are avoidable. In contrast, the pumps don't have improvement potential, as they account for less than 1% of the overall avoidable operating costs.

The required modifications should be focused to reduce the exergy destruction cost rate (even if this results in a higher investment cost), because the components that have the highest operating cost, also have the lowest exergoeconomic factor (less than 10%). These modifications could be done through changes to the operating conditions or through structural changes in the three components with the highest avoidable cost (D-102, E-103 and D-101) in order to increase their exergetic efficiency.

Figure 2 shows that an increase in the initial concentration of soluble solids in the extract from 18 w/w% to 26 w/w% causes a reduction of the avoidable exergy destruction cost rates of the D-101 and D-102, from 5.5 \$/h to 3.3 \$/h and 71.5 \$/h to 46.8 \$/h, respectively. Additionally, the avoidable investment cost in D-101 and D-102 are reduced from 4.5 \$/h to 3.7 \$/h and 8.6 \$/h to 7.1 \$/h, respectively.

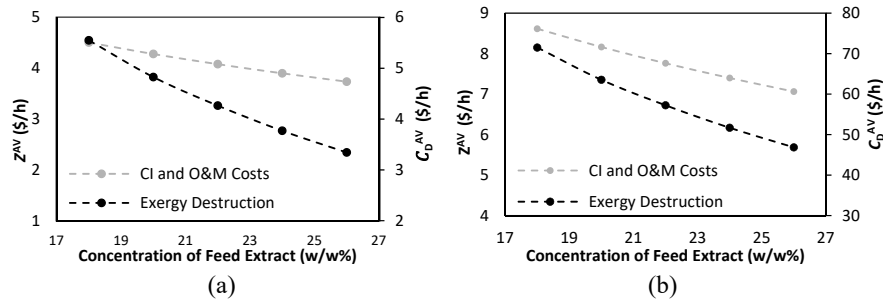


Figure 2. Effect of the initial concentration of soluble solids in coffee extract on the $\dot{C}_{D,k}^{AV}$ and \dot{Z}_k^{AV} of the components (a) D-101 and (b) D-102.

Figure 3 shows that an increment of the initial concentration of soluble solids increases the r_k value in both evaporators (D-101 y D-102), causing a reduction of the final product cost from 22931 \$/T to 21875 \$/T. Furthermore, f_k presents an increment in evaporators D-101 and D-102 of 3.9% and 5.6%, respectively. This means that an increment in the initial concentration of coffee extract could also improve the balances between the investment rate and exergy destruction cost rate.

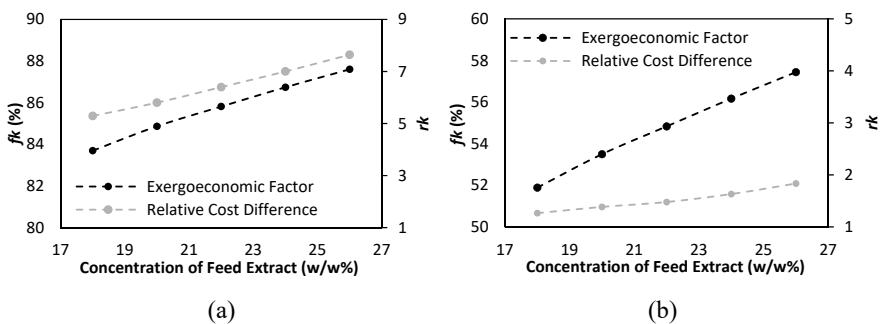


Figure 3. Effect of the initial concentration of soluble solids in the coffee extract on the r_k and f_k of the components (a) D-101 and (b) D-102.

4. Conclusions

The purpose of this study was to perform an advanced exergoeconomic analysis of the double effect evaporation process of coffee extract. The results of this research show that although B-201 is the component with the highest avoidable exergy destruction rate, it does not have a high avoidable cost rate. The components that have the highest potential for improvement are D-101, D-102 and E-103 because they have the highest avoidable exergy cost rate. Additionally, the initial concentration of soluble solids proved to be a significant parameter for the process, given that an 8 w/w% increase of the initial concentration of soluble solids reduced the avoidable exergy destruction cost by 15%.

These results suggest that a reduction of the avoidable exergy destruction rate in these components can be achieved and it is possible to have annual savings of $\$8.37 \times 10^5$ in the overall operating costs.

Further research is needed to optimize the solid-liquid extraction of coffee, in order to achieve a higher initial concentration of soluble solids in the coffee extract. Also extended parametric and structural change studies are recommended in the double effect evaporation process in order to maximize the exergetic efficiency and minimize the operational costs.

References

- Bejan, A., Tsatsaronis, G., Moran, M., 1996. *Thermal Design and Optimization*. Wiley, New York.
- Bühler, F., Nguyen, T. Van, Jensen, J.K., Holm, F.M., Elmegaard, B., 2018. Energy, exergy and advanced exergy analysis of a milk processing factory. *Energy* 162, 576–592.
- Coherent Market Insights, 2019. *Instant Coffee Market Size Share Forecast Report 2026*. URL <https://www.coherentmarketinsights.com/market-insight/instant-coffee-market-1458> (accessed 7.9.20).
- Gungor, A., Tsatsaronis, G., Gunerhan, H., Hepbasli, A., 2015. Advanced exergoeconomic analysis of a gas engine heat pump (GEHP) for food drying processes. *Energy Convers. Manag.* 91, 132–139.
- Liu, Z., Liu, Z., Cao, X., Luo, T., Yang, X., 2020. Advanced exergoeconomic evaluation on supercritical carbon dioxide recompression Brayton cycle. *J. Clean. Prod.* 256, 120537.
- Mojarab Soufiyan, M., Aghbashlo, M., 2017. Application of exergy analysis to the dairy industry: A case study of yogurt drink production plant. *Food Bioprod. Process.* 101, 118–131.
- Mojarab Soufiyan, M., Dadak, A., Hosseini, S.S., Nasiri, F., Dowlati, M., Tahmasebi, M., Aghbashlo, M., 2016. Comprehensive exergy analysis of a commercial tomato paste plant with a double-effect evaporator. *Energy* 111, 910–922.
- Morosuk, T., Tsatsaronis, G., 2019. Advanced exergy-based methods used to understand and improve energy-conversion systems. *Energy* 169, 238–246.
- Petrakopoulou, F., Tsatsaronis, G., Morosuk, T., Carassai, A., 2012. Advanced exergoeconomic analysis applied to a complex energy conversion system. *J. Eng. Gas Turbines Power* 134, 3.
- Tinoco-Caicedo, D.L., Lozano-Medina, A., Blanco-Marigorta, A.M., 2020. Conventional and Advanced Exergy and A Case Study of an Instant Coffee Factory in Ecuador. *Energies* 13, 21, 5622.

Energy-efficient membrane cascades for industrial separations

Jose Adrian Chavez Velasco^a, Zewei Chen^a, Radhakrishna Tumbalam Gooty^a, Mohit Tawarmalani^b and Rakesh Agrawal^{a,*}

^a*Davidson School of Chemical Engineering, Purdue University, 480 Stadium Mall Drive, West Lafayette 47907, USA*

^b*Krannert School of Management, Purdue University, 100 S. Grant Street, West Lafayette 47907, USA*
agrawalr@purdue.edu

Abstract

The energy requirement for the separation of a given mixture via a multistage membrane cascade depends on the choice of the cascade and its operating conditions. Identifying the optimal cascade along with its optimal operating conditions is challenging, since it requires the solution of a nonconvex mathematical program. To address the challenge, we propose novel Mixed Integer Nonlinear Programs (MINLPs) that are formulated such that they can be solved using off-the-shelf global optimization solvers. We illustrate the practicality of our models with two case studies: (1) separation of p-xylene from o-xylene (2) recovery of natural gas liquid (NGL) from shale gas. Further, for NGL recovery, we determine the target selectivity and permeability that will enable membrane technology to outcompete the conventional demethanizer. These target values provide guidance for experimental groups that are developing new membrane materials for NGL recovery.

Keywords: membranes, optimization, energy efficiency, shale gas

1. Introduction

For some separations, membrane technology has been demonstrated to be advantageous because of its operational simplicity, modular scale-up flexibility, and relatively low capital cost. Nowadays, membranes are used in important applications such as water desalination, natural gas sweetening, and nitrogen production. Owing to the commercial success in the above applications and some other applications, there has been an increased interest in expanding the use of membranes to other separations. Nevertheless, a majority of state-of-the-art membranes still have at most moderate permeability and selectivity values. This can be a limitation for separations that require both high purity and recovery of the component of interest; as a single membrane stage is typically insufficient to meet the specifications. One alternative to overcome the limitation is to use a multistage membrane cascade. It is well-known that a plethora of cascades exists even for the simplest case of a binary mixture (Agrawal, 1997). However, the energy requirement and the cost between alternative cascades can differ substantially. Moreover, even for a given cascade, the energy requirement and the cost can vary with the operating conditions. Identifying the optimal cascade along with its optimal operating conditions is challenging, since the equations governing the permeation process are nonconvex. Here, we develop two optimization models: (1) a Mixed Integer Nonlinear Program (MINLP) that is applicable for both liquid and gaseous binary mixtures (2) an MINLP that is applicable for a multicomponent gaseous mixture. For brevity, we highlight only the key elements

of our models, and refer the reader to Chavez Velasco et al. (2020); Chen et al. (2020) for further details.

While many optimization models have been proposed in the literature (e.g. Qi and Henson, 2000; Adi et al., 2016), our models differ in the following aspects. First, several formulations oversimplify the permeation process which leads to a significant reduction in model's fidelity. We, on the other hand, retain the essential physical aspects of the permeation process that are often ignored in the oversimplified models. Second, most of the studies rely on optimization solvers that do not always yield the global optimum. Here, we guarantee global optimality of the solutions by formulating the problem such that it can be solved using global optimization techniques, such as those implemented in BARON (Tawarmalani and Sahinidis, 2005; Kılınç and Sahinidis, 2018). The rest of the paper is outlined as follows. In section two, we describe the key elements of the proposed optimization models. In section three, we illustrate the practicality of our approach with two case studies. Finally, in section four, we present concluding remarks.

2. Two novel membrane cascade optimization formulations

2.1. Single-stage permeator model

The validity of the optimal solution depends on the fidelity of the single-stage permeator model. In this work, we use a permeator model that is reasonably accurate, and at the same time, is viable for global optimization. We use flux equations derived from the solution-diffusion theory (Wijmans and Baker, 1995) together with the crossflow model (Weller and Steiner, 1950) to model the permeation across a membrane. The governing equations constitute a system of differential algebraic equations (DAE) shown below:

$$\frac{dx_i}{df} = \frac{y_i - x_i}{f}, \quad y_i = \frac{S_i(x_i - \gamma_i y_i)}{\sum_{i=1}^n S_i(x_i - \gamma_i y_i)}, \quad x_i(f^{in}) = x_i^{in}, \quad i = 1, \dots, n. \quad (1)$$

Here, n denotes the number of components in the mixture, x_i and y_i denote the local mole fraction of component i on the retentate and permeate sides, respectively, f denotes local retentate flow, S_i denotes the selectivity of component i with respect to the component n , x_i^{in} denotes the mole fraction of component i in the inlet stream, and $\gamma_i = \{1/r$ for gases, $\exp(-V_i \Delta P / RT_0)$ for liquids} where r , ΔP , R , V_i and T_0 correspond to the transmembrane pressure ratio, transmembrane pressure difference, the universal gas constant, molar volume of component i , and the operating temperature, respectively. We solve the DAE system in Eq. (1) analytically for binary mixtures (see Chavez Velasco et al. (2020)). For a multicomponent mixture, we approximate the DAE system with a piecewise constant parametrization technique (see Chen et al. (2020)). Extensive numerical studies show that the approximated solution agrees well with the exact solution of the DAE system.

2.2. Proposed formulations

In this subsection, we briefly describe our optimization models. Figure 1 shows the superstructure we considered for binary liquid separations. The superstructure for binary gaseous mixtures has the same stream connections, but the pumps are now replaced with compressors. In either of these networks, the permeate stream from each stage can be recycled to any of the two immediate previous stages, or can be withdrawn from the cascade to be part of the final permeate product. Moreover, the feed can be located at any of the membrane stages. In the case of multi-component gaseous separations, currently,

we have considered a simpler cascade superstructure where the permeate stream can only be recycled to the previous stage.

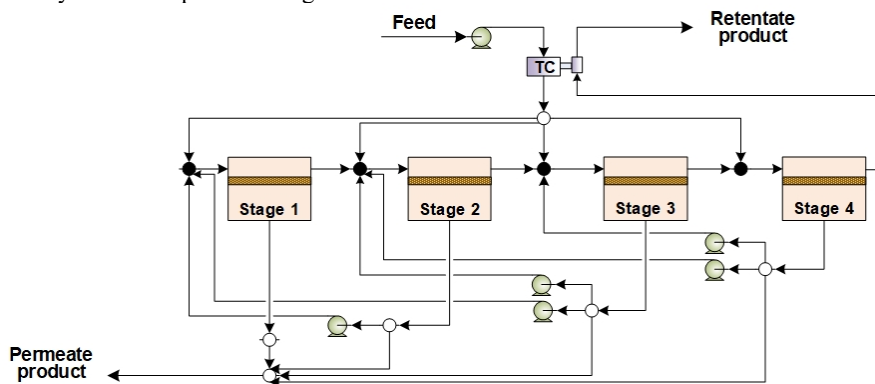


Figure 1: Cascade superstructure for binary liquid separations (Chavez Velasco et al., 2020). TC stands for turbocharger

The optimization problem for both binary and multi-component separations is summarized as: for a given feed, find the membrane cascade with at most N stages along with its operating condition that separate the given mixture into two product streams with a specified composition while minimizing the energy consumption. The optimization problem is formulated as MINLP. Here, binary variables are introduced to regulate the flows along the paths that connect various splitters and mixers. Whereas continuous variables model stream compositions and flows. Furthermore, in the case of binary separations, the pressure ratio (or trans-membrane pressure difference) was modeled as a continuous variable. Nevertheless, in the case of multicomponent gaseous separations, the pressure ratio is modeled as a discrete variable as it helps to expedite the convergence. The objective function computes the total power needed to operate the pumps (or compressors). The main constraints in the optimization model are: 1) overall and component mass balance constraints around mixers, splitters and each membrane, 2) single-stage permeator model for each of the membranes in the cascade, 3) constraints to restrict flows along specific paths.

2.3. The challenge of guaranteeing global optimality

To obtain global optimality certificate in a reasonable amount of time with BARON (Kılınç and Sahinidis, 2018), we derive additional cuts for the problem using physical insights and by exploiting mathematical properties of the governing equations. Some of these cuts, although implicit in the model, are not implied in the relaxation. Therefore, providing these cuts explicitly helps in expediting the convergence characteristics of BARON. We refer the reader to Chavez Velasco et al. (2020) for the details and numerical experiments demonstrating their potential.

3. Exploring the potential of membrane cascades

The energy requirement of a membrane cascade depends on the permselectivity of the chosen membrane. When evaluating the energy potential of a membrane-based process for a given separation, one of the following scenarios can be found: 1) a specific membrane with a specific permselectivity is available, 2) no specific membrane is available, and thus, the permselectivity is unknown. In the first scenario, the solution of

our proposed optimization model would allow to identify the optimum-energy membrane cascade along with its optimal operating condition. Then, the energy consumed by this cascade can be compared against the energy required by alternative separation options. If it happens that the optimum cascade consumes higher energy than that of alternative separation processes, or as in scenario 2, no specific membrane is available for the given separation task, it becomes relevant to determine what would be the target membrane permselectivity that is needed for the membrane-based process to make it more energy efficient. The following examples illustrate this targeting procedure. All membrane cascade optimization cases discussed in this section were solved to 5% of global optimality gap using BARON 18.5.8 (Kiliç and Sahinidis, 2018).

3.1. Separation of *p*-xylene/*o*-xylene

We consider the separation of a liquid mixture with 65 mol% *p*-xylene and rest *o*-xylene at 1 bar and 30 °C. Our goal is to recover *p*-xylene with 99 % recovery at 99.5 % purity. First, we consider a membrane with a *p*-xylene permselectivity of 50. For these conditions, the optimal four-stage membrane cascade identified by our approach consumes 52 kW per mole of feed for pumping the feed and the recycle streams¹. Membrane permselectivity values of 60, 77, and 106 would reduce the membrane cascade power by 10 %, 20 %, and 30 % respectively. The procedure illustrated here helps in guiding material researchers to determine the membrane target properties needed for a given application.

3.2. Natural gas liquid (NGL) recovery from shale gas

Here, the objective is to separate shale gas into a CH₄-rich stream and an NGL stream containing C₂H₆, C₃H₈, C₄H₁₀, and C₅H₁₂. Traditionally, this separation is performed using a cryogenic demethanizer (Yoon et al., 2017). However, the traditional technology is not suitable for modularization, and membranes are perceived as a potential alternative for modular plants. Nevertheless, even the state-of-the-art membrane materials have a low selectivity towards CH₄ (around 5 for C₂H₆/CH₄), so a single membrane stage is not sufficient to meet the desired product specifications. Most of the membranes designed for this separation have a reverse permselectivity, meaning that the heavy components, NGL, permeate faster than the light component, CH₄ (Merkel et al., 2000). For a comparative analysis between a cryogenic demethanizer and a membrane cascade, we consider 200 million standard cubic feet per day (MMSCFD) of shale gas from the Bakken Basin, which contains 60 % CH₄, 20 % C₂H₆, and 20 % C₃₊ (C₃H₈, C₄H₁₀ and C₅H₁₂ lumped as a single component). The feed and the retentate are assumed to be at 30 bar. Since the feed is already at high pressure, we send it directly to the cascade without any further compression. Furthermore, given that the separated retentate stream needs to be transported via pipeline, we also do not consider the expansion of this stream for the purpose of recovering energy from it. This eliminates the turboexpander shown in Figure 1. For the demethanizer, we simulated in Aspen Plus the process flowsheet proposed by Yoon et al. (2017). Our simulation shows that the demethanizer consumes 12,000 kW to produce CH₄ with 96 % recovery and 94 % purity. In our analysis of the alternative separation through a membrane cascade, we approximated the feed as three-component feed, which has 60 % CH₄, 20 % C₂H₆, and 20 % C₃H₆. Here all the C₃₊ components are lumped as a pseudo component with the same selectivity as C₃H₆. The selectivity of CH₄/C₂H₆ is 5 for a polymeric membrane (Starannikova et al., 2006) and we assumed a

¹ The efficiencies of the pumps and the turbocharger were assumed to be 75% and 80% respectively

selectivity of 100 for $\text{CH}_4/\text{C}_3\text{H}_8$. After solving our MINLP, we found that the optimum two-stage and three-stage membrane cascades would require a total energy of 84,181 kW and 28,456 kW respectively to produce CH_4 with the same recovery and purity as that obtained from the demethanizer². By performing sensitivities on $\text{C}_2\text{H}_6/\text{CH}_4$ permselectivity, we determined that a target selectivity of 25 and 20 would be needed for the two-stage and three-stage membrane cascade respectively in order to consume the same energy as the demethanizer.

It is important to mention that the viability of using either a cryogenic demethanizer or a membrane cascade not only depends on energy consumption, but also on economics. Therefore, we also calculate and compare the net present value (NPV) for the two processes. The NPV, which accounts for both capital and operating costs, was calculated assuming a plant-life of 15 years and 10 % discount rate. We note that although the membrane permeance does not affect the power consumption of a membrane cascade and the cost of the compressors, it strongly affects the required membrane area, and consequently the membrane cost. To identify the target combination of permeance and permselectivity that would be needed for the analyzed separation through a two-stage cascade, we generated the permeance-selectivity curve shown in Figure 2. This plot provides useful guidance to quickly determine if a two-stage membrane cascade would be economically attractive. To elaborate, if the membrane permselectivity and permeance falls above the corresponding curve in Figure 2a), an optimum two-stage membrane cascade would be more economical than the conventional cryogenic demethanizer. Different from the Robeson upper bound curve which reflects the challenge of improving both permeance and selectivity, the permeance-selectivity curve introduced here provides a lower bound on permeance and selectivity for an economically attractive process. Figure 2b) shows the net present value distribution for the circled point in the permeance-selectivity curve.

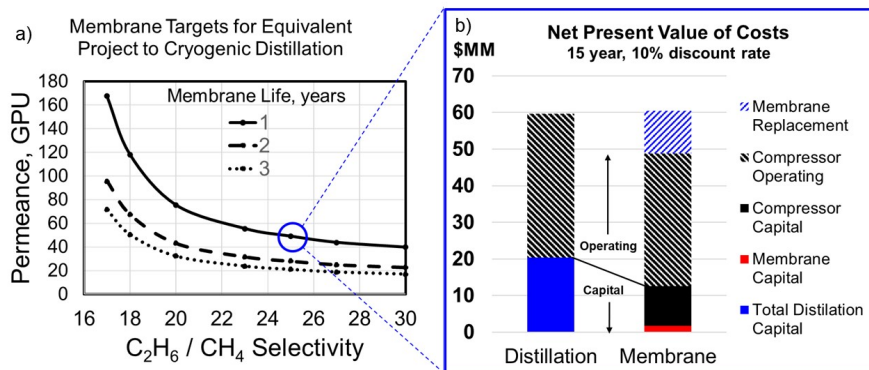


Figure 2: a) Net present value breakdown of the cryogenic demethanizer and the membrane cascade as function of selectivity and permeance. All points on each curve yield the same net present value, and require the same energy consumption as that for the cryogenic demethanizer. b) Cost distributions for the circled point in a)

² For each compressor in the membrane cascades we assumed an isothermal efficiency of 75%

4. Conclusions

In this article, we have briefly described two different membrane cascade optimization models that for a given mixture, either liquid or gaseous binary mixture, or multicomponent gaseous mixture, finds the optimum membrane cascade along with its corresponding operating conditions that minimize the energy input. To illustrate the utilization of our proposed MINLP models, we analyzed two separation examples. For these examples, we described a targeting procedure that for a given separation, helps to determine target membrane properties such as permselectivity that would be needed to outperform the energy efficiency of alternative separation options. Furthermore, for the analyzed separation of NGL from shale gas, we introduced an easy-to-use graphical tool that guides on which combination of permselectivity and permeance would be needed for a two-stage membrane cascade in order to be economically attractive as compared to a conventional demethanizer process.

5. Acknowledgement

We thank Gary Sawyer for providing us the economic analysis of the NGL case. Jose Adrian Chavez Velasco thanks COMEXUS-Fulbright, and CONACYT for providing him with financial support.

References

- V. S. Adi, M. Cook, L. G. Peeva, A. G. Livingston, B. Chachuat, 2016. Optimization of OSN Membrane Cascades for Separating Organic Mixtures. In: Z. Kravanja, M. Bogataj (Eds.), 26th European Symposium on Computer Aided Process Engineering. Vol. 38 of Computer Aided Chemical Engineering. Elsevier, pp. 379 – 384.
- R. Agrawal, 1997. A simplified method for the synthesis of gas separation membrane cascades with limited numbers of compressors. *Chemical Engineering Science* 52 (6), 1029 – 1044.
- J. A. Chavez Velasco, R. Tumbalam Gooty, M. Tawarmalani, R. Agrawal, 2020. Article in preparation. Available at <https://arxiv.org/pdf/2102.05226.pdf>
- Z. Chen, M. Tawarmalani, R. Agrawal, 2020. Article in preparation.
- M. R. Kılınç, N. V. Sahinidis, 2018. Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with baron. *Optimization Methods and Software* 33 (3), 540–562.
- T. C. Merkel, V. I. Bondar, K. Nagai, B. D. Freeman, I. Pinnau, 2000. Gas sorption, diffusion, and permeation in poly(dimethylsiloxane). *Journal of Polymer Science Part B: Polymer Physics* 38 (3), 415–434.
- R. Qi, M. A. Henson, 2000. Membrane system design for multicomponent gas mixtures via mixed-integer nonlinear programming. *Computers & Chemical Engineering* 24 (12), 2719 – 2737.
- L. Starannikova, Y. Yampolskii, K. Makovetskii, T. Golenko, 2006. A novel high permeability rubbery membrane material: cis-poly(pentenamer). *Desalination (Amsterdam)* 200 (1-3), 18–19.
- M. Tawarmalani, N. V. Sahinidis, 2005. A polyhedral branch-and-cut approach to global optimization. *Mathematical programming* 103 (2), 225–249.
- S. Weller, W. A. Steiner, 1950. Separation of gases by fractional permeation through membranes. *Journal of Applied Physics* 21 (4), 279–283.
- J. Wijmans, R. Baker, 1995. The solution-diffusion model: a review. *Journal of Membrane Science* 107 (1), 1–21.
- S. Yoon, M. Binns, S. Park, J.-K. Kim, 2017. Development of energy-efficient processes for natural gas liquids recovery. *Energy* 128, 768 – 775.

Techno-economic evaluation of Carbfix technology for carbon storage in the US

María Paz Etcheverry, Haoshui Yu*, Emre Gençer*

MIT Energy Initiative, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States
haoshuiy@mit.edu, egencer@mit.edu

Abstract

Carbfix is a technology to permanently store captured CO₂ and other sour gases from emission sources as rocks in the subsurface. CO₂ charged water is injected into the basalts to promote the carbonization of CO₂. The dissolution of CO₂ into the aqueous phase facilitates the mineralization of CO₂. It has been demonstrated that over 95% of CO₂ captured and injected was turned into a rock in the subsurface in less than two years. In the original project, this technology aims to store CO₂ from geothermal power plants in Iceland. However, it can be applied to the CO₂ capture from a combined-cycle natural gas power plant as well. This study would like to apply this technology to store the CO₂ from a combined-cycle natural gas power plant to achieve the near-zero emission target. Therefore, the CO₂ storage process has to be modeled and optimized based on this technology. In this process, the water consumption is determined by the depth of the injection well, the temperature, and the water's salinity. The capital costs (injection wells, monitoring wells, injection surface facilities) and operating costs (maintenance, electricity, water cost) interact with each other. In this study, a detailed techno-economic evaluation of Carbfix technology is performed in the context of a 500 MW combined cycle power plant in the US. The sensitivity analysis is carried out with respect to water price, electricity price, equipment cost, etc. The key factors are identified based on the tornado diagram. This methodology can be integrated with the power plant and CO₂ capture model to optimize the whole system simultaneously in future work.

Keywords: Carbfix technology, techno-economic evaluation, operating cost, capital cost, sensitivity analysis

1. Introduction

Increasing CO₂ concentration can largely be attributed to the usage of fossil fuels. Carbon capture and storage (CCS) play a fundamental role in achieving the goals of the Paris Agreement to limit global warming to 1.5-2°C (Snæbjörnsdóttir et al., 2020). It is reported that 37% of the total man-made greenhouse gas emissions are from the power sector (Selosse and Ricci, 2017). Therefore, the decarbonization of the power sector is the primary target on the way to a zero-emission world. However, large-scale deployment of CCS has yet to be achieved (Snæbjörnsdóttir et al., 2020). To accommodate the huge storage demand of CO₂ in the future, CO₂ storage technology has become of interest to both academia and industries. Carbfix technology, in which CO₂ is injected into reactive basaltic rocks, is one of the promising technologies for CO₂ storage because of the abundance of basaltic rocks on Earth's surface (Gislason and Oelkers, 2014). It has been demonstrated that over 95% of CO₂ captured and injected was turned into carbonate

minerals in the subsurface in less than two years in Carbfix technology. Carbonate minerals provide a long-lasting, thermodynamically stable, and environmentally benign CO₂ storage solution (Gislason et al., 2010). However, a huge amount of water is consumed to dissolve the CO₂ during the injection. Therefore, the cost of water and electricity to drive the pumps in the system has to be evaluated precisely before the implementation of the technology along with a natural gas power plant. Ragnheidardottir et al. (2011) assessed the costs of Carbfix technology for the Hellisheidi geothermal power plant. However, the geothermal power plant has its characteristics, and the cost analysis heavily depends on the application. This study aims to evaluate the techno-economic performance of the Carbfix technology to extend it from a geothermal power plant to natural gas combined cycle power plant located in the vicinity of North American Basalt Formations such as the Columbia Plateau basaltic aquifer in Washington state.

2. Problem statement and methodology

This study is novel in that it determines the cost per ton of CO₂ stored in USD and understands the cost drivers for implementing Carbfix technology in the US market. The cost model for the Carbfix system follows the framework allocation described below, and it is directly linked to the performance model which calculates the storage unit performance requirements based on the upstream technical assumptions. Performance and cost analysis including the profitability and sensitivity assessment are implemented in a MS Excel spreadsheet. There are three types of costs calculated by this model:

2.1 Site Costs

The site costs are those costs specific to the site itself. They are categorized into two categories:

Site screening: Involves the evaluation of regions within a larger area to identify selected areas that are potentially suitable for basalt mineralization and CO₂ storage. The site screening cost is based on the study of Smith et al. (2001), adjusted for inflation.

Permitting and licensing: Mainly associated with Environmental Impact Assessment, which is a crucial process to the successful permitting of the project. The licensing and permitting costs are estimates provided by the Carbfix team, converted to US dollars.

2.2 Capital Costs

The capital costs are fixed, one-time expenses incurred on the design, purchase of equipment, installation, and construction of the storage unit. The purchase of the land is excluded from capital costs. They are categorized into five categories:

Design: The cost associated with the conceptual, preliminary, and detailed engineering design of the whole system. It is calculated as 12% of the remaining capital costs.

Injection well: The cost associated with the engineering, procurement, and construction of a single injection well. It is calculated using a unit rate-based cost estimating as a function of depth. Quantities and rates in the breakdown were validated with the Carbfix team. Assumed diameter: 12", depending on the injected flow.

Monitoring well: The cost associated with the engineering, procurement, and construction of single monitoring well. The monitoring well cost is a lump sum estimating provided by the Carbfix team. Assumed diameter: 9".

Equipment: Includes pump compressors, valves, piping, and instrumentation, aside from the wells. The cost of equipment is scaled using the pilot project equipment cost (Ragnheidardottir et al., 2011) adjusted by inflation and the following formula:

$$\text{Capex Scaling Factor} = (\text{Flow}_{\text{scaled}}/\text{Flow}_{\text{pilot}})^{0.6}$$

Installation: The labor and machinery costs involved in the installation of equipment. It is scaled using the pilot project installation cost (Ragnheidardottir et al., 2011) adjusted by inflation and the formula above.

2.3 Operation & Maintenance (O&M)

The O&M costs are the costs incurred for maintaining and operating the whole system. They are categorized into four categories:

Fixed O&M: The cost of operating and maintaining the system regardless of the CO₂ tones stored. It is assumed as 2.5% of Capex.

Electricity: The cost associated with energy requirement for the injection, split up between pumping (7.5%) and auxiliary systems (92.5%). It does not include compression (CO₂ pressure provided by source). It is assumed that pumping energy requirements scale linearly with flow rates from the pilot case, and auxiliary system power consumption uses the power scaling factor (Ragnheidardottir et al., 2011). Electricity price from Washington state, USA (International Energy Agency, 2020).

$$\text{Flow Scaling Factor} = \text{Flow}_{\text{scaled}} / \text{Flow}_{\text{pilot}}$$

$$\text{Power Scaling Factor} = (\text{Flow}_{\text{scaled}} / \text{Flow}_{\text{pilot}})^{0.25}$$

Water: The cost associated with water consumption, based on the performance model requirements and the average industrial water price in the USA (Bunch et al., 2017).

Monitoring: it involves the cost of operating the monitoring well provided by Ragnheidardottir et al. (2011) adjusted by inflation, converted to US dollars.

3. Case study

The case study inputs are based on a combined-cycle natural gas power plant plus a thermal amine CO₂ capture unit (Table 1). However, both the power plant and the CO₂ capture unit are out of scope. The injection site is considered to be nearby the power plant, so the CO₂ injection pressure is equal to the CO₂ dispatch pressure from CO₂ capture unit. The target injection formation is assumed at 400-800 m depth consisting of basaltic lavas (Alfredsson et al., 2008). The reservoir capacity is such that it can store the power plant's annual emissions over a 30-year life (24Mt of CO₂).

Table 1. Upstream system characteristics

Nameplate Cap. (MW)	500
Capacity Factor	0.56
Generation (MWh/year)	2,452,800
Emissions (t/year)	833,158
Capture Eff.	95%
CO ₂ Pressure (bar)	30

Table 2. Storage unit performance details

CO ₂ Flowrate (kg/s)	25.1
H ₂ O Flowrate (l/s)	552.2
Injection Wells	6
Monitor Wells	1
Energy Req (kWh/year)	56,690,312
Lifetime (years)	30

By using the assumptions above and the referenced upstream system technical characteristics (Table 1), a performance model for CO₂ storage is executed (Table 2). The ratio of water to CO₂ required to dissolve the CO₂ stream at equilibrium is 22, defined as a function of temperature, pressure (or depth), and salinity of the water (Gislason et al., 2010). When at the injection well, both the CO₂ and water streams are injected separately until a depth in which the water reaches the same pressure as the CO₂ (around 350m for a 500m well, water level 100m), allowing the full dissolution while descending in the injection well. The number of injection wells is based on an assumed well injectivity of 92 l/s, and associated energy requirements are scaled utilizing a flow scaling factor of 35 and a power scaling factor of 4 from respective formulas (Ragnheidardottir et al., 2011). The ratio of monitoring wells to the injection well is 9 (National Petroleum Council,

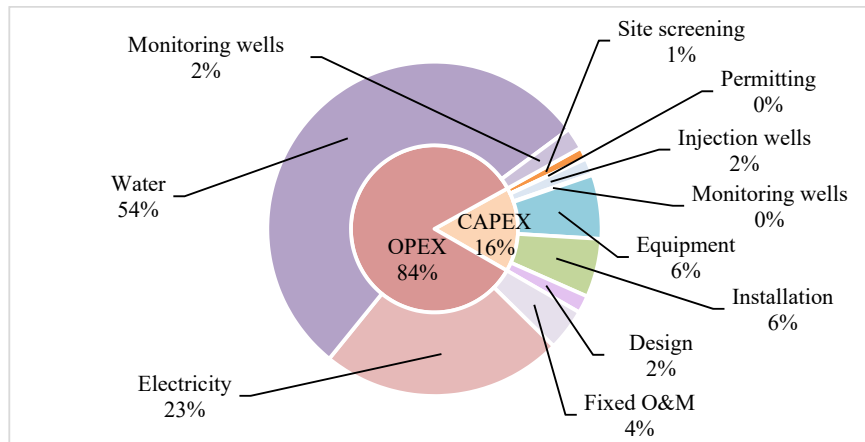
2019). Table 3 summarizes cost analysis based on the performance model. Figure 1 shows the component level contribution in annualized costs (site costs considered under capex). Once cost estimation is done, a profitability assessment optimizes the CO₂ storage price to reach an acceptable internal rate of return (IRR) for the project. Each of the cost estimation components described in Section 2 were assessed using a cash flow model with the assumptions in Table 4. Taxes and escalation are not considered at this point.

Table 3. Cost analysis results

Site Costs (USD)	\$ 2,118,500
Site screening	\$ 2,000,000
Permitting and licensing	\$ 118,500
Capital Costs (USD)	\$ 27,805,395
Design	\$ 3,089,488
Injection well x 6	\$ 2,834,559
Monitoring well	\$ 214,000
Equipment	\$ 11,304,703
Installation	\$ 10,362,644
O&M Costs (USD/year)	\$ 13,298,270
Fixed O&M	\$ 695,135
Electricity	\$ 2,993,248
Water	\$ 9,249,886
Monitoring	\$ 360,000

Table 4. Financial details

Working Capital	15%
DR	12%
IRR	15%
Equity Financing	30%
Loan Financing	70%
Loan Interest Rate	4.6%
Payback (years)	15
Inflation (O&M)	1.5%

Figure 1. Component level CO₂ storage price contribution for case study

The case study results in a required price of 21.9 USD/tCO₂. Subsequently, sensitivity analysis was applied to profitability, setting low and high inputs for all the components based on market prices across the US. The tornado in Figure 2 shows the analysis' results. In addition, the CO₂ storage price was evaluated by benchmarking within the US. According to the storage cost assessment performed by FE/NETL (Grant and Morgan, 2017), there are four saline formation regions with a storage cost of 15USD/tCO₂ and one region, North Central, with a cost of 22USD/tCO₂. In 2019, the NPC updated each region's storage costs using the FE/NETL model assumptions but modifying the

requirements for site screening and monitoring activities (National Petroleum Council, 2019). As a result, sites requiring strict monitoring would likely be excluded during initial site selection. These adjustments to the assumptions had the effect of reducing the cost of storage by approximately 50% compared with the FE/NETL results (Grant and Morgan, 2017). The US Carbfix price standing among typical US saline storage prices is shown in Figure 3. As noted, it is significantly higher than NPC prices. This is mainly because the NPC report considered a volume-weighted average cost only for regions with significant volumes of CO₂ storage, such as 11-135 Gt of CO₂.

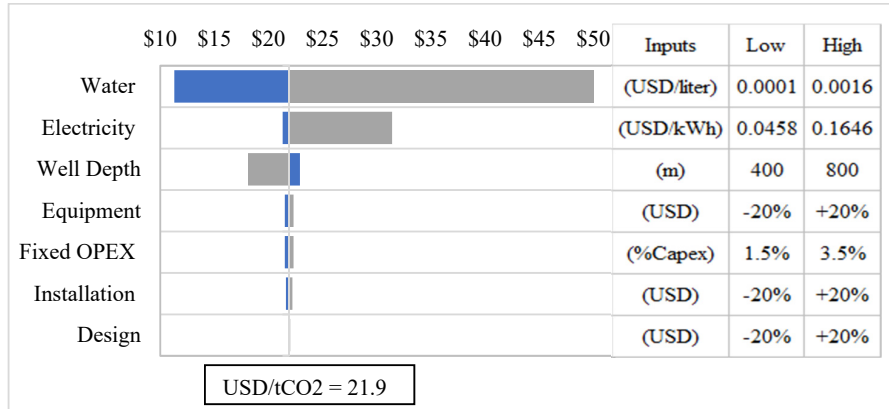


Figure 2. Tornado diagram for case study price

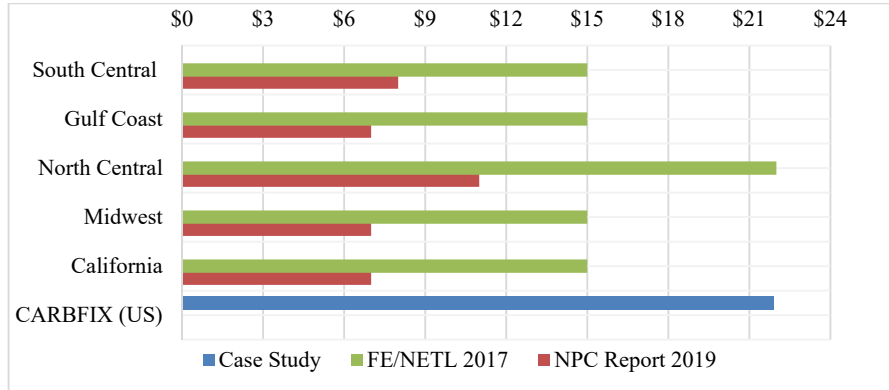


Figure 3. Saline formation CO₂ storage cost by region compared to US Carbfix cost

4. Conclusions

This paper reviews the cost analysis for a proposed case study in the US, based on the Carbfix pilot project and methodology. The cost analysis is used for the purposes of profitability and a sensitivity assessment.

The resulting cost is heavily weighted in operational costs, specifically water and electricity, which should be carefully examined for cost reduction strategies. More research should be carried out to study the water requirement to acquire the correct reactions and minimize water, and thus power usage. If minimum water quantity is achieved, using seawater or recycling water from the reservoir, once the CO₂ carbonated,

are the main alternatives to reduce costs. In terms of electricity, although the case study price for electricity is among the lowest in the US, there is some room for improvement by considering on-site generated renewable energy. It was also found that well borehole depth affects the cost significantly when varying depth from 400m to 800m. Although the single injection well cost increase with depth, water demand decrease, which creates important cost reductions. However, injectivity rates from a different formation than the one assumed in the case study can determine a different number of injection wells, altering the results.

The profitability assessment results in a price of 21.9 USD/tCO₂, which is higher than the current prices for saline formations in the US. However, Carbfix technology seems to be a good alternative when conventional storage formations are not available. By improving water demand and costs, the US Carbfix CO₂ storage price may become competitive. The tornado diagram shows that the total price of 11.3USD/tCO₂ can be achieved with a minimum cost of water.

In light of the findings, it is recommended to perform a study for different basalt formation locations around the US, getting specific data regarding the reservoirs and available utilities. In future work, this methodology will be integrated to SESAME framework, which is a system-scale energy analysis tool to assess the system-level greenhouse gas (GHG) emissions of today's changing energy system.

Acknowledgments

This publication has been funded by LCEC CCUS Center. The authors would like to thank colleagues at Carbfix and MIT Energy Initiative. Specifically, they would like to thank Kári Helgason and Bergur Sigfússon from Carbfix.

References

- H. Alfredsson, B. Hardarson, H. Franzson, S.R. Gislason, 2008, CO₂ sequestration in basaltic rock at the Hellisheidi site in SW Iceland: stratigraphy and chemical composition of the rocks at the injection site, *Mineralogical Magazine*, 72(1):1-5.
- S. Bunch, K. Cort K, E. Johnson E, D. Elliott D, K. Stoughton, 2017, *Water and Wastewater Annual Price Escalation Rates for Selected Cities across the United States*, Office of Energy Efficiency & Renewable Energy.
- International Energy Agency, 2020, *Average Price of Electricity to Ultimate Customers by End-Use Sector*.
- S.R. Gislason, E.H. Oelkers, 2014, Carbon storage in basalt, *Science*, 344(6182):373-4.
- S.R. Gislason, D. Wolff-Boenisch, A. Stefansson, E.H. Oelkers, E. Gunnlaugsson, H. Sigurdardottir, 2010, Mineral sequestration of carbon dioxide in basalt: A pre-injection overview of the CarbFix project, *International Journal of Greenhouse Gas Control*, 4(3):537-45.
- T. Grant T, D. Morgan, 2017, *CO₂ Saline Storage Cost Model: User's Manual*, NETL.
- National Petroleum Council, 2019, *Meeting the Dual Challenge A Roadmap to At-Scale Deployment of Carbon Capture, Use, and Storage*, Chapter 2: CCUS supply chains and economics.
- E. Ragnheidardottir, H. Sigurdardottir, H. Kristjansdottir, W. Harvey, 2011, Opportunities and challenges for CarbFix: An evaluation of capacities and costs for the pilot scale mineralization sequestration project at Hellisheidi, Iceland and beyond, *International Journal of Greenhouse Gas Control*, 5(4):1065-72.
- S. Selosse, O. Ricci, 2017, Carbon capture and storage: Lessons from a storage potential and localization analysis, *Applied Energy*, 188:32-44.

L.A. Smith, N. Gupta, B.M. Sass, T.A. Bubenik, C. Byrer, P. Bergman, 2001, Engineering and economic assessment of carbon dioxide sequestration in saline formations, National Energy Technology Lab, Pittsburgh, PA, and Morgantown, WV.

S.Ó. Snæbjörnsdóttir, B. Sigfússon B, C. Marieni, D. Goldberg, S.R. Gislason, E.H. Oelkers, 2020, Carbon dioxide storage through mineral carbonation, Nature Reviews Earth & Environment,1-13.

Computational Exploration of Adsorption Enhanced Haber-Bosch using MOFs and Ionic Liquid/MOFs

Amro M. O. Mohamed, ^{a*} Yusuf Bicer ^a

^aDivision of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Education City, Doha, Qatar

ammohamed@hbku.edu.qa

Abstract

In this work, molecular simulation is utilized to perform the adsorption based separation of the ammonia-hydrogen-nitrogen mixture on Metal-Organic Frameworks (MOFs) operating conditions relevant to enabling low-pressure Haber-Bosch. MOFs are nanoporous structures that possess several desirable features; among them is the tunability in which one can target specific molecules through replacement or functionalization of organic linkers, metal nodes, and finally, cage decoration. This study aims to provide an efficient ammonia separation to reduce operating pressure in the ammonia reactor and highlight MOFs' potential as material for novel applications in gas-gas separation. The pressure/temperature swing adsorption operational scheme is investigated, and the working delivered capacity and purity of ammonia are determined accordingly. Molecular simulation provides a way to examine nanomaterial potential in such applications inspecting a range of process conditions where the material is characterized based on loading, selectivity, and regeneration ability. Following an initial computational screening, Co₂Cl₂BBTA MOF is selected, and force field modifications have been done to fit experimental data. Besides analyzing the performance of Co₂Cl₂BBTA, encapsulation of ionic liquid (IL) [bmim][Tf₂N] effect was analyzed from structural and adsorption properties. Interestingly, an optimum IL loading is determined based on the performance objective. Purity factor reached 93.3% at IL loading of 0.372 weight fraction. When considering capacity and purity factor, IL composition in MOF corresponding to 0.165 weight fraction revealed the best performance. Implementation of adsorption enhanced Haber-Bosch is expected to reduce immensely the amount of electrical power utilized to recycle the unreacted syngas to the ammonia synthesis reactor and produce ammonia at high pressures desirable for urea synthesis.

Keywords: Molecular simulation, metal-organic framework, ammonia, separation processes, efficiency

1. Introduction

The Haber-Bosch process is the primary method in producing ammonia from nitrogen and hydrogen. Ammonia produced, utilized mainly as fertilizers, currently responsible for approximately 1.8% of carbon dioxide global emissions (The Royal Society, 2020). Efforts have been made to decarbonize the process, hoping to reduce the emission-intensive nature by implementing renewable energy sources. A critical parameter that drives current research investigations is to reduce the pressure and thus energy demand of ammonia production. One way to reduce pressure is by implementing

absorption/adsorption processes to target ammonia as it is produced. Palys et al. modeled an absorbent enhanced ammonia synthesis and optimize the operating conditions to minimize the cost (Palys et al., 2018). The work was motivated by introducing enhanced absorbent ammonia using a bed of supported alkali metal salt (Malmali et al., 2018; Wagner et al., 2017).

Metal-organic frameworks (MOFs) have been linked to ammonia in the literature due to MOFs' potential towards air purification of toxic chemicals (Wang et al., 2018). In 2018, a MOF was reported as the highest loading and capacity of ammonia (Rieth and Dincă, 2018). The top material showed 8.56 mmol/g capacity under personal protection equipment relevant conditions. The capacity value is higher than the previous state of the art HKUST-1 MOF by 27% (Rieth and Dincă, 2018). At equilibrium, on the other end, the adsorption at 1 bar reached 19.79 mmol ammonia per gram of the MOF (Cu₂Cl₂BBTA). Also, ammonia adsorption was studied in four MOFs (MIL-47, IRMOF-1, IRMOF-10, and IRMOF-16) (Yu et al., 2012). Ammonia adsorption on these MOFs indicated relatively weak ammonia-MOF interactions.

One way to functionalize a MOF for a particular application is evolved because of these materials' permanent porosity. The incorporation of organic materials inside MOFs' cages has been investigated in the literature, mostly for CO₂ capture. The most common molecules are ionic liquids (IL), which has several desirable and unique characteristics to bind to polar molecules. Han et al. constructed IL incorporated in MOF composite for increased ammonia adsorption capacity (Han et al., 2020). The materials thus possess several adsorption sites important for capturing ammonia in humid and dry conditions. The obtained structure of [BoHmim][Zn₂Cl₅]@MIL-101(Cr) was shown to have a record of 24.12 mmol/g at 1 bar and 298 K. Not only this excellent capacity is observed under dry conditions but also at conditions nearing the water-saturated ammonia solution as water provides additional sites for ammonia adsorption.

This work investigates MOFs' potential towards integration in the ammonia synthesis loop in low-pressure adsorption enhanced Haber-Bosch conditions using molecular simulation. The exploration is initiated by screening a small MOFs dataset followed by adsorption performance evaluation. Moreover, the influence of IL encapsulation in MOFs on ammonia capacity and selectivity is examined.

2. Methodology

2.1. Screening of MOFs and Force Field Modification

In this work, the structure set from (Rosen et al., 2019) was used for screening purposes and evaluating structure-property relationships. The dataset is based on experimental MOFs with various organic linkers and metal nodes. Structures in this dataset are open metal site MOFs which present a higher loading capacity of polar molecules. Grand Canonical Monte Carlo (GCMC) simulations have been throughout the study to compute the adsorption properties of structures using RASPA software (Dubbeldam et al., 2016). Structural parameters such as pore limiting diameter (PLD), largest cavity diameter (LCD), and pore volume have been evaluated using zeo++ code (Willems et al., 2012). Ammonia and nitrogen have been modeled using the TraPPE force field (Martin and Siepmann, 1998; Zhang and Siepmann, 2010), whereas hydrogen is modeled using Darkrim-Levesque three-site model. MOFs have been assumed as rigid structures and parameters for non-bonded interactions obtained from Universal Force Field (UFF) parameters (Rappe et al., 1992). Although the choice of force field impacts gas separation performance evaluation, Dokur and Keskin showed that MOFs' ranking using metrics is

not sensitive (Dokur and Keskin, 2018; McDaniel et al., 2015). The structures are first screened for ammonia capture properties using a combination of ideal selectivity of NH_3/N_2 using the ratio of Henry's constants at 298K (K_t^0) and the adsorption of ammonia at 1 bar and 298K (N_{ads,NH_3}). Factor f in Eq. (1) combines the importance of ammonia capacity while desiring selectivity.

$$f = \frac{K_{NH_3}^0}{K_{N_2}^0} \times N_{ads,NH_3} \quad (1)$$

As a result of the presence of open metal sites (OMS), computational of the adsorption figures are not precise due to the rise of unique interactions between adsorbates and the unsaturated metal centers (Daglar and Keskin, 2020). Here, we modify the MOF structure's UFF structure, altering its cohesive energy to match ammonia's adsorption isotherm before performance evaluation. This technique has been previously used to fit experimental adsorption properties data (Pérez-Pellitero et al., 2010).

2.2. Encapsulation of Ionic Liquid and Performance Evaluation of MOFs and IL@MOFs

In this work, IL pairs' insertion was accomplished in the canonical ensemble at various IL loading values as implemented previously in the literature (Mohamed et al., 2020, 2019). The LJ and point charges of atoms in [bmim][Tf₂N] IL were obtained from the literature (Androulaki et al., 2012). Computations towards adsorption properties were carried similar to the procedure aforementioned. The investigation of MOF and IL@MOF towards the enhanced Haber-Bosch process was to estimate adsorption of the studied structures at adsorption and desorption conditions. Adsorption conditions at 20 bar and 170 °C, whereas desorption conditions are at 13 bar and 500 °C. Ammonia mole fraction at the reactor's outlet is 4.5% (Palys et al., 2018). As a result, the adsorber's inlet has a mole composition of 23.9%, 71.6%, and 4.5% of N₂, H₂, and NH₃, respectively. The composition of the desorption conditions was determined per the equilibrium adsorption at the adsorption stage. Equilibrium absolute adsorption is determined in multicomponent GCMC simulations. Three critical indicators are investigated, the regenerability factor R% (Bae and Snurr, 2011), used here for ammonia adsorption as in Eq. (2), and a new parameter P that can be observed as the purity of the stream after the desorption step (Eq. (3)). Parameter P is similar to the working capacity indicator in previous literature studies. In order to combine purity with capacity, metric S is introduced in Eq. (4).

$$R\% = \frac{\Delta N_{NH_3}}{N_{ads,NH_3}} \times 100\% \quad (2)$$

$$P = \frac{\Delta N_{NH_3}}{\Delta N_{NH_3} + \Delta N_{N_2} + \Delta N_{H_2}} \quad (3)$$

$$S = P \times \Delta N_{NH_3} \quad (4)$$

3. Results and Discussion

The first part of the results is related to the initial screening of the selected set of structures. Equation 1 is used to highlight the best performing materials for ammonia capacity and best selective ones. The initial screening demonstrated that Co₂Cl₂BBTA

(BBTA: 1H,5H-benzo(1,2-d:4,5-d')bistriazole), Ni-BTP (1,3,5-tri(1H-pyrazol-4-yl)benzene) and $\text{Cu}_3(\text{DMTRZ})_2(\text{ox})_2$ are the highest f factor among the MOFs in the dataset while possessing required PLD for the particular application. The factor reached 12.3, 26.3, and 18.1 for the MOFs, respectively. It has to be noted that if one observes the ideal NH_3/N_2 selectivity, MOFs built from 1,3,5-benzenetricarboxylate (BTC) organic linker exhibit the highest selectivity behind $\text{Cu}_3(\text{DMTRZ})_2(\text{ox})_2$. However, the combination of capacity and selectivity is the goal behind this screening. Structural analysis showed that $\text{Co}_2\text{Cl}_2\text{BBTA}$ and Ni-BTP have LCD of 10.3 and 14.9 Å, respectively, which can be exploited via IL's insertion to increase materials' adsorption selective properties.

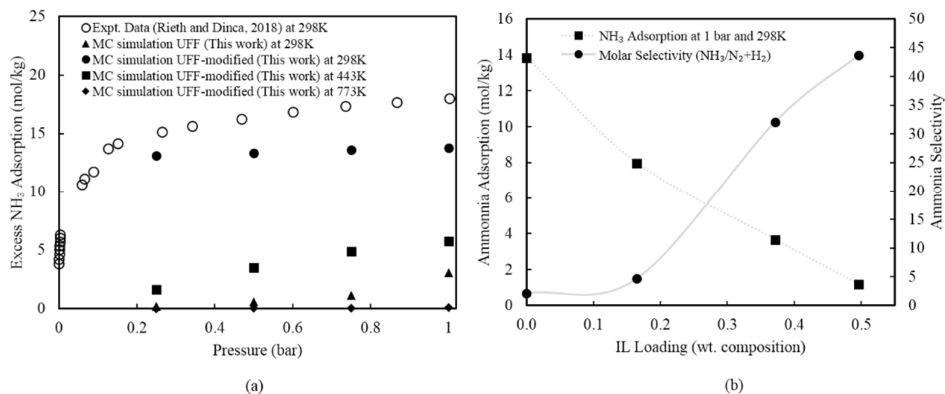


Figure 1. (a) Adsorption isotherms of NH_3 in $\text{Co}_2\text{Cl}_2\text{BBTA}$ at different temperatures and pressures (b) effect of IL loading composition on adsorption and selectivity of ammonia

To obtain precise performance evaluation, adsorption isotherms, especially ammonia, have to fit experimentally available data. As mentioned in the methodology section and because the generic force field does not resemble interaction between MOFs with OMS and adsorbates, further modification of the MOF's LJ parameters must be conducted. Among the top three candidate materials, $\text{Co}_2\text{Cl}_2\text{BBTA}$ has to be selected for this study as it has been studied experimentally and presented the highest ammonia capacity at the time of its publication. Several trials to modify the UFF towards a better agreement with experimental ammonia adsorption data, most notably between 0.5 and 1 bar. The depth of the potential well of the LJ potential ϵ was adjusted to 2.25 times the previous UFF, and relatively good agreement can be seen in Figure 1(a) to the experimental measurements. The modified parameters do not provide an accurate representation of the isotherm at the near zero-pressure limit; thus, further modification and incorporation of quantum chemistry computation are necessary. The adjusted value reflects the increase in dispersion-repulsion interactions cohesive energy of NH_3 . Figure 1(a) also reveals the influence of temperature on the adsorption isotherm using the modified force field. The effect of temperature is useful to better select adsorption and desorption conditions based on the sorbent.

Incorporation of IL inside MOFs' cavity can result in enhanced adsorption or selectivity depending on the studied mixture's components. The insertion of [bmim][Tf₂N] in $\text{Co}_2\text{Cl}_2\text{BBTA}$ was accomplished computationally at three different IL loading

(composition). The IL loading effect on structural properties such as free volume and surface area have been studied first. Then, adsorption of ammonia at 1 bar and 298K using the modified force field for the MOF. It can be observed that the presence of IL reduced the available volume significantly in the MOF. IL loading at 0.372 has reduced the free volume up to 64% relative to the pristine MOF. Besides, from Figure 1(b), adsorption of ammonia per kg of the total structure (MOF+IL) decreases as more ILs pairs are incorporated. The reduction in adsorption per mass can be explained as ILs covers essential favorable adsorption sites for ammonia. Meanwhile, selective adsorption is favored higher IL loading structures as a result of cavity volume reduction.

Co₂Cl₂BBTA and ILs incorporated structures are also investigated based on the conditions relevant to the low-pressure Haber-Bosch process. Relevant parameters have been calculated as described in the methodology section for the structures. The low vapor pressure of ammonia presents a challenge for high capacity and selectivity since the mole fraction of ammonia at the inlet is as low as 4.5%. The metrics results show that IL inclusion increases the purity parameter *P* which a factor of selectivity at adsorption and desorption conditions (Table 1). The inclusion of ILs reduces the free volume, which increases the selectivity significantly towards ammonia at adsorption conditions. The parameter *P* reaches a maximum at 0.372 weight fraction, while higher IL reduced the figure. However, when combining *P* with the working capacity of ammonia, lower IL loading composition at 0.165 shows the best at 1.533. The reason is the working capacity is slightly lower than pristine MOF while having better selectivity. The regenerability factor reached 72.68% for the pristine MOF. Reduction in regenerability factor is observed as ILs pairs increase in number as expected.

Table 1. Performance parameters for Co₂Cl₂BBTA MOF and IL encapsulated structures

IL weight composition	Ammonia adsorption (mol/kg) at adsorption condition	<i>P</i>	<i>S</i>	<i>R</i> %
Pristine MOF	3.311451	0.626	1.507	72.68
0.165	3.145765	0.761	1.533	64.03
0.372	1.496324	0.933	0.584	41.86
0.496	0.381627	0.915	0.084	24.13

4. Conclusions

Adsorption enhanced Haber-Bosch has the potential to reduce the pressure in the ammonia synthesis loop. High ammonia affinity at relatively lower partial pressure is necessary to achieve significant adsorption figures. In this work, MOFs have investigated the implementation of a low-pressure ammonia synthesis loop using molecular simulations. The study is initiated by screening synthesized MOFs small dataset, in which candidate materials were highlighted. Co₂Cl₂BBTA MOF was selected as one of the three top-performing materials, while experimental adsorption data is available. Modification of generic force to describe LJ parameters of MOFs was carried out to fit experimental data. Following that, performance evaluation of Co₂Cl₂BBTA and IL encapsulated MOF Co₂Cl₂BBTA was determined using several metrics and indicators, including purity and regenerability of adsorption sites. It has been shown that IL loading plays a role in increasing selectivity towards ammonia while reducing the capacity. An optimum composition is set given the objective function. From an overall point of view, these adsorbents increase the cost of the system's separation and may jeopardize the purity of the product. However, they provide a means to reduce the intensity of the ammonia

chemical plant. Therefore, a possible synergy between adsorption based and vapor-liquid separations could be a viable alternative.

Acknowledgment

This publication was made possible by an Award GSRA5-1-0513-18051 from Qatar National Research Fund (a member of Qatar Foundation). We are grateful to the High Performance Computing Center of Texas A&M University at Qatar for generous resource allocation.

References

- Androulaki, E., Vergadou, N., Ramos, J., Economou, I.G., 2012. Structure, thermodynamic and transport properties of imidazolium-based bis(trifluoromethylsulfonyl)imide ionic liquids from molecular dynamics simulations. *Mol. Phys.* 110, 1139–1152.
- Bae, Y.S., Snurr, R.Q., 2011. Development and evaluation of porous materials for carbon dioxide separation and capture. *Angew. Chemie - Int. Ed.*
- Daglar, H., Keskin, S., 2020. Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations. *Coord. Chem. Rev.* 422, 213470.
- Dokur, D., Keskin, S., 2018. Effects of Force Field Selection on the Computational Ranking of MOFs for CO₂ Separations. *Ind. Eng. Chem. Res.* 57, 2298–2309. <https://doi.org/10.1021/acs.iecr.7b04792>
- Dubbeldam, D., Calero, S., Ellis, D.E., Snurr, R.Q., 2016. RASPA: Molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* 42, 81–101.
- Han, G., Liu, C., Yang, Q., Liu, D., Zhong, C., 2020. Construction of stable IL@MOF composite with multiple adsorption sites for efficient ammonia capture from dry and humid conditions. *Chem. Eng. J.* 401, 126106.
- Malmali, M., Le, G., Hendrickson, J., Prince, J., McCormick, A. V., Cussler, E.L., 2018. Better Absorbents for Ammonia Separation. *ACS Sustain. Chem. Eng.* 6, 6536–6546. <https://doi.org/10.1021/acssuschemeng.7b04684>
- Martin, M.G., Siepmann, J.I., 1998. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* 102, 2569–2577.
- McDaniel, J.G., Li, S., Tylaniakis, E., Snurr, R.Q., Schmidt, J.R., 2015. Evaluation of Force Field Performance for High-Throughput Screening of Gas Uptake in Metal–Organic Frameworks. *J. Phys. Chem. C* 119, 3143–3152.
- Mohamed, A.M.O., Krokidas, P., Economou, I.G., 2020. Encapsulation of [bmim +][Tf₂N –] in different ZIF-8 metal analogues and evaluation of their CO₂ selectivity over CH₄ and N₂ using molecular simulation. *Mol. Syst. Des. Eng.* 5, 1230–1238.
- Mohamed, A.M.O., Moncho, S., Krokidas, P., Kakosimos, K., Brothers, E.N., Economou, I.G., 2019. Computational investigation of the performance of ZIF-8 with encapsulated ionic liquids towards CO₂ capture. *Mol. Phys.* 1–15. <https://doi.org/10.1080/00268976.2019.1666170>
- Palys, M., McCormick, A., Cussler, E., Daoutidis, P., 2018. Modeling and Optimal Design of Absorbent Enhanced Ammonia Synthesis. *Processes* 6, 91.
- Pérez-Pellitero, J., Amrouche, H., Siperstein, F.R., Pirngruber, G., Nieto-Draghi, C., Chaplais, G., Simon-Masseron, A., Bazer-Bachi, D., Peralta, D., Bats, N., 2010. Adsorption of CO₂, CH₄, and N₂ on zeolitic imidazolate frameworks: Experiments and simulations. *Chem. - A Eur. J.* 16, 1560–1571.
- Rappe, A.K., Casewit, C.J., Colwell, K.S., Goddard, W.A., Skiff, W.M., 1992. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* 114, 10024–10035.
- Rieth, A.J., Dincă, M., 2018. Controlled Gas Uptake in Metal–Organic Frameworks with Record Ammonia Sorption. *J. Am. Chem. Soc.* 140, 3461–3466.
- Rosen, A.S., Notestein, J.M., Snurr, R.Q., 2019. Structure–Activity Relationships That Identify Metal–Organic Framework Catalysts for Methane Activation. *ACS Catal.* 9, 3576–3587.

- The Royal Society, 2020. Ammonia: zeo-carbon fertiliser, fuel and energy store. London.
- Wagner, K., Malmali, M., Smith, C., McCormick, A., Cussler, E.L., Zhu, M., Seaton, N.C.A., 2017. Column absorption for reproducible cyclic separation in small scale ammonia synthesis. *AIChE J.* 63, 3058–3068.
- Wang, L., Xia, M., Wang, H., Huang, K., Qian, C., Maravelias, C.T., Ozin, G.A., 2018. Greening Ammonia toward the Solar Ammonia Refinery. *Joule* 2, 1055–1074.
- Willems, T.F., Rycroft, C.H., Kazi, M., Meza, J.C., Haranczyk, M., 2012. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* 149, 134–141.
- Yu, D., Ghosh, P., Snurr, R.Q., 2012. Hierarchical modeling of ammonia adsorption in functionalized metal–organic frameworks. *Dalt. Trans.* 41, 3962–3973.
- Zhang, L., Siepmann, J.I., 2010. Development of the trapezoidal force field for ammonia. *Collect. Czechoslov. Chem. Commun.* 75, 577–591.

Optimization of a Continuous Fluidized Bed Process for the Separation of Enantiomers by Preferential Crystallization

Nadiia Huskova^a, Michael Mangold^b, Andreas Seidel-Morgenstern^a

^a*Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany*

^b*Bingen University of Applied Sciences, Berlinstraße 109, 55411 Bingen, Germany*
huskova@mpi-magdeburg.mpg.de

Abstract

Enantiomers are molecules found in two variants being non-superimposable mirror images of each other. Chemical synthesis typically produces a 50:50 % mixture of both enantiomers. The separation of such a mixture is crucial for many applications, because enantiomer and counter-enantiomer often have completely different effects on biological organisms or the human body. However, the separation is a challenging task, as the physical properties of both enantiomers are identical.

The presented work consists of two parts, which are based on model optimization. The first part of this paper describes the process of fitting the model to experimental data described in (Gänsch et al., 2020). The focus of the second part is on optimization considering two conflicting objectives: total harvested product mass and purity of the product. A sensitivity analysis presented in (Gänsch et al., 2020) is used as a basis for the selection of optimization variables. Some parameters have a significant impact, and their optimization allows to control not only the key characteristics as productivity and purity, but also the mean crystal size. The genetic algorithm CMAES (Hansen, 2006) and parallel computing techniques are applied for optimization.

Keywords: enantiomers, separation, mathematical modelling, population balance equation, genetic optimization

1. Introduction. Enantiomers and process setup

Enantiomers are optically active stereoisomers being non-superimposable mirror images of each other. They can exist as levorotatory (-) and dextrorotatory (+) molecules and are denominated usually as L and D. While possessing identical physical properties, enantiomer and counter enantiomer may have different effects on biological organisms and the human body. Based on the physical identity of the enantiomers, separation becomes a challenging task that requires right technique and operation conditions. In the concept of the preferential crystallization (PC), a desired enantiomer is separated from a racemic solution by providing its single seeds. Cooling crystallization technology is applied to promote crystal growth in the solution. Detailed explanation of the PC concept is provided in (Coquerel, 2006). The setup of the fluidized bed crystallizer considered in this work and a proof of principle were given in (Temmel et al, 2020).

Super-saturated solution enters the bottom of a tubular crystallizer (see Fig. 1) and causes growth of seeding crystals inside the crystallizer. Small crystals move with the liquid flow to the crystallizer top and are recycled in a feed tank. Larger crystals sink to the bottom, are crushed in a mill and sent back to the crystallizer as new feed. The product is withdrawn at an outlet on the side of the crystallizer. This is done in a periodic pulsewise manner. To describe the periodic product removal, the triggering of an upper photo sensor (position h_s , Fig. 1) has to be modeled. It is assumed that the product withdrawal is initiated by the volume fraction of the liquid phase at the sensor position h_s falling below a pre-specified volume fraction $\epsilon_{desired}$. In a next step the withdrawal of a segment of the suspension is described. In the model, this is assumed to happen instantaneously, i.e. the suspension located within the height segment Δh_p located above the product outlet, H_{prod} , is taken out as the harvested product. The removed volume in the height segment Δh_p is equal the volume between the two sensors, Δh_s . During product collection the suspension above $H_{prod} + \Delta h_p$ moves downwards over the length Δh_p . In a last modeling step, the upper part of the crystallizer, which is now “empty”, has to be replenished. This is done by filling up the empty top segment of the crystallizer with pure liquid possessing the racemic inlet composition.

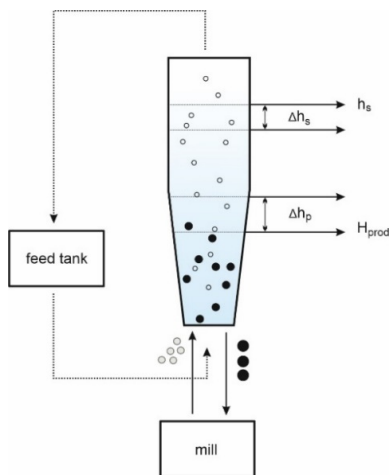


Figure 1. Scheme of a single fluidized bed crystallizer for continuous preferential enantioselective crystallization.

Process model and variables

Crystallization processes are complex and difficult to model and control, e.g. they are multicomponent systems, multiphase systems, there are property, space and time dependent variables. However, when information of the solubility and kinetics are available, modeling and further control of the separation process is possible. For understanding and designing the process, it is not enough to rely only on experimental studies. The model used should be sufficiently detailed, and at the same time allow computations to be performed with reasonable effort.

A model, which provides a reasonable compromise between these conflicting objectives, has been published in (Mangold et al., 2017). It could be shown that that model describes the interaction between crystal growth, particle transport, and fluid dynamics in the isothermal case with reasonable accuracy. A refined process model from (Gänsch et al., 2020) distinguishes between the L and the D enantiomer. This allows to evaluate the purity of the withdrawn product, which is important when productivity of chiral substance increases. Further, it is well known that crystallization kinetics crucially depend on the temperature, for this reason an energy balance equation is added. Detailed description of the model assumptions and equations is provided in (Gänsch et al., 2020).

A few of the main model assumptions are listed below (for further details, see Gänsch et al., 2020). The crystal population is described by a number density function that depends on a single space coordinate h in direction of the liquid flow, a single property coordinate L representing a characteristic crystal size, and on time t . Plug flow conditions are assumed for the liquid. The particle velocity, which differs from the liquid velocity, is described by the classical model by Richardson and Zaki (Richardson and Zaki, 1954). Prediction of the crystal shape is based on available experiment data and it is described by the sphericity parameter ψ (Binev et al., 2015). The only asymmetry between the two enantiomers comes from the initial conditions. Crystal growth is modeled by experimentally validated kinetics (Temmel et al., 2020). Nucleation kinetics of a similar structure as the growth kinetics are assumed in order to assess the purity of the product. Instead of assuming a constant product flow rate as in (Mangold et al., 2017), the periodic, pulse-wise withdrawal of product crystals is modeled. As a simple model of the disperser, we use a standard population balance model with breakage.

The developed model allows for studying the process and making predictions in terms of productivity, yield and mean product crystal size.

Laboratory experiments described in (Gänsch et al., 2020) are used as nominal case for the presented study. It could be shown that the model agrees reasonably well with experimental data. However, a systematic adaptation of parameter values not accessible to direct measurements has not been done, yet. This is one objective of this contribution with the aim to provide a valid and efficient process model. The second objective is to maximize productivity and purity by finding optimal operation conditions in numerical optimizations. Both tasks are non-trivial due to the high system order of the process model, the inherent nonlinearities of the process, and the periodic operation regime.

Model parameters

Based on the sensitivity analysis presented in (Gänsch et al., 2020), it was determined that not all parameters equally affect the key characteristics of the crystallization process. For a better understanding of the process, all the model parameters are divided into two groups. The first group of parameters has confirmed values obtained from separate measurements made in the laboratory. The values of the second group are based on the preliminary study.

Due to the availability of laboratory experiments, only those parameters are fitted in the following chapter that do not have experimentally confirmed values. Four model parameters belong to the second group and are chosen as optimization variables:

x_1 - critical breakage size L_c , [μm]. The critical particle size L_c , above which breakage may occur, is chosen as the gap between rotating blade and stationary parts of the disperser.

x_2 - desired fraction of the fluid $\varepsilon_{desired}$ at the sensor position, which in the model determines the triggering of a new product removal pulse.

x_3 - number of daughter crystals p in a break-up event inside the mill

x_4 - sphericity ψ . It is defined as the ratio of the surface area of a sphere, which has the same volume as the particle, to the surface area of the particle.

Asparagine monohydrate is used as an example, corresponding data on solubility and kinetic parameters were obtained experimentally and presented in (Temmel et al.,2020).

2. Optimization of the process model

Parameter fitting

The main idea of this chapter is to minimize the difference between the values of the objective functions OF obtained by experimental and simulation methods using a genetic algorithm CMAES (Hansen, 2006).

The focus of this optimization round is to achieve productivity close to the nominal case described in (Gänsch et al., 2020). Desired productivity as well as results of performed simulations presented in Table 1.

$$OF_1 = (f_{sim}^1 - f_{exp}^1)^2, \quad (1)$$

where f^1 - productivity $Pr, [g \cdot l^{-1} \cdot h^{-1}]$

Table 1. Results of the productivity fitting

f_{exp}^1	33.1						
Iteration GA	1	2	3	4	5	6	7
f_{sim}^1	41.07	38.4	37.4	37.1	36	35.82	31.33
OF_1	63.52	28.09	18.49	16	8.41	7.4	3.13

The use of the CMAES algorithm allows us to get closer to the experimental productivity. However, achieving the desired productivity leads to a decrease in other process characteristics. In order to find compromise between conflicting criteria, the extended objective function OF_2 defines a weighted square error of four criteria f^k :

f^1 - productivity $Pr, [g \cdot l^{-1} \cdot h^{-1}]$;

f^2 - time between 2 withdrawals after reaching cyclic steady state $\Delta t, [h]$;

f^3 - mean crystal size of the product $\bar{L}, [\mu m]$;

f^4 - standard deviation of the product size distribution $\sigma, [\mu m]$;

$$OF_2 = w_1 \cdot (f_{sim}^1 - f_{exp}^1)^2 + w_2 \cdot (f_{sim}^2 - f_{exp}^2)^2 + w_3 \cdot (f_{sim}^3 - f_{exp}^3)^2 + w_4 \cdot (f_{sim}^4 - f_{exp}^4)^2 \quad (2)$$

Since the criteria f^k have due to their nature a different scale, they are weighted with weighting coefficients w_k , which assure similar impact of each criterion on the objective

function OF_2 . For this the values $w_1 = 0.08, w_2 = 26, w_3 = 0.07, w_4 = 0.0028$ were chosen.

The results of the parameter identification are summarized in Table 2. The parameter fitting process shows an improvement of the criteria.

Table 2. Results of the model fitting

Criteria	Experiment	Simulation	1	2	3	4	5
f^1	33.1		35	48.54	49	47.88	47.6
f^2	0.65	1.7	1.22	1.16	1.08	1.06	
f^3	271.4	291	269.88	273.14	271.34	271.02	
f^4	58.2	16.50	13.63	12.76	12.53	12.31	
OF		60.71	33.19	32.87	28.12	27.18	

Optimization of the operating parameters

After finding the reference solution and achieving the quantitative agreement between the simulation and experiment results, we return to sensitivity analysis to determine parameters for further optimization. It was confirmed (Gänsch et.al, 2020) that operating parameters such as inlet liquid concentration c_{in}^i , and inlet volume flow rate V_{in} play a significant role in the crystallization process. The focus of this section is on the periodical steady state solution of the crystallization process and its dependence on the operating parameters.

The bi-criterial optimization problem is converted to a single objective problem by introducing a Pareto coefficient w varying from 0 to 1. The resulting optimization function is presented in the form:

$$OF_3 = (1 - w) \cdot \text{purity} + w \cdot \text{product} \quad (3)$$

The aim of the simulation is to find optimal operating parameter values to increase the productivity and the purity of the product enantiomer.. When using the coefficient $w=0$, the purity of desired enantiomer is maximized, when using the coefficient $w=1$, emphasis is on the amount of harvested mass. As an initial estimate for new points on the Pareto front, the maximum value of the objective function calculated of the previous step is used.

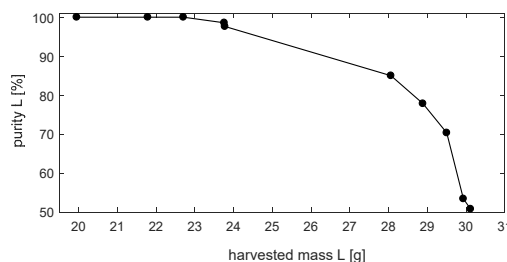


Figure 1. Pareto Front, obtained by using the CMAES algorithm. Number of individuals inside population $\lambda=88$ for the first iteration, $\lambda=33$ for subsequent iterations. Initial conditions for the inlet volume flow rate and the inlet liquid concentration were used from

the reference case (Gänsch et.al, 2020). Operation condition corresponding to the depicted points are given in the Appendix.

As a result of operating parameters optimization, suitable values for the inlet liquid concentration and the inlet volume flow rate were determined depending on the goal set.

Conclusions

Improved parameter fitting made it possible to provide a model that simulates key process characteristics as productivity, time window between harvestings, mean crystal size, and standard deviation of the withdrawal product. Subsequent optimization of essential operating parameters has demonstrated the potential of the process with respect to achieve high purity and productivity. The evolutionary algorithm CMAES showed high efficiency and is promising for other studies. In view of the big data and the complexity of calculations, parallelization techniques should be applied.

References

Gänsch, J.; Huskova, N.; Kerst, K.; Temmel, E.; Lorenz, H.; Mangold, M.; Janiga, G.; Seidel-Morgenstern, A. Continuous Enantioselective Crystallization of Chiral Compounds. *Chemical Engineering Journal* 2020/424 (submitted).

N. Hansen, The CMA evolution strategy: a comparing review, in: J. A. Lozano, Larrañaga, P., Inza, I., Bengoetxea, E. (Eds.), *Towards a new evolutionary computation. Advances on estimation of distribution algorithms*, Springer, 2006, pp. 75-102.

Coquerel, G., *Preferential Crystallization*. In *Novel Optical Resolution Technologies*; Springer, 2006; Vol. 269, 416 *Topics in Current Chemistry*, pp. 1–51

Temmel, E.; Gänsch, J.; Seidel-Morgenstern, A.; Lorenz, H. Systematic investigations on continuous fluidized bed crystallization for chiral separation. *Crystals* 2020, 10

M. Mangold, D. Khlopov, E. Temmel, H. Lorenz, A. Seidel-Morgenstern, Modelling geometrical and fluid-dynamic aspects of a continuous fluidized bed crystallizer for separation of enantiomers. *Chem. Engng. Sci.*, 160 (2017) 281-290.

J. Richardson, W. Zaki, The sedimentation of a suspension of uniform spheres under conditions of viscous flow, *Chem. Engng. Sci.*, 3 (1954) 65-73.

D. Binev, A. Seidel-Morgenstern, H. Lorenz, Study of crystal size distributions in a fluidized bed crystallizer, *Chem. Engng. Science*, 133 (2015) 116-124.

Appendix

Table A1. Operating conditions

$c_{in}^i [kg \cdot m^{-3}]$	0.045	0.0458	0.046	0.052	0.054	0.0541	0.047	0.05	0.048	0.045
$V_{in} [l \cdot h^{-1}]$	7.56	7.7	7.6	7.6	8.1	8.4	11.0	11.3	11.7	14.3
$m_L [g]$	30.1	29.92	28.45	28.9	28.06	23.8	23.76	22.7	21.7	19.953
pur_{ity_L}	0.5	0.53	0.7	0.703	0.78	0.85	0.91	0.976	0.99	0.99

Methanol and power production from waste tyres: a process synthesis approach

Athi-enkosi Mavukwana^{a,*}, Ephraim Llane^b, James Fox^b, Baraka Celestin Sempuga^b

^a*Department of Civil and Chemical Engineering, College of Science, Engineering and Technology, University of South Africa (UNISA), c/o Christiaan de Wet & Pioneer Avenue, Florida Campus 1710, Johannesburg, South Africa*

^b*Institute for the Development of Energy for African Sustainability, College of Science Engineering and Technology, University of South Africa (UNISA), c/o Christiaan de Wet & Pioneer Avenue, Florida Campus 1710, Johannesburg, South Africa.*

Abstract

Waste tyres are part of solid waste materials that are toxic to the environment but have the potential to be converted to something of value. South Africa contributes approximately 11 million tyres to the global 1.5 billion tyres produced per annum. With waste tyres identified as a potential energy source due to its high carbon content and heat value, it is not surprising that the government of South Africa has invested in the creation of waste tyre management plan intending to recover products and energy from this waste material. In this paper, a waste tyre to methanol and electricity process is assessed from a thermodynamic and environmental perspective. Two key factors are analysed, carbon efficiency and chemical potential efficiency. The purpose of this study is to contribute to this research by finding the limit of performance for the conversion of waste tyres to methanol and power from a fundamental thermodynamic perspective. For the system developed a carbon efficiency of 54.6% and the chemical potential efficiency of 63% were achieved. Equally the system produces 2.42 GJ/ ton of waste tyre and has the potential to generate a revenue of \$ 620/ton of waste tyre. The study shows converting tyres to methanol is achievable, however, 45% of the carbon resource will end up as carbon dioxide.

Keywords: gasification, waste-to-methanol, electricity, carbon emissions, efficiency

1. Introduction

Solid waste materials such as biomass, waste tyres, and municipal solid waste, have been shown to contain adequate energy content to have the potential to substitute fossil fuels in the production of power and useful chemicals. Many existing technologies currently used for solid fossil fuels can be utilised easily to accommodate these waste materials, thus making the transition to alternate fuels achievable. Much of research in energy has largely focused on the optimization of technologies such as gasification, pyrolysis, combustion and liquefaction to estimate the optimum operating conditions to efficiently convert solid waste materials to useful products and power.

Developing countries such as South Africa have seen an increase in energy needs as the economy grows to include the participation of previously disadvantaged persons. However, South Africa's energy needs have been in dire straits in the last decade with the state producer Eskom losing most of its available capacity plunging the country into severe rolling blackouts, requiring upwards of 6000MW of power demand to be removed from the electrical grid to prevent a catastrophic failure of the nation's power network. Equally, the country saw the cost of liquid fuels also increasing due to the weakening currency and fluctuations in global crude oil prices, and demand. In 2020 the government announced plans to diversify South Africa's energy sources which will embrace new entrants and capacity into the energy space. This will allow generation for own use and securing power from independent power producers (IPP), creating easy access to cheap and clean energy (South Africa, 2020).

In terms of waste generation, the country has identified 38 waste streams that need to be diverted from landfills, and one such waste materials is waste tyres. South Africa contributes approximately 11 million tyres to the global 1.5 billion tyres produced per annum. Waste tyre recycling is identified in the National Waste Management Strategy (NWMS) as an area which can contribute towards South Africa's economic growth and the creation of green jobs (Godfrey and Oelofse, 2017). Waste tyres have a high carbon content and heat value, much of the research on waste tyres have been dedicated to recover products and energy from this waste material.

Methanol is a promising clean-burning fuel with a high octane number, and is an antecedent to the production of more complex chemical structures, such as dimethyl ether, methylamine, acetic acid, methyl tertiary butyl ether, etc (Dalena et al., 2018). The conversion of waste tyres to methanol and electricity would contribute to the country's energy mix. Studies by (Borgogna et al., 2019; Rivarolo et al., 2016) show that there is an environmental benefit to converting waste into methanol than burning it for energy as the CO₂ emissions reduced by an average of 14%. Whereas (Iaquaniello et al., 2017, 2018) studies show that converting refuse-derived fuel to methanol leads to about 40% and 30-35% reduction in greenhouse gas emissions for methanol production from fossil fuels and bio-resources, respectively. Waste to methanol process CO₂ emissions are half the emissions produced by waste to energy route.

The purpose of this study is to contribute to this research by finding the limit of performance for the conversion of waste tyres to methanol and power from a fundamental thermodynamic perspective.

2. Methodology

In this work, the thermodynamic efficiency of the process converting waste tyre to methanol and power is measured using two basic metrics, mainly; carbon efficiency and chemical potential efficiency. The properties are described as follows:

$$\text{Carbon Efficiency } (C_{eff}) = \frac{\text{moles of C in the desired product}}{\text{moles of C in the feed}} \quad (1)$$

$$\text{Chemical Potential Efficiency } (\eta_{cp,eff}) = \frac{\text{Gibbs free energy of combustion of products}}{\text{Gibbs free energy of combustion of feed}} \quad (2)$$

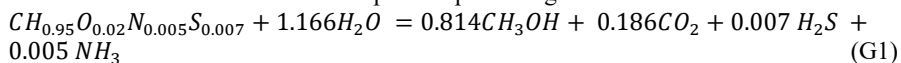
Equation 1 measures how much of the carbon in the feed stream ends up in the desired products. Less than 100% carbon conversion leads to carbon dioxide emissions. Equation 2 measures how much of the chemical potential stored in the feed material is translated to the desired products during chemical transformation (Sempuga & Yao, 2017). The thermodynamic properties of waste tyres utilised in this work can be found in (Mavukwana et al., 2020).

3. Result and discussion

3.1. Methanol production from tyres

The process starts with the gasification of tyres with steam to produce syngas. Steam is selected as the gasification agents since it yields a higher composition of H₂/CO compared to oxygen. The syngas from gasification is cleaned of acid gases and its H₂/CO ratio is adjusted in the water gas shift reactor to meet the required ratio for methanol in the synthesis reactor. The recommended ratio of H₂/CO is 2. The adjusted syngas is fed to the methanol synthesis reactor to produce methanol and water.

The overall material balance for the process producing methanol is:



$$\Delta H = 73.99 \text{ kJ}, \quad \Delta G = 105.36 \text{ kJ}$$

The performance of this process in equation G1 is as follows:

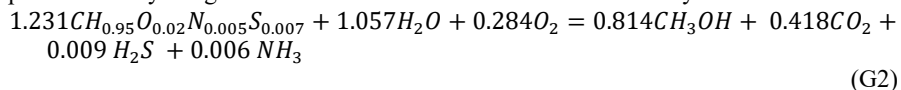
$$\text{Carbon efficiency} = 81.4\%$$

$$\eta_{cp} = 86.3\%$$

It is assumed that all the available sulfur leaves as hydrogen sulphide, since chemicals in the tyre rubber, such as zinc oxides, prevent sulfur-oxidising.

Equation (G1) represent the ideal process or the maximum achievable conversion of waste tyres to methanol. This process represents a case where all the operating units are energy integrated, with all the units that are exothermic sending their energies to the endothermic gasification step. That is, the water-gas shift, and the methanol synthesis will send all their energy to the gasification step, to try and achieve an adiabatic process. However, the overall mass balance of the process shows that the system still requires energy input. If this energy is not added the process would thus be infeasible. One strategy for supplying this energy is to combust a portion of the tyre feed.

Figure 1 shows the system that is fuelled by burning an additional number of tyres. The process is fully integrated and the overall material balance for the system is:



$$\Delta H = -29.9 \text{ kJ}, \quad \Delta G = 0 \text{ kJ}$$

The performance of this process in figure 1 equation G2 is as follows:

$$\text{Carbon efficiency} = 66\%$$

$$\eta_{cp} = 70.1\%$$

Therefore, supplying the required energy by burning an additional number of tyres increases the carbon emissions and thereby reduces the carbon efficiency to 66% and the chemical potential efficiency decreases to 70%. However, the process in figure 1 is not possible with conventional technology. Since the gasification step operates at higher

temperatures than the other units, the transfer of energy between the various process units would have to be done against a temperature gradient and practically that might require expensive equipment (Fox & Stacey, 2019).

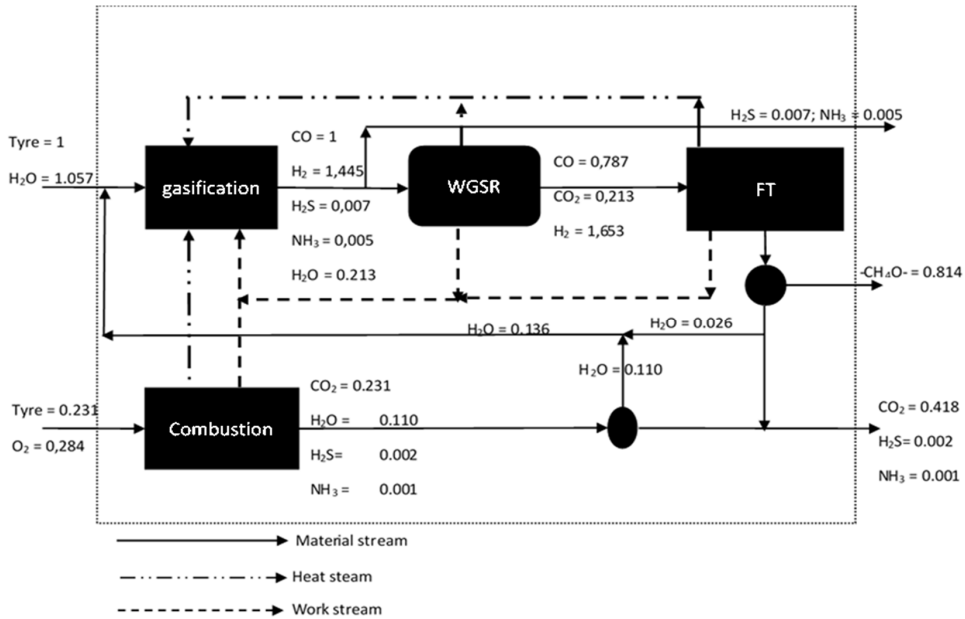
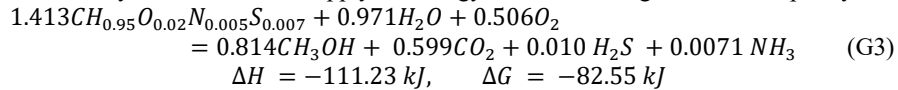


Figure 1: Fully integrated methanol process

Due to practical difficulties in implementing Figure 1, let us consider the approach of partial integrating the process, by supplying the energy needed for the gasification step only.

Figure 2 and equation G3 shows the overall material balance to produce methanol when additional tyres are burned to supply the energy needed for the gasification step only.



The performance of this process in Figure 2 and equation G3 is as follows:

$$\text{Carbon efficiency} = 57.6\%$$

$$\eta_{cp} = 70\%$$

The decision to not fully integrate the G3 process means more tyres need to be burned, since the energy from the water gas shift and methanol synthesis is no longer available. This leads to higher carbon dioxide emissions and thus the carbon efficiency of the process reduces to 57%. The system also produces excess work, if not recovered the system will lose it to the environment. The excess energy is recovered by applying heat engines at the different units. This leads to the production of 82.55 kW of shaft work which can be converted to electricity as shown in Figure 2.

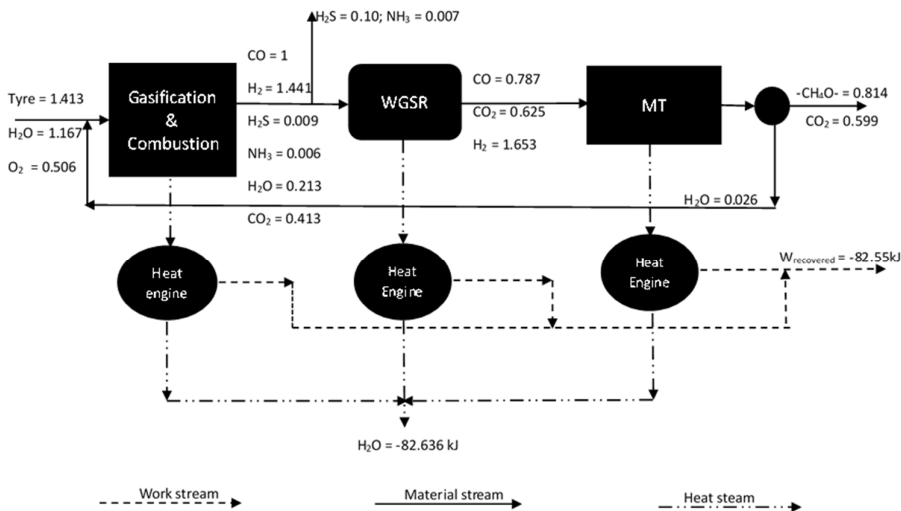


Figure 2: Partially integrated methanol process

3.2. Effect of temperature

The process is described on Figure 2 and equation 3 is assumed to be operating at the Carnot temperatures. The Carnot temperature can be viewed as the “ideal” temperature to operate a process, since neither heat nor work is being wasted. However, Carnot temperatures often prove to be impractical. Reaction kinetics are a common limitation to the use of Carnot temperature which necessitates a deviation from the ideal. For instance, gasification operates at temperatures above 1000°C to ensure high conversion of the carbon sources whereas the water gas shift and methanol synthesis reactor must operate at temperatures befitting the catalyst used. 1300°C is recommended for gasification of waste tyres (Mavukwana et al., 2020), 450°C for water gas shift and 220°C for the methanol synthesis. Operating at a temperature other than the Carnot temperature results in the loss of some of the potential work. To compensate for the lost work, more feed is burned. However, this increases the CO₂ emissions even more. The operating temperature for the water gas shift reactor is greater than the Carnot temperature, which means the unit will require further work input since operating at higher temperatures than Carnot temperature recovers more work than the system can provide. Using the notion of heat engines, the work produced by the process reduces from 82kW to 48 kW, which is equivalent to 2.42 GJ/ ton of waste tyre. An additional amount of 0.492 mol/s of tyres is required to make the process feasible. However, the carbon efficiency decreases to 54.6% and the chemical potential efficiency decreases to 63%. This, therefore, means that converting tyres to methanol is achievable, however, 45% of the carbon resource will end up as carbon dioxide. Furthermore, by operating the gasification unit at elevated pressures and then expanding the gaseous products, extra work will be recovered.

3.3. Revenue potential

The revenue potential of the process is obtained by a difference of the product selling price and of the cost of feed material. The analysis excludes the capital costs associated with each pathway. When electricity is sold at \$ 0.098/kWh and methanol sold at \$

480/ton (Global petrol prices, 2019), a waste tyre to methanol process has the potential to generate a revenue of \$ 620/ton of the tyre.

4. Conclusion

This work presented a method for process analysis at the “systems” level. This method allows a “big-picture” perspective. Such an elevated perspective further allows for the possibilities and limits to be investigated before significant time and resources are expended in more details-orientated methods. The analysis presented here focused on the performance limits of converting waste tyres to methanol and power. A notable result of this method shows that high process efficiency can only be achieved through a high degree of process integration and with the practice of co-production. Highly specialized processes that focus on producing only a few products, will always make a sacrifice in overall process efficiency. This study showed that 45% of the carbon feed is lost to carbon dioxide when tyres are converted to methanol and power. However, the process has a revenue of approximately \$ 620/ton of waste processed. Therefore, from an economic perspective, converting waste tyres to methanol is preferable to landfilling, as it reduces the volume of a hazardous waste stream by converting it to a useful transportation fuel and power. However, another pollutant, CO₂ is produced.

References

- Borgogna, A., Salladini, A., Spadacini, L., Pitrelli, A., Annesini, M. C., & Iaquaniello, G. (2019). Methanol production from Refuse Derived Fuel: Influence of feedstock composition on process yield through gasification analysis. *Journal of Cleaner Production*, 235, 1080–1089. <https://doi.org/10.1016/j.jclepro.2019.06.185>
- Dalena, F., Senatore, A., Marino, A., Gordano, A., Basile, M., & Basile, A. (2018). Methanol Production and Applications: An Overview. In *Methanol: Science and Engineering*. Elsevier B.V. <https://doi.org/10.1016/B978-0-444-63903-5.00001-7>
- Fox, J. A., & Stacey, N. T. (2019). Process targeting: An energy based comparison of waste plastic processing technologies. *Energy*, 170, 273–283. <https://doi.org/10.1016/j.energy.2018.12.160>
- Global petrol prices*. (2019). <https://www.globalpetrolprices.com>. https://www.globalpetrolprices.com/diesel_prices/
- Iaquaniello, G., Centi, G., Salladini, A., & Palo, E. (2018). Waste as a Source of Carbon for Methanol Production. In *Methanol: Science and Engineering*. Elsevier B.V. <https://doi.org/10.1016/B978-0-444-63903-5.00004-2>
- Iaquaniello, G., Centi, G., Salladini, A., Palo, E., Perathoner, S., & Spadaccini, L. (2017). Waste-to-methanol: Process and economics assessment. *Bioresource Technology*, 243, 611–619. <https://doi.org/10.1016/j.biortech.2017.06.172>
- Mavukwana, A., Fox, J. A., & Sempuga, B. C. (2020). Waste tyre to electricity : Thermodynamics analysis. *Journal of Environmental Chemical Engineering*, 8, 103831. <https://doi.org/10.1016/j.jece.2020.103831>
- Rivarolo, M., Bellotti, D., Magistri, L., & Massardo, A. F. (2016). Feasibility study of methanol production from different renewable sources and thermo-economic analysis. *International Journal of Hydrogen Energy*, 41(4), 2105–2116. <https://doi.org/10.1016/j.ijhydene.2015.12.128>
- Sempuga, B. C., & Yao, Y. (2017). CO₂ hydrogenation from a process synthesis perspective: Setting up process targets. *Journal of CO₂ Utilization*, 20, 34–42. <https://doi.org/10.1016/j.jcou.2017.05.004>
- South Africa. (2020). *The South African Economic Reconstruction And Recovery Plan*. 1(October), 1–32. https://www.gov.za/sites/default/files/gcis_document/202010/south-african-economic-reconstruction-and-recovery-plan.pdf

Analysing the robustness of multi-stage bioseparations to measurement errors

Robert Milton, Emma Chandler and Solomon F. Brown

University of Sheffield, UK
r.a.milton@sheffield.ac.uk

Abstract

Protein extraction is essential to the design and manufacture of bioproducts. A benign method of increasing academic and commercial interest is aqueous two-phase extraction (ATPE) (Iqbal et al., 2016), wherein a phase forming additive (in this case polyethylene glycol with phosphate buffer) separates an aqueous mixture into a target-rich top phase and a contaminant-rich bottom phase. This paper concerns a multistage separation, performed by repeating the ATPE iteratively to achieve a target protein yield or purity (Rosa et al., 2009). A McCabe-Thiele diagram is used to compute the number of stages required to achieve the target.

Unfortunately, the protein phase equilibrium curves used to construct the McCabe-Thiele diagram are subject to substantial variation depending on the precise composition and temperature of the mixture being separated. In order to assess the reliability of a multi-stage ATPE, we examine the robustness of the McCabe-Thiele diagram to variations in the phase equilibrium curves. It is seen that error propagation is weak, in that the same number of stages are required to achieve target under a variety of error scenarios. When this is not the case, a novel method of predicting the number of extra stages required is presented.

Keywords: protein extraction, multi-stage extraction, ATPE, model uncertainty, error analysis

1. Introduction

The extraction and purification of biomolecules is central to the manufacture of a vast range of bio-pharmaceuticals and speciality biochemicals (Doran, 2012). Aqueous Two-Phase Extraction (ATPE) is an efficient, benign and environmentally-friendly, low-cost and easily scalable bioseparation technique (Azevedo et al., 2009), suitable for batch or continuous processing. A solute is fractionated between two aqueous phases, e.g two polymer solutions or a polymer and a salt, accumulating preferentially in one of the phases. ATPE has been implemented with a diverse range of solutes (Iqbal et al., 2016), including proteins, enzymes (Kula et al., 1982), nucleic acids, antibodies (Rosa et al., 2013), viruses, metals and metal ions.

A multistage fractionation, repeating the same ATPE, is often needed to meet the exacting requirements for purity and yield demanded in bio-processing applications (Chandler et al., 2021; Rosa et al., 2009). A Solute is extracted from lean phase x to rich phase y ,

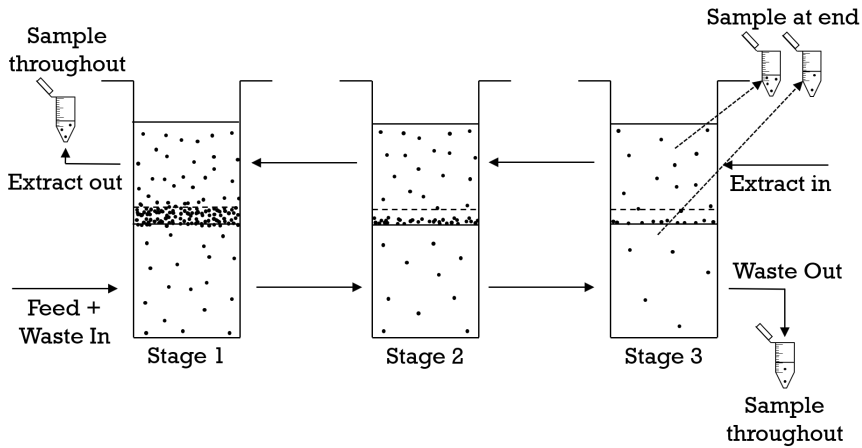


Figure 1: A schematic experimental design for a 3-stage ATPE.

iterating through stages according to

$$a_{y,n+1} - a_{x,n} = -A \leq 0 \quad (1)$$

where a denotes amount of solute (in mol or kg), of which A is lost during each stage. Starting from $a_0 := a_{x,0} > a_{y,0} := 0$, N stages are required to reach a desired goal

$$Xa_0 \geq a_{x,N} = a_{y,N+1} + A \quad (2)$$

Where the goal is removal of $X \times 100\%$ solute (contaminant), x is the extract phase. Where the goal is recovery of $(1 - X) \times 100\%$ of solute (product), x is the waste phase. The process is illustrated in Figure 1.

The iteration through stages is classically expressed in the McCabe-Thiele diagram Figure 2(a), stepping between the operating line $a_y = a_x - A$ and the phase equilibrium curve $a_y = a_y(a_x)$ lying above it. Experimentally, however, the total amount $a := a_x + a_y > 0$ is measured more reliably and accurately than either phase, as protein is known to precipitate into the interface between phases (Chandler et al., 2021; Andrews and Asenjo, 1996). Therefore we work with phase equilibrium curves written as

$$f(a) := a_y = a - a_x \quad (3)$$

and an operating line (which is usually not straight) depicting

$$f_{n+1} = a_{x,n} - A = a_n - f_n - A \quad (4)$$

as shown in Figure 2(b). The midpoint between the equilibrium curve and the operating line is the straight line $2f = a - A$. The goal Equation (2) becomes $f_{N+1} \leq Xa_0 - A$.

This paper studies the robustness of the iteration to measurement errors in the construction of the McCabe-Thiele diagram, essential to planning multistage ATPEs effectively. Such analyses have only been conducted for liquid-vapour fractionations (Medina et al., 1974),

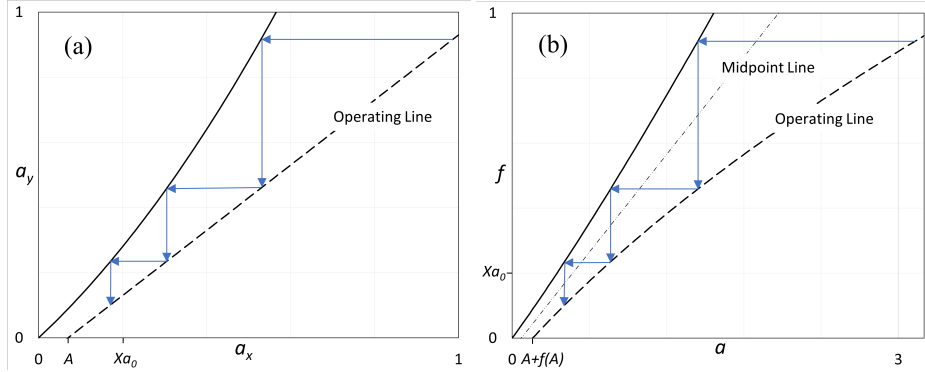


Figure 2: A 3-Stage McCabe-Thiele iteration depicted (a) classically and (b) in total amount terms.

exploiting the particular algebraic form of equilibrium curve in these systems. Section 2 of this article analyses the propagation of errors through McCabe-Thiele iteration. The robustness of the number of stages N required to achieve X with respect to errors in the phase equilibrium curve $f(a)$ is described in detail. This analysis is applied to the extraction of haemoglobin using 13% w/w PEG 1500 (polyethylene glycol, a polymer) and 11.2% w/w potassium phosphate (at pH 8.0) in Section 3. This is a particularly awkward extraction to perform experimentally, and so an ideal test and application of theory. Discussion is provided in Section 4.

2. Error Analysis

We are primarily concerned with errors in the equilibrium curve f , whose value differs from truth \hat{f} by $\delta f_n := f_n - \hat{f}_n$.

This error propagates through the amount a_n , shifting it from its true value \hat{a}_n by $\delta a_n := a_n - \hat{a}_n$

The total error of equilibrium evaluation linearises to

$$e_n := f_n - \hat{f}(\hat{a}_n) = \delta f_n + f'_n \delta a_n \quad (5)$$

where $e_1 = 0$. Introducing $g := (1/f' - 1) > 0$, substitution in iteration Equation (4) yields

$$e_{n+1} = g_n e_n - (g_n + 1) \delta f_n \quad (6)$$

An easy induction on Equation (6) provides the goal error

$$e_X := \frac{e_{N+1}}{a_0} = - \sum_{n=1}^N \frac{(g_n + 1) \delta f_n}{a_0} \prod_{k=n+1}^N g_k \quad (7)$$

This formula is easily adjusted to incorporate (less significant) errors in the initial con-

centration (δa_0) and the stagewise loss (δA_n)

$$e_X = \frac{\delta a_0 - \delta A_0}{(g_0 + 1)\hat{a}_0} \prod_{k=0}^N g_k - \frac{\hat{a}_{x,N}}{\hat{a}_0} \frac{\delta a_0}{\hat{a}_0} - \sum_{n=1}^N \frac{(g_n + 1)\delta f_n + \delta A_n}{\hat{a}_0} \prod_{k=n+1}^N g_k \quad (8)$$

Restating these results for classical McCabe-Thiele iteration is simply a matter of replacing f with y and

$$g := \left(\frac{1}{f'} - 1 \right) = \left(\frac{\partial a_y}{\partial a_x} \right)^{-1} \quad (9)$$

When error invalidates apparent succes at stage N , the goal is not truly achieved until

$$\hat{f}_{N+J+1} \leq Xa_0 - A < \hat{f}_{N+1} \quad (10)$$

Linearising an arbitrary approximant f about A

$$f_n := f_A + f'_A(a_n - A) \quad \forall n > N \quad (11)$$

such that

$$\sum_{j=1}^J g_A^{-j} \delta f_{N+j} \leq \sum_{j=1}^J g_A^{-j} \delta f_A \quad (12)$$

eventually recasts Equation (10) as

$$Xa_0 - A - \frac{(g_A + 1)}{(g_A - 1)} \hat{f}_A \geq \left(\hat{f}_{N+1} - \frac{(g_A + 1)}{(g_A - 1)} \hat{f}_A \right) g_A^J \quad (13)$$

The goal is therefore achievable if and only if

$$(g_A - 1)\hat{f}_{N+1} < (g_A + 1)\hat{f}_A \quad (14)$$

It would appear one has some freedom to choose f , provided Equations (12) and (14) are satisfied. The other quantities in this formula should be opaque: $\hat{f}_A := \hat{f}(A)$ can be measured carefully in advance, and \hat{f}_{N+1} is the result given by any failed extraction.

3. Results

This Section analyses errors in the extraction of haemoglobin using 13% w/w PEG 1500 and 11.2% w/w potassium phosphate (at pH 8.0), experiments described in Chandler et al. (2021). This system is one of the more difficult ATPes, because a significant amount of haemoglobin precipitates into a solid interface between the top and bottom phases, where it joins contaminant solids. Such interface effects are a common problem in protein ATPes (Andrews and Asenjo, 1996), but usually less dramatic. The interface is variable and hard to measure, but may be inferred from the top and bottom phases. A compromise must be struck between recovering target precipitate and rejecting clogging contaminant solids. The compromise in Chandler et al. (2021) took the extract phase y as 87% of the top phase, plus the interface whenever the total haemoglobin a exceeded 0.458mg in

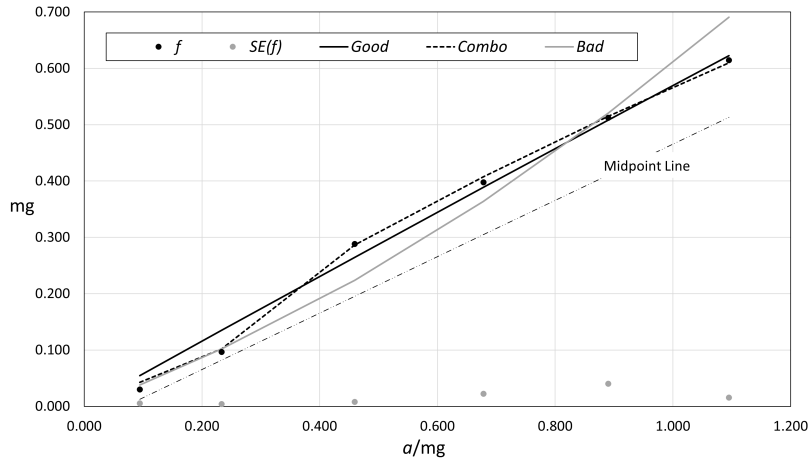


Figure 3: Phase equilibrium curves fitting various quadratics to experimental data f .

a 1.555ml system. Otherwise the interface was added to the waste phase x , along with the bottom phase and the remainder of the top phase. Figure 3 shows the results of OLS fitting of quadratic phase equilibrium curves

$$a^{-I} f(a) = B_2 a^{2-I} + B_1 a^{1-I} \quad (15)$$

The Good equilibrium curve fits the extract phase with $I = 0$. The Bad equilibrium curve fits the extract phase with $I = 1$, overemphasizing the goodness of fit at small a . The Combo equilibrium curve takes $I = 0$ to fit the top, bottom and interface curves separately, then allocates the interface phase post hoc by combining these three quadratics.

In analysing errors in the equilibrium curve, we must accommodate another difficulty with this system, which is the proximity of the equilibrium curves to the midpoint line (and therefore the operating line(s)) in Figure 3. For the purpose of analysis we use the Good equilibrium curve as f , and take in turn each other curve as truth \hat{f} . Included in

f	B_2	B_1	e_2	e_4	J	Predicted X
Good	-0.012	0.582				
Combo (I in W)	-0.122	0.465	0.029	0.012	0	
Combo (I in E)	-0.105	0.672				
Bad	0.225	0.384	-0.020	-0.132	3	0.04
Good-	0.155	0.450	-0.007	-0.079	1	0.14
Good-	0.071	0.516	-0.004	-0.038	0	
Good+	-0.095	0.647	0.005	0.033	0	
Good++	-0.179	0.713	0.010	0.057	-1	

Table 1: Results of McCabe-Thiele iteration, using the experimental values $A = 0.069\text{mg}$, $a_0 = 0.5\text{mg}$ and $X = 0.1$, corresponding to 90% recovery of haemoglobin. The first row is used as f , yielding $N + 1 = 4$, the remaining rows used successively as \hat{f} to assess errors. The final column gives the X achieved in Equation (13).

the latter are versions of the Good equilibrium curve in which the coefficient B_1 is shifted up (+) or down (-) by its regression standard error once or twice, setting B_2 to obey the expectation $\bar{f} = B_2\bar{a}^2 + B_1\bar{a}$. This is designed to invite failure – the slowest equilibrium curves to reach the goal will be those true ones whose gain in each step is squeezed by proximity to the midpoint (and therefore operating) line.

The results are collected in Table 1. The number of stages required is quite robust, despite the fact that errors accumulate quite dramatically between e_2 and e_4 , an effect clearly exacerbated as the equilibrium curve is lowered towards the midpoint line. Shifting B_1 up or down by two standard deviations only subtracts or adds 1 stage, lesser shifts having no effect.

Equation (13) is seen to be an excellent predictor of extra stages, although it narrowly fails to achieve the goal when the Good– curve represents truth. The failure is small (an extra stage will fix it), and Good– skirts very close to the operating line, so one expects errors to creep in.

4. Conclusions

In this short paper we have presented and tested an analysis of error propagation in the McCabe-Thiele method applied to multi-stage ATPE. It is hoped the analysis provided in Section 2 will be applied to many other ATPE processes. Equations (7) and (8) might provide useful error bounds, though these remain somewhat elusive. More valuable are Equations (13) and (14) giving the requirements for extra steps. These strongly encourage experiments to refine the phase equilibrium curve around the goal $f(Xa_0)$ and the stagewise loss $f(A)$ (which will often be close to $f(0) = 0$). The comparison with experimental results shows excellent agreement, given that we have chosen one of the most difficult ATPEs imaginable to perform this research on.

References

- B. Andrews, J. Asenjo, oct 1996. Protein partitioning equilibrium between the aqueous poly(ethylene glycol) and salt phases and the solid protein phase in poly(ethylene glycol)-salt two-phase systems. *Journal of Chromatography B: Biomedical Sciences and Applications* 685 (1), 15–20.
- A. M. Azevedo, P. A. Rosa, I. F. Ferreira, M. R. Aires-Barros, apr 2009. Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. *Trends in Biotechnology* 27 (4), 240–247.
- E. Chandler, J. Cordiner, S. Brown, feb 2021. Accounting for interface behaviour in multi-stage aqueous two-phase extraction. *Chemical Engineering Science* 230, 116172.
- P. M. Doran, 2012. *Bioprocess Engineering Principles*. Elsevier LTD, Oxford.
- M. Iqbal, Y. Tao, S. Xie, Y. Zhu, D. Chen, X. Wang, L. Huang, D. Peng, A. Sattar, M. A. B. Shabbir, H. I. Hussain, S. Ahmed, Z. Yuan, oct 2016. Aqueous two-phase system (ATPS): an overview and advances in its applications. *Biological Procedures Online* 18 (1).
- M.-R. Kula, K. H. Kroner, H. Hustedt, 1982. Purification of enzymes by liquid-liquid extraction. In: *Reaction Engineering*. Springer Berlin Heidelberg, pp. 73–118.
- A. Medina, C. McDermott, N. Ashton, dec 1974. On the effect of experimental error on the calculation of the number of stages for a given distillation separation. *Chemical Engineering Science* 29 (12), 2279–2281.
- P. Rosa, A. Azevedo, I. Ferreira, S. Sommerfeld, W. Bäcker, M. Aires-Barros, dec 2009. Downstream processing of antibodies: Single-stage versus multi-stage aqueous two-phase extraction. *Journal of Chromatography A* 1216 (50), 8741–8749.
- P. A. J. Rosa, A. M. Azevedo, S. Sommerfeld, M. Mutter, W. Bäcker, M. R. Aires-Barros, jan 2013. Continuous purification of antibodies from cell culture supernatant with aqueous two-phase systems: From concept to process. *Biotechnology Journal* 8 (3), 352–362.

Improvements in Methyl Ethyl Ketone Production Through Intensified Processes

Eduardo Sánchez-Ramírez^{a,*}, Heriberto Alcocer-García^a, Ana Gabriela Romero-García^a, Gabriel Contreras-Zarazua^a, Juan Gabriel Segovia-Hernandez^a

^a*Departamento de Ingeniería Química, Universidad de Guanajuato, Noria Alta s/n, Guanajuato, Gto., 36050, México.*

*Corresponding author. Email: eduardo.sanchez@ugto.mx

Abstract

Methyl ethyl ketone (MEK) is a widely used chemical in the industry. Most of the MEK uses are directly related to industrial application; however, recent studies suggest the use of MEK as a possible fuel for spark-ignition engines. The industrial process to produce MEK is performed through the hydration of butylene and the dehydration of that alcohol. Butylene is produced from oil cuts. However, petroleum is a recognized finite resource and is pointed out as a major cause of environmental issues. An alternative to improve a process is process intensification (PI). In this work, an intensified alternative to produce MEK from 2,3- Butanediol is proposed. 2,3-Butanediol might be produced by a fermentation process based on renewable biomass. The intensified alternatives consist of a reactive distillation column, an extractive distillation column, and three conventional distillation columns. The novel alternatives produce and purify MEK, 1,2-BD, 2MPL, and water, from 2,3-BD using calcium phosphate as catalyst were modelled in Aspen Plus. To evaluate the intensified alternative, it was considered the energy requirements per kilogram of MEK, conversion, selectivity as well as an environmental impact index, the greenhouse gas emissions. The RD column resulted in a conversion of 99.86 % and a selectivity for MEK of 44 %. After evaluation, the direct scheme resulted as the most promising one, with energy requirements of 2790 kcal/kg_{MEK} (11.6 MJ kg_{MEK}). Regarding the environmental impact, the direct scheme reported 7.07 tCO₂/h of greenhouse gas emissions.

Keywords: Methyl-Ethyl Ketone Production, Process Intensification, Biofuels Production, Hybrid Processes, Downstream Process.

1. Introduction

Political and governmental organizations have promoted agreements to mitigate the environmental effects generated using fuels derived from petroleum. An important sector to achieve the agreed objective is transportation, which in fact contributes with 23% of total CO₂ emissions. Even though electricity has been highlighted as a promising alternative for CO₂ reduction, biofuels are designated as responsible for reducing CO₂ emissions. Recently, methyl ethyl ketone (MEK) has been screened as a possible biofuel. MEK, also named 2-butanone, is a common industrial compound. Its use is mainly oriented as a solvent. Currently, MEK production is generated primarily as an oil-derived cut. However, Hoppe et al. (2016) carried out interesting studies that prove the feasibility of using MEK in internal combustion engines. Hoppe et al. (2016) concluded that the use of MEK generates greater heat of combustion, lower emissions

of greenhouse gases and better cold ignition. In a green process framework, MEK can be produced directly by fermentation, however, the process yields are quite low, $0.004 \text{ g}_{\text{MEK}}/\text{g}_{\text{glucose}}$ (Yoneda, Tantillo and Atsumi, 2014). On the other hand, an interesting process with the highest yield potential is using 2,3-Butanediol (2,3-BD) as an intermediary. 2,3-BD can also be produced by fermentation of lignocellulosic material, but unlike MEK fermentation, the yield in the production of 2,3-BD is high and the theoretical limit is quite close, $0.5 \text{ g}_{2,3\text{-BD}}/\text{g}_{\text{glucose}}$ (Syu, 2001). Once 2,3-BD is produced, it can be dehydrated and converted directly to MEK. As mentioned, the final part of the process is dehydration of 2,3-BD. This process has been addressed on multiple occasions. The use of zeolites, clays, calcium phosphate, etc has been reported; and most of the cases with selectivity between 90-95%. However, despite the large number of papers published in this regard, none of these papers addresses production kinetics in detail. In an interesting work reported by Song (2016), a set of detailed equations and reaction kinetics were reported for the dehydration of 2,3-BD to 1,3 Butadiene and Methyl Ethyl Ketone over a catalyst of Calcium Phosphate. In subsequent work, Song et al. (2017) presented an alternative to produce and purify MEK from 2,3-BD. In their proposal, they presented a production scheme based on a reactor followed by a series of separation columns and decanter, making a total of 10 separation units and a reactor (See Figure 1). Most of all separation units are distillation columns.

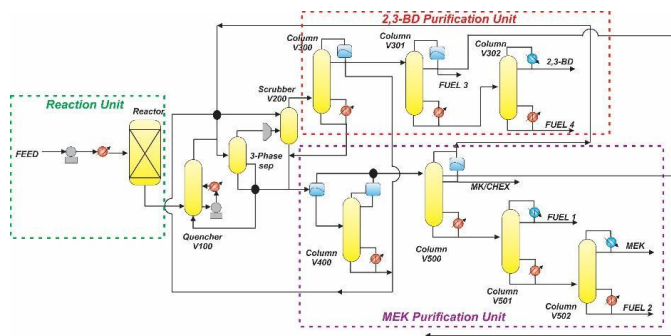


Figure 1. Conventional process to produce MEK from 2,3-BD

An alternative to improve a process is the intensification of the process (PI). PI is distinguished by five characteristics: reduced size of equipment, the increased performance of the process, reduced equipment inventory, diminishment in using utilities and raw materials, and increased efficiency of process equipment (Ponce-Ortega, 2012). Considering the background mentioned, the aim of this work is to propose intensified alternatives based on intensified schemes such as reactive distillation and extractive distillation to produce and purify MEK in order to improve energy requirements and the number of equipment, having as reference the current technology to produce MEK. For comparison purposes between the intensified alternatives, it will be used the energetic requirements per kilogram of MEK, the CO₂ global emissions and an environmental index (the eco-indicator 99). So, the target is to generate new and feasible technology to improve the current MEK production.

2. Intensified Alternative and Kinetics

Song et al. (2017) presented a proposal for MEK production. In the reaction stage, all reaction can be summarized as follow:

Table 1. Chemical reactions and kinetic parameters for MEK production

Reaction	Activation energy	Kinetic constant
1) $C_4H_{10}O_2(2,3 - BD) \xrightarrow{r^1} C_4H_8O(3B2OL) + H_2O$ $C_4H_{10}O_2(2,3 - BD) \xrightarrow{r^1} C_4H_8O(3B2OL) + H_2O$	2.33 E+05	7.45E-04
2) $C_4H_8O(3B2OL) \xrightarrow{r^2} C_4H_6(1,3 - BD) + H_2O$ $C_4H_8O(3B2OL) \xrightarrow{r^2} C_4H_6(1,3 - BD) + H_2O$	2.82E+05	4.41E-04
3) $C_4H_{10}O_2(2,3 - BD) \xrightarrow{r^3} C_4H_8O(MEK) + H_2O$ $C_4H_{10}O_2(2,3 - BD) \xrightarrow{r^3} C_4H_8O(MEK) + H_2O$	1.93E+05	6.64E-04
4) $C_4H_{10}O_2(2,3 - BD) \xrightarrow{r^4} C_4H_8O(2MPL) + H_2O$ $C_4H_{10}O_2(2,3 - BD) \xrightarrow{r^4} C_4H_8O(2MPL) + H_2O$	1.66E+05	1.27E-04

Being 2,3-BD, 3B₂OL, 1,3-BD, MEK and 2MPL, 2,3-Butanediol, 3-Buten-2-ol, 1,3-Butadiene, Methyl Ethyl Ketone, and 2-Methylpropanal respectively. In the reference case of Figure 1, Song et al. (2017) considered a reasonable number of decanters which result in many waste streams and consequently only 1,3-BD and MEK are partially recovered. The reason for such design is the thermodynamic complexity of the stream to be purified. In the mixture of five components, there are 4 azeotropes: 3 heterogeneous between 1,3-BD and H₂O, MEK-H₂O, and 2MPL-H₂O respectively, and 1 azeotrope homogeneous between 3B₂OL-and H₂O. As shown in Figure 2, the alternative presented in this work is designed considering a reactive distillation column and an extractive distillation column, using glycerol as extractant. The intensified alternative accomplish a recovery constraint of at least 98% wt for all components, and a purity constraint of 99.5% wt for MEK, 99% wt for 3-Buten-2-ol, 99% wt for 2MPL, 99% for 1,3-butadiene and 99.99% for glycerol, the purity of MEK is set in such number since 99.5% wt is the lowest purity for being considered as fuel (Penner et al., 2017). As highlighted by Huang et al. (2005) there is no general and structured methodology for the optimal design of reactive distillation columns. Therefore, in this work, the heuristics proposed by Subawalla and Fair (1999) were considered. This heuristic is summarized in an algorithm that can be used to estimate parameters such as column pressure, location of the reactive zone, theoretical stages of the column, reflux ratio, column diameter, etc. These variables were varied with the objective of increasing the yield and production of MEK. The size of the holdup was not established arbitrarily, to find an appropriate value the guide proposed by Barbosa and Doherty (1988) was followed. The methodology initially consisted of specifying the compositions of the products and the value of the reflux ratio (obtained with the strategy of Subawalla and Fair; in this way, the volume of the plates could be found so that the chemical reaction was carried out. The implementation of the algorithm proposed by Subawalla and Fair

(1999) is simple but does not guarantee the optimal design. Esta no es la única metodología, disponible, algunas ideas basadas en la integración de building block. These building blocks can be associated with different process phenomena (Demirel et al., 2017) Based on the results of the short methods, they were rigorously simulated in compliance with the MESH equations using the Radfrac module of Aspen Plus. In order to warranty the minimal energy consumption, several sensitivity analysis were performed to identify the correct design parameters.

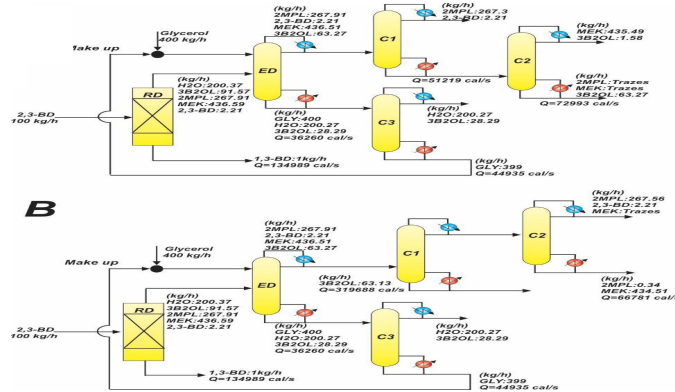


Figure 2. Intensified process to produce MEK from 2,3-BD

3. Performance Evaluation Indices

A major drawback in the operation of distillation columns is the energy requirements. The first performance index is the energy requirements. The main idea is to realize the amount of energy invested in the process which is further represented in units like MJ/Kg_{MEK}. The second performance index is the eco-indicator 99 (EI99). The eco-indicator 99 is calculated as follow (Goedkoop, M., & Spriensma, 2000):

$$EI99 = \sum_b \sum_d \sum_{k \in K} \delta_d \omega_d \beta_b \alpha_{b,k} \quad (5)$$

Where β_b represents the chemical b released per unit of reference flow due to direct emissions, $\alpha_{b,k}$ is the damage caused in category k per unit of chemical b released to the environment, ω_d is a weighting factor for damage in category d , and δ_d is the normalization factor for damage of category d . Finally, the third performance index is greenhouse gas emissions, which are indeed related to the energy requirements. The CO₂ emission is calculated as follows (Gadalla et al., 2005):

$$[CO_2]emiss = \left(\frac{Q_{fuel}}{NHV} \right) \left(\frac{C\%}{100} \right) \alpha \quad (6)$$

Where $\alpha = 3.67$ is the ratio of molar masses of CO₂ and C, while NHV (kJ/kg) represents the net heating value of fuel with a carbon content of C%. For CO₂ emission calculation, we assume that the energy involved in both processes, reaction-separation and intensified process comes from burning CH₄ gas. The target is to generate as little as it is possible CO₂ emissions.

3. Results

Once the intensified was designed and through a sensitivity analysis a minimal energy requirement obtained, the conversion and selectivity for the MEK production were

calculated, obtaining a conversion of 99.86% and a selectivity of 44%. The energy consumption of the intensified scheme was 1218797 kcal/h (5073 MJ/h). Thus, considering the total production of MEK, energy consumption would be 2786 kcal/kg_{MEK} (11.6 MJ/kg_{MEK}). Equally weighting all compounds, MEK, 1,3-BD, and 2MPL; The energy consumption for the entire product would be 1218 kcal/kg_{PROD} (5.07 MJ/kg_{PROD}). Moreover, in the hypothetical scenario where all MEK produced as fuel was burned, with 31.45 MJ/kg as energy density, the energy profit would be 8601 MJ. The design parameters for the intensified alternative are shown below.

Table 2. Design parameter for the intensified alternative

	RD	ED	C1	C2	C3
Number of stages	80	50	80	46	30
Reactive Stages	10-80	---	---	---	---
Solvent (kg h ⁻¹)	---	400	---	---	---
Reflux ratio	2.95	0.5	5.488	4.712	0.039
Feed stage	76	5, 46	23	25	14
Operative pressure (kPa)	101.353	101.353	101.353	101.353	101.353
Distillate flowrate (kg h ⁻¹)	999	770	270	437.6	230
Reboiler duty (kcal h ⁻¹)	485962	130535	184390	256145	161765
CO2 Emissions (ton h ⁻¹)			7.07		
Eco-Ind (points y ⁻¹)			1.825E+06		

The energy involved in the upstream process in the scheme reported by Song et al. (Figure 1) (Song, Yoon, and Lee, 2017) is 11140000 kcal/h (46640.96 MJ/h) to produce 4385 kg/h of MEK. Thus, the energy consumption per kilogram of MEK is 2540.5 kcal/kg_{MEK} (10.64 MJ/kg_{MEK}). So far, the scheme presented by Song et al. (2017) seems interesting, however, when it is considered the production of the other purified by-product (3340 kg of 2,3-BD) the picture changes completely. If the production of MEK and 2,3-BD is considered together, the energy requirements change to 1142.1 kcal/kg_{PROD} (6.04 MJ/kg_{PROD}). In other words, when jointly considering the production of MEK and all the by-products generated in the process, the intensified scheme proposed in this article overwhelmingly exceeds that presented by Song et al. (2017) by an average amount of 76kcal/kg_{PROD} (0.97 MJ/kg_{PROD}). It should be noted that, due to the lack of data published by Song et al. (2017) in this analysis, the energy consumption of the reactor is not being considered. However, it can be inferred that the energy consumption of the scheme presented by Song et al. (2017) would increase since as discussed Song et al. (2017), the reactor operates at a temperature of 360 °C. On the contrary, in the energy analysis, the energy consumption of the reactive distillation column was considered. In this way, we can assume a favorable scenario for the intensified scheme in terms of energy consumption. Probably one of the reasons why the intensified process energetically exceeds the conventional scheme presented by Song et al. (2017), is the number of decanters present in the whole process. Note, for example, that in the scheme presented in Figure 1, several streams called “fuel” are presented, which can be inferred as a MEK stream but with a lower purity to consider MEK as fuel for ignition purposes. The above is assumed due to the presence of a stream named specifically MEK. Additionally, note that there is at least one waste stream in which MEK is involved (MEK/CHEX stream) which shows that the recovery

of MEK may not be high. Unfortunately, in the work of Song et al. (2017), a complete mass balance is not presented to verify these assumptions.

4. Conclusions

In this chapter, the intensified technology proposal to produce MEK was analyzed, having 1,2-BD, 2MPL, and water as by-products. 2,3-BD was used as a raw material in a reactive distillation column, having previously studied calcium phosphate as a catalyst. Once the reactive distillation column was implemented, it was possible to obtain conversion of 99.86% and a MEK selectivity of 44%. At the outlet of the reactive distillation column, an effluent with the compounds was obtained and purified using an extractive distillation column and three conventional distillation columns. After evaluation, the intensified alternative resulted with energy requirements of 2790 kcal/kg_{MEK} (11.6 MJ/kg_{MEK}). Regarding environmental impact the greenhouse gas emissions, that scheme A reported 7.07 tonCO₂/h. Even though, there is not a direct and fair comparison; previous work has proposed a conceptual design using a reactor and a set of 10 separation units accomplishing to purify MEK and 1,3-BD. In this work, the intensified scheme can purify all products obtained in the reactive distillation, with less equipment and higher selectivity for MEK. So, we consider most promissory this intensified alternative in comparison with the conventional alternative previously presented.

5. References

- Barbosa, D., & Doherty, M. F., 1988. Design and minimum-reflux calculations for single-feed multicomponent reactive distillation columns. *Chemical Engineering Science*, 43(7), 1523-1537.
- Demirel, S. E., Li, J., & Hasan, M. F., 2017. Systematic process intensification using building blocks. *Computers & Chemical Engineering*, 105, 2-38.
- Figueirêdo, M. F. De et al., 2011, 'Optimal design of extractive distillation columns — A systematic procedure using a process simulator', *Chemical Engineering Research and Design*, 89, pp. 341–346.
- Huang, K., Iwakabe, K., Nakaiwa, M., & Tsutsumi, A., 2005. Towards further internal heat integration in design of reactive distillation columns—part I: the design principle. *Chemical engineering science*, 60(17), 4901-4914.
- Gadalla, M. A. et al., 2005, 'Reducing CO2 emissions and energy consumption of heat-integrated distillation systems', *Environmental Science and Technology*, 39(17), pp. 6860–6870.
- Goedkoop, M., & Spriensma, R., 2000, 'Eco-indicator 99 Manual for Designers', PRE' Consultants, Amersfoort, The Netherlands.
- Hoppe, F. et al., 2016, 'Tailor-Made Fuels from Biomass: Potentials of 2-butanone and 2-methylfuran in direct injection spark ignition engines', *FUEL*. Elsevier Ltd, 167, pp. 106–117.
- Penner, D. et al., 2017, 'Conceptual Design of Methyl Ethyl Ketone Production via 2,3-Butanediol for Fuels and Chemicals', *Industrial and Engineering Chemistry Research*, 56(14), pp. 3947–3957.
- Ponce-Ortega, J. M., Al-Thubaiti, M. M. and El-Halwagi, M. M., 2012, 'Process intensification: New understanding and systematic approach', *Chemical Engineering and Processing: Process Intensification*. Elsevier B.V., 53, pp. 63–75.
- Song, D., 2016, 'Kinetic Model Development for Dehydration of 2, 3- Butanediol to 1, 3-Butadiene and Methyl Ethyl Ketone over an Amorphous Calcium Phosphate Catalyst', *Industrial & Engineering Chemistry Research*, 55(45), pp. 11664–11671.
- Song, D., Yoon, Y. and Lee, C., 2017, 'Chemical Engineering Research and Design Conceptual design for the recovery of 1, 3-Butadiene and methyl ethyl ketone via a 2,

3-Butanediol-dehydration process', *Chemical Engineering Research and Design*, 123, pp. 268–276.

Subawalla, H., & Fair, J. R., 1999. Design guidelines for solid-catalyzed reactive distillation systems. *Industrial & engineering chemistry research*, 38(10), 3696-3709.

Syu, M. J. ,2001, 'Biological production of 2,3-butanediol', *Applied Microbiology and Biotechnology*, 55(1), pp. 10–18. doi: 10.1007/s002530000486.

Yoneda, H., Tantillo, D. J. and Atsumi, S.,2014, 'Biological production of 2-butanone in *Escherichia coli*', *ChemSusChem*, 7(1), pp. 92–95.

Flexibility Assessment of Spirits Distillation Processes: Focus on the Armagnac Distillation

Alessandro Di Pretoro^a, Michel Meyer^a, Marie-Claude Ségur^b, Xavier Joulia^{a*}

^a*Laboratoire de Génie Chimique, Université de Toulouse, CNRS/INP, Toulouse, France*

^b*Bureau National Interprofessionnel de l'Armagnac, 11 place de la Liberté, Eauze, France*

xavier.joulia@ensiacet.fr

Abstract

In recent years, process intensification and design under uncertainty have been topics of major concern in the Process Systems Engineering (PSE) domain. Flexible unit design resulted of particular interest when applied to bio-processes since biomass shows an intrinsically unstable nature with respect to the year's seasons and their location. Spirits distillation can be seen as a millenary bioprocess. It is based on well-established techniques that were, at first, mainly empirical and relied on the expertise of the distillers. During the last decades, available technologies allowed a deeper knowledge of the process impact on the final product properties by means of a detailed modelling of the aroma compounds and process units performances. Therefore, the flexibility analysis focused on an Armagnac continuous distillation process whose detailed modelling and simulation were carried out in collaboration with the French National Interprofessional Office of the Armagnac (BNIA) by means of ProSimPlus[®] simulator. The Armagnac continuous distillation, performed this way since the 14th century, is a highly energy integrated system since the feed stream is used as a coolant for the column condenser and wood or gas combustion as the reboiler duty. The heat integration has been included in the analysis and showed to substantially constrain the process flexibility. In particular, the preheater was detected as the most critical unit from and possible design alternatives have been proposed. In conclusion, the presented flexibility analysis allows a more detailed understanding of the Armagnac distillation process and a better analysis of the perturbations impact on the product properties. Moreover, it sets the basis for future studies on advantages and drawbacks of process intensifying alternatives from a flexibility perspective.

Keywords: flexibility, energy integration, distillation, Armagnac, ProSimPlus[®].

1. Introduction

In recent years, bioprocesses have seen a renewed interest in the Process Systems Engineering due to the sustainability goals imposed by the EU. However, in the food and beverage industry, agricultural feedstocks exploitation is at the basis of a relevant number of centuries old processes and related well-established procedures. In this list of bioproducts, spirits can be certainly classified among the eldest ones. Their production processes have been widely studied from the fermentation up to the ageing phase and the resulting methodologies give rise to thousands of different product varieties according to the specific way they are performed. As the majority of processes based on agricultural feedstocks, they are affected by uncertain raw material properties along the year's seasons

and according to their geographic location. The impact of these uncertainties on the thermodynamic feasibility constraints in water-alcoholic mixtures have been indeed discussed in detail by Di Pretoro et al. (2020a).

This research work focuses on an Armagnac distillation case study. Armagnac is one of the main spirits produced in France, in particular in the south-western region. In order to meet the French AOC label requirements, it should be produced by continuous distillation in a specific tray column, during a specific time of the year and with fixed product properties in terms of alcoholic composition (JORF, 2015). All these product standards are thus coherent with previous remarks about bio-processes characteristics. Moreover, the “alambic armagnacais” is a highly process-intensified unit since the feed stream is preheated by the top product and, at the same time, it serves as a condenser coolant as later discussed in more detail. On the other hand, wood, sometimes replaced by gas fuel, is used as heat duty in the reboiler. The constraining impacts of process intensification from a flexibility point of view was introduced in recent studies by Di Pretoro et al. (2021) and the same procedure based on the flexibility indicators is used in this work.

For all these reasons, the presented case study is totally worth a flexibility assessment both to highlight the impact of uncertainties on bioprocesses with very strict product standards and to discuss the effects process intensification from a flexibility perspective.

2. Flexibility assessment

In the PSE domain, the flexibility assessment tool is based on well-established procedures aimed at quantifying the ability of a system to accommodate a set of uncertain parameters. This analysis is carried out by means of the several flexibility indices proposed in literature during the last decades.

These indicators can be mainly classified in two categories, namely deterministic and stochastic, and the former will be explored in this research work. In particular, the class of deterministic indices estimates the maximum magnitude of the disturbance load, of one or more uncertain parameters at the same time, that can be withstood by the system without becoming infeasible regardless of its probability.

The best established and most commonly used deterministic flexibility index is the one proposed by Swaney and Grossmann (1985), hereafter called F_{SG} .

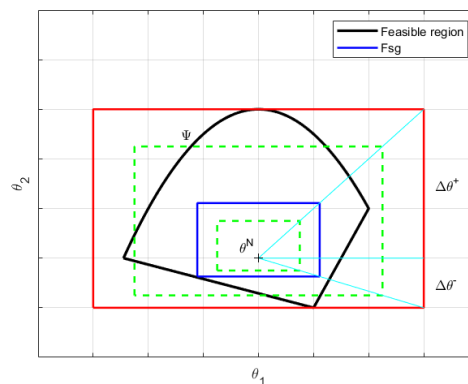


Figure 1 – Swaney & Grossmann deterministic flexibility index

It is defined as the solution of the following flexibility index optimization problem:

$$F_{SG} = \max \delta \quad (1)$$

$$s. t. \max_{\theta \in T(\delta)} \min_z \max_{j \in J} f_j(d, z, \theta) \leq 0 \quad (2)$$

where θ refers to the uncertain variables, d to the design parameters, z to the control variables. Finally, δ is the scaling factor of the hyperrectangle:

$$T(\delta) = \{\theta: \theta^N - \delta\Delta\theta^- < \theta < \theta^N + \delta\Delta\theta^+\} \quad (3)$$

represented in Figure 1.

These indices were then used to perform the flexibility analysis of the Armagnac distillation case study that is presented in detail in the following section.

3. The Armagnac distillation case study

As previously discussed, this research work will focus on the Armagnac spirit. It is produced in the south-western region of France and, by law, the distillation should take place between the end of the grapes harvest and the 31st march of the following year.

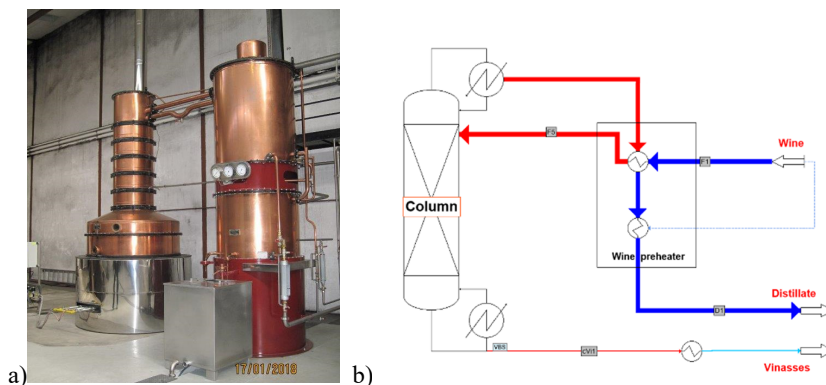


Figure 2 – a) Armagnac distillation layout and b) ProSimPlus[®] process simulation diagram

The “alambic armagnacais”, shown in Figure 2a, is an atmospheric tray distillation column. The wash, that in this case is wine, is produced by fermentation of white grapes and its alcohol (ethanol) by volume concentration (hereafter ABV) at 20°C should range between 7.5% v/v and 12.0% v/v. Before ageing, the distillate ABV is fixed between 52.0% and 72.4 % v/v according to the last regulation (JORF, 2015). The commercial product is obtained by reducing the aged distillates to a minimal ABV of 40% v/v.

A 15 trays (condenser and reboiler included) column was simulated (cf Figure 2b) considering an ideal behaviour for the vapor phase and using the NRTL activity coefficient model to describe the liquid phase non-ideality (Puentes et al., 2018a). Since the purpose of this study is to define the feasibility boundaries of the operation, mainly water and ethanol were accounted for in the distillation unit since volatile aroma do not affect mass and energy balances due to their low concentrations (Sacher et al., 2013). The employed binary interaction parameters are those obtained by Puentes et al. (2018b).

Process parameters	Name	Value	Unit
	Wine inlet temperature	13.1	° C
Process specifications	Alcohol loss (bottom)	1	%
	Distillate temperature	18	°C
	Distillate ABV	≥ 52	%
Deviation	Wine ABV	0.5	%
	Wine flowrate	0.05	m ³ /h

Table 1 – Process parameters and specifications

For this flexibility assessment an innovative approach was used. The wine ABV was selected as the first uncertain variable since it is the parameter that mainly vary according to the available wine. Then, the feed flowrate was included as a second uncertain variable. When operating the particular “alambic armagnacais” column, the feed flowrate is the second manipulated variable together with the reboiler heat duty. Therefore, this flexibility analysis is carried out from a unit design perspective, with an external perturbation and a manipulated variable as uncertain parameters. The column layout is fixed as well as the alcohol losses in the vinasses and the distillate temperature after the wine preheater. The distillate ABV cannot be fixed but the AOC constraints should be respected. The reboiler heat duty and the preheater surface area sizing are calculated for the variable operating conditions, i.e. ABV and wine flowrate. Finally, process parameters and specifications are resumed in Table 1.

4. Results

The “alambic armagnacais” flexibility assessment was then performed according to the well-established procedure proposed by Di Pretoro et al. (2019) for distillation units. In particular, the feed ABV uncertain interval ranged from 7,5 - 12 % according to the actual legislation, while its flowrate uncertain domain varied from 0.6 to 1 m³/h that is the capacity range of the studied alambic unit. The required heat transfer surface area as well as the distillate ABV value trends will be discussed over the entire uncertain domain and coupled with the flexibility index. It is finally worth remarking that the following are intentionally very conservative results since they are obtained for a distillate temperature at the outlet of the preheater equal to 18° C that is the desired value and not the upper boundary (that is about 25 °C).

4.1. Sensitivity analysis

Figure 3a shows the trend of the distillate ABV variable. For the given specifications, it achieves higher value if the feed ethanol content is lower. This phenomenon is due to the fact that, since the alcohol losses are fixed at 1 %, in case of higher ethanol fraction in the feed a more severe vaporization is required and, thus, a higher amount of water is withdrawn with the distillate that results in a more diluted mixture. For the same reason, the ABV value increases along with the feed flowrate due to the less effective preheating and the lower amount of water in the top vapor in absence of a rectifying section.

On the contrary, the required heat transfer surface area (Figure 3b) increases for higher ABV in the feed due to the more severe separation conditions related to the product streams adjustments in order to keep low the ethanol losses in the bottom.

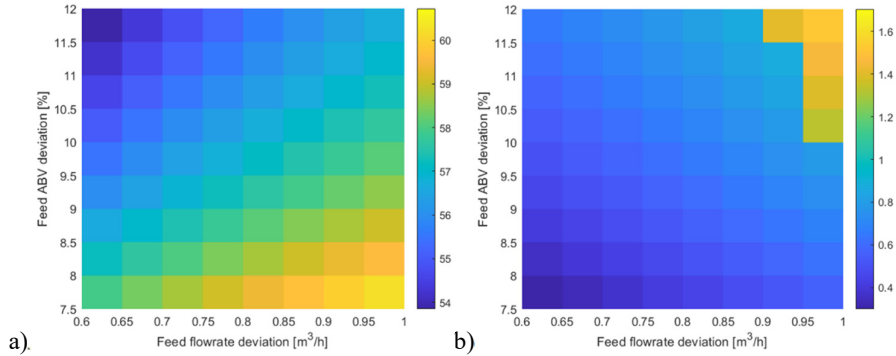


Figure 3 – a) Distillate ABV % and b) preheater heat transfer surface area (log₁₀-scale)

Moreover, considerable inlet flowrates have also an impact on the preheater size due to the higher thermal capacity of the process stream. These two phenomena, i.e. poor ethanol losses allowed combined with a more heat to be exchanged, result in an exponential growth of the preheater heat transfer surface area.

4.2. Flexibility results

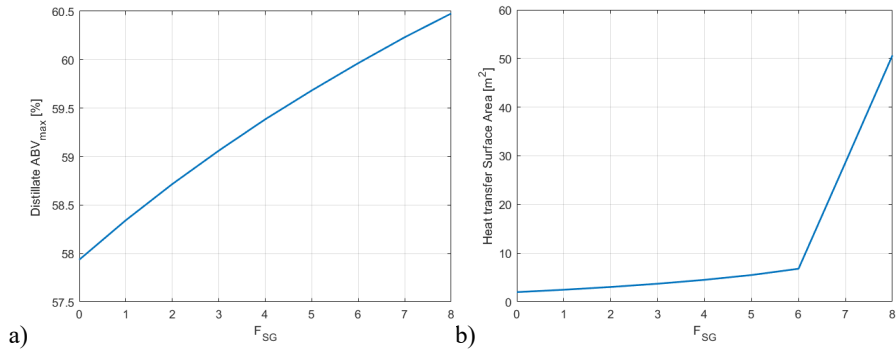


Figure 4 – a) Distillate ABV and b) preheater heat transfer surface area vs flexibility

The outcome obtained by coupling the flexibility index with the previous results is shown in Figure 4. The plots should be read as follows: F_{SG} equal to δ corresponds to δ times the $\Delta\theta$ deviation both for the value of the feed ABV (0.5 %) and for its flowrate (0.05 m³/h) and both positive and negative. Thanks to the domain convexity it was possible to simplify the nonlinear optimization solution algorithm by using the vertex analysis.

On the one hand, the maximum ABV that can be obtained increases with the deviation magnitude. All the values over the uncertain interval fall in the required ABV range despite the very constraining specifications selected. On the other hand, the required heat transfer surface slightly increases for a flexibility index between 0 and 6, while an exponential growth can be detected for higher deviation magnitude. Besides the impact of the higher capacity, this trend is the expected consequence of process intensification as already highlighted by Di Pretoro et al. (2021).

Therefore, from the flexibility assessment results, a few important aspects can be deduced. First of all, the ABV plot shows that the operation feasibility is not compromised in the operation range. Then, the flexible design for the preheater not only

allows quantify the unit oversizing but also highlights the process limitations from a flexibility perspective. Since, in the alambic distillation common practice, the feed flowrate is used as manipulated variable to manage the top ABV specification, the flexibility analysis outcome suggests not to approach flowrates higher than 0.9 m³/h. Indeed, poor heat integration would be performed under those operating conditions.

5. Conclusions

The alambic simulation and flexibility analysis were successfully performed over the selected uncertain domain. The obtained results allowed to identify the criticalities of the operation and the most constraining design parameters of the distillation unit that represent the bottleneck from a flexibility point of view.

In particular, the desired distillate temperature can be achieved with a reasonable preheater sizing for moderate deviation magnitude while it becomes more and more difficult to attain if higher deviations are expected. Even for this unit then, the consequences of the process intensification result to be evident and critical for the operation. Therefore, in these cases, suitable design solutions can be applied. One of them could be the addition of a coil fed by water on the utility side in order to enhance the distillate cooling down to the desired temperature. Although already adopted in the common practice, further studies and a detailed flexibility assessment could allow to analyse this additional process integration solution and outline in detail its limitations.

References

- Di Pretoro, A., Montastruc, L., Manenti, F., Joulia, X., 2019. Flexibility analysis of a distillation column: Indexes comparison and economic assessment. *Computers & Chemical Engineering* 124, 93–108.
- Di Pretoro, A., Montastruc, L., Manenti, F., Joulia, X., 2020a. Exploiting Residue Curve Maps to Assess Thermodynamic Feasibility Boundaries under Uncertain Operating Conditions. *Ind. Eng. Chem. Res.* 59, 16004–16016.
- Di Pretoro, A., Montastruc, L., Joulia, X., Manenti, F., 2021. Robust Optimization and Control for Sustainable Processes, in Ren, J. (ed.) *Methods in Sustainability Science*, 24. Elsevier.
- JORF, Cahier des charges de l'appellation d'origine contrôlée Armagnac homologué par le décret n°2014–1642 du 26décembre 2014, *Journal Officiel de la République Française*, 2015; Paris.
- Puentes, C., Joulia, X., Vidal, J.-P., Esteban-Decloux, M., 2018a. Simulation of spirits distillation for a better understanding of volatile aroma compounds behavior: Application to Armagnac production. *Food and Bioproducts Processing* 112, 31–62.
- Puentes, C., Joulia, X., Athes, V., Esteban-Decloux, M., 2018b. Review and thermodynamic modeling with NRTL model of vapor-liquid equilibria (VLE) of aroma compounds highly diluted in ethanol - water mixtures at 101.3 kPa, *Industrial & Engineering Chemistry Research*. 57, 10, 3443-3470.
- Sacher, J., García-Llobodanin, L., López, F., Segura, H., Pérez-Correa, J.R., 2013. Dynamic modeling and simulation of an alambic pear wine distillation. *Food and Bioprod. Process.* 91, 447–456.
- Swaney, R.E., Grossmann, I.E., 1985. An index for operational flexibility in chemical process design. Part I: Formulation and theory. *AIChE Journal* 31, 621–630.

OUTDOOR – An open-source superstructure construction and optimization tool

Philipp Kenkel^{a*}, Timo Wassermann^a, Celina Rose^a, Edwin Zondervan^b

^a*Advanced Energy Systems Institute, Enrique-Schmidt Straße 7, 28359 Bremen, Germany*

^b*Department of Chemical Engineering, Twente University, 7522NB Enschede, Netherlands*

kenkel@uni-bremen.de

Abstract

To enhance the availability of superstructure optimization as a tool in process design for industries as well as science this work presents a fully *Open sUperstrucTure moDeling and OptimizatiOn fRamework* (OUTDOOR). This framework is written in Python using object-oriented programming in combination with algebraic modeling utilizing the PYOMO modeling language. In addition, an Excel-based data preparation tool, called Excel-Wrapper is presented. It provides an intuitive way to prepare process data and generate superstructures, which are ready-made for solving using open-source as well as commercial optimization solvers.

Keywords: Superstructure Optimization, Python, Open Source

1. Introduction

Superstructure optimization is a powerful tool to perform preliminary process design. Superstructure models map all possible flowsheets to produce a product from a specified set of raw materials. Such models are defined as mixed integer (non)-linear mathematical programming (MI(N)LP) models and can be solved for different objective functions, e.g. minimal production costs using open-source and commercial optimization solvers. Superstructure models have been applied to different areas, from biomass-based economy up to power-to-x processes (Galanopoulos et al., 2019; Kenkel et al., 2020). However, they can be formulated in different ways leading to unnecessary complexity and ambiguity. The common practice is that research groups develop their own models and tools, often using commercial software, e.g. GAMS. Only a few software solutions for generic superstructure optimization have been presented, none of them being fully open source (Mencarelli et al., 2020). In addition, the application of superstructure models in industry is limited by its complicated algebraic formulation and data provision. Generally, straight forward graphical solutions are preferred over complex mathematical models and extensive programming codes.

This work proposes an approach to solve both problems, presenting an *Open sUperstrucTure moDeling and OptimizatiOn fRamework* (OUTDOOR). This framework is setup in Python using a combination of modular object-oriented programming and

algebraic model formulation in PYOMO. Its focus is to construct deterministic MILP models to solve superstructure design problems.

2. Programming and modeling fundamentals

The presented framework is a tool for intuitive construction of superstructure models. It utilizes different concepts such as superstructure optimization modeling, object-oriented programming as well as the open-source optimization modeling language (PYOMO). The combined effort leading to OUTDOOR is briefly discussed in the following sections.

2.1. Superstructure optimization and algebraic modeling

Superstructure optimization is a methodology which is used for process design. It is based on mathematical models derived from mass- and energy balances in combination with cost calculation of capital expenditures and operational costs to optimize process flow sheets for different objective functions (Quaglia et al., 2015)]. These models are formulated as depicted in Eq. (1). Here \mathbf{c} and \mathbf{b} are vectors of known constant parameters, while \mathbf{x} is a vector or variables (continuous as well as integer), and A is matrix of known parameters.

$$\begin{aligned} & \text{maximize } \mathbf{c}^T \mathbf{x} \\ & \text{s. t. } A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0 \end{aligned} \tag{1}$$

Recurring equations are often formulated using indexed variables and parameters. For example, the electricity demand E_u^{EL} of a unit operation u is dependent of the total inlet flow F_u^{IN} and a specific factor p_u (kWh/t_{input}). If this holds for each unit operation u in a set of U , the equations can be written as shown in Eq. (2).

$$E_u^{EL} = p_u \cdot F_u^{IN} \quad \text{with } u \in U \tag{2}$$

The fixed parameters of p_u are then supplied to the model as an indexed list of a set U inside of a data file.

2.2. Object-oriented programming

Object-oriented programming, in contrast to procedural programming, uses classes and objects as containers to allocate attributes and methods. These objects can be used to store data related to unit operations (such as the presented specific electricity demand p_u) directly in the process unit object. This way process-data can be built up around the process classes and objects, presenting a more intuitive way to handle and store data compared to indexed lists.

2.3. PYOMO Modeling language

PYOMO is a Python-based, open-source optimization language with a diverse set of capabilities to solve algebraic optimization models, while providing access to optimization solvers such as CPLEX or IPOPT. (Hart et al., 2017) The user benefits from PYOMO implemented direct access to solver configurations and solver output regarding calculation procedure, infeasibilities or other occurrences.

3. OUTDOOR

OUTDOOR combines the concepts that were discussed in the earlier sections. It provides an intuitive workflow for superstructure model creation and solving. Figure 1 depicts the main concepts and workflow of the application of OUTDOOR, and will be discussed in the following sections.

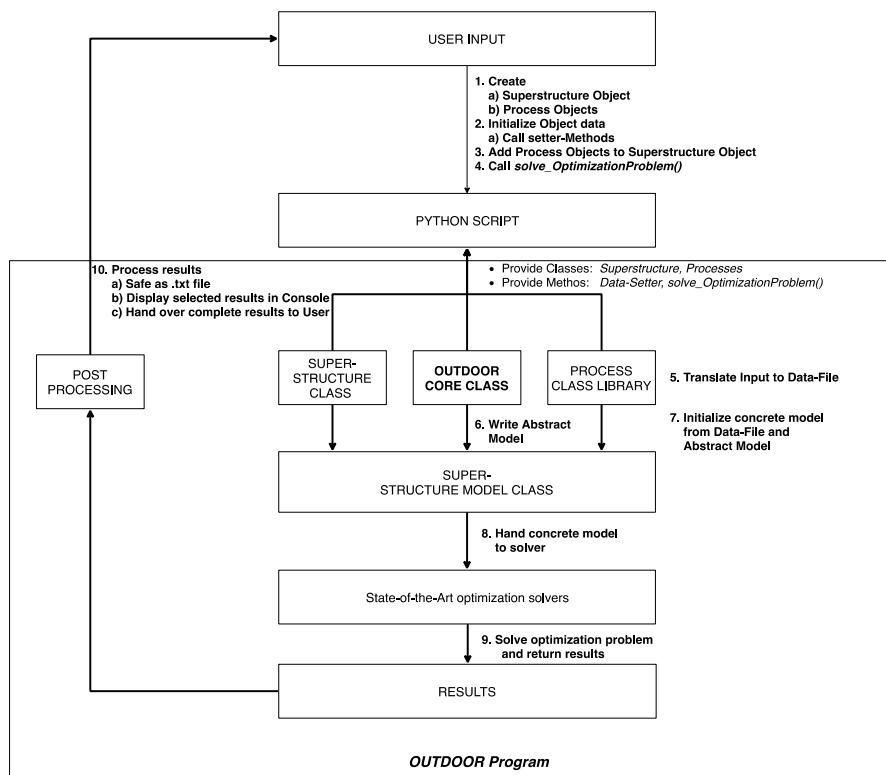


Figure 1: Visualization of the OUTDOOR workflow.

3.1. User input

As a first step user input is required. This input of data can be divided into two categories. The first category is the superordinate system data. An object of the provided *Superstructure-Class* is created and initialized using Python scripting. This object demands important attributes such as the costs for electricity, the occurring chemical compounds or reactions or the desired product and objective. The second category is the input of considered unit operations. Different objects of the *Process-Classes* such as stoichiometric reactors or stream splitters are created and initialized with their respective data such as the required electricity demand or stoichiometric coefficients. The initialization of data of the different objects (step 2) is performed by calling predefined setter-methods, which are provided by the Classes.

Next, (step 3) the process objects are added to the superstructure creating a complete superstructure system with all unit operations and external parameters such as electricity

prices. An input of initial guesses is not required at this state, however it could be necessary in the future to investigate more complex problems. When all inputs are provided, the user calls the `solve_OptimizationProblem()` method from the OUTDOOR module (step 4) to start the OUTDOOR program and solve the optimization problem.

3.2. OUTDOOR program

OUTDOOR processes the inputs provided by the user in a series of steps. First (step 5) the data in the process and superstructure objects is translated to a Python dictionary. Generally, two different types can be generated. The first one comprises easy convertible parameters and the second one comprises complex convertible parameters.

The specific electricity demand p_u (kWh/ t_{input}) is an illustration of a parameter of the first type. It stores only a numeric value which has to be converted into the right format while being assigned to the right unit operation.

Other relationships however, demand more complex translation methods. One example is the allocation of the referred reference flow of the electricity demand calculation. This flow can be the total inlet flow as depicted in Eq. (2). It could also be the outlet flow, or even just specific components of one of these flows. Hence, a method which translates correlations like “outlet flow” or “inlet flow” into algebraic understandable parameters is required. The procedure of these methods will be explained exemplary for this electricity demand reference flow in the following. The program needs two different information. The delivery of these information is still conducted during the input-phase of the user. One of the already mentioned setter-methods of the process class asks if “FIN” or “FOUT” has to be considered. Another setter-method demands the input of a Python list of components or the string “all components”.

During the calculation, the program calls a translation method which transforms the input of the user into process specific $\kappa_{EL,u,i}^1$ and $\kappa_{EL,u}^2$ parameters. The first one is a binary parameter being 1 if the component i is to be considered for electricity in unit operation u . The second parameter determines if the reference flow is the inlet (Value = 1) or the outlet flow (Value = 0). These parameters are integrated into the algebraic model using $\kappa_{EL,u,i}^1$ as equation parameter and $\kappa_{EL,u}^2$ as a conditional factor in construction of the constraint (cf. Figure 2).

```
def ElectricityBalance_1_rule(self,u):
    if self.kappa_2_ut[u,'Electricity'] == 1:
        return self.REF_FLOW_EL[u] == sum(self.FLOW_IN[u,i] \
            * self.kappa_1_ut[u,'Electricity',i] for i in self.I)
    elif self.kappa_2_ut[u,'Electricity'] == 0:
        return self.REF_FLOW_EL[u] == sum(self.FLOW_OUT[u,i] \
            * self.kappa_1_ut[u,'Electricity',i] for i in self.I)
    else:
        return self.REF_FLOW_EL[u] == 0
```

Figure 2: Python code example for calculation of references flow in electricity consumption

Subsequent to the translation of input into the right format, the program creates the generic equations of the model (step 6). This model is written as an object of the PYOMO *AbstractModel Class*, which means it contains empty parameters which can be filled at a later stage using a data file.

Step 7 of the program uses the created Python dictionary (data file) and the abstract model to create a concrete model, depicting the algebraic formulation of the defined superstructure. This model is then handed over to the selected optimization solver using the capabilities of the PYOMO modeling language (Step 8) and optimized for the defined objective function (Step 9)

3.3. User output

After the solver found a solution to the optimization problem, the complete raw results are returned to the user. It is possible to check the solutions for the variables and constraints utilizing the Python console. However, to simplify the results access, the OUTDOOR program also processes them (step 10). This step includes the presentation of selected results in the Python console as well as the saving of the main results as a .txt file into a chosen folder. The main results include the net production costs and CO₂ emissions, the cost breakdown and the energy consumption details including heat integration details.

3.4. Extensions and development

OUTDOOR depends on the user input data in form of a Python script using the definition of objects as well as calling setter-methods and the *solve_OptimizationProblem()* - method. This is however, not really user friendly for non-programmers. Therefore, two extensions are under development. The first module is an *Excel-Wrapper*. This tool enables user to insert data via a predefined Microsoft Excel template, Afterwards the user has to start the *Excel-Wrapper* tool, which itself creates the objects, calls the solve-function and saves the results. The second module presents a graphical user interface in which processes can be created by drag and drop, and data can be provided by popup tables. Other features of interest which are under development include more detailed and customized error handling and output, as well as capabilities for multi-criteria decision making and robust decision making taking stochastic programming into account.

4. An example: Power-to-methanol process

The goal of this case study is to investigate renewable methanol production with three different CO₂ sources as well as five hydrogen supply technologies and two offgas utilization methods. Two objective functions were considered in the superstructure optimization: i) Minimal net production costs and ii) Minimal net production CO₂ emissions. Detailed outcomes of the study are presented in Kenkel et al. (2020).

To construct and solve the superstructure model, OUTDOOR is applied in combination with the Excel-Wrapper. All relevant data was entered into the predefined Excel sheet provided by OUTDOOR. Data on considered unit operations was implemented in single Excel sheets. Figure 3 depicts an excerpt of this user input in form of the specific energy demand, reference flow and component for one unit operation. Afterwards, OUTDOOR automatically translated the input into a readable data file, formulated the general model including mass- and energy balances as well as cost and emission functions. The generated model is automatically handed to the Gurobi solver and optimization results in terms of chosen technologies, net present costs and emissions, as well as costs and energy break-down are presented in the Python console as well as in a .txt. file. The cpu time of the data collection using the excel-wrapper is 0.016 seconds. The construction of the superstructure objects takes 2.3 seconds, resulting in a MILP with 29411 variables (5746 binary) and 43581 constraints. The cpu time for solving this problem and returning the results using Gurobi is 11.1 seconds.

Energy Related Factors			
Index	Electricity	Cooling	Heating
Specific Energy demand (MWh/t)	54,3	-20,97	
Reference Flow	FOUT	FOUT	
Considered Component	H2	H2	
Considered Component			
Considered Component			
Considered Component			

Figure 3: Excerpt of Excel-based parameter input for OUTDOOR

5. Conclusions

The main programming concepts and workflow of the *Open sUperstrucTure moDeLing and OptimizatiOn fRamework* (OUTDOOR) were presented in this work. This framework enables the initialization of data as unit operation related attributes using object-oriented programming, while still writing the actual model as algebraic model utilizing the PYOMO optimization language. To connect both of these concepts, different methods in terms of so-called translation functions were developed. To further strengthen the applicability of the tool an Excel-Wrapper was developed which uses data stored in predefined Excel-Templates as well as the Class-specific methods to create the superstructure, solve it and display and save the results. First results were derived using a beta-version of the framework. These results were presented in Kenkel et al. (2020).

Funding

Funding of this research by the German Federal Ministry of Economic Affairs and Energy within the KEROSyN100 project (funding code 03EIV051A) is gratefully acknowledged.

References

- Galanopoulos, C., Kenkel, P., Zondervan, E., 2019. Superstructure optimization of an integrated algae biorefinery 130. <https://doi.org/10.1016/j.compchemeng.2019.106530>
- Hart, W.E., Laird, C.D., Watson, J.-P., Woodruff, D.L., Hackedbeil, G.A., Nicholson, B.L., Siirola, J.D., 2017. Pyomo — Optimization Modeling in Python, Springer Optimization and Its Applications. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-58821-6>
- Kenkel, P., Wassermann, T., Zondervan, E., 2020. Design of a Sustainable Power-to-methanol Process: a Superstructure Approach Integrated with Heat Exchanger Network Optimization. *Comput. Aided Chem. Eng.* 48, 1411–1416.
- Mencarelli, L., Chen, Q., Pagot, A., Chemical, I.G.-C.&, 2020, U., 2020. A review on superstructure optimization approaches in process system engineering. Elsevier.
- Quaglia, A., Gargalo, C.L., Chairakwongsa, S., Sin, G., Gani, R., 2015. Systematic network synthesis and design : Problem formulation , superstructure generation , data management and solution. *Comput. Chem. Eng.* 72, 68–86. <https://doi.org/10.1016/j.compchemeng.2014.03.007>

Probabilistic Modeling for Optimization of Bioreactors using Reinforcement Learning with Active Inference

Ernesto C. Martínez^{a,b,*}, Jong Woo Kim^b, Tilman Barz^{b,c}, Mariano N. Cruz Bournazou^{b,d}

^a *INGAR (CONICET-UTN), Avellaneda 3657, S3002GJC Santa Fe, Argentina*

^b *KIWI-biolab, Bioprocess Engineering, TU Berlin, Ackerstrasse 76, 13355 Berlin, Germany*

^c *AIT Austrian Institute of Technology GmbH, Giefingasse 2, 1210 Vienna, Austria*

^d *DataHow AG, Vladimir-Prelog-Weg 1, 8093 Zurich, Switzerland*

* ecmarti@santafe-conicet.gov.ar

Abstract

The open-ended complexity of abiotic conditions in bioreactors means that it is generally infeasible to model its dynamic behaviour comprehensively. Learning optimization-oriented probabilistic models encoding a parsimonious representation is far more efficient for bioprocess development and optimization. In this work, active inference is integrated with reinforcement learning to demonstrate that useful probabilistic models for bioreactor optimization can be learned by balancing optimization-oriented and information-seeking objectives. The baker's yeast bioprocess is used as a case study. For online Bayesian update of model parameter distributions, simulation results demonstrate that highly informative data can be sampled by minimizing the variational free energy of the expected future. The resulting probabilistic model is thus biased towards bioreactor optimization.

Keywords: active inference, optimization, probabilistic models, reinforcement learning.

1. Introduction

Even though is not yet accepted nor recognized, abstract (e. g., macroscopic or cybernetic) models used for bioreactor optimization are too shallow to account for the complexity of switching in metabolic pathways when responding to changes in the abiotic conditions (Richelle et al., 2014). A challenge in modeling for optimization is how such models can be learned from designed experiments given (i) the rich complexity of profiling operating conditions, and (ii) the circular dependence of model learning and information content of sampled data, which may lead to suboptimal performance. As most bioreactor models are not a veridical representation over a wide region of operating conditions, it is infeasible to achieve parametric precision comprehensively, let alone design optimally informative experiments for this objective (Martínez et al., 2013; Luna and Martínez, 2017). Probabilistic models are more efficient to deal with the process-model structure mismatch and hidden states present (Daunizeau et al., 2014). In this work, we illustrate how ideas from active inference (Tschantz et al., 2020) can be integrated with reinforcement learning (Sutton and Barto, 2018) to unify optimization-oriented and information-seeking

objectives in modeling for optimization using the variational free energy of the expected future (Parr and Friston, 2019) over the shrinking horizon of a dynamic experiment.

2. Probabilistic modeling and active inference

A probabilistic (causal) model of a bioreactor is defined by a joint probability distribution over the following set of stochastic variables:

- x ; y : the $n \times n_t$ hidden states time-series; the $p \times n_t$ observations (sampled data),
- u : the $n_u \times n_t$ manipulated (controlled) inputs time-series,
- θ ; φ : the $n_\theta \times 1$ evolution parameters; the $n_\varphi \times 1$ observation parameters,
- α : the state noise precision (structural errors),
- σ : the measurement noise precision (analytical and sensor calibration errors).

From sample to sample in a dynamic experiment, these variables follow the equations:

$$\begin{aligned} x_t &= f(x_{t-1}, \theta, u_{t-1}) + \eta_t; \quad \eta_t = N(0, \alpha^{-1}\mathbf{I}), & \text{(hidden) state evolution} \\ y_t &= g(x_t, \varphi) + \varepsilon_t; \quad \varepsilon_t = N(0, \sigma^{-1}\mathbf{I}), & \text{observation (samples)} \end{aligned} \quad (1)$$

where f (resp. g) is the first-principles model (observation model), and η_t (observation ε_t) is the state (resp. measurement) modeling errors (noise). A probabilistic model m of a bioreactor is completed by specifying the (initial) Gaussian prior distributions for its parameters θ, φ . Also, Gamma distribution priors are defined for the precision hyperparameters α, σ . Given these priors, the left part of Eq. (1) induces a (so-called semi-Markovian process) prior density on the trajectory of hidden states x . Similarly, the right part of Eq. (1) yields a likelihood function which measures how plausible an observation y is when the bioreactor content is sampled at time t . Uncertainties from noisy observations and model imperfections are thus taken explicitly into account.

In the variational Bayesian framework, model identification (or inversion) entails the estimation of the marginal likelihood or evidence of a bioreactor model, that is a probabilistic description of the main (causal) metabolic mechanisms by which sampled data are generated. Probabilistic Bayesian treatment of an experiment data makes full usage of prior assumptions regarding the statistical distributions for initial conditions, evolution/observation parameters and state/measurement noise (Daunizeau et al., 2014). Inverting a probabilistic model m requires approximating the conditional density $p(\xi|y, m)$ of the unknown hidden states and parameters $\xi = \{x, x_0, \theta, \varphi, \alpha, \sigma\}$ given a data set of sampled measurements y and computing the model evidence $p(y|m)$. Nonlinearities in the probabilistic model prevent exact analytical solutions to the model inversion problem which can approximately be solved using Bayesian variational approaches such as active inference. Active inference is a normative theory that unifies observation, external stimulus and model learning under a single imperative—the minimization of the variational free energy (Parr and Friston, 2019). More specifically, probabilistic model learning is posed as the maximization of a free-energy lower bound $F(q)$ for the model evidence with respect to an approximate density $q(\xi)$:

$$F(q) = \langle \ln p(\xi|m) + p(\xi|y, m) - p(\xi) \rangle_q = \ln p(y|m) - \mathcal{D}_{KL}(p(\xi); p(\xi|y, m)) \quad (2)$$

where \mathcal{D}_{KL} is the Kullback-Leibler divergence between two distributions and the expectation $\langle \circ \rangle_q$ is taken under the approximate posterior distribution q . As can be deduced from Eq. (2), maximizing the functional $F(q)$ with respect to q drives the Kullback-Leibler divergence between $p(\xi)$ and the exact posterior $p(\xi|y, m)$ to zero. The reader is referred to the work of Daunizeau et al. (2014) for methodological details and the therein described VBA Toolbox for variational Bayesian analysis.

Active inference proposes that the modeler’s goal or intent are encoded in the probabilistic model as a prior preference for favourable observations (e. g., higher biomass productivity or protein expression). Thus, active inference demands that parameter distributions in the probabilistic model are based on data sampled from favourable operating conditions. This can be achieved by means of online experimental redesign aiming at minimizing the free energy of the expected future (Parr and Friston, 2019), which corresponds to the trajectory of hidden states that is expected to occur from applying the optimal sequence of decisions (sampling times and manipulated inputs) that maximise the model evidence in the desired region of operating conditions.

Let $z_{t:T}$ denote a sequence of variables through time, $z_{t:T} = \{z_t, \dots, z_T\}$, and let define a policy as a sequence of actions $\pi = \{u_t, \dots, u_{T-1}\}$. In “modeling for optimization,” the specific aim is to minimize the free energy of the expected future \tilde{F}_π , which is defined as:

$$\tilde{F}_\pi = \mathcal{D}_{KL}(q(y_{t:T}, x_{t:T}, \Phi|\pi) \| p^*(y_{t:T}, x_{t:T}, \Phi)); \Phi = (\theta, \varphi, \alpha, \sigma) \quad (3)$$

where $q(y_{t:T}, x_{t:T}, \Phi|\pi)$ models the probability distribution for future trajectories in a dynamic experiment under a given policy π , whereas $p^*(y_{t:T}, x_{t:T}, \Phi)$ defines the joint probability distribution for the optimal trajectory of the hidden states, model parameters and preferred observations. Thus, when \tilde{F}_π is driven to zero, the policy π becomes the (probabilistic) optimal policy. Notice that by minimizing \tilde{F}_π , the surprise $-\ln p(y_{t:T}|m)$ is also minimized, which implies that the Bayesian model evidence is also maximized.

3. Adaptive optimization-oriented experimental design

In this section, we describe an efficient implementation of the proposed objective function for online experimental redesign in the context of reinforcement learning (Sutton and Barto, 2018). To select actions for purposely biasing the model, we iteratively optimise a policy π at each time step t using different samples from the distributions of model parameters. A pseudocode for the proposed algorithm for adaptive experimental design is shown in Fig. 1. The internal loop has a forward pass where the density $q(y_{t:T}, x_{t:T}, \Phi|\pi)$ is increasingly converted into a posterior density upon simulated data using a stagewise greedy redesign procedure based on Thompson sampling (Russo et al., 2018) of the posterior distribution $q(\Phi)$ for model parameters. At each stage, an action is chosen by solving an optimization problem for the expected look-ahead reward based on the predicted impact on the biased model evidence $r_t(\hat{y}_{t+1}) = \ln p(r(\hat{y}_{t+1})|\Phi, u_t)$ of simulated data \hat{y}_{t+1} for alternative redesign decisions in a bounded compact set Ω :

$$u_t^* = \operatorname{argmax}_{u \in \Omega} r_t(\hat{y}_{t+1}) \quad (4)$$

where r is a reward (or desirability) function for observations and their underlying hidden states. The predicted observation \hat{y}_{t+1} and the action u_t^* are then used to generate the posterior distribution for model parameters $\hat{q}(\Phi_k)$ to be used at $t+1$. Thompson sampling is again applied and using Eq. (4) the action u_{t+1}^* is calculated and the corresponding simulated observation \hat{y}_{t+2} is computed. This forward pass is finished when u_{T-1}^k is

computed and the k th policy $\pi_k = \{u_t^k, \dots, u_{T-1}^k\}$ is completely defined. For each iteration of the outer loop, a different policy π_k is computed. The generated sequences are then ranked based on their corresponding cumulative rewards \mathcal{R}_k over the planning horizon $[t, \dots, T]$. From the top ranked policy π_t^* , only the first action u_t^* is used for redesigning the dynamic experiment in the bioreactor and, at the next sampling time $t+1$ the observation y_{t+1} is obtained. Using experimental data (u_t^*, y_{t+1}) , the joint posterior distribution $q(\Phi)$ for the parameters of the probabilistic model is updated using variational Bayesian inference (based on the Laplace approximation) and the master (external) loop begins re-estimating the optimal policy π_{t+1}^* from $t+1$ until the end of the experiment at time T .

```

Inputs:  $T, K, x_0$ , prior  $q(\Phi)$ , state evolution and observation functions  $f, g$ 
> For  $t = 1$  to  $T - 1$ 
  Infer current state  $\hat{x}_t$  using  $u_{t-1}^*$ ,  $\hat{x}_{t-1}$  and Thompson Sampling of prior  $q(\Phi)$ 
  > For  $k = 1$  to  $K$ 
     $\hat{q}(\Phi_k) = q(\Phi)$ 
    While  $t < T$  (Forward Pass)
      Thompson Sampling of the prior  $\hat{q}(\Phi_k)$ :  $\phi_k = \text{TS}[\hat{q}(\Phi_k)]$ 
       $u_t^k = \text{argmax}_{u \in \Omega} r_{t+1}(\hat{y}_{t+1} | \phi_k, \hat{x}_t)$ 
      Simulate redesign using  $u_t^k$  and predict  $r_{t+1}(\widehat{y}_{t+1}^k | u_t^k, \hat{x}_t)$ 
      Update prior:  $\hat{q}(\Phi_k) \leftarrow \hat{q}(\Phi_k | u_t^k, y_{t+1}^k)$  using  $(u_t^k, y_{t+1}^k)$ 
      Accumulate reward:  $\mathcal{R}_k = \mathcal{R}_k + r_{t+1}(\widehat{y}_{t+1}^k | u_t^k)$ 
    End while
    Define the policy:  $\pi_t^k = \{u_t^k, \dots, u_{T-1}^k\}$  with its corresponding  $\mathcal{R}_k$ 
  > End for
  Rank policies  $\pi_t^k, k = 1, \dots, K$ , using  $\mathcal{R}_k$ 
  Select the best policy  $\pi_t^* = \{u_t^*, \dots, u_{T-1}^*\}$  with the highest  $\mathcal{R}_k$ 
  Redesign the experiment using only  $u_t^*$  and measure  $y_{t+1}$  at  $t + 1$ 
  Update prior:  $q(\Phi) \leftarrow q(\Phi | u_t^*, y_{t+1})$  using experimental data  $(u_t^*, y_{t+1})$ 
> End for
Outputs:  $\pi^* = \{u_1^*, \dots, u_{T-1}^*\}$ ,  $\mathbf{y} = \{y_1, \dots, y_T\}$ ,  $\mathbf{r} = (r_1, \dots, r_T)$ ,  $q(\Phi)$ 

```

Figure 1. Adaptive optimization-oriented experimental design

4. Case study: Baker's yeast

The macroscopic model proposed by Richelle et al. (2014) for fed-batch baker's yeast production process in which the nitrogen and glucose consumptions are coordinated is used to test the adaptive experimental procedure in Fig. 1. The model includes a reaction

in which the nitrogen and α -ketoglutarate are consumed to produce biomass. Also, the inhibition effect on glucose consumption by the accumulation of α -ketoglutarate is accounted for. The model has 15 parameters that define the reaction rates in biomass and ethanol production and consumption of glucose, nitrogen, and α -ketoglutarate. The concentration of α -ketoglutarate is a hidden state which is inferred using the model. The inputs which are redesigned in each experiment are the feeding rate profiles for glucose and nitrogen. The duration of each modelling experiment is fixed to 20 hours, bioreactor initial conditions are known, and its content is sampled every hour to measure the concentration of biomass, ethanol, glucose, and nitrogen. Sample processing time is assumed equal to 30 min, leaving a maximum of 30 min to compute the redesign decision to be applied at the next time step. The parameter K in the algorithm (Fig. 1) is set to 10.

The aim is to purposefully bias model identification towards operating conditions that maximize the total amount of biomass that can be obtained at the end of the culture. Thus, the reward function r_k is defined to achieve a steady increase in the biomass concentration for consecutive samples. Results obtained are summarized in Fig. 2 and Fig. 3. Notice that due to initial uncertainty (prior distributions), for the first experiment exploration is significantly high. Also note that final biomass concentrations for modeling run #1 and #2 in Table 1 are quite high despite they are not “one-shot” optimized feeding profiles using the probabilistic model with the updated parameter distributions. After two modeling experiments, as shown in Table 2, the parameter distributions have been updated to make the probabilistic model biased towards the most profitable region of operating conditions (Richelle, et al., 2014) where biomass production is higher.

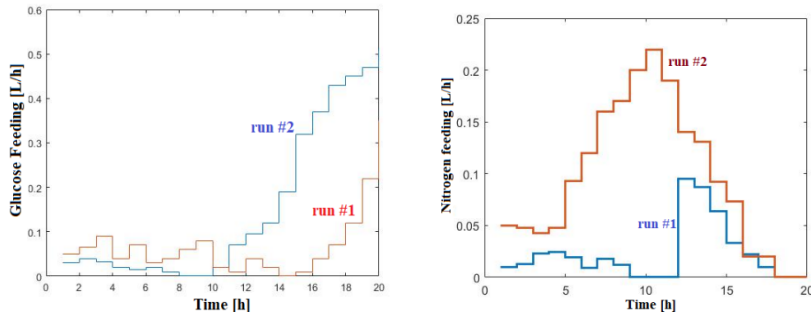


Figure 2. Substrate feeding profiles in run #1 and #2

Table 1. Biomass final concentration

Modeling Run #	Biomass [g/L]
1	24.11
2	32.64
Richelle et al. 2014	32.00 (Exp. measurement)

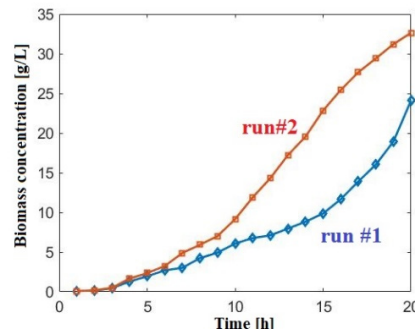


Figure 3. Biomass production in run #1 and #2.

Table 2. Prior and posterior Normal distributions $N(\mu, \sigma^2)$ of the model parameters

Param.	Units	Prior		Post.		Run 2	
		μ	σ	μ	σ	μ	σ
k_1	gX/gG	0.5431	0.250	0.6232	0.112	0.7012	0.010
k_2	gX/gG	0.0612	0.020	0.0578	0.011	0.0851	0.006
k_3	gX/gE	0.8929	0.085	0.8500	0.005	0.8091	0.002
k_4	gE/gG	0.2647	0.080	0.2450	0.009	0.2350	0.002
k_5	gA/gX/gG	0.2589	0.080	0.0189	0.005	0.2162	0.037
k_6	gX/gN	1.0150	0.150	0.9733	0.088	0.8817	0.044
μ_{Omax}	gG/gX/h	0.4445	0.125	0.4210	0.039	0.4198	0.028
μ_{Gmax}	gG/gX/h	2.5364	0.200	2.6472	0.145	2.7116	0.009
μ_{Nmax}	gN/gX/h	1.1903	0.150	1.2279	0.067	1.2009	0.011
K_G	gG/L	0.1524	0.030	0.1208	0.025	0.0989	0.010
K_I	gE/L	3.1817	0.050	3.2011	0.443	2.9442	0.031
K_N	gN/L	2.9370	0.050	2.9674	0.632	3.3598	0.278
K_A	gA/gX/L	9.0014	2.000	9.4569	1.227	10.106	0.583
K_{IA}	gA/gX/L	5.5981	0.500	5.8919	0.389	6.2991	0.267
K_{IA2}	gA/gX/L	5.5737	0.210	6.131	0.201	5.8404	0.113

5. Concluding remarks

A novel probabilistic method for modeling the dynamic behaviour of bioreactors in the most profitable region of operating conditions is proposed. Based on simulation data, a dynamic experiment is redesigned online through active inference. Reinforcement learning is used to maximize the Bayesian model evidence, that is, minimize surprise.

References

- J. Daunizeau, V. Adam, L. Rigoux, 2014, VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data, *PLoS Comput Biol* 10, 1, e1003441.
- M. F. Luna, E. C. Martínez, 2017, Iterative modeling and optimization of biomass production using experimental feedback, *Comput. Chem. Eng.*, 104, 151–163.
- E. C. Martínez, M. D. Cristaldi, R. J. Grau, 2013. Dynamic optimization of bioreactors using probabilistic tendency models and Bayesian active learning. *Comput. Chem. Eng.*, 49, 37–49.
- T. Parr, J. K. Friston, 2019, Generalised free energy and active inference, *Biological Cybernetics*, 113, 495–513.
- D. J. Russo, B. Van Roy, A. Kazerouni, Ia. Osband, Z. Wen, 2018, A Tutorial on Thompson Sampling, *Foundations and Trends in Machine Learning*, 11, 1, 1-96.
- A. Richelle, P. Bogaerts, 2014, Off-line optimization of baker's yeast production process, *Chem. Eng. Sci.*, 119, 40–52.
- R. S. Sutton, A. G. Barto, 2018, Reinforcement learning: An introduction, Second Edition, MIT Press, Cambridge, MA.
- A. Tschantz, A. K. Seth, C. L. Buckley, 2020, Learning action-oriented models through active inference. *PLoS Comput. Biol.*, 16, 4, e1007805.

SPH Simulation of the Rising of a Kerosene Bubble in a Water Column: Effect of the Injection Condition

M. Robles-Santacruz,^a C. E. Alvarado-Rodríguez,^{a,b,*} F.I. Gómez-Castro,^a A. R. Uribe-Ramírez^a

^a*Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Universidad de Guanajuato, Noria Alta S/N, Guanajuato, 36050, México.*

^b*Consejo Nacional de Ciencia y Tecnología, Crédito Constructor, Alcaldía Benito Juárez, Ciudad de México, 03940, México*

Abstract

In this work, the numerical simulation of bubbles of kerosene in a column with water is performed by using the Smoothed Particles Hydrodynamics (SPH) method. The SPH method allows a good representation and monitoring of the interface between two or more miscible or immiscible fluids, advantage that can be used to study the hydrodynamic of bubbles in an immiscible environment). In this study, the contribution of the surface tension and wetting model in a biphasic flow of bubbles of kerosene along a column filled with water is analysed. The simulations are performed using a modify version of the free access code DualSPHyics. The validation is performed by coding the case of an ascending bubble and comparing the numerical results with the data reported in the literature. Several 2D simulations of kerosene injection in a column with water for three different flow rates, contact angles and injector diameters are performed, bearing a total of 27 cases of study. The center of mass and rise velocity in the validation case have good agreement with the data reported the literature. In the cases of injection, it has been determined that the length of injection is the principal parameter affecting the size of the bubbles. In conclusion, the SPH method can properly and naturally represent the creation of the bubbles due to its purely Lagrangian description. The surface tension model allows a stable surface drop during the rise of the bubble; moreover, with this model it is possible to study the rise of drops for different regimes.

Keywords: Multiphase flow, SPH, CFD, Rising Bubble, Injection of Bubble.

1. Introduction

The problems that exist in the injection of bubbles for processes of absorption, distillation and design of reactors make their study complex, due mainly to the wide variety of configurations related to the spatial distribution of both phases in the column. Many authors in the literature had studied the rise of drops and bubbles. Hysing et al. (2008) propose reference configurations for quantitative validation and comparison of incompressible interfacial flow codes, which model two-dimensional bubbles that rise in columns of liquid. The reference quantities: circularity, centre of mass and mean elevation velocity are defined and measured to control convergence towards a reference solution. Zhang et al. (2015) analyse several cases of individual bubbles rising through viscous

fluids, and the SPH results are validated by both experimental data and other numerical results in the literature. Furthermore, bubble coalescence phenomena are simulated in both vertical and horizontal directions and the results agree well with the experimental data. The bottom pressure in the equation of state was found to be essential to keep the multiphase interface smooth and stable when the Bond number is relatively small. In the study by Vahabi & Kamkari (2019) the increase and deformation of a single bubble in a viscoelastic fluid that obeys the Giesekus model is numerically simulated by a modified version of weakly compressible smoothed particle hydrodynamics (WC-SPH). In this work, the numerical simulation of the injection of a kerosene bubble in a water column is presented using the standard method of Smoothed Particles Hydrodynamics (SPH) which has increased its application for studies of computational dynamics of fluids. The numerical model used to simulate interfacial tension is based on a pairwise force of interaction between particles reported by Tartakovsky and Panchenko (2016). The simulations were performance using a modify version of the free code DualSPHysics v3.2 (Crespo et al., 2015). In the section 2 is presented the complete SPH numerical model used in the simulations. In the section 3 is presented a validation test of the rising bubble to prove the accuracy and convergence of the numerical results. In the section 4 are reported several simulations where different values of the inlet velocity, the injection diameter and the contact angle are evaluated as the injection conditions. Finally in the section 5 is reported the conclusions.

2. The SPH Numerical Model

For viscous incompressible flows, the governing equations are given by the Navier–Stokes equation expressed in SPH form as

$$\frac{d\mathbf{v}_a}{dt} = - \sum_b m_b \left(\frac{p_a + p_b}{\rho_a \rho_b} + \Gamma \right) \nabla_a W_{ab} + \mathbf{g} \quad (1)$$

where ρ is the density, p the pressure, \mathbf{v} the velocity field, W is the kernel function, \mathbf{g} is the gravity acceleration and Γ is the viscosity dissipative term. In a standard SPH formulation, where the pressure is given as a function of the density, local variations of the pressure gradient may induce local density fluctuations in the flow. Therefore, the flow is modelled by an artificial fluid that is approximately incompressible. The mass of a fluid element remains constant and only its associated density fluctuates. Such density fluctuations are calculated by solving the continuity equation expressed in the SPH in Eq. (2). The dynamical pressure p_d , which for simplicity we shall denote by p , is calculated using the relation expressed in Eq. (3) (Becker and Teschner, 2007).

$$\frac{d\rho_a}{dt} = \sum_{b=1}^N m_b (\mathbf{v}_a - \mathbf{v}_b) \cdot \nabla_a W_{ab}, \quad (2)$$

$$p = p_0 \left[\left(\frac{\rho}{\rho_0} \right)^\gamma - 1 \right] \quad (3)$$

where $\gamma=7$, $p_0 = c_0^2 \rho_0 / \gamma$, ρ_0 is a reference density, and c_0 is the sound speed at the reference density. This equation enforces very low density fluctuations since the speed of sound can be artificially slowed with accurate results in fluid propagation. The viscous term in equation (1) is modelled using the artificial viscosity model proposed by Monaghan (1992) due to the simplicity and numerical stability. Finally, the surface tension is modelled according to Tartakovsky and Panchenko (2016) where a Pairwise-Force is inserted in the Eq. (1)

$$\frac{d\mathbf{v}_a}{dt} = - \sum_b m_b \left(\frac{p_a + p_b}{\rho_a \rho_b} + \Gamma \right) \nabla_a W_{ab} + \mathbf{g} + \frac{\mathbf{F}_a}{m_a} \quad (4)$$

where \mathbf{F}_a is calculated from the interaction particles using the Eq. (6) .

In Eqs. (4) the pressure gradient is written in SPH form using the symmetric representation proposed by Colagrossi and Landrini (2003), which ensures numerical stability at the interface between two media with large density differences, while the surface tension term F_a is calculated according to the sum of the F_{ab} formulations given by Tartakovsky and Panchenko (2016) in Eq. (6).

$$\Gamma = \begin{cases} \frac{-\alpha \bar{c}_{ab} \mu_{ab}}{\rho_{ab}} & \mathbf{v}_{ab} \cdot \mathbf{r}_{ab} < 0 \\ 0 & \mathbf{v}_{ab} \cdot \mathbf{r}_{ab} \geq 0 \end{cases}, \quad (5)$$

where $\mu_{ab} = h \mathbf{v}_{ab} \cdot \mathbf{r}_{ab} / (\mathbf{r}_{ab}^2 + \eta^2)$, $\bar{c}_{ab} = 0.5(c_a + c_b)$, where c_a y c_b are the speed of sound for particle a and b respectively, $\eta^2 = 0.01h^2$ and α is a free parameter that can be adjusted according to the simulation case.

$$F_{ab} = \begin{cases} S_{ab} \cos\left(\frac{1.5\pi}{3h} |r_b - r_a|\right) \frac{r_b - r_a}{|r_b - r_a|}, & |r_b - r_a| \leq h \\ 0, & |r_b - r_a| > h \end{cases}, \quad (6)$$

where the S_{ab} parameter can be calculated from the surface tension fluid, the smoothing length the kernel function and the contact angle between the wetting fluid and the bound as is reported by Tartakovsky and Panchenko (2016).

No-slip boundary conditions are implemented at the walls of the vessel using the method of dynamic boundary particles developed by Crespo et al. (2007). In this method, a linear distribution of uniformly-spaced particles is placed at the walls of the enclosure, with separations of $\approx h/1.42$. This external particles are used to cope with the problem of kernel deficiency outside the computational domain. The wall particles are updated using the same loop as the inner fluid particles and so they are forced to satisfy Eqs. (9) and (10). However, they are not allowed to move according to Eqs. (14) and (15) so that their initial positions and velocities ($\mathbf{v}_w = \mathbf{0}$) remain unchanged in time. In this way, the presence of the wall is modelled by means of a repulsive force, which is derived from the source term of the momentum Eq. (1) and includes the effects of compressional, viscous, and gravitational forces.

3. Simulation of ascending bubble.

For the validation case the numerical approach and convergence of the numerical SPH model was verify. One bubble of air in a column filled with water was set according to the parameters reported by Hysing et al. (2009). The initial condition of the simulation is shown in the Figure 1 where the radio of the bubble is 2.5 centimetres, the simulation was performance in 2D considering the bubble as a circular shape. The ratio between the density and viscosity of the disperse phase (ρ_1) and the continuous phase (ρ_2) is $\rho_1/\rho_2=0.1$. The value of gravity acceleration is 1 m/s^2 , the surface tension is 24.5 and the Re and Eo dimensionless numbers are 35 and 10 respectively. The evolution of the center of mass of the bubble was compared with the numerical data reported by Hysing et al. (2009). Moreover, four simulations with different number of particles (1250, 5000, 20000 and 40000) were performance to evaluate the convergence of the numerical results. In the Figure 2 is shown the numerical convergence and error ($\varepsilon=RMSE$) of the SPH data with the numerical results reported by (Hysing et al., 2009).

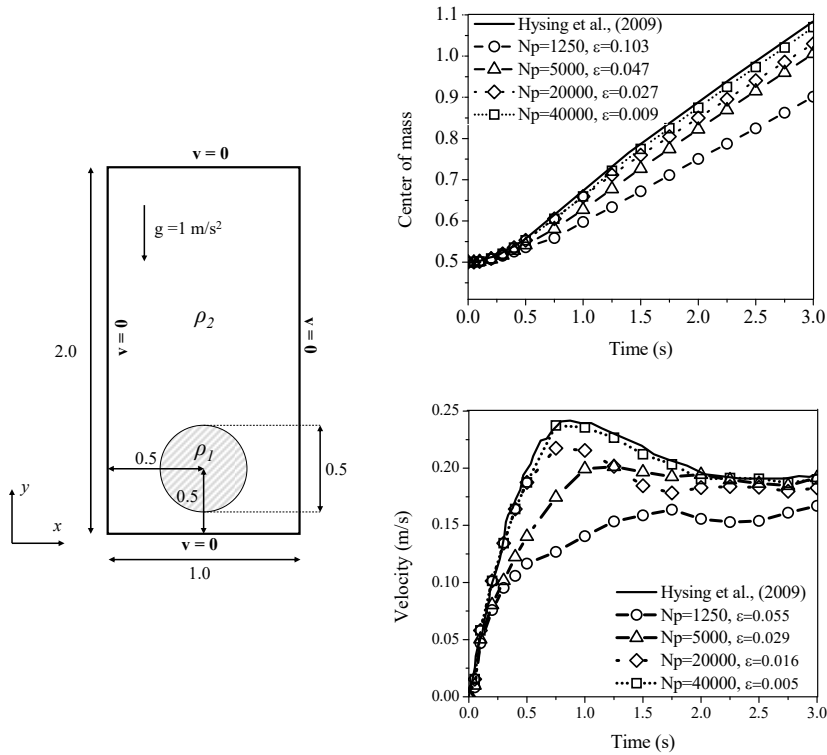


Figure 1. Left frame: Initial condition and dimensions of the simulation of an air bubble in water used in the validation case. Right frame top: Center of mass. Right frame bottom: Rise velocity.

4. Influence of Injection Conditions.

A 2D simulation of a kerosene flow with different injector diameters was carried out. Bubbles do not, in truth, have the capacity to exist in two dimensions. However, it is possible to create an environment in which they behave as though they existed only in two dimensions for example in a "sandwiching" the foam between two sheets of glass. So long as the separation of the sheets of glass is much smaller than the cell size of the bubbles, the foam can be treated in two dimensions. In two dimensions, the foams still follow the Young-Laplace equation for thin films, but the cell walls are only circular arcs instead of a three dimensional curve. It was chosen to use three different flows ($q = 1, 2, 3 \text{ cm}^3/\text{min}$), three injection length ($L=1, 1.5, 2 \text{ mm}$) and three contact angles ($\theta = 90^\circ, 50^\circ, 130^\circ$) to carry out the analysis of injection conditions in the creation of bubbles. In this way, they were made 27 study cases. The fluid and SPH parameters used for the simulations are shown in Table 1 and the geometry is shown in Figure 2. In the left frame of the Figure 3 is shown the evolution of the bubble formed by injection of Kerosene in the column filled with water. The variation of the area of the bubbles formed due to different injections conditions is reported in the right frame of the Figure 4. The numerical results show that the length of injections is the principal parameter that modify the size

of bubbles, moreover, when the contact angle increase the size of bubble decrease but to a lesser extent.

Table 1. Initial parameters for the SPH simulations in the injection study case.

Parameter	Amount	Parameter	Amount
Water density	1000 kg/m ³	Total simulation time	4 s
Kerosene density	890 kg/m ³	Gravity	9.8 m/s ²
Water viscosity (α)	0.05	Total particles (Np)	145961
Kerosene viscosity (α)	0.05	Surface tension	47 - 49 mN/m

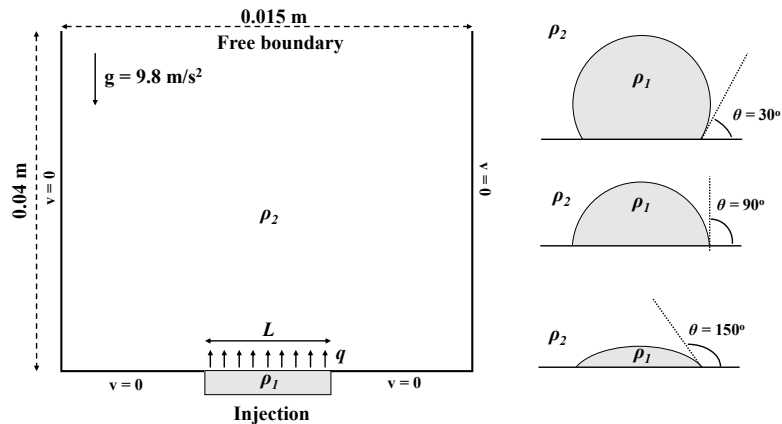


Figure 2. Left frame: Initial condition and dimensions for the simulations of analysis of injection conditions. Right frame: Contact angles for the fluid injected.

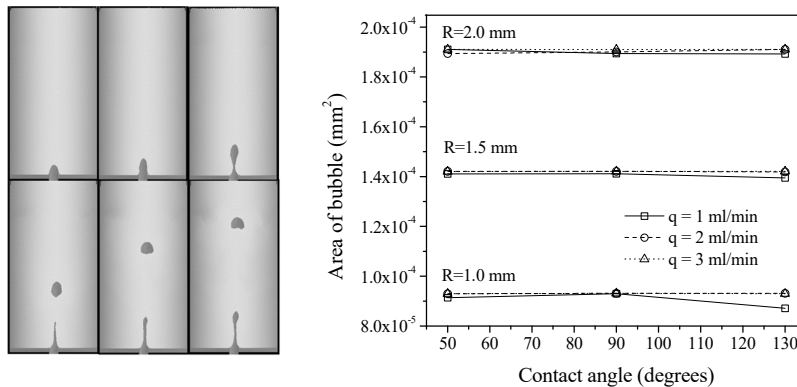


Figure 3. Left frame: Example of the evolution of one bubble. Right frame: Area of the bubbles obtained from numerical results.

5. Conclusions.

In conclusion, the SPH method can properly and naturally represent the creation of bubbles due to its purely lagrangian description. The model reported in the methodology and solved with the SPH method generates representative results with good precision as is prove with the numerical results obtained in the validation test. The surface tension model allows a stable surface drop during the rise of the bubble; moreover, with this model it is possible to study the rise of drops for different regimes. According to the parameters used in simulations, the length of injections is the principal parameter that modify the size of bubbles. The size of bubbles don't be affected by the flows used in the simulations. The 2D simulation is enough to know the area and speed of the bubbles, however, it is not able to predict well the shape of the bubbles. So, a 3D simulation is preferable for cases where is necessary to know about the shapes, movement and other important events.

References

- M. Becker, M. Teschner, 2007, Weakly compressible SPH for free surface flows, Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on computer animation. ACM Press, p. 209–17.
- S. Hysing, S. Turek, D. Kuzmin, N. Parolini, E. Burman, S. Ganesan, L. Tobiska, 2009, Quantitative benchmark computations of two-dimensional bubble dynamics, *Int. J. Numer. Meth. Fluids*, 60, 1259–1288.
- A. Zhang, P. Sun, F. Minga, 2015, An SPH modeling of bubble rising and coalescing in three dimensions, *Comput. Methods Appl. Mech. Engrg.*, 294,189–209.
- M. Vahabi, B. Kamkari, 2019, Simulating gas bubble shape during its rise in a confined polymeric solution by WC-SPH, *European Journal of Mechanics / B Fluids*, 75, 1–14.
- A. M. Tartakovsky, A. Panchenko, 2016, Pairwise Force Smoothed Particle Hydrodynamics model for multiphase flow: Surface tension and contact line dynamics, *Journal of Computational Physics*, 305, 1119–1146.
- A. J. C. Crespo, J. M. Domínguez, B. D. Rogers, M. Gómez-Gesteira, S. Longshaw, R. Canelas, R. Vacondio, A. Barreiro, O. Garcia-Feal, 2015, DualSPHysics: Open-source parallel CFD solver based on Smoothed Particle Hydrodynamics (SPH), *Com. Phy. Comm.*, 187, 204–216.
- J. Monaghan, 1992, Smoothed particle hydrodynamics, *Ann. Rev. Astron. Astrophys.*, 30, 543–574.
- A. Colagrossi, M. Landrini, 2003, Numerical simulation of interfacial flows by smoothed particle hydrodynamics, *Journal of Computational Physics*, 191, 448–475.
- A. J. C. Crespo, M. Gómez-Gesteira, R. A. Dalrymple, 2007, Boundary conditions generated by dynamic particles in SPH methods, *Computers, Materials & Continua* , 5, 173–184.

Novel method for properties prediction of pure organic compounds using machine learning

Nattasinee Chorbngam^a, Rathachai Chawuthai^b, Amata Anantpinijwatna^{a*}

^a*Department of Chemical Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand*

^b*Department of Computer Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand*

amata.an@kmitl.ac.th

Abstract

In classical thermodynamic, the estimation method of pure compounds properties was based on Newtonian physics, which required experimental data. It is proven to be inadequate for the growing demand of the novel chemical synthesis. There were several studies on the prediction of the pure compound properties based on QSPR methods. However, the conventional group-contribution based methods predictive capability was limited by the available measured data. Therefore, this study aims to approach the property prediction with a novel statistical-based method. The proposed method is derived using supervised machine learning algorithms. The experimental data used to train and validate the models were collected from the published literature. These data set are composed of the alkanes, alkenes, and alkynes derivatives containing 1-12 carbon atoms. The results show the improved accuracy of the model prediction compare to the conventional method in terms of root mean square error (RMSE) and mean absolute percentage error (MAPE).

Keywords: Properties prediction, Machine learning, Organic compounds

1. Introduction

Normal boiling point is defined as the temperature at which the vapor pressure of a liquid is equal to the atmospheric pressure. This thermo-physical property is important to design and calculation unit equipment of chemical engineering such as storage tank, vessel, flash drum, distillation column and so on. Typically, normal boiling point can be obtained as a result of the experiment. The boiling point is usually measured by small-scale simple distillation and microscale method (Shriner et al., 2004); however experimental measurements are often difficult, time-consuming, expensive, and sometimes hazardous. When the property is not accessible, the prediction methods are alternatives to determine the normal boiling point.

In general, normal boiling point of a compound is related to its molecular structure. Different models were used to correlate the boiling points of organic compounds with the different methods i.e. Constantinou and Gani (1994) develop a group contribution for estimating properties of pure compounds; Cordes et al. (2002) proposed a group contribution method for estimation of normal boiling point of organic compounds;

Al-Malah (2013) proposed a correlation between molecular weight and carbon atomic fraction to the normal boiling point of hydrocarbons; and Ghasemitarbar and Movagharnajad (2016), develop a new group contribution model to estimate the normal boiling point of organic compounds. These estimation methods can use to estimate normal boiling point for organic compounds, but it limited in nonetheless. In recent years, the large amount of information leading to the development of predictive techniques as machine learning which can find correlation from the dataset. Díaz et al. (2018) successfully used to predict the viscosity of ionic liquids with artificial neural network (ANN) models.

In this work, the proposed method is developed a novel method by applying machine learning techniques. The normal boiling point of organic compounds is expressed as a function of the simplified molecular-input line-entry system (SMILES).

2. Methodology

2.1. Data set

The accuracy and precision of prediction models depend on the number of experimental data and the comprehensiveness of the applied data set. In this work, the data set for the normal boiling point consists of 560 compounds. The experimental data are taken from ProCAPE software and the CRC Handbook (2010). The database of hydrocarbon compounds includes the following categories:

- (1) Normal paraffin: n-alkane.
- (2) Non-normal paraffin: example: iso-alkane, methyl-alkane, and ethyl-alkane.
- (3) Naphthene: cyclo-alkane.
- (4) Olefin: example: alkene, methyl-alkene, and ethyl-alkene.
- (5) Diolefin: example: alkadiene, methyl-alkadiene, and ethyl-alkadiene.
- (6) Cyclic olefin: cyclo-alkene, methyl-cylo-alkene, and ethyl-cyclo-alkene.
- (7) Alkyne: example: acetylene, propyne, and butyne.

The smile structure used to train and validate the models was collected from the published literature. To preparing data, the attribute set consists: the number of carbon atom 'C', the number of double bonds '=', the number of triple bonds '#', the number of branches '(', and the number of rings 'numeral' was created. The data set that consists of these 5 attributes were divided into a training set (460 compounds) and a test set (100 compounds).

2.2. Supervised machine learning algorithms

Supervised learning is a machine learning task for the predictions of unknown values by learning a model from labeled training data consisting of a set of training example input-output pairs (Russell and Norvig, 2010). The training set use to train the model by selected learning algorithms. To measure the accuracy of the learned model, the performance of the resulting model was measured on a test set of examples that are distinct from the training set. If the results are not satisfactory then changing learning algorithms is available. The options of supervised machine learning are classification and regression when the output is a finite set of values and a number, respectively. In this work, the regression methods are used to predict normal boiling point of compounds. A flowchart of the supervised machine learning model is displayed in Figure 1.

2.3. *Model development*

To select the appropriate regression methods, need to find correlations between each attribute and the experimental value. The attributes mentioned in the previous section, the number of carbon atoms has the best tendency to predict the normal boiling point properties as shown in Figure 2. Then we choose the number of carbon atoms as the basis of the prediction, the effect of this attribute is presented in equation (1). However, the values obtained from the first prediction were very different from experimental data. Thus, using the other attributes are used to improve the approximate value to the actual value. To find an effect of double bonds, triple bonds, branches, and rings. The regression methods were used to estimate as weight. The combination of first and second prediction is expressed in equation (2).

$$\beta_i^0 = \alpha x_i^\beta \quad (1)$$

$$y_i = \beta_i^0 + \sum \beta_i x_i \quad (2)$$

where β_i^0 is a first predicted value based on the number of carbon atoms, β_i is a second predicted value based on double bonds, triple bonds, branches, and rings, x_i is an explanatory variable, y_i is a dependent variable, α and β are constants from prediction.

2.4. Data analysis

The performance analysis of predicted normal boiling point of this method used root mean square error (RMSE) and mean absolute percentage error (MAPE).

3. Results and discussion

The novel supervised machine learning method was evaluated by using a data set composed of 560 compounds belonging to the most chemical families of compounds. The number of carbon atoms ranges from 1 to 12. The predicted value of normal boiling point using the novel method in comparison with the experimental data are presented in Figure 3. The values of the predicted normal boiling point as well as the status in the

Figure SEQ Figure * ARABIC 1 Supervised learning flowchart

Figure SEQ Figure * ARABIC 2 Plot of the experimental normal boiling point versus the number of carbon atom

method (the training set or the test set) are presented as supplementary information. The Statistical parameter of the method is presented in Table 1. For this method, the RMSE and MAPE are 8.340 and 0.016, respectively.

As a result, it seems possible to use the novel machine learning method to predict the chemical properties with this attributed label. Nonetheless, the problem with using this method is the inability to clearly specify the difference between similar compounds, examples are shown in Table 2. These six compounds have the different SMILES structures and boiling points; however, their set attributes are the same with 7 number of carbon atoms, 1 number of double bonds, and 1 number of rings. Therefore, the predicted boiling points of these 6 compounds turned out to be identical.

3.1. *Comparison to other prediction methods*

To demonstrate the accuracy of the novel method with other works developed for the similar purpose. The prediction result will be compared with those of MG (Marrero and Gani, 2001), CG (Constantinou and Gani, 1994), and JR (Joback and Reid, 1987). The comparison results are presented in Table 3.

Comparing to MG method, this novel method exhibits far lower error in terms of both RMSE and MAPE for every chemical category. Similarly, this method exhibits an improved accuracy of the prediction over JR method. Although the CG method made a better prediction to the alkanes and alkenes groups; the novel method still exhibits lower error overall.

Figure SEQ Figure * ARABIC 3 Plot of the predicted normal boiling point versus the experimental normal boiling point

Table SEQ Table * ARABIC 1 Statistical parameter of the obtained method.

Table SEQ Table * ARABIC 2 Prediction of the normal boiling point with the similar attributed label

Table SEQ Table * ARABIC 3 Comparison of proposed for functional compounds ; Where is the experimental value of normal boiling point (K), is the predicted value of normal boiling point (K), and is the number of data point.

4. Conclusions

The novel supervised machine learning method has been applied for the prediction of chemical properties. The aim of this development is to simplify the input; as well as improve the prediction accuracy. Normal boiling point has been selected for the demonstration of the method. In the first regression step, the number of carbon atoms shows the best tendency with the predicted property; therefore, it is used as the primary attribute. Then the other attributes are used to improve the approximation value accuracy with weighted regression.

The results of the estimated normal boiling points of 560 chemical compounds in 7 categories show acceptable error for both the training and the test sets. Around 1.6% error of prediction in term of mean absolute error has been achieved. Comparing to the conventional methods, this novel method exhibits far lower error than the

Marrero-Morejon and Gani method, a bit improvement over Joback and Reid method, and a comparable result to Constantino and Gani method. However, it is worth noted that the mentioned conventional methods are developed for the prediction of a bigger data set of chemical compounds. Therefore, the error may increase over the expansion of the data set involved, which need to be addressed in further study.

In conclusion, the novel supervised machine learning method has been proven to be useful for the prediction of normal boiling points of chemical compounds with simplify input. It should be further applied for the prediction of various other properties; as well as further applied for larger group of compounds. Moreover, the supervised machine learning method can also be used for the improvement of the conventional group-contribution methods by re-parameterization the groups attributes.

References

- Al-Malah, K.I., 2013. Prediction of Normal Boiling Points of Hydrocarbons Using Simple Molecular Properties. *J. Adv. Chem. Eng.* 3, 1–9. <https://doi.org/10.4303/jace/235654>
- Constantino, L., Gani, R., 1994. New group contribution method for estimating properties of pure compounds. *AIChE J.* 40, 1697–1710. <https://doi.org/10.1002/aic.690401011>
- Cordes, W., Rarey, J., 2002. A new method for the estimation of the normal boiling point of non-electrolyte organic compounds. *Fluid Phase Equilib.* 201, 409–433. [https://doi.org/10.1016/S0378-3812\(02\)00050-X](https://doi.org/10.1016/S0378-3812(02)00050-X)
- Díaz, I., Rodríguez, M., González-Miquel, M., González, E.J., 2018. COSMO-derived descriptors applied in ionic liquids physical property modelling using machine learning algorithms. *Comput. Aided Chem. Eng.* 43, 121–126. <https://doi.org/10.1016/B978-0-444-64235-6.50023-1>
- Ghasemitabar, H., Movagharnejad, K., 2016. Estimation of the normal boiling point of organic compounds via a new group contribution method. *Fluid Phase Equilib.* 411, 13–23. <https://doi.org/10.1016/j.fluid.2015.11.029>
- Joback, K.G., Reid, R.C., 1987. Estimation of Pure-Component Properties from Group-Contributions. *Chem. Eng. Commun.* 57, 233–243. <https://doi.org/10.1080/00986448708960487>
- Lide, D.R., Haynes, W.M.M., Baysinger, G., Berger, L.I., Roth, D.L., Zwillinger, D., Frenkel, M., Goldberg, R.N., 2010. *CRC Handbook of Chemistry and Physics*, 90th ed.
- Marrero, J., Gani, R., 2001. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* 183–184, 183–208. [https://doi.org/10.1016/S0378-3812\(01\)00431-9](https://doi.org/10.1016/S0378-3812(01)00431-9)
- Russell, S.J., Norvig, P., 2010. *Artificial Intelligence A Modern Approach*, 3rd ed. Pearson Education, Inc.
- Shriner, R.L., Hermann, C.K.F., Morrill, T.C., Curtin, D.Y., Fuson, R.C., 2004. *The Systematic Identification of Organic Compounds*, 8 th. ed, John Wiley & Sons. Inc.

Self-learning surrogate models in superstructure optimization

Julia Granacher^a, Ivan Daniel Kantor^{a,b}, Michel Lopez^a, Francois Maréchal^a

^aEcole Polytechnique Fédérale de Lausanne - CH-1015 Lausanne

^bHES-SO Valais-Wallis - CH-1950 Sion

Abstract

In this contribution, we propose an algorithm for replacing non-linear process simulation integrated in multi-level optimization of an energy system superstructure with surrogate models. With our approach, we demonstrate that surrogate models are a valid tool to replace simulation problems in multi-stage optimization frameworks and enable the improvement of their computational performance. Furthermore, we want to show that the quality of the results is not penalized and flexibility is provided to the optimization.

It is desired to keep the amount of labeled data samples needed to create the surrogate model to a minimum, since their creation is computationally expensive.

In our algorithm, sampling methods are used to create an initial set of data points in the input domain in the decision variables of the simulation model to be replaced.

ANNs are trained on the initial training set. Using Dropout as a Bayesian approximation for quantifying the uncertainty of a prediction, the predictions can be qualified. New data points are continuously labelled and added to the training set based on the achieved prediction quality, until a minimum quality of the model is met. When applied in the optimization superstructure, the ANN can only be used when the prediction quality for the given data point is satisfying. Integrating these surrogate models in an optimization framework of an energy system will allow to only access the computationally expensive simulation when the quality of the prediction of the surrogate model is not sufficient. Simultaneously, a continuous improvement of the surrogate model will be achieved by using the created simulation results to parallelly refine the surrogate model by adding the created data points to the training set and therefore improve the model's validity range.

It is found that the methodology of continuously adding the data points based on the prediction uncertainty improves the quality of the surrogate model. Initial results indicate that when applied in the optimization framework, the suggested methodology holds potential to improve computational time and flexibility.

Keywords: Energy System Optimization, Process Design, Surrogate Models, Artificial Neural Networks, Active Learning

1. Motivation and Background

One of the main challenges we are facing as society is moving away from fossil energy carriers, searching for alternatives to sustainably and safely provide energy. Industries converting woody biomass hold a great potential of supporting the energy transition by valorizing their resources in such a way they can on the one hand help to provide storage options of volatile energy carriers, and on the other hand reduce their emissions and thus increase their carbon efficiency. Selecting and integrating the biorefinery processes in a complex industrial energy system requires systematic approaches like computer simulation, mathematical modelling and superstructure optimization. These techniques lead to highly complex, large scale mixed integer nonlinear optimization models that face some computational challenges (Cozad et al., 2014). High computational cost, noisy

function evaluation of simulation-based optimization models as well as solving issues due to complexity and non-convexity of rigorous first-order process influence the efficient solution generation (Fahmi and Cremaschi, 2012). In an attempt to reduce a model's complexity and to address the issues described above, the interest in developing surrogate models has increased over recent years. Surrogate models are especially interesting for application to multi-objective optimization problems that include superstructure consideration. These types of models are computationally expensive - due to the presence of many possible alternative configurations and operating conditions that need to be taken into account (Teske, 2014) - and hence, calculating the model at each iteration point is not practical. With surrogate models, new calculations or estimations can be performed in less computational time than with the original model. Furthermore, they are able to generalize the model result without running into convergence issues.

Especially artificial neural networks (ANNs) have been widely applied in chemical process modeling (Nascimento et al., 2000; Teske, 2014; Fahmi and Cremaschi, 2012; Tock and Maréchal, 2014). It was reported that neural networks have the ability to fill gaps in the search grid caused by the absence of analytical solutions (Nascimento et al., 2000) and that integrating surrogate models of certain processes in the optimization reduces the computational time considerably without penalizing the power plant operation quality (Tock and Maréchal, 2014).

Apart from ANNs, other approaches such as polynomial- and Kriging-based techniques were proved to be valid for surrogate modeling in process design (Caballero and Grossmann, 2008; Henao and Maravelias, 2011; Pedrozo et al., 2020).

Alternatively to using static data sets for training a surrogate model, one can let the machine decide which data to add best at the current state. Active learning is especially valuable when labelling data is computationally expensive, or where the purpose is to update the model by data that arrives continuously (Settles, 2012). Active learning is closely linked to hyperparameter optimization, as there the goal of the both is to choose data and model parameters in the most efficient way (Jablonka et al., 2020; Lookman et al., 2019). The objective of the research presented is to develop an approach for designing surrogate models of chemical process units with an active learning strategy. The design shall help to replace complex simulation of chemical processes in optimization frameworks with more flexible design and avoid non-convergence issues in the master optimization. Correlations between the inputs and outputs of the complex non-linear relations should be detected and replaced in the global optimization framework. Since labelling data is computationally expensive; it is anticipated to leave the number of data points at a minimum, and the surrogate model shall be able to continuously improve itself.

2. Methodology

2.1. Energy system design and optimization strategy

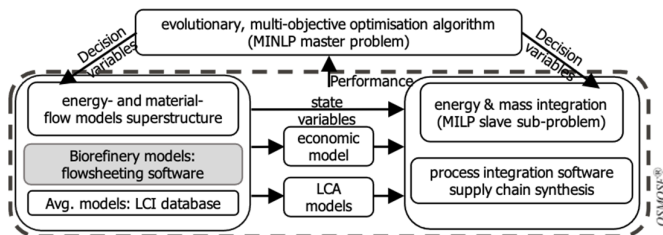


Figure 1: simplified OSMOSE methodology (Kermani, 2018)

Systematic modelling and optimization approaches are required to design complex energy systems. The mass and energy integration and optimization model of an industrial sector used in this study is built in the OSMOSE platform, which is continuously developed in the Industrial Process and Energy Systems Engineering Group (PESE) at EPFL (Figure 1). OSMOSE follows a two-stage decomposition approach, where the slave problem consists of a mass and energy integration framework developed as a mixed-integer linear programming (MILP) problem. Depending on the case study, it contains non-linear process unit simulations modelled in commercially available software. An evolutionary, multi-objective optimization algorithm is used in the master problem to solve a mixed-integer nonlinear programming (MINLP) problem considering economic and environmental objectives (Gassner and Maréchal, 2009).

2.2. Adaptive Surrogate model design

The relationship between the design variables and the system responses of complex thermochemical processes currently approximated by flowsheeting software in the overall energy system superstructure (see grey box in Figure 1) shall be investigated. For this purpose, a reliable presentation of the respective process with as few label generations as possible is anticipated. Integrating the developed surrogate model in the energy system superstructure shall allow for an estimation of the quality the surrogate model achieves for each provided prediction, and respectively enable the model to improve itself continuously.

2.2.1. Surrogate model design

For addressing the requirement of as few labelling points as possible when creating the model, we propose the algorithm shown in Figure 2.

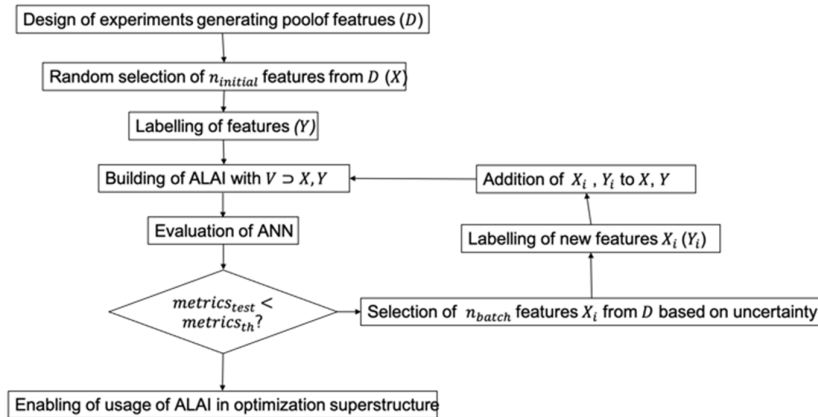


Figure 2: Proposed algorithm for designing ALAI

Design of experiments

The design space D is created by drawing samples for the design variables of the process that shall be replaced in an initial design of experiment (DoE).

Initial feature and label generation

For a randomly selected feature subset X of samples from D of the size $n_{initial}$, the respective labels Y are created by calling the simulation software with samples present in X . *Artificial neural network design and evaluation*

ANNs are used to replace the flowsheeting process models in the presented approach. Features X and labels Y are transformed to distributions centered around zero and unity variance by applying standard scaling. The model is built and trained on a share of the provided data V , the remaining share is used for testing. Depending on the performance of the ANN, the Active Learning Artificial Intelligence (ALAI) instance is accepted to be used in the OSMOSE superstructure instead of the flowsheeting software. Hereby, the performance is determined by the metrics achieved for evaluating the test set. In case the metrics is not satisfying, the ALAI instance is further improved by adding more datapoints to V .

Adaptive model improvement

For adding new datapoints to V , the uncertainty of the prediction ALAI makes on the remaining unlabelled features in D is measured. We label the sample points with the highest obtained uncertainty and add them to V . For approximating the uncertainty of a prediction, we use Dropout layers in the ANN as it is demonstrated by Gal and Ghahramani (2016). According to Gal and Ghahramani (2016), when applying Dropout, layer nodes are randomly deactivated following a Bernoulli distribution during training. A neural network with Dropout applied before every weight layer is mathematically equivalent to a Bayesian approximation of the Gaussian process (Gal and Ghahramani, 2016). To apply this concept, a second ANN including Dropout layers is built with the same hyperparameters as the first one and added to the ALAI instance. For each sample in D , we obtain a range of corresponding predictions from the ANN with Dropout. With this range of predictions per sample, the standard deviation can be computed and applied as a measure of uncertainty (Gal and Ghahramani, 2016). The unlabeled samples with the highest uncertainty are labelled and the resulting new features x_i and labels y_i are added to the training and testing data. This procedure is repeated i times until the resulting test metrics of the ALAI instance satisfies the quality demands.

2.2.2. Surrogate model application and continuous improvement

After the designed ALAI instance has achieved high enough quality for being integrated in the energy system superstructure defined, it can replace the call of the flowsheeting software. When OSMOSE calls for the flowsheeting unit that has been replaced by ALAI, the labels \hat{y}_i for the features x_i given by the master optimization are predicted with the

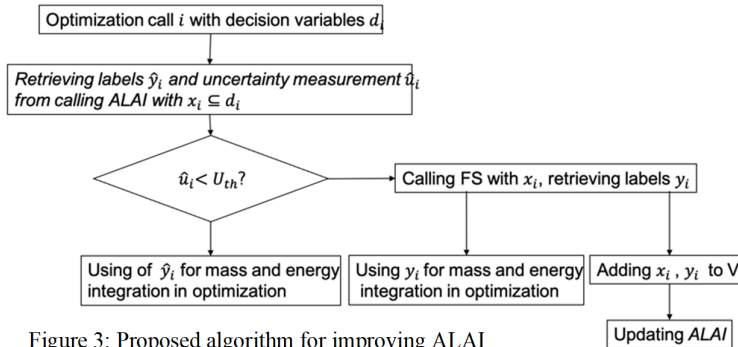


Figure 3: Proposed algorithm for improving ALAI

the ANN without dropout. Apart from the predicted labels, ALAI returns the predicted uncertainty \hat{u}_i by using the ANN with dropout layers. The results of the ANN are used in the master optimization if the uncertainty of the prediction is lower than a specified threshold U_{th} . However, if ALAI returns a high uncertainty for the given prediction, the call is discarded and instead, OSMOSE is calling the original flowsheeting software to label the samples and continue the optimization call. Parallel to completing the optimization call with the generated labels y_i , x_i and y_i are added to dataset V and ALAI is updating itself with the new set of training data.

3. Case study

The overall energy system used for validation of the suggested approach consists of a Fischer Tropsch (FT) production unit modelled with the flowsheeting software Belsim Vali, a steam network superstructure and various utilities and service that interact with each other. For validating the suggested approach, the flowsheet for the FT fuel production is approximated with an ALAI instance. The decision variables the master optimization can chose include 7 variables (operating conditions) that affect the FT production, and 63 outputs (operating conditions, mass and energy flows) are predicted by the flowsheet that are needed by OSMOSE. For starting the ALAI design, a design of experiments with 10000 sampling points of the 7 features is performed, out of which $n_{initial} = 1000$ random samples are labelled and added to V with which the initial ANNs are trained. The regression metrics for evaluating the models performance chose is the mean squared error (MSE) The maximum accepted MSE for testing with the initial ANN design MSE_{th} is set to 15%. The size of data batches added to V in each iteration i is set to $n_{batch} = 300$. For predicting the uncertainty of generated labels \hat{y}_i , the ANN with Dropout is evaluated 200 times. For being accepted in for being used in the optimization, ALAI needs to predict an uncertainty smaller $\sigma_{th} = 10\%$ for respective decision variables.

4. Preliminary results and discussion

When initially applying the proposed methodology for training the model adaptively, the test MSE reaches below 15% after 25 iterations, or 8200 labelled datapoints. We decided to stop the adaptive training here and work on further model improvements, so that the performance of ALAI can be improved before adding more labelled data. Using the current version of ALAI with OSMOSE results in 72% OSMOSE calls using ALAI, while the rest is calling the flowsheeting software. The call of an ALAI instance from OSMOSE takes approximately 30% less computational time than calling a the original flowsheeting software.

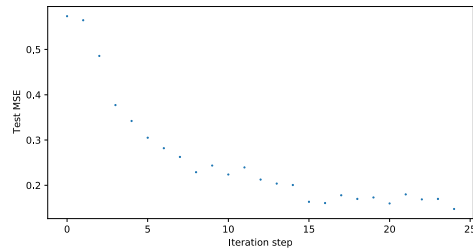


Figure 4: Mean squared error development

However, when validating the results OSMOSE computes with ALAI to the ones that are achieved with the flowsheeting software, considerable deviations in objective functions

of 23% are observed, which means the ALAI performance has to improve considerably before being applied. Future work will include the optimization of the ALAI's hyperparameters, in order to achieve the required metrics with fewer labelled datapoints. Furthermore, combining multiple ANNs for prediction and using combinations of other algorithms such as Random Forests and Kriging interpolation shall be investigated for performance improvement. The effects of different threshold for metrics and uncertainty will be investigated, and the results of the optimization will be further validated by comparing them to the results achieved with applying flowsheeting instead of ALAI.

5. Acknowledgments

This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No SEP-210500079.

References

- J. A. Caballero, I. E. Grossmann, 2008. An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal* 54 (10), 2633–2650.
 URL <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.11579>
- A. Cozad, N. V. Sahinidis, D. C. Miller, 2014. Learning surrogate models for simulation-based optimization. *AIChE Journal* 60 (6), 2211–2227.
 URL <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.14418>
- I. Fahmi, S. Cremaschi, Nov. 2012. Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. *Computers & Chemical Engineering* 46, 105–123.
 URL <http://www.sciencedirect.com/science/article/pii/S0098135412001822>
- Y. Gal, Z. Ghahramani, Oct. 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. arXiv:1506.02142 [cs, stat]ArXiv: 1506.02142.
 URL <http://arxiv.org/abs/1506.02142>
- M. Gassner, F. Maréchal, Mar. 2009. Methodology for the optimal thermo-economic, multi-objective design of thermochemical fuel production from biomass. *Computers & Chemical Engineering* 33 (3), 769–781.
 URL <http://www.sciencedirect.com/science/article/pii/S0098135408001981>
- C. A. Henao, C. T. Maravelias, May 2011. Surrogate-based superstructure optimization framework. *AIChE Journal* 57 (5), 1216–1232.
 URL <https://aiche.onlinelibrary.wiley.com/doi/full/10.1002/aic.12341>
- K. M. Jablonka, D. Ongari, S. M. Moosavi, B. Smit, Jan. 2020. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. arXiv:2001.06728 [cond-mat]ArXiv: 2001.06728.
 URL <http://arxiv.org/abs/2001.06728>
- M. Kermani, 2018. Methodologies for simultaneous optimization of heat, mass, and power in industrial processes.
 URL <https://infoscience.epfl.ch/record/257269?p=Kermani>
- T. Lookman, P. V. Balachandran, D. Xue, R. Yuan, Feb. 2019. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Computational Materials* 5 (1), 1–17, number: 1 Publisher: Nature Publishing Group.
 URL <https://www.nature.com/articles/s41524-019-0153-8>
- C. A. O. Nascimento, R. Giudici, R. Guardani, Oct. 2000. Neural network based approach for optimization of industrial chemical processes. *Computers & Chemical Engineering* 24 (9), 2303–2314.
 URL <http://www.sciencedirect.com/science/article/pii/S0098135400005871>
- H. A. Pedrozo, S. B. R. Reartes, Q. Chen, M. S. Diaz, I. E. Grossmann, Jul. 2020. Surrogate-model based MILP for the optimal design of ethylene production from shale gas. *Computers & Chemical Engineering*, 107015.
 URL <http://www.sciencedirect.com/science/article/pii/S0098135420303434>
- B. Settles, Jun. 2012. Active Learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 6 (1), 1–114, publisher: Morgan & Claypool Publishers.
 URL <https://www.morganclaypool.com/doi/abs/10.2200/S00429ED1V01Y201207AIM018>
- L. S. Teske, 2014. Integrating Rate Based Models into a Multi-Objective Process Design & Optimisation Framework using Surrogate Models. Ph.D. thesis, EPFL.
- L. Tock, F. Maréchal, Feb. 2014. Process design optimization strategy to develop energy and cost correlations of CO₂ capture processes. *Computers & Chemical Engineering* 61, 51–58.
 URL <http://www.sciencedirect.com/science/article/pii/S0098135413003335>

Data-driven Modeling of an Industrial Ethylene Oxide Plant: Superstructure-based Optimal Design for Artificial Neural Networks

Hasan Sildir,^{a*} Sahin Sarrafi,^b Erdal Aydin^c

^a*Department of Chemical Engineering, Gebze Technical University, Kocaeli 41400, Turkey*

^b*Socar Turkey, R&D and Innovation Co., Aliaga, Izmir, 35800, Turkey*

^c*Department of Chemical Engineering, Bogazici University, Bebek, İstanbul 34342, Turkey*

hasansildir@gtu.edu.tr

Abstract

Optimum selection of input variables, number of hidden neurons and connections between the network elements delivers the best configuration of an artificial neural network (ANN), resulting in reduced over-fitting and improved performance. In this study, a superstructure-oriented ANN design and training algorithm is suggested and implemented on an industrial Ethylene Oxide (EO) plant for the prediction of product related variables (i.e. EO production rate). Proposed formulation is a mixed integer nonlinear programming problem (MINLP), which takes the existence of inputs, neurons and connections of the network into account by binary variables in addition to continuous weights of existing connections. Investigations show that almost 90% of the connections are removed compared to the fully connected ANN (FC-ANN) with 50% decrease in the number of inputs of the ANN, approximately. The modified ANN delivers a better prediction performance over FC-ANN, which suffers from over-fitting.

Keywords: machine learning; artificial neural networks; superstructure optimization; process modelling; mixed integer nonlinear programming

1. Introduction

Ethylene oxide (EO), a captive product used for the production of ethylene glycols (MEG, DEG, and TEG), is produced through selective oxidation of ethylene and oxygen in the presence of an Ag supported α -Al₂O₃ catalyst in the EO/EG plant located in the Socar Turkey/Petkim Petrochemical plant in Turkey. The EO process takes place in two parallel fixed-bed multi-tubular reactors at a temperature and pressure range of 240–260 °C and 17–18 bar, respectively. Both ethylene epoxidation and EO combustion occur in the reactors and only the former is desirable. Side reactions reduce the EO yield and an increase in CO₂ emission from the plant. From both safety and optimum control perspectives, the monitoring of the concentrations of the streams, which are measured through two separate online gas chromatography (GC) instruments, is of great importance. The former is designed for hydrocarbons and permanent gases whereas the latter focuses on the chlorine-based compounds. The operating strategy of the reactors is

to maximize the EO yield under various operational constraints. The simplified process flow diagram of the process is shown in Fig. 1.

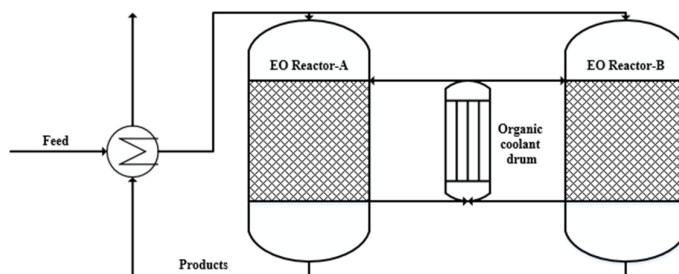


Figure 1: Simplified process flow diagram of EO process.

High number of criteria is considered during EO production at the desired levels under the continuous and sluggish catalyst deactivation. These criteria lay the foundation of the features of information matrix exploited throughout this study as detailed in Table.1. All the feed compositions are measured before the reactor (i.e., after EO and CO₂ absorption section). Fresh VCM feed is introduced into the reactor inlet stream right before the reactor. Real-time optimization and automation of such a complex process is a challenging task, requiring significant effort in first principles and mechanistic modeling. Alternatively, Artificial Neural Networks (ANNs) are promising empirical models to estimate the product related variables from easily measurable variables (i.e. temperatures and pressures).

Table.1 Descriptions of the features and inputs.

TAG Description	Unit	TAG Description	Unit
Lean absorbent flow rate	t/h	Fresh C ₂ H ₄ feed	t/h
CO ₂ absorber overhead pressure	kg/cm ² g	Fresh CH ₄ feed	kg/h
Feed composition (C ₂ H ₄)	% (v/v)	CH ₃ Cl (methyl chloride) in the recycle	ppm
Feed composition (O ₂)	% (v/v)	VCM (Vinyl chloride monomer) in the recycle	ppm
Feed composition (N ₂)	% (v/v)	C ₂ H ₅ Cl (ethyl chloride) in the recycle	ppm
Feed composition (Ar)	% (v/v)	Reactor A coolant pressure	kg/cm ² g
Feed composition (CH ₄)	% (v/v)	Reactor A inlet pressure	kg/cm ² g
Feed composition (C ₂ H ₆)	% (v/v)	Flow rate to reactor A	t/h
Feed composition (CO ₂)	% (v/v)	Fresh VCM feed to reactor A	kg/h
Feed composition (EO)	% (v/v)	Product composition (EO)	% (v/v)
Fresh O ₂ feed	t/h		

2. Methodology: Optimal Superstructure-based Design of ANNs

The standard application of ANNs includes using fully connected networks, where all inputs, neurons and outputs are entirely connected. Fully connected ANN architectures (FC-ANNs) have high number of parameters. Typically, as the dimensions get larger, higher number of connections and parameters are introduced. It is usually expected that the increase in the number of parameters provides higher capability of fitting to the training data. Nevertheless, it is worthwhile here to mention that this increase may easily result in overfitting, and therefore poor prediction capability of the ANN model. Introducing more data is usually not a satisfactory effort as new data do not carry

additional statistical information unless they are collected from a different location in the plant. A feedforward ANN model is generally expressed as:

$$y = f_1(A \cdot f_2(B \cdot u + C) + D) \quad (1)$$

where f_1 and f_2 are output and hidden layer activation functions respectively. A typical activation function is the hyperbolic tangent function, which is used in this study as well. A and B are weight matrices; C and D are bias vectors; u is the input vector and y is the output vector. Related ANN parameters are all continuous, and theoretically unbounded, and their dimensions depend on the number of inputs, outputs, and number of neurons which is determined manually before training. Identification issues of these ANN parameters may result in poor testing performance. A typical solution to this problem is to include a regularization term to penalize the large values of ANN parameters, which unfortunately cannot regularize the hyper parameters ([1]). Pruning is another method to reduce the number of connections in ANNs ([2-3]). Dua proposed using a general mixed-integer optimization formulation to eliminate overfitting by detecting the optimal configuration of ANNs ([4]). Both number of neurons and existence of the interconnections are included in the objective function. Yet, this formulation does not consider selection of optimum input variables. Moreover, resulting formulations are either mixed-integer linear programs with fixed parameter weights or small scale MINLPs with fixed structures, number of nodes and/or interconnections, exhibiting poor flexibility. On the other hand, this significant study shows that much better test, and thus prediction performance can be achieved with fewer neurons and connections.

In this study, a novel MINLP (mixed-integer nonlinear programming) formulation is developed for the design and training of an optimal architecture feedforward ANN (OA-ANN), by modifying traditional ANN equations. The MINLP formulation introduces additional binary variables to the traditional ANN equations to represent the existence of network elements. The resulting superstructure formulation also takes the selection of the input variables into account in addition to the number of hidden neurons and connections, which in turn maximizes the overall flexibility and strength of the formulation. Accordingly, the proposed MINLP problem, whose objective is to minimize the number of connections of the ANN, is given by:

$$\begin{aligned} & \min_{A_{binary}, B_{binary}, C, D, N_{binary}, U_{binary}} \sum_{i,j} A_{binary,ij} + B_{binary,ij} \\ \text{s. t. } & F = \sum_{i=1}^N \|f_1((A \circ A_{binary}) \cdot \text{diag}(N_{binary}) \cdot f_2((B \circ B_{binary}) \cdot \text{diag}(U_{binary}) \cdot u_i + C) + D) - y_i\| \\ & F \leq \varepsilon \\ & A_{binary,ij} \leq N_{binary,j} \\ & B_{binary,ij} \leq U_{binary,j} \\ & A_{i,j}, B_{i,j}, C, D \in [-4,4] \\ & A_{binary}, B_{binary}, N_{binary}, U_{binary} \in \{0,1\} \end{aligned} \quad (2)$$

where \circ is the Hadamard product operator; u_i and y_i are the i^{th} input and output sample respectively. N is the number of samples used for the training. $A_{binary,ij}$ and $B_{binary,ij}$ are matrices with binary values representing the existence of connections. N_{binary} and U_{binary} are the binary vectors defining the existence of a neuron and an input, respectively. The existence of a particular connection between a neuron and an input is defined by the binary variable $B_{binary,ij}$. A_{ij} is the continuous weight parameter of the connection between the j^{th} neuron and the i^{th} output. Similarly, B_{ij} represents the connection between an input and the corresponding neuron. ε is the upper bound for the overall training error. This

way, a trade-off is included into the training, which in turn is expected to increase the test performance after implementing the proposed formulation. Please note that suggested formulation has the flexibility to result in a fully connected network once ε is set to a small value. In other words, a fully connected ANN is the upper bound of the suggested formulation. Problem (2) is a non-convex, and generally a large-scale MINLP, which is quite challenging to solve to the global optimum. In this work, an adaptive and evolutionary algorithm is used to solve the resulting non-convex MINLP problem. This method has a similar idea with [5] and it decomposes the original MINLP into an integer program (IP) and a nonlinear program (NLP). IP's only include binary decision variables that can be adjusted during optimization, whereas NLP's only involve continuous decision variables. The IP stands on the outer loop and is solved via the genetic algorithm based IP solver of Matlab while the inner loop NLP is solved by using IPOPT ([6]). Two problems are solved sequentially until the tolerance value of the original problem objective value or the maximum wall clock time is reached.

Please note that IPOPT is a local solver and might add randomness to the outer problem as the IP solution is based on a black-box genetic algorithm. Thus, we do not propose any integer cuts to be provided during the iterations. In addition to that, sensitivity analysis of the proposed method might be of interest to fully assess the regularization of the optimal ANNs using heuristic solutions, which is left as a future study. A global NLP solver can also be used for the inner problem to deal with the aforementioned challenges. On the other hand, investigations show that suggested method usually provides a sufficient heuristic solution in less than 3 min.

3. Results and Discussion

The proposed MINLP based approach is tested on the industrial EO plant. Corresponding data cover hourly measurements of 105 randomly selected days. We tested different training ratios, e.g. 20%, 50%, and 70%, to demonstrate the impact of the approach. OA-ANN architecture from the training with 20% of the data is shown in Fig. 2 as a typical demonstration. In the design of OA-ANN with 20% of the data for training, 9 inputs, out of 20, are selected from which connections exist to hidden layer as shown in Fig. 2. The thickness of the connections represents the absolute magnitude of the corresponding weight. Note that only significant connections are maintained, eliminating the ineffective connections and thereby tightening the non-linear training optimization problem.

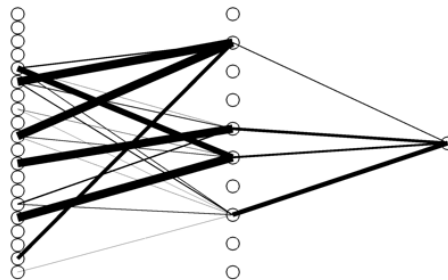


Figure 2: OA-ANN for 20% training data

In addition, the OA-ANN architecture, with only 4 hidden layer neurons, does not have a fully connected architecture as the selected inputs do not have connection to all hidden neurons. As a result, a more efficient information flow is obtained through the network.

On the other hand, there are many different architectures which may deliver a similar performance since the MINLP problem is highly non-convex and inherently contains input multiplicity both due to input variable correlations and parameter dependencies. Accordingly, the heuristic solution method and even the optimization algorithm tuning parameters might result in different architectures. The selected inputs include the lean absorbent flow rate, absorber overhead pressure, feed composition of N₂ and Ar, feed CO₂ composition, fresh O₂ feed, fresh CH₄ feed, Vinyl chlorine monomer and ethyl chloride feed. Note that some input variables show significant correlation and there are many subset combinations to deliver a similar performance. Thus, the selected inputs, likewise the structure in Fig. 2, is not a unique solution; different initial guesses and MINLP solution approaches would result in a different optimal subset. Fig. 2 also reveals the impact of the input variables on the network, which is represented by the thickness of the connection and scaled by the corresponding weight. For instance, N₂ composition in the feed contributes to the network calculations in smaller scale compared to the chlorine compounds, which have a significant impact on the process control. A similar observation is also valid for O₂ feed, as well.

Detailed training and test performance comparison of FC-ANN and OA-ANN at various training ratios are presented in Table 2 which includes some common statistical measures.

Table 2: OA-ANN and FC-ANN detailed performance comparison

Train ratio	20%				50%				70%			
	OA-ANN		FC-ANN		OA-ANN		FC-ANN		OA-ANN		FC-ANN	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	0.003	0.004	0.0001	0.0140	0.003	0.0035	0.0004	0.031	0.0045	0.005	0.0004	0.047
CV	0.046	0.061	0.008	0.147	0.051	0.061	0.018	0.25	0.065	0.093	0.017	0.296
MAE	-	0.424	-	2.481	-	0.49	-	4.53	-	0.37	-	1.03

Table 2 includes root mean square error (RMSE), coefficient of variation (CV) and maximum absolute error (MAE) as a comparison metric. Due to the eliminated overfitting, the training and the test performances of OA-ANN show a good agreement at all training ratios. Although FC-ANN demonstrates significantly lower training error with high number of connections, it suffers from larger test error compared to OA-ANN. Similar insight can also be obtained from CVs. In all cases, test performance of OA-ANN is superior to FC-ANN despite higher training error. The architecture comparisons are given in Table 3. The network dimensions of OA-ANN are significantly less than FC-ANN in all training ratios, providing additional computational advantages when ANN update is necessary. Overfitting causes poor test performance as shown in Fig. 3. This issue stems from high number of connections in FC-ANN, whose weights could not be identified based on the statistical content of the current training data. Both FC-ANN and OA-ANN test results are demonstrated in addition to normalized measurements in Fig. 3. Note that the OA-ANN has a better capability to predict a different operating condition, as shown between sample numbers 400-500.

Table 3: OA-ANN and FC-ANN architecture comparison

	OA- ANN			FC-ANN
	20%	50%	70%	-
Neurons	4	3	5	10
Connections	22	19	26	210
Inputs	9	9	8	20

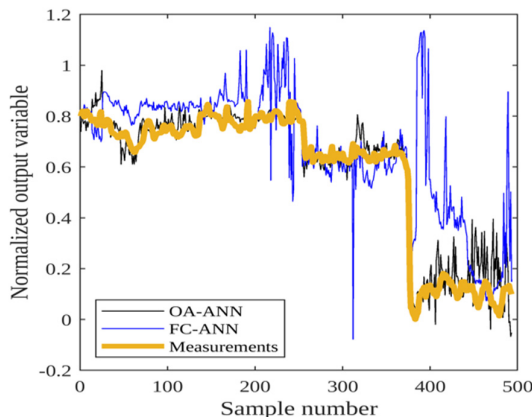


Figure 3: Test data performance of OA-ANN and FC-ANN.

4. Conclusion

The design and synthesis of ANNs is a challenging and important task. Ineffectively large networks provide significant test errors due to over-fitting. Test performance of the proposed approach is significantly better compared to the FC-ANN, despite using approximately 10% of the connections and half of the inputs only. Reduced input space is advantageous for the real-time operation when model update or sensor failures are under consideration. Furthermore, the OA-ANN is superior in various training ratios, more robust to extrapolation and prediction of different operating conditions. The heuristic MINLP solution algorithm provides a satisfactory network architecture and still has theoretical potential for development. Our current focus includes the implementation to actual plant.

5. Acknowledgements

This publication has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C245). However, the entire responsibility of the publication belongs to the owner of the publication.

6. References

- [1] M. Mangård, J. Kronqvist, and J. M. Böling, "Structural learning in artificial neural networks using sparse optimization," *Neurocomputing*, vol. 272, pp. 660–667, 2018.
- [2] Y. Zhou, G. G. Yen, and Z. Yi, "A Knee-Guided Evolutionary Algorithm for Compressing Deep Neural Networks," *IEEE Trans. Cybern.*, pp. 1–13, 2019.
- [3] Xie, H. Zhang, J. Wang, Q. Chang, J. Wang, and N. R. Pal, "Learning Optimized Structure of Neural Networks by Hidden Node Pruning with L1 Regularization," *IEEE Trans. Cybern.*, vol.50, no. 3, pp. 1333–1346, 2020.
- [4] V. Dua, "A mixed-integer programming approach for optimal configuration of artificial neural networks," *Chem.Eng. Res. Des.*, vol. 88, no. 1, pp. 55–60, 2010.
- [5] Pintaric, Z.N.; Kravanja, Z. The two-level strategy for MINLP synthesis of process flowsheets under uncertainty. *Comput. Chem. Eng.* 2000, 24, 195–201.
- [6] Wächter, A., and L. T. Biegler. On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming, volume 106. 2006.

Novel Tool for Selecting Surrogate Modeling Techniques for Surface Approximation

Bianca Williams and Selen Cremaschi*

Department of Chemical Engineering, Auburn University, Auburn, AL, United States

selen-cremaschi@auburn.edu

Abstract

Surrogate models are used to map input data to output data when the actual relationship between the two is unknown or computationally expensive to evaluate for several applications, including surface approximation and surrogate-based optimization. Many techniques have been developed for surrogate modeling; however, a systematic method for selecting suitable techniques for an application remains an open challenge. This work compares the performance of eight surrogate modeling techniques for approximating a surface over a set of simulated data. Using the comparison results, we constructed a Random Forest based tool to recommend the appropriate surrogate modeling technique for a given dataset using attributes calculated only from the available input and output values. The tool identifies the appropriate surrogate modeling techniques for surface approximation with an accuracy of 87% and a precision of 86%. Using the tool for surrogate model form selection enables computational time savings by avoiding expensive trial-and-error selection methods.

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

1. Introduction

Surrogate models are simplified approximations of more complex, higher-order models. They are used to map input data to outputs when the actual relationship between the two is unknown or computationally expensive to evaluate. Surrogate models are of particular interest where expensive simulations are used or when the fundamental relationship between the design variables and output variables is not well understood, such as in the design of cell manufacturing processes (Williams et al., 2020). Surrogate models can also be constructed for surrogate-based optimization when a closed analytical form of the relationship between input data and output data is not available or is not conducive for use in conventional gradient-based optimization methods. Several techniques have been developed for surrogate modeling, requiring a systematic approach for selecting which technique may be appropriate for an application.

Current standard practices for selecting which surrogate model form is appropriate rely on process-specific expertise. Numerous studies have been conducted to compare surrogate modeling techniques (Davis et al., 2017). However, most of these only evaluate a few models on a limited number of functions or applications. Recently, progress has been made in generalizing the process for selecting a surrogate model to approximate a surface by using meta-learning approaches to build selection frameworks (Cui et al., 2016; Garud et al., 2018). These frameworks provide “best” recommendations for surrogate modeling techniques based on the attributes calculated

from the data being modeled and avoiding expensive trial-and-error methods. Few of the developed meta-learning tools take model complexity into account, which can lead to overfitting, or consider that multiple models might perform similarly to the one identified as best.

This work aims to comprehensively investigate the performance of several different surrogate modeling techniques for approximating smooth, continuous functional relationships and to link that performance to the characteristics of the data being modeled. The performance metric used for evaluating how well the surrogate modeling techniques approximate surfaces is the adjusted R^2 , which considers both model accuracy and complexity. Simulated data was generated using a suite of optimization test functions. Data attributes were calculated based only on input and output values for each dataset to represent its overall behavior. Attributes that have the most influential relationships for predicting the adjusted R^2 were selected using feature reduction. These attributes were used as inputs to construct a Random Forest based tool to make predictions on the surrogate models' performance and provide recommendations for which surrogate modeling technique(s) may be most accurate for the dataset.

2. Computational Experiments

2.1 Test Functions

The test functions used to simulate data for constructing the surrogate models and the recommendation tool are from the Virtual Library of Simulation Experiments optimization test suite (Surjanovic & Bingham, 2013). The functions are divided by their shapes, which include the categories: multi-local minima (29 functions), bowl-shaped (31 functions), plate-shaped (9 functions), valley-shaped (12 functions), and other-shaped (18 functions) that do not fit into the other four categories. Functions with two (29 functions), four (20 functions), six (17 functions), eight (17 functions), and ten (16 functions) inputs were used.

2.2 Surrogate Model Performance Comparison

Input-output pairs were generated from each test function using three different space-filling sampling methods: Halton Sequence Sampling, Sobol Sequence Sampling, and Latin Hypercube Sampling (LHS). Data was generated at seven different sample sizes (50, 100, 400, 800, 1200, and 1600 samples), producing 693 total datasets. Eight surrogate modeling techniques were used for comparison: multivariate adaptive regression splines (MARS);(Friedman, 1991), random forests (RF);(Breiman, 2001) single hidden layer feed-forward artificial neural networks (ANN);(Haykin, 2009), extreme learning machines (ELM);(Haykin, 2009), Gaussian process regression (GP);(Rasmussen & Williams, 2005), support vector machines (SVM);(Drucker et al., 2002), Automated Learning of Algebraic Models using Optimization (ALAMO);(Cozad et al., 2014) and radial basis function networks (RBFN);(Gomm & Yu, 2000). Surrogate models were trained using the input-output pairs with each of the surrogate modeling techniques for the test functions. This process yielded 16,632 trained models. When necessary, the hyperparameters of each surrogate modeling technique (such as the number of hidden neurons for the neural network-based models and the number of trees in RF models) were optimized before training the models using ten-fold cross-validation. After the surrogate models were trained, the adjusted- R^2 values were calculated for each modeling technique-dataset pair.

2.3 Recommendation Tool Construction

Cui et al. (2016) and Garud et al. (2018) extract information from the datasets for use in their recommendation frameworks in the form of attributes. The attributes include common statistical measures, such as mean and standard deviation, gradient-based attributes, and attributes related to the extrema of the output values. We have defined additional attributes, including the first four statistical moments of the determinants of the estimated Hessian matrices of the datasets, and as the number of data points in the dataset, to use as potential inputs for predicting the model performance with the recommendation tool, resulting in a total of 40 attributes. The attributes aim to capture the overall behavior of the underlying model that generated the dataset. They were calculated for the datasets generated from the 99 test functions and used to construct the surrogate model recommendation tool.

A RF model was trained for each surrogate modeling technique to predict its adjusted- R^2 value using the identified attributes as inputs. Random forests are decision tree-based machine learning models, where the final output of the model is the average of the value predicted by every decision tree in the forest. Feature reduction was performed to determine which attributes had the most influence on the predicted output value for each modeling technique. Feature reduction techniques included linear and rank correlations (Zou et al., 2003) between the adjusted- R^2 value and the attributes, and the built-in feature selection method in RF models. In RFs, features are selected based on how well they improve the data separation at each decision node in each decision tree in the RF (Brieman, 2001). For each dataset, based on the adjusted- R^2 values, each of the surrogate modeling techniques was classified as either being recommended or not recommended for both the predicted and actual metric values. These classifications were compared and used to evaluate the quality of the selection recommendations.

3. Performance Metrics

The adjusted- R^2 value is used to assess the surrogate models' performance for surface approximation. The formula for calculating adjusted- R^2 (\hat{R}^2) is shown in Eq. (1).

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

In Eq. (1), R^2 is the R-squared regression coefficient, n is the number of data points in the training set, and k is the number of model parameters (or hyperparameters). The adjusted- R^2 takes into account both the surrogate model accuracy and complexity (Miles, 2005). Taking complexity into account is essential in ensuring that the model is not overfit as overfit models do not generalize well to new data. \hat{R}^2 values typically fall between zero and one, with an \hat{R}^2 of one indicating a perfect fit. However, with the adjustment for model size, adjusted- R^2 values can become negative.

The metrics used to evaluate the performance of the recommendation tool (i.e., the classification of surrogate modeling techniques given a dataset) are accuracy, precision (Sokolova & Lapalme, 2009), and the hit ratio (Cui et al., 2016). The accuracy is the percentage of recommendations that are correct. Precision is the probability that a model classified as recommended should actually be recommended. The hit ratio is the percentage of the time the model with the highest calculated adjusted- R^2 is included in the set of recommended models. All three performance metrics range from 0 to 100%. Monte Carlo cross-validation was used to evaluate the performance of the

recommendation tool with 100 Monte Carlo trials. Each trial had a test set size of 75, which was about 11% of the total simulated data.

4. Results and Discussion

4.1 Surrogate Model Performance

The surrogate modeling technique that yielded the model with the highest adjusted- R^2 value was selected as the “best” one. For each shape category, the number of times a technique was selected as best was tabulated. These tabulated values were divided by the total number of datasets in the category to calculate the fraction of datasets for which each surrogate modeling technique was selected as the best performing (Fig. 1). There was no significant difference in the adjusted- R^2 values among the three sampling methods. Therefore, only results for Sobol sequence sampling are shown here. For valley, bowl, and other-shaped functions, GP models provide the highest adjusted- R^2 . However, ALAMO and MARS models produce the highest adjusted- R^2 most frequently for bowl and multi-local minima-shaped functions, respectively. These results indicate that the underlying function shape has an effect on which surrogate modeling technique may be most appropriate for approximating a dataset. While in general, GP models may provide the most accurate approximation, specific shape characteristics may lead to another technique’s being more appropriate. It should be noted that although these results only reflect a single technique being selected with the highest adjusted- R^2 , in many cases, there were multiple techniques with values that were not significantly different than that of the highest adjusted R^2 value.

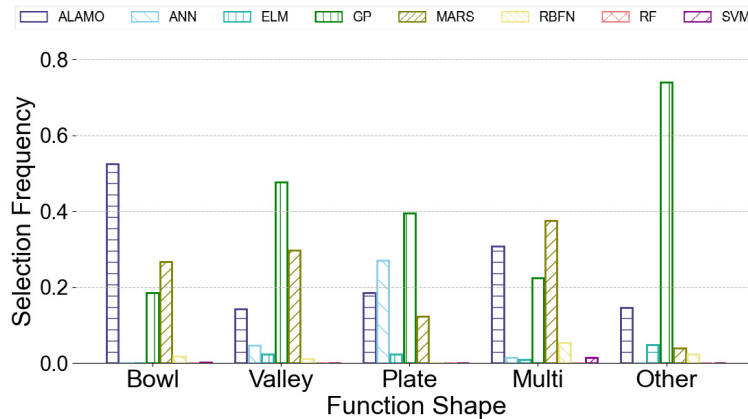


Figure 1- Percentage of datasets grouped by function shape for which each surrogate modeling technique had the highest adjusted- R^2

4.2 Attribute Selection for Adjusted- R^2 Prediction

From the comparison results, we can conclude that there is a relationship between the underlying shape of the surface being modeled and the performance of each of the surrogate modeling techniques. The minimum Mahalanobis distance (De Maesschalck et al., 2000) between any two points in the simulated dataset was moderately correlated to the calculated adjusted- R^2 of RF models, with a linear correlation coefficient of -0.58 and a rank correlation coefficient of -0.71. The position of the data points in the dataset and how close they are to each other may be correlated to the approximation of RF

models due to the need for the models to partition the design space of the surface when determining the decision nodes in each tree in the RF model. Data points that are closer together lead to smaller partitions and more accurate predictions.

For the feature selection by RF models, each technique had a different set of selected attributes for prediction. For ALAMO, ANN, RBFN, and SVM models, 18 different attributes were selected as important. Random forest models selected 19, 11, 20, and 17 attributes as important for ELM, GP, MARS and RF models, respectively. The attribute most commonly selected as being important for predicting the adjusted- R^2 was the minimum Mahalanobis distance between training points. Other commonly selected features include those related to the distributions of output values, specifically the relative size of the output distribution tails and the output distribution skewness, and the ratios of the average estimated gradient to the minimum and maximum estimated gradients for all of the data points in the dataset. These results suggest that the distribution and location of the sample points and the relative steepness and smoothness of the surface have a high level of influence on how well each of the surrogate models is able to approximate that surface.

For all of the neural network-based models (ANN, ELM, and RBFN) and RF models, the attribute selected with the highest importance was the percentage of the simulated data points that were located in the upper tail of the output distribution. The closely related attributes of the ratio of the upper and lower tail sizes and the skewness of the output value distribution were selected as most important for GP and MARS models, respectively. These attributes may have an effect on the accuracies of all these techniques as having data unevenly concentrated (or sparse) at the extreme values may skew models to predict more accurately in areas of data concentration and less so for other areas of the design space. For example, in the case of RF models, uneven tails could cause decision nodes in the model trees to split more frequently at the extremes of the output values while more finely split partitions are really needed elsewhere, such as where the gradients are steeper. For the neural network-based models, the on-off nature of the hidden layer nodes may make them more suitable for making accurate predictions for surfaces where large areas of the design space have similar output values, creating flat or nearly flat areas. The coefficient of variation (COV) was selected as the most important feature for the prediction of the performance of SVM models. The COV is inversely related to the signal-to-noise ratio of a surface (Wang et al., 2013). This attribute may be important for SVM model performance as the support vectors fitted in the model construction can easily become sensitive to noise as they are only dependent on a small set of the data used to train the model (Sabzekar et al., 2011). For ALAMO models, all of the selected attributes had roughly equal amounts of importance.

4.3 Recommendation Tool Performance

The selected attributes were used as inputs to train a RF model for the eight techniques to predict the adjusted- R^2 for a given dataset. Based on the predicted adjusted- R^2 value, the recommendation tool then classifies each of the surrogate models as being recommended or not for that dataset. This recommendation scheme allows for multiple similarly performing surrogate modeling techniques to be suggested for use in surface approximation. The selection tool identified which techniques should be recommended for the simulated datasets with an accuracy of 87%. The precision, or the probability that a recommended technique should actually be recommended, was 86%. The hit ratio, the percentage of time techniques that had the highest adjusted- R^2 for a dataset were included in its set of recommended models, was 80%.

5. Conclusions

Selecting an appropriate surrogate modeling technique depends on the characteristics of the dataset being modeled. We identified attributes of datasets that are appropriate for use in predicting the adjusted-R² value. Using these attributes, we have constructed a tool that can recommend surrogate modeling techniques for approximating a dataset with 87% accuracy and 86% precision. Future work on the tool will include expanding it to surrogate-based optimization recommendations and investigation of additional attributes and machine learning techniques to improve recommendation quality.

References

- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Cozad, A., Sahinidis, N. V., & Miller, D. C. (2014). Learning surrogate models for simulation-based optimization. *Aiche Journal*, 60, 2211-2227.
- Cui, C., Hu, M. Q., Weir, J. D., & Wu, T. (2016). A recommendation system for meta-modeling: A meta-learning based approach. *Expert Systems with Applications*, 46, 33-44.
- Davis, S., Cremaschi, S., & Eden, M. (2017). Efficient Surrogate Model Development: Optimum Model Form Based on Input Function Characteristics. In A. Espuna, M. Graells & L. Puigjaner (Eds.), *27th European Symposium on Computer Aided Process Engineering (ESCAPE 27)* (Vol. 40, pp. 457-462). Barcelona, Spain: Elsevier.
- De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50, 1-18.
- Drucker, H., Shahrory, B., & Gibbon, D. C. (2002). Support vector machines: relevance feedback and information retrieval. *Information Processing & Management*, 38, 305-323.
- Friedman, J. H. (1991). Multivariate Adaptive Regression Splines - Rejoinder. *Annals of Statistics*, 19, 123-141.
- Garud, S. S., Karimi, I. A., & Kraft, M. (2018). LEAPS2: Learning based Evolutionary Assistive Paradigm for Surrogate Selection. *Computers & Chemical Engineering*, 119, 352-370.
- Gomm, J. B., & Yu, D. L. (2000). Selecting radial basis function network centers with recursive orthogonal least squares training. *Ieee Transactions on Neural Networks*, 11, 306-314.
- Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.
- Miles, J. (2005). R Squared, Adjusted R Squared. In *Encyclopedia of Statistics in Behavioral Science*: John Wiley & Sons Ltd.
- Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning*, 1-247.
- Sabzevar, M., Yazdi, H. S., & Naghibzadeh, M. (2011). Relaxed constraints support vector machines for noisy data. *Neural Computing & Applications*, 20, 671-685.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45, 427-437.
- Surjanovic, S., & Bingham, D. (2013). *Virtual Library of Simulation Experiments*. In (Vol. 2018). Simon Fraser University.
- Wang, B., Goodpaster, A. M., & Kennedy, M. A. (2013). Coefficient of variation, signal-to-noise ratio, and effects of normalization in validation of biomarkers from NMR-based metabolomics studies. *Chemometrics and Intelligent Laboratory Systems*, 128, 9-16.
- Williams, B., Lobel, W., Finklea, F., Halloin, C., Ritzenhoff, K., Manstein, F., Mohammadi, S., Hashemi, M., Zweigerdt, R., Lipke, E., & Cremaschi, S. (2020). Prediction of Human Induced Pluripotent Stem Cell Cardiac Differentiation Outcome by Multifactorial Process Modeling. *Front Bioeng Biotechnol*, 8, 851.
- Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and simple linear regression. *Radiology*, 227, 617-622.

ProREFD: Tool for Automated Computer-Aided Refrigerant Design, Analysis, and Verification

Kornkanok Udomwong^a, Arisa Robin^a, Nichakorn Kuprasertwong^a, Orakotch Padungwatanaroj^a, Anjan Kumar Tula^b, Lingyu Zhu^c, Liyang Zhou^d, Binhui Wange, Shuhua Wang^e, Rafiqul Gani^{f*}

^a*PSE for SPEED Company Limited, 294/65 RK Office Park, Romklao Rd., Bangkok, 10520, Thailand*

^b*College of Control Science and Engineering, Zhejiang University, Hangzhou, China*

^c*College of Chemical Engineering, Zhejiang University of Tehnology, Hangzhou, China*

^d*Juhua Group Corporation, Quzhou, China*

^e*Zhejiang Juhua Novel Material Research Institute Co. Ltd, Hangzhou, China*

^f*PSE for SPEED, Skyttemosen 6, Allerød, DK-3450, Denmark*

* *Corresponding author: rgani@pseforspeed.com*

Abstract

ProREFD is a computer-aided software for refrigerant design, analysis and verification, which consists of six tools that are suitable for different aspects of refrigerant design and verification. The six tools are the following: three databases plus a search engine; pure component properties estimation; mixture properties estimation; refrigerant selection and/or design; refrigeration cycle simulation; and modelling. The database tool consists of a database of measured pure component properties of especially selected refrigeration related compounds; a database of measured vapor-liquid equilibrium data of binary mixtures; and a database of measured binary azeotrope data. The property estimation tools employ a collection of especially developed and tested property models suitable for refrigeration processes. The refrigerant selection and/or design tool helps the user to select and/or design a refrigerant (single compound or mixture), given a set of target properties. The refrigeration cycle simulation performs simulation for the selected refrigerant and specified cycle operation conditions. The modelling tool is used to revise and/or add new property or process models. The paper presents the different implemented tools in ProREFD and highlights its use through several illustrative examples.

Keywords: Refrigerant, Refrigerant design, Properties, Refrigerant cycle simulation

1. Introduction

Refrigeration systems, whether for industrial or home applications, are very common since they perform the important operation of maintenance of temperature at a desired value. The selection and testing of any refrigerant involve many steps, such as identifying a suitable chemical (single compound or mixture) through evaluation of its properties, testing its efficiency in the actual refrigeration operation, matching of the desired operational specifications and many more. Moreover, due to increased health, safety and environmental concerns, the refrigerants selected for application must satisfy environmental, safety and sustainability constraints. Therefore, there has been a continued interest in finding new refrigerants that are significantly better than the currently available ones (Wei, 2007; Sahinidis et al., 2003). In this case, a computer-aided tool as a virtual reality that can systematically guide the engineer through the steps of

refrigerant selection and/or design as well as verification would be very useful, especially, if it can quickly and reliably identify promising candidates and thereby reduce time and cost for refrigerant development by focusing on experiments only at the last step to verify the performance of a small set of promising candidates. With these objectives in mind, the ProREFD software has been developed as a computer-aided hybrid data-model based tool suitable for refrigerant design, analysis and verification by extracting already developed tools from ProCAPD (Kalakul et al., 2018) and OptCAMD (Liu et al., 2019) but including especially developed property models for refrigerant molecules and their associated measured data as well as refrigeration process model. The goal of ProREFD is to mimic the steps typically employed in reality mode for refrigerant selection and verification. It should also be able to provide typical analysis tools such as databases, property estimation (pure component and mixture properties) tools that help with the virtual selection and verification steps as well as virtual analysis of available refrigerant candidates. The objective is not to replace experiments but to guide and focus them. This paper gives an overview of the ProREFD software, highlighting the work-flow and data-flow of its main functions as well as a short description of its six main tools. The use of the different tools of ProREFD is highlighted through illustrative examples.

2. Overview of ProREFD and its embedded tools

ProREFD is a systematic computer-aided tool, which combines the main steps of any refrigeration system study by providing the necessary functions (tools) and their associated models and sub-tools. The six tools of ProREFD, highlighted in Figure 1 (overview) and Figure 2 (data-flow and work-flow) are briefly described below.

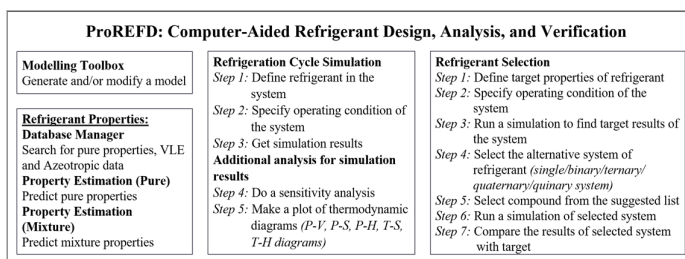


Figure 1: Overview of ProREFD

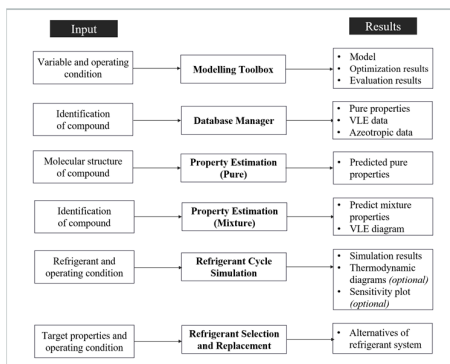


Figure 2: Information flow of each feature in ProREFD

2.1 Tools of ProREFD

Tool-1: Modelling Toolbox (MoT) is a mathematical modelling solver, which consists of in-house algebraic solvers, integrators, ordinary differential equation solvers and optimizers. MoT (Fedorova et al., 2015) is an equation-based simulation software that allows users to perform simulation processes without having to write any source code and has an option to convert models into COM-objects for use with external software.

Tool-2: As shown in Fig 1, the Database Manager is a tool for searching and retrieving of the refrigerant properties (pure compounds and/or mixtures). The search engine allows forward search (given a compound identification, show all available properties) and an advanced reverse search (given a set of target properties, show all compounds that match the property targets). This feature is very useful in the selection of a refrigerant based on measured properties. The database tool also includes a collection of measured binary mixture data in terms of PTxy-data from phase equilibrium experiments as well as a collection of binary azeotropic data. Table 1 lists the available data related statistics for the three databases.

Table 1: Statistics of available data in the Database Manager

Pure properties statistic		VLE data statistics				Azeotropic data statistics		
Compound	Known refrigerants	Binary pairs	Data sets	Data points	Compounds	Binary pairs	Data points	Compounds
1766	306	1340	9304	113887	214	795	2130	238

Tool-3: Property Estimation (Pure), is a property estimation toolbox for pure compounds (see Fig 1). This toolbox allows users to predict the properties of 60 different pure component properties. As refrigerant compounds are usually small molecules, the well-known group contribution property estimation methods that are suitable for larger molecules have been found to be not suitable. Therefore, a new set of property models for a basic set of properties (normal boiling point and critical properties) have been developed (Kuprasertwong et al, 2021). Using these properties, other pure component properties, such as, heats of vaporization, heat capacities, densities, etc., are computed through validated models. As environmental properties of refrigerants are important, properties such as ozone depletion potential, global warming potential, auto-ignition temperature, etc., are also estimated. This toolbox is integrated with the modelling toolbox so that model parameters can be fine-tuned with available measured data.

Tool-4: Property Estimation (Mixture) is the toolbox (see Fig. 1) that provides the predicted properties of refrigerant (organic) mixtures that are needed for refrigeration system design and analysis. For a specified binary or multicomponent mixture, this toolbox predicts, as highlighted in Fig 2, phase equilibrium properties (Vapor Liquid Equilibrium), saturated conditions (bubble/dew points) as well as properties for a given phase such as density, viscosity, liquid or vapor enthalpies and heats of vaporizations. Also, for this toolbox, a collection of validated in-house mixture property models is available (correlations, cubic equations of state, PC-SAFT, etc.).

Tool-5: Refrigeration Cycle Simulation is a toolbox that allows users to perform simulations of a refrigeration cycle comprising of evaporator, compressor, condenser and expansion valve units connected in a single-loop (see Fig 1 for more details). The refrigerant can be a single molecule or a mixture of up to 5 compounds. Before simulation, the tool checks if the given operating conditions and the properties of the refrigerant satisfy a set of consistency rules. This toolbox also computes, employing in-house

equations of state, thermodynamic diagrams (such as pressure-enthalpy diagrams) and embeds the cycle operation into these diagrams. Another useful feature of this toolbox is a sensitivity analysis option, where the sensitivity of compound properties or operational variables to the calculated Coefficient of Performance (COP) are checked. More details of the process models are given in (Kuprasertwong et al., 2021).

Tool-6: Refrigerant Selection is the toolbox that helps to design, select and optimize refrigerants (single molecules or mixtures) for a specific application based on user defined requirements (target properties), as highlighted in Fig 1). Users can select and compare the feasible alternatives based on their target properties, additional properties (not used as targets) as well as refrigeration cycle performance. More details on the refrigerant selection method is given in (Kuprasertwong et al, 2021). Note that exact matches of all Set-1 target properties are possible, while for Set-2 target properties they are satisfied within the defined bounds.

3. Application Examples

In this section, two main toolboxes of ProREFD, which are Refrigerant Cycle Simulation and Refrigerant Selection including comparison of calculations with external software are highlighted through illustrative examples.

3.1. Application example 1: Verify the refrigeration cycle operation of R-134a

In this case, 1,2,2-Tetrafluoroethane (R-134a) is used in a refrigeration cycle operating with room temperature and ambient temperature of 298.15 and 303.15 K, respectively. The evaporator pressure and condenser pressure are set to be 3.5 and 15 bar with 0.9 isentropic efficiency of the compressor. The R-134a mass flow rate is 180 kg/h. The expansion valve is assumed to be an isenthalpic system. Using the properties of R-134a (retrieved from the database), the consistency rules are checked and are found to be satisfied. Next, the cycle simulation is performed and the COP is determined. Next, the thermodynamic diagram is calculated and the operation of the cycle conditions is embedded into it. Finally, by perturbing the operating conditions (such as condenser pressure or evaporator pressure), the improvement of COP is checked. Figure 3 shows several screen shots from ProREFD. On the left-hand side, the refrigeration cycle simulation is shown together with the specified and calculated values. On the right-hand side (top), the thermodynamic diagram is shown and on the right-hand side (below), the sensitivity analysis plots are shown. From the thermodynamic diagram, it can be seen that decreasing the condenser pressure or increasing the evaporation pressure can increase the COP. The sensitivity analysis shows that T_b (normal boiling point) is the most sensitive property of the refrigerant and the critical pressure, although used to check the consistency rules, is not sensitive to COP.

3.2. Application example 2: replace R-134a

In this example the replacement of R-134a with other refrigerants (single, binary mixture, and ternary mixture) are highlighted. For verification, the refrigeration cycle is considered to operate at the same conditions as given in section 3.1. Results for selected refrigerant alternatives are highlighted in Figure 4 (screenshots from ProREFD). Figure 4 shows that the Refrigerant Selection toolbox allows users to replace a specified refrigerant with other

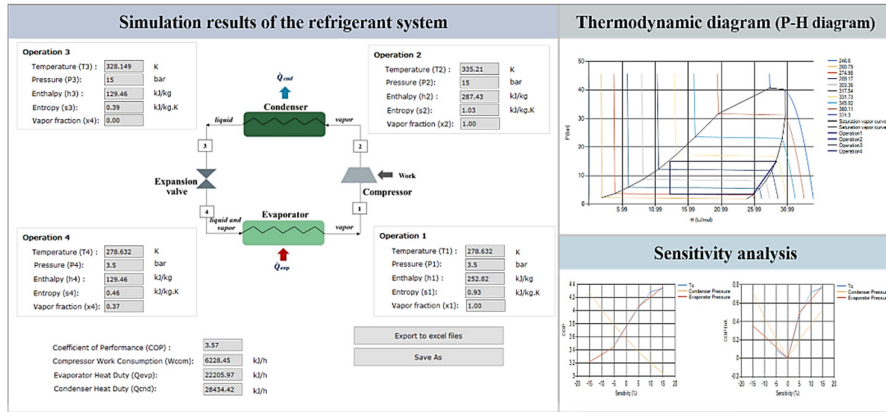


Figure 3: Results from Refrigeration Cycle Simulation function

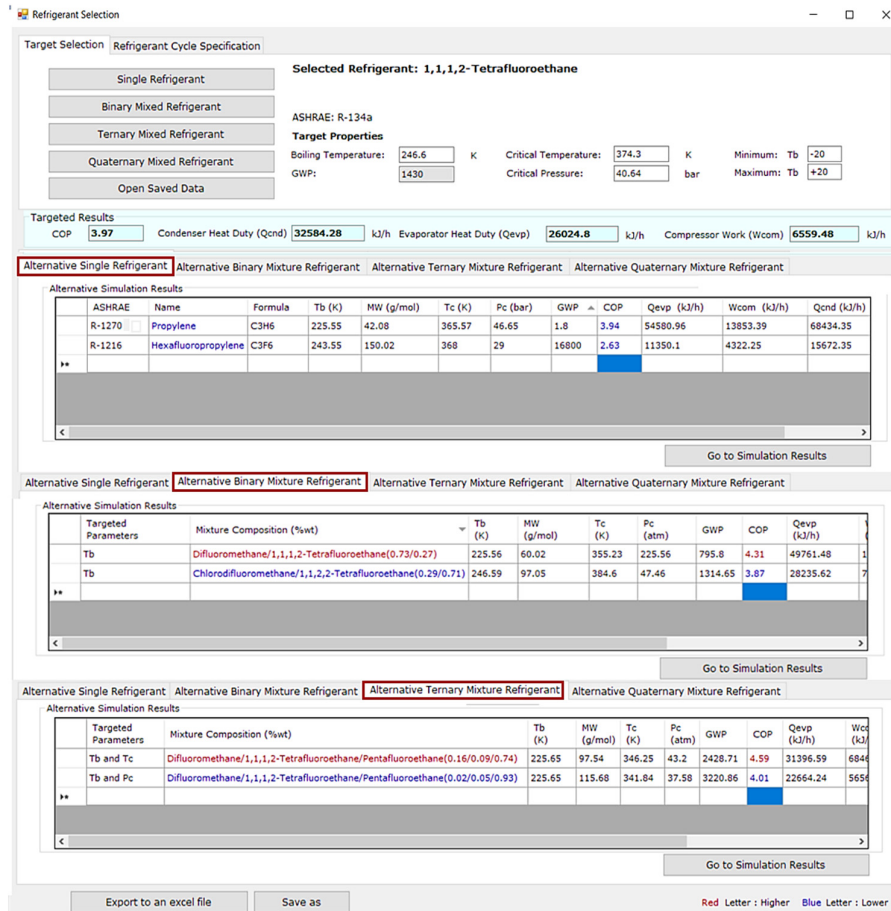


Figure 4: Results from refrigerant selection function

single or mixture refrigerants. Here, the reverse database search and/or molecular-mixture design methods (Kalakul et al., 2018) are employed to determine the alternatives. Furthermore, users can compare each designed alternative and optimize it to achieve the most efficient refrigeration cycle. See also Sahinidis et al., (2003) for a similar example.

3.3. Application example 3: ProREFD refrigeration cycle model validation

The simulated refrigeration cycle results with R-134a as the refrigerant, which are given in section 3.1, are compared and cross-validated with simulations of the same refrigeration cycle at the same conditions and for the same refrigerant with two external process simulators (PRO/II and AspenPlus). The simulated results given in Table 2 confirm that at least the three software tools give approximately similar results.

Table 2: Comparison of results from ProREFD with commercial simulators

Parameters	Unit	Calculation Source		
		ProREFD	External Simulator-1	External Simulator-2
COP		3.57	3.59	3.60
Evaporator Heat Duty (Q_{evap})	kJ/h	22,206	22,061	22,132
Compressor Work (W_{comp})	kJ/h	6,228	6,148	6,154
Condenser Heat Duty (Q_{cond})	kJ/h	28,434	28,210	28,286

4. Conclusion

The capabilities of ProREFD have been highlighted through illustrative examples. The Refrigerant Selection toolbox has the potential to find appropriate candidates that may not exist in any database of refrigerants, especially mixed refrigerants. For Refrigeration Cycle Simulation, the features of embedding the simulation results on the thermodynamic diagrams help to identify how to change the operating conditions to improve the process performance. The sensitivity analysis helps to identify the parameters of the refrigerant that can influence the process operation the most. Hence, ProREFD is a promising software tool that can play a leading role in refrigeration process development by helping to find promising candidates and more sustainable solutions rapidly, reliably and efficiently. ProREFD is unique as no other tool that have all the functions integrated in one software could be found. Due to unavailable property data, currently, the number of compounds with boiling points below 250 K is not many. Consequently, the number of new refrigerants that can be generated is also limited. To overcome this issue, current and future work is increasing the number of small molecules with 1, 2 and 3 carbon atoms so that truly novel alternative refrigerants can be found.

References

- M. Fedorova, G. Sin, R. Gani, 2015, *Computers & Chemical Engineering*, 83, 232-247
- S. Kalakul, L. Zhang, Z. Fang, H.A. Choudhury, S. Intikhab, N. Elbashir, M.R. Eden, R. Gani, 2018, *Computers & Chemical Engineering*, 116, 37-55.
- Q. Liu, L. Zhang, L. Liu, J. Du, A.K. Tula, M. Eden, R. Gani, 2019, *Computers & Chemical Engineering*, 124, 285-301.
- N. Kuprasertwong, O. Padungwatanaroj, A. Robin, K. Udomwong, A. K. Tula, L. Zhu, L. Zhou, B. Wang, S. Wang, R. Gani, 2015, *Computer Aided Refrigerant Design: New developments*, ESCAPE-31 Proceedings (submitted).
- N. Sahinidis, M. Tawarmalani, M. Yu, 2003, *AIChE Journal*, 49, 1761-1775.
- J. Wei, *Product Engineering: Molecular Structure and Properties*, Oxford University Press, New York, USA, 2007

Event driven analysis to enhance model calibration of experiments with high offline sampling rates

M. Adnan JOUNED^{a,b}, Julian KAGER^a, Christoph HERWIG^a, Tilman BARZ^b

^a ICEBE, TU Wien, Gumpendorfer Straße 1a 166/4, 1060 Wien, Austria

^b AIT Austrian Institute of Technology GmbH, Giefinggasse 2, 1210 Wien, Austria

Abstract

The use of autosamplers connected to high throughput analytical devices allows for a high sampling frequency and analytics with reduced manual labor, leading to better process characterization (Maurer et al., 2015; Hofer et al., 2020). Increased sampling often leads to a significantly increased information content in the generated data. However, in combination with miniaturized or lab-scale reactors, the effect of volume change by the frequent sampling becomes challenging.

Sampling leads to fast, almost instantaneous volume changes in the reactor. This process represents a discontinuous behavior in the continuous-time kinetic model. A commonly applied so-called “time stepping” method ignores the discrete behavior and relies on the solver’s local error estimator to solve continuous-time differential equations. Therefore, in regions where discontinuities of the solution or its derivative occur, the method may fail to deliver an accurate solution. An alternative is the so-called “event driven” method, which explicitly accounts for discontinuities in the model. During the solution (integration) of the model, the method accurately locates time points, where discontinuities occur (event detection), e.g., volume changes and continues the solution process (Dieci and Lopez, 2012). It is well-known that proper handling of discontinuities can significantly increase the models' accuracy and reduce simulation runtime (Alsoudani, 2016). Still, bioprocess developers often ignore or tolerate discontinuities when implementing models in simulation software. Hence, this contribution highlights the importance of a proper handling of discontinuities in a relevant common case study for bioprocess development. Results are presented for the determination of kinetic parameters of Sonnleitner and Käppeli’s (1986) *Saccharomyces cerevisiae* growth model on a lab-scale fed-batch process with fast volume changes caused by frequent sampling. It turns out that the “time stepping” method misses several volume changes. In contrast, the “event driven” method does not. Accordingly, the “event driven” method yields accurate model predictions which are not affected by the reactor volume’s prediction error and thus improves the model calibration, lowers parameters uncertainty, and supports a robust convergence to the best fitting model parameters.

Keywords: parameter estimation, kinetic modelling, event driven modelling

1. Introduction

The acceleration of bioprocess development for decreasing time to market (TTM) of biopharmaceuticals has not only been known since the COVID-19 pandemic. For this purpose, mechanistic models represented by systems of ordinary differential equations (ODE’s) are indispensable tools for bioprocess design, monitoring, and control (Narayanan et al., 2020). However, the challenge of underfitting these models is present when an insufficient number of observations is used for model calibration. This problem

usually leads to correlated model parameters, which could hinder clear interpretations of the results. Autosamplers and automated analytical devices allows for tackling this challenge by increasing the sampling frequency without additional manual labor (Maurer et al. 2015; Hofer et al. 2020). Frequent sampling inevitably leads to significant volume changes in the reactor, which besides being a bottleneck in miniaturized systems, needs to be properly considered during modelling steps.

Volume changes of an ideal stirred tank reactor are usually modeled by considering mass balance equations. The changes in the volume over time is calculated as the difference between input and output flow rates (Doran, 2012). It is widely accepted to consider sampling volume F_{Sampling} as a part of the flows that are leaving the reactor (Rocha, 2003; Callewaert and De Vuyst, 2000; Kager et al., 2020). Hence, volume changes are written usually as (where V : volume, F : liquid mass flow rates):

$$\frac{dV}{dt} = F_{\text{in}} - F_{\text{out}} = F_{\text{Feed}} + F_{\text{Base}} + F_{\text{Acid}} - F_{\text{Sampling}} \quad (1)$$

In usual fed-batch fermentation, the measured flow rates (except F_{Sampling}) usually show comparatively smooth curves. These curves are represented by discrete signals and can be transformed to smooth functions with relatively little effort, e.g., by applying a smoothing filter and by subsequent interpolation using piece-wise spline interpolation. Avoiding discontinuities on the right-hand-side of eq. 1 can significantly improve the efficiency and accuracy of its solution (Alsoudani, 2016). F_{Sampling} represents strongly discontinuous curves defined by (negative) pulse signals. A transformation of these sharp peaks by smoothing is not a viable solution. The result is highly nonlinear terms, which would require an (inefficient) dense time grid for accurate integration of eq. 1.

In the “time stepping” method (TSM), which is implemented by using standard ODE solvers, the sampling volume is calculated by the integration of eq. 1 considering a (negative) pulse signal. The method relies on the solver error estimator to determine the step size (Dieci and Lopez, 2012). Hence, in regions where discontinuities of the solution or its derivative occur, i.e., sampling timepoints, there is a probability to miss certain events such as sampling volume, causing inaccurate volume calculation. This is exemplarily shown in Figure 1. This results in a wrong volume mass balance and therefore affects subsequent calculations. The extent of the errors resulting from improper handling of samples volumes depends on the ratio of the sampled volume to the reactor volume. The errors are expected to have a bigger influence on smaller platforms, e.g., miniaturized bioreactor systems (< 0.3 L), where often no reactor volume measurement is available.

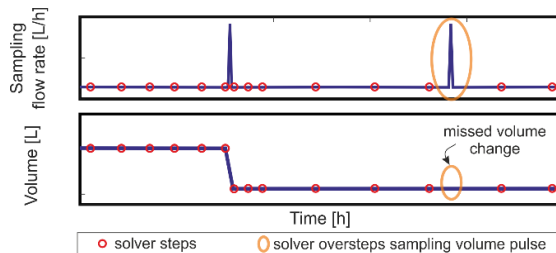


Figure 1: Modeling of volume changes due to sampling. Sampling volumes are considered by a “sampling” flow rate represented by a flow pulse signal. The time stepping method (TSM) does not accurately track the sudden changes in the flow pulse signal. The ODE solver steps miss the second pulse (overstepping problem), which has an immediate effect on the volume.

In contrast, the “event driven” method (EDM) accurately locates time instances (so-called events) where instantaneous sampling happens. At these time points, the integration is stopped, the volume is updated, and the integration is restarted with the updated volume as initial condition. By this, the EDM efficiently prevents missing any sample, leading to a correct and robust volume balance during model simulations.

It is noted that established simulation software like *gProms* and *Modelica/Dymola* use built-in routines to automatically detect discontinuities, locate events, and restart integration (Process Systems Enterprise Limited, 2013; Dynamic Modeling Laboratory, 2004). However, in low-level modeling languages such as the frequently used *MATLAB* (*ODE Suite*, 2020) or *Python* (*SciPy* package, 2020), the proper handling of discontinuities needs tailored solutions and special programming efforts as the ODE solvers provide only the basic functionalities by the so-called “event functions”.

2. Materials and Methods

2.1. Cultivation process and Sampling

A *S. cerevisiae* fermentation process is considered as an experimental case study. Samples were withdrawn from the fermentation medium at irregular time intervals either by-hand (10 samples of roughly 20 ml per sample) or automatically using (Numerica from Securecell) autosampler (20 samples of roughly 7 ml per sample) and distributed along the time of the experiment. The reactor's initial volume was 1.5 liter. The experiment consists of a batch and a fed-batch phase with different feed regimes.

2.2. *S. cerevisiae* fermentation model

The model from (Sonnleitner and Käppeli, 1986) considers the growth on glucose and ethanol substrates. It describes fermentative and oxidative growth based on all forms of biomass specific intake ($q_s^{oxidative}$, $q_s^{reductive}$, $q_{ethanol}$) using ($Y_{biomass/glucose}^{oxidative}$, $Y_{biomass/glucose}^{reductive}$, $Y_{biomass/ethanol}$) yields parameters. Total growth is written as:

$$\mu_{total} = Y_{biomass/glucose}^{oxidative} \cdot q_s^{oxidative} + Y_{biomass/glucose}^{reductive} \cdot q_s^{reductive} + Y_{biomass/ethanol} \cdot q_{ethanol} \quad (2)$$

Mass balances equations are (x: biomass, s: glucose, e: ethanol, F: flow rate)

$$\begin{aligned} \frac{dC_s}{dt} &= -(q_s^{reductive} + q_s^{oxidative}) \cdot C_x - \frac{F_{in}}{V} \cdot C_s + \frac{F_{Feed}}{V} \cdot C_{s,in} \\ \frac{dC_e}{dt} &= (q_s^{reductive} - q_e) \cdot C_x - \frac{F_{in}}{V} \cdot C_e \\ \frac{dC_x}{dt} &= \mu_{total} \cdot C_x - \frac{F_{in}}{V} \cdot C_x \end{aligned} \quad (3)$$

2.3. Time-stepping method (TSM) versus event-driven method (EDM)

Eq. 1 is used to model volume changes in both methods. In EDM, $F_{Sampling}$ is omitted from eq. 1, and an external algebraic equation is used instead to account for volume changes. Figure 2 illustrates the working principles of the TSM and EDM. In the EDM, to accurately locate each sampling timepoint t_s , the simulation time interval is separated into k sub-intervals, where for each sampling point 1, 2, 3 ... K, the volume is corrected outside the ODE system using the additional algebraic equation; $V = V - \Delta V$. Following

the integration is restarted with the updated volume. In TSM, the sampling volume rate is integrated into eq. 1 without special treatment. The same initial conditions applied to both methods.

2.4. Convergence analysis of parameter estimation

To highlight the negative effects of improper handling of sampling volume, the model predictions are fitted to the experimental data. The performance of the fitting is assessed by applying either TSM or EDM. The following parameters

$\theta = [q_s^{max}, q_{O_2}^{max}, Y_{biomass/glucose}^{oxidative}, Y_{biomass/glucose}^{reductive}, Y_{biomass/ethanol}]$ are selected for estimation based on the local parameter sensitivities and

an identifiability analysis based on importance ranking (López et al., 2013). The robustness of both methods is assessed through a Monte Carlo approach. 500 uniform distributed initial parameter guesses are generated over a +/- 25% interval around the literature's nominal values (Sonnleitner and Käppeli, 1986). For each parameter realization, an estimation problem was solved using *MATLAB R2017b* "ODE15s" solver and the nonlinear fitting algorithm "lsqnonlin". Normalized residual sum of squares (NRSS) between model predictions and measurements was used as an objective function.

3. Results and Discussion

Figure 3 (right) shows a visual comparison of one simulation run to highlight the differences in the calculated volume using both methods. It can be noticed that TSM oversteps certain sampling times and consequently doesn't update volume correctly. The wrong volume affects the other model states (concentrations) described in eq. 3 and displayed in Figure 3 (right) for two identical model simulations. The error becomes more pronounced towards the end of the simulation as its effect accumulates over time.

In TSM, to account for sampling volume changes correctly, solver steps must exactly locate sampling times. However, TSM locates solver steps based on the integrator's local error estimator, which is a tool to control approximation error at each step. If the solver oversteps a sampling time interval, the local error estimator does not indicate an approximation error. Consequently, sampling instances are located by chance. Therefore, by evaluating the model at slightly different initial conditions or with slightly different parameter values, the solver may overstep very different sets of sampling timepoints. The accuracy of model simulations might be strongly affected. This is also critical for model parametrization, as the deviations in model volume and predictions are different at each optimization run, which adds artificial noise to the optimization problem. This is shown in Figure 3 (left).

In contrast, EDM accurately locates all sampling events and uses an external algebraic equation to correct the volume and to restart the integration at each sampling timepoint. By doing this, it suppresses the noise and provides accurate and reproducible model simulations. This behavior is illustrated in Figure 3 (left), which shows the optimization

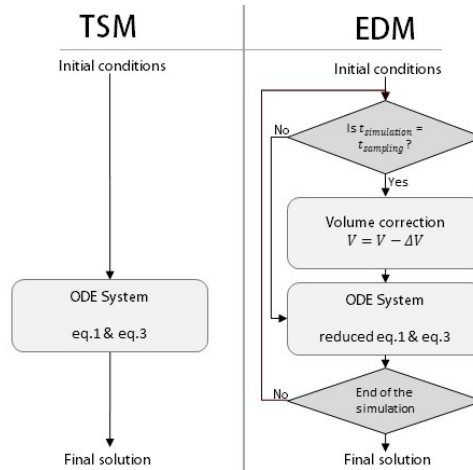


Figure 2: TSM and EDM for handling instantaneous volume changes because of sampling.

surface for two selected parameters. The noisy nature of the surface in TSM reduces the effectiveness of the gradient-based optimizer to seek an optimum. Tables 1 & 2 highlight this observation by comparing the means of prediction errors of the calibrated models obtained by both methods. The two-tailed p-value of the t-test is less than 0.0001, indicating a strong statistical significance. The results clearly indicate a higher probability to obtain better predictions using EDM and to converge to the best fitting parameter estimates. Reducing the solver tolerance in TSM might reduce these effects but not completely avoid them, as the error estimator still does not directly address the underlying reason of the problem, aside from increasing the computational cost.

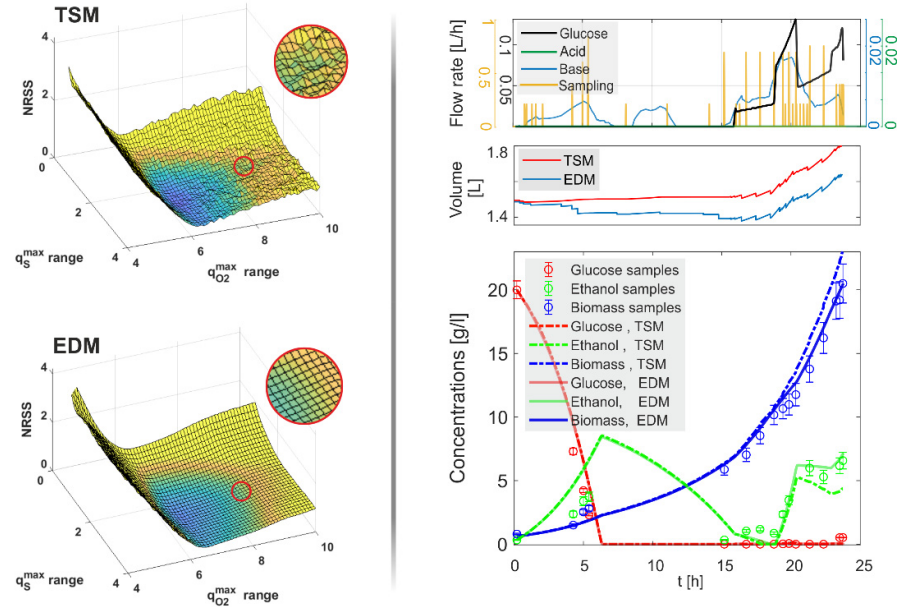


Figure 3: Left) States, flow rates and volumes calculated by the same model and parameters using EDM and TSM. Using TSM certain sampling times are overstepped. EDM delivers accurate results by accurately locating sampling times. Right) Optimization (NRSS) surfaces are smoother in case of EDM compared to standard TSM, allowing for a better convergence and efficiency of gradient-based optimization algorithms.

Table 1 & 2: Results of Monte Carlo procedure ($n=500$). EDM shows a statistically significant lower mean of the model prediction errors.

		<i>TSM</i>	<i>EDM</i>			
<i>Mean of prediction error</i>		3.1	2.3			
<i>Standard deviation (σ)</i>		1.65	1.35			
		θ_1	θ_2	θ_3	θ_4	θ_5
<i>Mean of Estimated. θ</i>	TSM	3.28	6.81	0.46	0.05	0.70
	EDM	3.30	6.56	0.44	0.05	0.69
<i>95% confidence interval</i>	TSM	0.97	2.50	0.16	0.018	0.24
	EDM	0.92	2.25	0.14	0.016	0.24

4. Conclusion

Using EDM to consider sampling volume in *S. cerevisiae* fermentation models delivers more reproducible and accurate model predictions. Moreover, for model calibration, the quantitative analysis reveals a significantly improved convergence of the parameter estimation algorithm. Accordingly, EDM results are obtained with a lower prediction error and parameters uncertainty. The reason is a smoother optimization surface that leads to more robust convergence to the best fitting model parameters.

EDM is ready-for-use in high-level symbolic modelling languages such as *gProms* and *Modelical/Dymola*. In low-level languages, e.g., *MATLAB* or *Python*, EDM can be implemented with a reasonable effort using available initial value solvers with event detection. Hence, this method is simple to implement for fermentation processes with high sampling rates and platforms where explicit weight measurements are crucial but not available or hard to obtain, e.g., mini multi-bioreactor systems. This will allow for more consistent results, resulting in reduced iterations in bioprocess development and, therefore, a decreased time to market (TTM) of biopharmaceuticals.

Acknowledgment

This work was partially funded by the Austrian Research Funding Association (FFG) within the program Bridge 1 in the project "AdaMo" (No. 864705).

References

- Alsoudani. 2016. *Discontinuities in Mathematical Modelling : Origin , Detection and Resolution*. University College London Department.
- Callewaert & De Vuyst. 2000. "Bacteriocin Production with *Lactobacillus Amylovorus* DCE 471 Is Improved and Stabilized by Fed-Batch Fermentation." *Applied and Environmental Microbiology* 66 (2): 606–13.
- Dieci & Lopez. 2012. "A Survey of Numerical Methods for IVPs of ODEs with Discontinuous Right-Hand Side." *Journal of Computational and Applied Mathematics* 236 (16): 3967–91.
- Doran. 2012. *Bioprocess Engineering Principles: Second Edition*. Bioprocess Engineering Principles: Second Edition. Vol. 9780080917. Academic Press.
- Dynamic Modeling Laboratory. 2004. "Dymola." *Dynasim AB*. Lund, Sweden: Dynasim AB.
- Hofer, Kroll, Barmettler, & Herwig. 2020. "A Reliable Automated Sampling System for On-Line and Real-Time Monitoring of CHO Cultures." *Processes* 8 (6): 637.
- Kager, Tuveri, Ulonska, Kroll, & Herwig. 2020. "Experimental Verification and Comparison of Model Predictive, PID and Model Inversion Control in a *Penicillium Chrysogenum* Fed-Batch Process." *Process Biochemistry* 90 (March): 1–11.
- López, Barz, Peñuela, Villegas, Ochoa, & Wozny. 2013. "Model-Based Identifiable Parameter Determination Applied to a Simultaneous Saccharification and Fermentation Process Model for Bio-Ethanol Production." *Biotechnology Progress* 29 (4): 1064–82.
- Maurer, Skerker, Arkin, Miller, Biksacky, & Huether-franken. 2015. "Automated Bioreactor Sampling – Process Trigger Sampling for Enhancing Microbial Strain Characterization."
- Narayanan, Luna, von Stosch, Cruz Bournazou, Polotti, Morbidelli, Butté, & Sokolov. 2020. "Bioprocessing in the Digital Age: The Role of Process Models." *Biotechnology Journal*. Wiley-VCH Verlag.
- Process Systems Enterprise Limited. 2013. "Model Developer Guide." London, UK. <http://www.psenterprise.com>.
- Rocha. 2003. "Model-Based Strategies for Computer-Aided Operation of a Recombinant E . Coli Fermentation." Braga: Escola de Engenharia Universidade do Minho.
- Sonnleitner & Käppeli. 1986. "Growth of *Saccharomyces Cerevisiae* Is Controlled by Its Limited Respiratory Capacity: Formulation and Verification of a Hypothesis." *Biotechnology and Bioengineering* 28 (6): 927–37.

Process System Engineering Tool Integration in the Context of Industry 4.0

Manuel Rodríguez Hernández*, Ismael Díaz Moreno

Universidades Politécnica de Madrid, José Gutierrez Abascal 2, Madrid 28006

manuel.rodriguez@upm.es

Abstract

The objective of this work is to have a model (or series of models) of a process plant that is available during the whole lifecycle of the system (process + product). To achieve this goal, we use a Model Based Engineering approach. In this work we will show how very different tools can be integrated and applied for a single process, we will illustrate (as a proof of concept) the integrations of the following tools (and languages): Aspen, Excel, MATLAB, and surrogate models. This will be demonstrated using Phoenix Integration software that provides mechanisms to automate the execution of different simulation and modelling tools in a single simulation workflow. In this integration, besides the tool integration, we will show how a superstructure optimization using an Aspen model (rigorous model) can be implemented and the simulation and optimization of a process composed of some rigorous units along with some surrogate units. The process used is the production of ethylbenzene from benzene and ethylene.

Keywords: Tool integration, model based engineering, process integration

1. Introduction

Industry is in constant evolution. Nowadays we are in what is called the fourth industrial revolution, Industry 4.0 (Schwab, 2016 and Kagermann et al., 2013). Its target is to make more efficient and flexible plants, reduce times and costs of projects and products lifecycle. This target has a strong basis in the use of digital twins. In the case of process plants there are a lot of specific software packages that are used by different agents (operators, maintenance workers, engineer, etc) in all the phases of the plant from planning to decommissioning this generates numerous documents and data that are prone to have inconsistencies (Fillinger et al., 2019) which can have an important impact on the plant in several aspects like safety or economic performance (Herzig, 2014). Several existing initiatives address the data transfer between tools, like Cape Open Laboratories Network (CO-LaN, 2018) for process simulators or the Namur initiative for data exchange between P&IDs and process control systems (NAMUR, 2018). Most of these approaches lack of a holistic view and focus on specific phases in the lifecycle. A Model Based Systems Engineering approach is needed to cover all the development phases, and to guarantee the traceability and consistency along the complete plant time span (Friedenthal, 2014). Several software frameworks exist for MBSE like Syndeia provided by Intercax (Bajaj et al., 2016) or ModelCenter provided by Phoenix Integration. In this work we use the latter to show tool integration in different applications following a Model Based Engineering approach, the ongoing work is to extend this approach to cover from the requirements of the process/product to its conception, design and operation.

2. Phoenix Integration ModelCenter

ModelCenter is a software developed by Phoenix Integration. It is an environment for Model Based Engineering (ModelCenter 2020). It is a vendor-neutral software which allows the integration of different tools in a single automated workflow. Several analyses can be conducted on the created workflow as design of experiments, optimization or trade studies.

The software also can create Response Surface Models (surrogate models) that can be used instead of or combined with original models in order to reduce computational burden or search for a global optimization solution. Finally, an important feature of the software (not presented in this work) is the enabling of Model Based Systems Engineering in order to have a complete framework to develop a process/product including the whole lifecycle and guaranteeing the achievement of the established requirements.

ModelCenter includes some plug-ins for some tools (like MATLAB, Maple PTC, Ansys, etc.), although almost any external program can be wrapped and included in a library of available models. The simulation/analysis workflow is created by dragging and dropping wrapped applications from the library and combining them using if-then branches, loops or other flowchart-like constructs. Any program that can use a text file as an input and output file is easily wrapped, basically ModelCenter operates writing user values into the input file and reading the results from the output file. The wrapped models can be distributed in different machines with different operating systems. In this paper we will show the integration of MATLAB, Aspen and Excel and the use of surrogates in simulation and optimization.

3. Tool integration

3.1. Process description

Benzene and ethylene are fed to an isothermal Continuous Stirred Tank Reactor (CSTR) where ethylbenzene (EB) is produced along with some diethylbenzene (DEB) as side product. The effluent of this CSTR is fed to another reactor (in this case it is adiabatic) where transalkylation of the DEB with benzene to EB is produced. The outlet stream is fed to a separation section with two distillation columns, the first one separates non reacted benzene as distillate that is recycled to the first reactor and the second one obtains the EB product at the top and non-reacted DEB at the bottom that is recycled to the second reactor. Figure 1 shows the developed Aspen simulation file of the process.

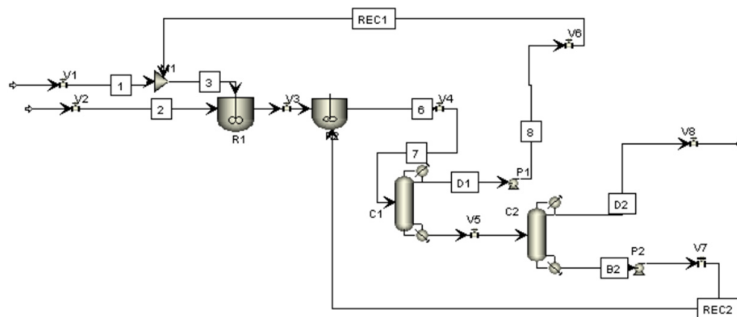


Figure 1. Ethylbenzene production process

3.2. Tool integration, an extra value to software communication

This section shows how different tools can be integrated in a single workflow and automate a simulation of the process. The ethylbenzene process is simulated according to the data provided by (Luyben, 2011). There is a tradeoff in the reactors volume and the amount of B recycled in order to minimize costs. A case study has been created to include an Excel file where the initial data and results are stored, an aspen file with the simulation and a MATLAB file that computes the Total Annualized Costs. Fig 2 (left) shows the wrapped models, the aspen model is *etilb-master* and has been created writing to the text input file (.inp file) and reading to the report text file (.rep file). Fig 2. (right) shows how the costs can be also calculated in Excel and in separate files for different equipment costs and then combined to get the overall amount. This allows to work by different users without interference in different parts of the process. This integration provides the mechanisms to establish initial requirements, perform the simulation and assess the achievement of the requirements, guaranteeing consistency between different views or parts of the model and traceability of the whole process.

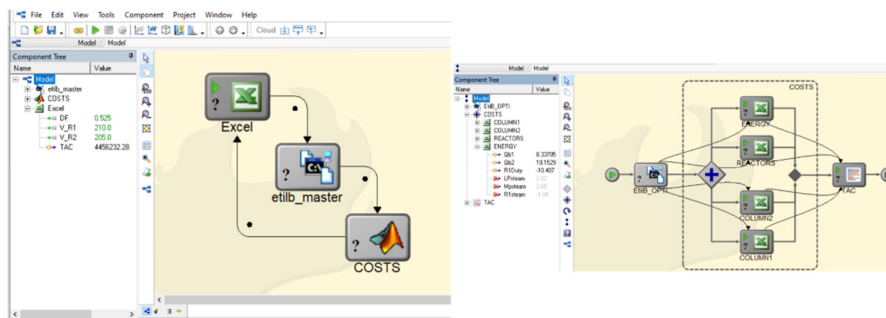


Figure 2. Tool integration: MATLAB, Aspen and Excel integration

4. Model Integration and applications

4.1. MATLAB inside Aspen

Sometimes, the available models in Aspen cannot represent a complex unit or a model of a unit has already been developed for some other purpose. In this case it would be desirable to integrate the external model into the Aspen process flow diagram. Model reuse and interoperability has been an important research field and some important initiatives have been developed (CO-Lan, 2018).

Using this framework, the integration is straightforward, in Aspen the Reactor is removed from the model and an input stream is added to the flowsheet. The original input is fed to the MATLAB model that calculates the output from the reactor and sends it to Aspen which calculates the flowsheet and obtains a new stream that should be equal to the MATLAB input. ModelCenter includes a module called Converger that changes the input to MATLAB until convergence is achieved. Basically, it is the convergence of a tear stream as it happens in a recycle or design spec in Aspen. A model developed with any other software or programming language could be integrated in the same way or even more than unit could have been replaced (for example one reactor developed with MATLAB and the other one with Python).

4.2. Superstructure optimization inside Aspen

In the design of a process several alternatives have to be considered. In order to get the optimal process, a superstructure optimization is usually created. This is usually created with a simplified model as the equations (and, overall, the physical properties) are not available. Still, in case of implementing the equations it takes an important amount of time while the Aspen model is many times available and its development is usually easier. The optimization problem is the same outlined previously, the tradeoff between reactors volume and recycled benzene but in this case the costs of the different routes are included in the optimization. In this case, we show how a superstructure can be created using the rigorous Aspen model. Fig 3. shows the superstructure created with two possible reactors (one CSTR and one PFR) and two possible separations (a flash and a distillation column). Splitters have been added to the simulation to allow for the different process paths. New feeds to the column and the flash have been added with a negligible amount, this is to avoid initial convergence problems with the simulation for the non-selected route. The flowchart (bottom left of the figure) optimizes the process sending different Boolean values to the splitters.

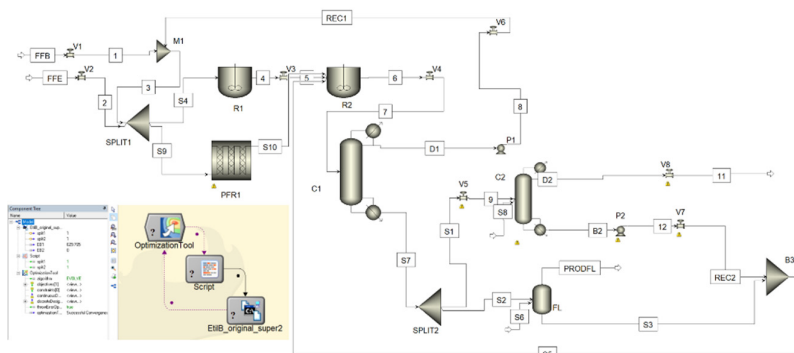


Figure 3. Superstructure optimization using Aspen rigorous model

The result obtained is the original flowsheet, CSTR and distillation column. This case can be combined with the previous one commented in section 4.1 and some of the units could be external models developed with other software than Aspen.

4.3. Using Surrogates

The use of surrogate models is gaining importance lately, if well designed, they can have similar performance than the original models (in the range of values of its design) allowing for faster execution and global optimization. ModelCenter has a module to create Response Surface Models. In this case we have created a surrogate model of the ethylbenzene process and a surrogate model of the process and the costs. ModelCenter provides two methods to create the surrogates, Polynomial and Kriging, and different Design of Experiment techniques (Full factorial, Orthogonal Array, Central Composite, Box-Behnken, Latin-Hypercube, etc). Several methods can be run for a DoE.

In this case Orthogonal Array supplemented with Latin-Hypercube was used to get the surrogate model with the Kriging approximation. Fig 4. shows the validation of the two surrogate models against the original one (left). The original process is in the middle and

the two surrogated alternatives are up and down, the results are compared in order to validate the surrogates. The implementation of the optimization using the surrogated model of the process is shown in Fig 4 right.

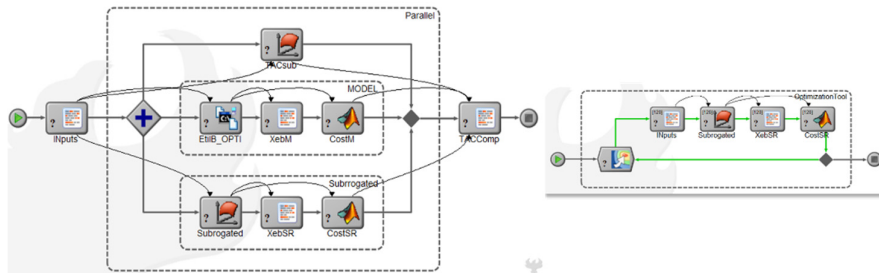


Figure 4. Surrogate modelling implementation, simulation (left) and optimization (right)

Next table shows the optimization results obtained using Aspen and using both surrogate models.

Table 1. Surrogate modelling performance comparison

	Simulation	Process Surrogate	Complete Surrogate
Reactor volume (m3)	248.4	248.4	248.4
Recycled Benzene (kmol/h)	1513	1522	1510
TAC (\$/y)	3.94M	3.98M	3.92M
Execution time	40min	3min	1min

4.4. Future applications

The next applications that the authors are working on are:

Multiscale modelling. Our research focus is not only on the lifecycle of the process but also in considering all the scales needed in a model from molecular modelling to macro scale modelling. Molecular modelling is approached with the integration of COSMOThermX, used for properties prediction of complex substances, with process simulation tools (such as Aspen Plus). The macro scale model and its impacts are being done integrating LCA tools, in this case SimaPro. The final target is to have an automated multiscale process for decision making, obtaining the optimal components and operating conditions to get the most sustainable process (considering the three dimensions economic, environmental and societal).

Operational integration. The multiscale modelling is mostly used in the design phase, in the operational phase the integration of standard engineering applications is important to guarantee a good plant performance. P&ID integration into the operating model, alarm recognition or safety assessment are some of the functionalities to be integrated with the central model.

Model Based Systems Engineering. The last application the authors are working on is the integration of PSE tools with standard MBSE languages (SysML) and tools (like IBM Rhapsody) in order to guarantee the traceability and consistency from the requirements phase to the commissioning phase. The idea is to have a core single model that will be used from the concept phase (requirements) to the final phase (decommission).

5. Conclusions

Nowadays we are in a digital transformation where models are gaining each day more importance. To have a framework that allows for vertical (whole lifecycle) as well as transversal (different tools and applications in each phase) integration using models that represent the actual process (and product) in a multiscale dimension is the overall goal of this work which is completely aligned with the current digital transformation. In this work we have presented a model-based engineering approach integrating different applications for different uses. This work has presented the potential uses and benefits of having a single model built with different standard software tools and the future steps towards having a model for a holistic view of the process/product to be developed and used.

References

- M. Bajaj, D. Zwemer, R. Yntema, A. Phung, A. Kumar, A. Dwivedi and M. Waikar, 2016, MBSE++ — Foundations for Extended Model-Based Systems Engineering Across System Lifecycle. INCOSE International Symposium, 26: 2429-2445
- CO-LaN, 2018, CAPE-OPEN Laboratories Network, Jouy-en-Josas. www.colan.org
- S. Fillinger, E. Esche, G. Tolksdorf, W. Welscher, G. Wozny and J. Repke, 2019, Data Exchange for Process Engineering – Challenges and Opportunities. *Chemie Ingenieur Technik*, 91: 256-267.
- S. Friedenthal, A. Moore, R. Steiner, 2014. A Practical Guide to SysML, 3rd Edition. Morgan Kaufmann.
- S. J. Herzig and C. J. Paredis, 2014, A conceptual basis for inconsistency management in model-based systems engineering, *Procedia CIRP*, vol. 21, pp. 52 – 57, 24th CIRP Design Conference.
- Kagermann, H., Wahlster W., Helbig, J., 2013, Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group.
- Luyben, W.L., 2011, Design and Control of the Ethyl Benzene Process. In *Principles and Case Studies of Simultaneous Design*, W.L. Luyben (Ed.)
- ModelCenter, 2020, “Introduction to Model Based Engineering with ModelCenter”, Phoenix-int.com. Retrieved 10 October 2020.
- Schwab, K., 2016, The Fourth Industrial Revolution. In *World Economic Forum*.
- NAMUR , 2018, Standardized NAMUR Interface for Data Exchange Between CAE Systems for Process Design and CAE Systems for PCT Hardware Planning, 2nd, Recommendation NE 159.

Benchmarking of Surrogate Models for the Conceptual Process Design of Biorefineries

Nikolaus I. Vollmer,^{a*} Resul Al,^b Gürkan Sin^a

^a *Process and Systems Engineering Research Center, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søtofts Plads 228A, 2800 Kgs. Lyngby, Denmark*

^b *Novo Nordisk A/S, Bagsværd, Denmark*

*nikov@kt.dtu.dk

Abstract

Surrogate models are an efficient method to expedite the process design by superstructure optimization. For their application in biorefineries' process design, several surrogate models are benchmarked in a case study regarding their validation metrics and their performance in a reference superstructure optimization problem. Despite good validation metrics for most surrogate models, their prediction quality in the superstructure optimization does not reflect this. For the use of surrogate models in superstructure optimization, the need for a profound assessment of options, and the possible use of dynamic sampling strategies become evident.

Keywords: Biorefinery, Process Design, Superstructure Optimization, Surrogate Modelling, Machine Learning.

1. Introduction

A key approach in promoting the transition towards a bio-based economy is the conceptual design and implementation of integrated second-generation biorefineries. A conceptual design approach for these biorefineries is Superstructure Optimization (SSO), which yields an optimal candidate process topology, but which is inherently limited by the initial search space and the fidelity of the models (Mencarelli et al., 2019). This contrasts highly with the complexity of fermentation processes and disregards advances in synthetic biology to optimize cell factories (Straathof et al., 2019). However, high-fidelity models, as, e.g., mechanistic models, which resemble the underlying physical, chemical, and biological phenomena with suitable mathematic descriptions, can describe this complexity well.

To facilitate the computational tractability of the SSO with high-fidelity models, surrogate models, which are commonly machine learning or other statistical models, can be used instead in the superstructure formulation (Mencarelli et al., 2020). Regarding the use of surrogate models in SSO, we hence propose and benchmark five different alternatives to determine the most suitable surrogate alternative for a process design task. The first four alternatives comprise 1) a Gaussian Process Regressor, 2) a Support Vector Machine Regressor, 3) a Regression Tree Ensemble, and 4) an Artificial Neural Network

as a surrogate model, which are respectively fitted to flowsheet-wide simulation data over the design space. The fifth alternative serves as a comparison to the prior four and is a regression model based on linear regression in multiple dimensions. Each model is validated, and relevant metrics are assessed and compared. With the respective surrogate models for each flowsheet option in the superstructure, the underlying optimization problem of the SSO, which is classically a mixed-integer nonlinear program, can be reformulated either as a series of nonlinear programs for the first four models or as a mixed-integer linear program for the last surrogate. The optimization problem's solution with each surrogate and validation with the original flowsheet model serves as the benchmarking purpose in this study.

2. Methods

2.1. Gaussian Process Regression

Gaussian Process Regression (GPR) is a prominent type of machine learning model with a plethora of applications. The regression is based on a stochastic process. In this stochastic process, a kernel function correlates any point in the input space to the points used to fit the GPR (Al et al., 2019). GPRs are extremely versatile regarding their application and can interpolate very well with small amounts of data available; however, they work best for low-dimensional problems, steady design spaces, and moderately sized datasets (McBride and Sundmacher, 2019).

2.2. Artificial Neural Network

Artificial Neural Networks (ANN) are another prevalent type of machine learning model with applications reaching from standard machine learning and deep learning tasks up to artificial general intelligence. One of the simplest representations of an ANN is a multilayer perceptron (MLP) with one input layer, at least one hidden layer, and one output layer, where each hidden layer contains a certain number of nodes that use a specific transfer function to calculate the node's outputs to all its inputs (Al et al., 2019). The hyperparameters of the network architecture, e.g., number of layers, neurons, types of the transfer function, and training algorithms, can be screened in the fitting process using a grid search approach to improve the model quality. ANNs are especially suitable for highly nonlinear and high-dimensional systems; however, their functionality is hindered by low amounts of data unless shallow ANNs are used (McBride and Sundmacher, 2019).

2.3. Support Vector Regression

Another heavily used representative of machine learning models for surrogate modelling are support vector machines. The general idea of support vector machines is to fit hyperplanes in the n -dimensional input space to maximize the distance of the points in the input space to the hyperplane by utilizing kernel functions. The fitted hyperplanes then serve as the basis for the regression (Williams and Cremaschi, 2019).

2.4. Regression Tree Ensemble

The last introduced machine learning concept is a so-called regression tree ensemble (RTE), belonging to the class of ensemble methods, where several decision trees are fitted to the data, and the weighted combination of these trees serves as the basis for regression. The concept of a single decision tree is visually very well depicted by dividing input data based on several input criteria onto different branches of the tree. Tree ensembles are

quite powerful as they can effectively fit sparse datasets with exceptional predictive capacities (Thebelt et al., 2020).

2.5. Delaunay Triangulation Regression

In addition to all the introduced machine learning models, the idea behind Delaunay triangulation regression (DTR) as a surrogate model is the extension of the concept of a piecewise-linear function in multiple dimensions (Franke, 1979). Triangulation as a mathematical topology is a homomorphism to the n -dimensional design space $X \in R^n$, consisting of a set of n -simplices. An n -simplex consists of $n + 1$ vertices, which are set to be the sampling points $p_x \in X$ in the design space. Any point x within the simplex can be described as a linear combination of the vertices.

2.6. Surrogate Model Validation

As sampling strategy for each surrogate model, Latin Hypercube sampling is chosen. Alternatively, Sobol sequences can be used with similar results (Williams and Cremaschi, 2019). Each surrogate model type is then fitted to the set of sampled points, while the fit is optimized respectively with a corresponding routine. All machine learning models are cross-validated, employing a k -fold cross-validation. The cross-validation methodology is altered to preserve the convex hull for the DTR surrogate model, as the DTR model intrinsically has no extrapolative capabilities. The calculated validation metrics are the coefficient of determination R^2 and the root mean squared error $RMSE$ for the testing and training datasets with $k = 5$.

2.7. Superstructure Optimization

The superstructure for superstructure-based process design can be transferred into an optimization problem to solve it mathematically. While there are different ways of postulating a superstructure, e.g., as state-task-network or state-equipment network, the underlying optimization problem is commonly a mixed-integer problem (MIP):

$$MIP: z = f(x, y) \quad s.t. \quad g(x, y) \leq 0 \quad h(x, y) = 0 \quad x \in X, \quad y \in [0,1] \quad (1)$$

With z being the objective to be minimized, $f(x, y)$ being the functional relation of respectively one of the surrogate models, $g(x, y)$ representing inequality constraints, $h(x, y)$ representing equality constraints and y defining binary decision variables. If either of the functional relationships is nonlinear, the MIP is denoted as a mixed-integer nonlinear program (MINLP). Due to the strenuous computational tractability for state-of-the-art solvers to solve MINLPs, two different reformulation strategies are chosen: The first is the reformulation to a series of nonlinear programs (NLP), eliminating all integer variables by iterating over each flowsheet alternative. The second option is the reformulation to a mixed-integer linear program (MILP), where all nonlinear functional relations in the model are linearized.

3. Results & Discussion

3.1. Case Study

As a case study, previously presented by Vollmer et al. (2020), a base-case biorefinery setup is chosen, with wheat straw as substrate and xylitol as the product. For this work, one reference flowsheet with fixed topology out of the entire superstructure is chosen. The reference flowsheet involves a pretreatment unit, an evaporator as an upconcentration unit, a fermentation unit for converting the hemicellulosic sugars to xylitol, an evaporation unit, and two crystallization units in the downstream processing (Vollmer et

al., 2020). For all unit operations, mechanistic models based on mass and energy balances and kinetic equations for the respective process are employed. All models are implemented in MATLAB. All models are analyzed regarding their robustness by a comprehensive Monte Carlo-based uncertainty and a variance-based sensitivity analysis. The selected variables for the input space t are the acid concentration for the pretreatment C_{ac} , the fermentation time t_{FX} , the inoculum concentration for the fermentation process C_{inoc} , the vapor fraction of the upconcentration unit v_{UCH} and the evaporation unit v_{EX} . The other variables are set to fix values within their operational ranges. The calculated output variables are the produced mass of xylitol M_{Xyo} , the concentration of two inhibitors (5-HMF and acetic acid) after the final crystallization step C_{Hmf} and C_{AAC} , as well as a ratio φ , indicating the mass of produced CO_2 by steam generation for providing process heat in the upconcentration and evaporation unit per mass of produced xylitol.

3.2. Surrogate Model Creation and Validation Results

The input space with the five variables is sampled with $N = 125, 250, 500,$ and 1000 samples. For the GPR, the SVM, and the RTE, the Statistics & Machine Learning toolbox functionalities of MATLAB are used. An internal routine respectively optimizes the hyperparameters for all three in MATLAB. For the ANN, the functionalities of the Deep Learning toolbox of MATLAB are used. The ANN's fitting is performed by grid search functionalities implemented within the easyGSA toolbox (Al et al., 2019). The DTR model's triangulation is obtained by the Delaunay triangulation functionality of the scipy library for Python. The parity plots in Figure 1 show that the GPR surrogates fit best for all the sampling sizes. The SVM surrogates show a significant amount of outlier for increasing sampling sizes, and the RTE surrogates to a lesser amount. However, the ANN surrogates also show good fitting qualities with a larger variance than the GPR surrogates.

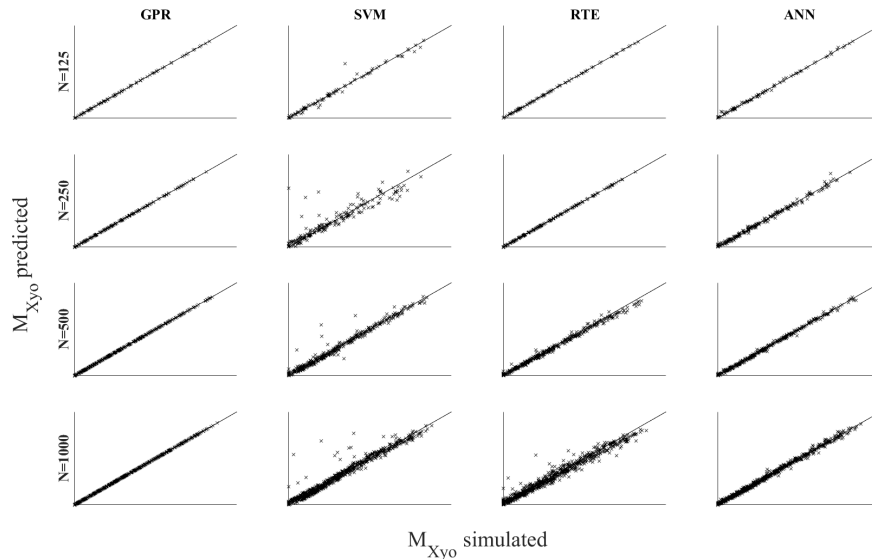


Figure 1: Parity plots of the fit for the GPR, SVM, RTE, and ANN surrogates with all data points.

The validation results for $N = 1000$ samples are shown in Table 1. The metrics for all machine learning surrogates show a good fit. The DTR surrogate predicts unseen data poorly, worsening the quality metrics. Similar metrics result for all other sample sizes.

Table 1: Metrics for the k-fold cross-validation of all surrogates with $N=1000$ samples.

N=1000	GPR	SVM	RTE	ANN	DTR
R^2_{train}	1	0.95	0.99	0.995	1
R^2_{test}	0.981	0.937	0.921	0.985	0.68
$RMSE_{train}$	0.017	2.858	1.057	0.961	0
$RMSE_{test}$	1.818	3.382	3.721	1.602	7.788

3.3. Superstructure Optimization Results

The input variables for the optimization problem are considered in the following range: the acid concentration $C_{ac} = [0.5, 2]$ g/100g biomass, the fermentation time $t_{FX} = [12, 48]$ h, the inoculum concentration $C_{inoc} = [0.5, 3]$ g/L, the vapor fraction of the upconcentration unit $v_{UCH} = [0.4, 0.6]$ and the vapor fraction of the evaporation unit $v_{EX} = [0.99, 0.998]$. Furthermore, the concentration of 5-HMF and acetic acid respectively in the last unit operation cannot exceed $C_{Hmf}, C_{Aac} \leq 0.5$ g/L. Equally, not more than 10 kg of CO₂ can be produced per kg of xylitol produced, so $\varphi \geq 0.1$. The objective is chosen to be the produced amount of xylitol $z = M_{Xyo}$. For all NLPs, a surrogate is fitted to the input samples and, respectively, one output variable and solved with the MATLAB solver finincon (SQP algorithm) and a multi-start method to seek global optimality. For the MILP, one DTR surrogate is fitted to the input samples, capturing all four outputs, implemented in Pyomo and solved with the GUROBI solver. The following table shows the optimization results and the corresponding validation simulation with the original model.

Table 2: Results for the objective function and the constraints from the NLP/MILP (left) and the validation simulation (right).

125	GPR		ANN		SVM		RTE		DTR	
M_{Xyo}	55.9	52.6	54.8	54.2	64.8	0.00	55.2	0.00	47.1	48.2
C_{HMF}	0.50	0.90	0.5	0.52	0.5	2.82	0.34	8.82	0.50	0.50
C_{Aac}	0.03	0.03	0.03	0.02	0.04	0.18	0.04	0.31	0.02	0.02
φ	0.12	0.12	0.13	0.13	0.17	0.00	0.11	0.00	0.11	0.11
250	GPR		ANN		SVM		RTE		DTR	
M_{Xyo}	52.6	49.2	85.3	0.00	62.7	47.0	43.0	37.2	48.7	48.7
C_{HMF}	0.50	0.90	0.50	14.0	0.50	0.31	0.48	2.28	0.42	0.43
C_{Aac}	0.01	0.03	0.11	0.27	0.00	0.02	0.04	0.07	0.02	0.02
φ	0.12	0.11	0.15	0.00	0.18	0.11	0.09	0.09	0.11	0.11
500	GPR		ANN		SVM		RTE		DTR	
M_{Xyo}	53.9	53.9	53.3	53.8	84.4	0.0	47.9	2.80	50.0	51.1
C_{HMF}	0.50	0.45	0.50	0.34	0.50	3.76	0.25	2.80	0.50	0.45
C_{Aac}	0.02	0.02	0.02	0.02	0.07	0.13	0.10	0.13	0.02	0.02
φ	0.12	0.13	0.12	0.13	0.61	0.00	0.11	0.01	0.12	0.12
1000	GPR		ANN		SVM		RTE		DTR	
M_{Xyo}	54.3	51.0	52.1	53.1	56.1	0.00	47.2	4.24	49.2	50.4
C_{HMF}	0.50	0.75	0.50	0.34	0.5	2.82	0.44	4.74	0.50	0.35
C_{Aac}	0.02	0.03	0.02	0.02	0.04	0.13	0.05	0.15	0.02	0.02
φ	0.13	0.12	0.12	0.12	0.12	0.00	0.11	0.01	0.11	0.12

For the GPR, the predictions are constantly on a high accuracy level independent of the sample size. However, the GPR tends to overpredict the objective function value and to underpredict the concentration constraints. The ANN shows good predictions of the objective functions and no bound infringements except for the case with $N=250$. Both the SVM and the RTE mostly fail to predict the objective value accurately due to poor fitting. Lastly, the DTR model predicts the objective function value accurately for all cases without bound violations despite the impaired validation metrics.

4. Conclusion

It becomes evident that despite the brilliant predictive qualities of most machine learning surrogates, this is not reflected in the SSO. The DTR model performs best in the benchmark, indicating a promising potential as surrogate modelling technique for optimization applications. Possibly, the predictive quality of all models could be improved with dynamic sampling strategies, reducing the inconsistencies in prediction over differently sized datasets. Overall, the crucial role of cross-validation and case-dependent analyses of different surrogate models becomes evident, as sampling size and strategy clearly play a role in the resulting model quality and the optimization results, thus application strategies can hardly be generalized.

Acknowledgments

The research project is part of the Fermentation-Based Biomanufacturing Initiative and is funded by the Novo Nordisk Foundation (NNF17SA0031362).

References

- R. Al, C.R. Behera, A. Zubov, K.V. Gernaey and G. Sin. 2019. Meta-modeling based efficient global sensitivity analysis for wastewater treatment plants – An application to the BSM2 model. *Computers & Chemical Engineering*. Vol. 127. 233-246.
- R. Al, C.R. Behera, K.V. Gernaey and G. Sin. 2020. Stochastic simulation-based superstructure optimization framework for process synthesis and design under uncertainty. *Computers & Chemical Engineering*. Vol. 143. 107118.
- K. McBride and K. Sundmacher. 2019. Overview of Surrogate Modeling in Chemical Engineering. *Chemie Ingenieur Technik*. Vol. 91(3). 228-239.
- L. Mencarelli, Q. Chen, A. Pagot and I.E. Grossmann. 2020. A review on superstructure optimization approaches in process system engineering. *Computers & Chemical Engineering*. Vol. 136. 106808.
- R.A. Franke. 1979. A critical comparison of some methods for interpolation of scattered data. Naval Postgraduate School Monterey CA, 1979.
- A.J.J. Straathof, S.A. Wahl, K.R. Benjamin, R. Takors, N. Wierckx and H.J. Noorman. 2019. Grand Research Challenges for Sustainable Industrial Biotechnology. *Trends in Biotechnology*. Vol. 37(10). 1042-1050.
- A. Thebelt, J. Kronqvist, R.M. Lee, N. Sudermann-Merx and R. Misener. 2020. Global Optimization with Ensemble Machine Learning Models. *Computer Aided Chemical Engineering*. Vol. 48. 1981-1986.
- N.I. Vollmer, C.K. Yamakawa, K.V. Gernaey, S.I. Mussatto and G. Sin. 2020. Surrogate Modelling Based Uncertainty and Sensitivity Analysis for the Downstream Process Design of a Xylitol Biorefinery. *Computer Aided Chemical Engineering*. Vol. 148. 1663-1668.
- B.A. Williams and S. Cremaschi. 2019. Surrogate Model Selection for Design Space Approximation And Surrogate-based Optimization. *Computer Aided Chemical Engineering*. Vol. 47. 353-358.

Predicting the Estrogen Receptor Activity of Environmental Chemicals by Single-Cell Image Analysis and Data-driven Modeling

Hari S. Ganesh^a, Burcu Beykal^a, Adam T. Szafran^b, Fabio Stossi^{b,c}, Lan Zhou^d, Michael A. Mancini^{b,c,e,f}, Efstratios N. Pistikopoulos^{a,g,*}

^aTexas A&M Energy Institute, Texas A&M University, College Station, TX, United States of America

^bMolecular and Cellular Biology, Baylor College of Medicine, Houston, TX, United States of America

^cGCC Center for Advanced Microscopy and Image Informatics, Houston, TX, United States of America

^dDepartment of Statistics, Texas A&M University, College Station, TX, United States of America

^eTexas A&M University Institute for Bioscience and Technology, Houston, TX, United States of America

^fPharmacology and Chemical Genomics, Baylor College of Medicine, Houston, TX, United States of America

^gArtie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, United States of America
stratos@tamu.edu

Abstract

A comprehensive evaluation of toxic chemicals and understanding their potential harm to human physiology is vital in mitigating their adverse effects following exposure from environmental emergencies. In this work, we develop data-driven classification models to facilitate rapid decision making in such catastrophic events and predict the estrogenic activity of environmental toxicants as estrogen receptor- α (ER α) agonists or antagonists. By combining high-content analysis, big-data analytics, and machine learning algorithms, we demonstrate that highly accurate classifiers can be constructed for evaluating the estrogenic potential of many chemicals. We follow a rigorous, high throughput microscopy-based high-content analysis pipeline to measure the single cell-level response of benchmark compounds with known *in vivo* effects on the ER α pathway. The resulting high-dimensional dataset is then pre-processed by fitting a non-central gamma probability distribution function to each feature, compound, and concentration. The characteristic parameters of the distribution, which represent the mean and the shape of the distribution, are used as features for the classification analysis *via* Random Forest (RF) and Support Vector Machine (SVM) algorithms. The results show that the SVM classifier can predict the estrogenic potential of benchmark chemicals with higher accuracy than the RF algorithm, which misclassifies two antagonist compounds.

Keywords: Predictive modeling, estrogen receptor activity, classification analysis, big-data analytics, image analysis.

1. Introduction

Advancements in machine learning algorithms combined with computational power and storage capacities of modern computers have enabled data-driven modeling as a promising approach for prediction and decision-making in biological and environmental systems (Yin et al., 2015). Specifically, data-driven algorithms enable pathways to understand similarities between complex chemicals (Onel et al., 2019, Onel et al., 2018) and facilitate rapid detection and evaluation strategies for predicting the endocrinological activity of unknown chemicals created during environmental emergencies (Mukherjee et al., 2020). In such catastrophic situations (i.e., hurricanes, earthquakes, etc.), many unknown chemicals are mobilized, increasing the risk of exposure and long-term adverse biological effects on living organisms (Cooper and Wardropper, 2020). Understanding the effects of these novel chemicals on a key transcription factor that regulates the most fundamental biological processes in the human body, such as the Estrogen Receptor- α (ER α), is crucial for diagnosing potential health hazards.

The toxic compounds created during and after environmental emergencies are capable of binding to the ER α and initiating a biological response as agonists or interfere with the receptor function as antagonists (Warfvinge et al., 2020). The agonistic and antagonistic behaviors of chemicals can be measured using cell-based high content/high throughput microscopy imaging, which records numerous intensity and morphology features of treated cells (Szafran et al., 2017). Recently, we have shown that this multidimensional imaging data with population-averaged samples can be used to classify agonist and antagonist benchmark chemicals with high accuracy using linear and nonlinear models (Mukherjee et al., 2020). In this work, we use cell-level data (instead of the population-averaged data) to assess the predictive capabilities of nonlinear classifiers, namely the Random Forest (RF) and the Support Vector Machine (SVM) algorithms. To sustain cell-level information in the dataset, we fit non-central gamma probability distribution functions to the data and use their corresponding distribution shape parameters (i.e., alpha, beta, delta) as the descriptors in our predictive models. We further follow a rigorous training and testing procedure to build and assess the predictive capabilities of our data-driven models.

2. Experimental Data Gathering: High Content Image Analysis

We used high-throughput microscopy and image-analysis pipelines to measure the estrogenic activity of chemicals using an engineered biosensor, a multi-parametric cell-based *in vitro* assay as described in Ashcroft et al. (2011). Biosensor cells were treated for 2 hours with different concentrations of 45 reference endocrine-disrupting chemicals (EDCs) identified by the EPA to be either ER α inactive, agonist, or antagonist. The 45 reference chemicals with known estrogenic activity were supplemented with 8 additional known ER α antagonists to partially balance the number of known agonists and antagonists tested. The cell-level response was imaged and recorded, and 70 size, shape, and intensity features per cell were extracted using custom-built software (Szafran et al., 2014). These features describe several key mechanistic steps of the ER pathway including, ER α expression levels, its subcellular/subnuclear distribution, and DNA binding (Szafran et al., 2017). A single screen of the aforementioned analysis generates 60 GB of image data, from which approximately 50 million object-based data points are extracted. This high-dimensional dataset is reduced prior to the classification model building step using a technique described in the following section.

3. Data Pre-Processing: Dimensionality Reduction

Each complete screen of the EDCs and reference chemicals consists of five 384-well plates with each unique chemical-concentration combination having a minimum of 4 in-screen replicate sample wells. A total of 5 complete screens were completed. For classification model development, a subset consisting of the highest tested concentration of each active (either agonist or antagonist) chemical was generated. This subset contains 32 agonists and 15 antagonist chemicals and 39 cell-level features (the discarded 31 features were of the medium) were utilized. To sustain the cell-level information while making the dataset amenable for classification model building, a non-central gamma distribution with 3 parameters (i.e., alpha, beta, and delta) was fitted for each sample. This reduced the size of the dataset to 0.15 % (by a factor of approximately 3/2000) of its original size. Figure 1 shows example histograms and the fitted non-central gamma distribution for representative compounds 100 μ M Genistein, 100 μ M Fenarimol, and 10 μ M 4-Nonylphenol, and features array total pixel intensity (PI), array mean PI, and array circularity, respectively. In these plots, alpha, beta, and delta, the parameters of the fitted gamma distributions, are shown in the legends. These characteristic shape parameters were calculated for every feature, compound, and concentration.

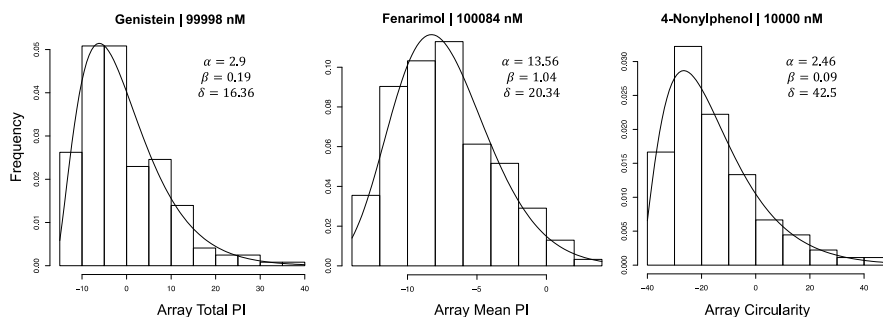


Figure 1. Non-central gamma distribution fit on the cell-level data for reference compounds. The values of the characteristic shape parameters alpha, beta, and delta are listed on the legends.

Once the parametrization of the dataset is complete, the outliers in each feature-parameter were removed. Following, the dataset was column-wise scaled using the minimum-maximum scaling procedure. Then, for each feature, the data for alpha was plotted against beta, and beta was plotted against delta for all the compounds and concentrations to visualize the distribution of agonist and antagonist compounds in the parametrized space. Figure 2 shows a representative set of plots of alpha versus beta and beta versus delta for the “Nucleoplasm Pixel Intensity (PI) Variance” feature, which quantifies the statistical variance of GFP-ER α pixel intensity in the nucleoplasm of the cell. The visualization results show that agonist and antagonist data points mostly separate into two distinct groups. Yet, several data points belonging to the opposite classes are very close to each other, indicating that a model using linear classification algorithms may not result in high accuracy prediction.

Similar trends were observed for almost all the features. Therefore, the elimination of less indicative features by this qualitative assessment could not be made. To follow this observation with quantitative analysis, we performed Principal Component Analysis (PCA). The PCA results showed that 19 components out of 117 accounted for 90 % of the variance in the original data and 40 components out of 117 accounted for 99 % of the

variance in the data, indicating that the variability across many features in the dataset was similar and not distinct. This result is different from our previous study for the population-averaged data where 5 top informative features led to satisfactory linear and nonlinear classification results (Mukherjee et al., 2020). Note however that the dataset and pre-processing strategies used in the previous study were different from those in this study. Nevertheless, the qualitative difference is noteworthy.

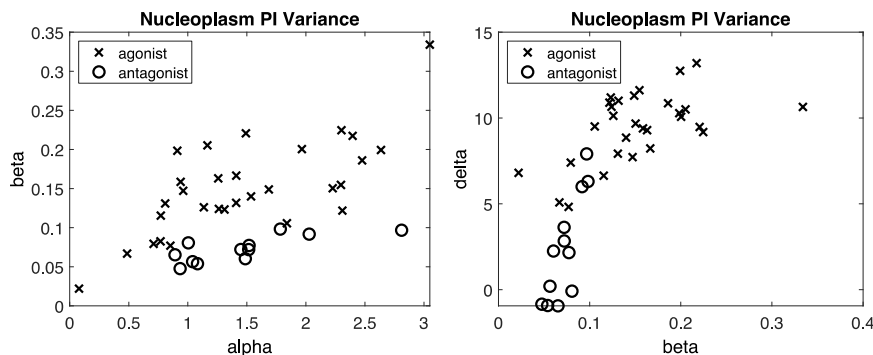


Figure 2. Alpha vs. beta and beta vs. delta plots to visualize the grouping of compounds per feature per concentration based on their estrogenic potentials.

4. Data-Driven Model Building: Nonlinear Classification Algorithms

Once the pre-processing step was completed, the following supervised learning models were built to predict the estrogenic potential of chemicals as agonists or antagonists: (1) The RF classifier, (2) 1-class nonlinear SVMs for predicting agonist or antagonist classes, and (3) 2-class nonlinear SVM. RF fits several decision trees on various sub-samples of the data and uses averaging to improve the accuracy of the classifier prediction and control over-fitting. 2-class SVM finds the optimal separating hyperplane to maximize the width of the gap between the two categories. In 1-class SVM, the width is maximized with respect to the category used for training and the rest are regarded as anomalies.

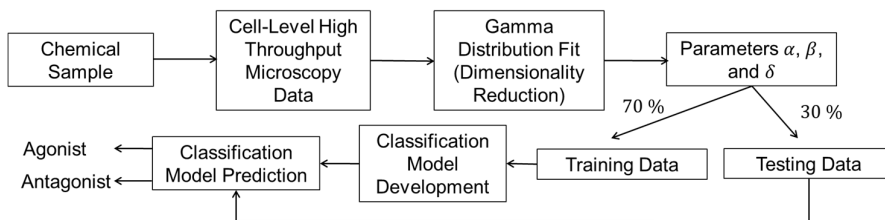


Figure 3. Algorithm for classification model building.

We follow the steps outlined in Figure 3 for building our classification models using the experimental image analysis data. 70% of the dataset per replicate was used for training and the remaining 30% was used for testing. After dimensionality reduction, in RF, a recursive addition algorithm was followed for simultaneous feature selection and model building. In this method, an RF model using all the features was built and the features were ranked according to their importance in model prediction. Following, the accuracies of model predictions using recursive addition of features in the model were recorded. The

model (and thus features) with the maximum accuracy was selected. In both 1-class and 2-class SVMs, the radial basis function kernel was used, and the model parameters were tuned to maximize the accuracy of prediction of the testing dataset using grid search.

5. Results and Discussions

Table 1. Testing performance of the RF, 1-class SVM, and 2-class SVM classification models.

	Accuracy	Precision	Sensitivity	Specificity	F-Score
RF	0.875	1	0.6	1	0.75
1-class SVM (agonist)	0.763	1	0.763	-	0.865
1-class SVM (antagonist)	0.72	1	0.72	-	0.837
2-class SVM	0.925	0.952	0.8	0.982	0.869

Table 1 lists the model prediction summary statistics for all the model types. For the RF model, the recursive feature addition step resulted in 61 features to be selected out of 117 to yield the maximum predictive accuracy. The final RF model predicts the test data with an accuracy of 87.5%. It correctly classifies 55/55 agonist samples, whereas 10 out of 25 antagonist samples are misclassified as agonists where the prediction sensitivity is 0.6 and the precision is 1. The two antagonist compounds, namely ICI 182 780 and AZD 9496, are misclassified in all 5 replicate datasets. Interestingly, these two chemicals are unique in the tested set of ER α antagonists in that they are known to destabilize ER α expression via the same pathway activated by ER α agonists. As these two chemicals mechanistically resemble agonists, the classification model cannot distinguish them as antagonist chemicals.

Furthermore, the 1-class SVM predicts the agonist class with an accuracy of 76% and the antagonist class with an accuracy of 72%, which are lower than the RF model. The 2-class SVM model, predicts the outcome of the test data correctly with an accuracy of 92.5%, the highest among all analyzed classification models whereas the specificity of the 2-class SVM is reported to be less than the RF model. However, it is observed that the 2-class SVM is more sensitive towards classifying antagonist compounds correctly where ICI 182 780 is correctly classified as antagonist for all replicates by the 2-class SVM. On the other hand, the RF is more precise when classifying agonist compounds correctly where 7,4'-Dihydroxyisoflavone, a weak agonist is misclassified as an antagonist by the 2-class SVM. Therefore, if the dataset has a considerable number of antagonist compounds, then the 2-class SVM will provide higher predictive accuracy, else, the RF will outperform.

6. Conclusions

High-throughput microscopy and single-cell image analysis followed by machine learning were used to predict the ER α activity of environmentally relevant chemicals. Dimensionality reduction was performed on the experimental big-data by fitting a non-central gamma probability distribution function. The scaled characteristic shape parameters of the distribution were used as the features for the classification analysis. Among three different nonlinear classifiers tested, namely the Random Forest (RF) algorithm, the 1-class Support Vector Machine (SVM), and the 2-class SVM, the 2-class SVM predicted the outcome of the test data with the highest accuracy of 92.5%, compared to 87.5% for the RF, and 73.3% and 72% for the 1-class SVMs, respectively. Overall, it

is observed that cell-level information increases the complexity of data handling, analysis, and model building steps. We believe that this study will encourage researchers to investigate further the trade-offs between obtaining additional information from cell-level measurements and increasing the complexity of big-data handling and classification model development. This research was funded by the U.S. National Institutes of Health (NIH) grant P42 ES027704.

References

- F.J. Ashcroft, J.Y. Newberg, E.D. Jones, I. Mikic, M.A. Mancini, 2011, High content imaging-based assay to classify estrogen receptor- α ligands based on defined mechanistic outcomes. *Gene*, 477(1-2), 42-52.
- C.M. Cooper, C.B. Wardropper, 2020, Environmental, public health, and economic development perspectives at a Superfund site: AQ methodology approach. *Journal of Environmental Management*, 111571
- O. Lanvin, S. Bianco, N. Kersual, D. Chalbos, J.M. Vanacker, 2007, Potentiation of ICI182, 780 (Fulvestrant)-induced estrogen receptor- α degradation by the estrogen receptor-related receptor- α inverse agonist XCT790. *Journal of Biological Chemistry*, 282(39), 28328-28334.
- R. Mukherjee, B. Beykal, A.T. Szafran, M. Onel, F. Stossi, M.G. Mancini, D. Llyod, F.A. Wright, L. Zhou, M.A. Mancini, E.N. Pistikopoulos, 2020, Classification of Estrogenic Compounds by Coupling High Content Analysis and Machine Learning Algorithms. *PLoS Computational Biology*, 16(9), e1008191.
- A. Nardone, H. Weir, O. Delpuech, H. Brown, C. De Angelis, M.L. Cataldo, ..., C. Nagi, 2019, The oral selective oestrogen receptor degrader (SERD) AZD9496 is comparable to fulvestrant in antagonising ER and circumventing endocrine resistance. *British journal of cancer*, 120(3), 331-339.
- M. Onel, B. Beykal, K. Ferguson, W.A. Chiu, T.J. McDonald, L. Zhou L, J.S. House, F.A. Wright, D.A. Sheen, I. Rusyn, E.N. Pistikopoulos, 2019, Grouping of complex substances using analytical chemistry data: A framework for quantitative evaluation and visualization. *Plos One*, 14(10), e0223517.
- M. Onel, B. Beykal, M. Wang, F.A. Grimm, L. Zhou, F.A. Wright, T.D. Phillips, I. Rusyn, E.N. Pistikopoulos, 2018, Optimal Chemical Grouping and Sorbent Material Design by Data Analysis, Modeling and Dimensionality Reduction Techniques, *Computer Aided Chemical Engineering*, 43, 421-426.
- A.T. Szafran, M.A. Mancini, 2014, The myImageAnalysis Project: A Web-based Application for High-content Screening. *Assay and Drug Development Technologies*, 12(1), 87-99.
- A.T. Szafran, F. Stossi, M.G. Mancini, C.L. Walker, M.A. Mancini, 2017, Characterizing properties of non-estrogenic substituted bisphenol analogs using high throughput microscopy and image analysis, *PloS one*, 12(7), e0180141.
- K. Warfvinge, D.N. Krause, A. Maddahi, J.C. Edvinsson, L. Edvinsson, K.A. Haanes, 2020, Estrogen receptors α , β and GPER in the CNS and trigeminal system-molecular and functional aspects. *The Journal of Headache and Pain*, 21(1), 1-16.
- S. Yin, O. Kaynak, 2015, Big data for modern industry: challenges and trends [point of view]. *Proceedings of the IEEE*, 103(2), 143-146.

Data-driven prognostics for Lithium-ion battery health monitoring

Sukanya G^a, Resmi Suresh^{b,*}, Raghunathan Rengaswamy^{a,**}

^a Dept. of Chemical Engineering, IIT Madras, 600036, India

^b Dept. of Chemical Engineering, IIT Guwahati, 781039, India

* resmis@iitg.ac.in, ** raghur@iitm.ac.in

Abstract

Li-ion batteries are a popular choice of rechargeable battery for use in many applications like portable electronics, automobiles as well as stationary applications for providing uninterruptable power supply. State of Charge (SoC) and State of Health (SoH) are important metrics of a Li-ion battery that can help in both battery prognostics and diagnostics for ensuring high reliability and prolonged lifetime. The ML algorithms available in the literature for SoC and SoH prediction involves use of various derived features rather than directly measurable features making it difficult for industrial applications. In this work, we use battery data obtained from different batteries to develop supervised models that can be used for the on-line estimation of SoC and SoH. This work involves two parts: a) developing a classifier based on SoH b) dynamic prediction of battery SoC given the past operational data of current, voltage, and temperature of the battery which are easily measurable. Random forest algorithm is used for battery site classification based on the SoH data available from the manufacturer. The battery SoC estimation is performed using a random forest algorithm and Neural network-based NARX model.

Keywords: Li-ion battery prognostics, SoH, SoC, Random forest, NARX.

1. Introduction

Lithium-ion batteries are widely used rechargeable batteries in electric cars, smartphones, laptops, and telecommunication base stations due to their high reliability, high energy density, high efficiency and long cycle life. Since their invention in the late 20th century, Li-ion batteries have established themselves as the frontrunner of rechargeable batteries in providing power solutions across several fields such as military, aerospace, medical, telecommunication, residential, and industrial applications. To ensure a prolonged lifetime as well as optimum performance, it is necessary to continuously evaluate the health condition of the battery. Battery health prognostics refers to the task of assessing the current health of the battery and forecast degradation and detect the occurrence of any abnormalities. For developing reliable techniques for battery health prognostics, accurate estimation of State of Charge (SoC) and State of Health (SoH) plays a major role. SoC is the ratio of available charge in a battery to the maximum attainable charge in that battery at its current health. Although there exist numerous techniques for SoC estimation, there is no technique that meets all the requirements for accurate estimation of SoC online. SoH is an overall battery health indicator that quantifies the deviation of battery behavior from its ideal behavior. Unlike SoC that is well-defined mathematically (though difficult to accurately estimate), there is no unique mathematical representation for SoH as health is a broader concept than what can be represented using a single numerical value.

SoH is an important diagnostic and prognostic measure for battery operation. At any instant of operation, the knowledge of the battery's SoH will help the user to make informed decisions. It is undesirable to operate a battery below a threshold value of SoH as it may result in short-circuiting, thermal runaway and explosion. If a user can a-priori know the SoH, it is possible to predict when the battery may fail.

The objective of this work is to build models using actual battery data collected from four different batteries that could be used for two purposes: develop a classifier based on SoH and perform dynamic prediction of battery SoC given the past operational data (current, voltage, and temperature) of the battery. An online data-based SoC estimator can help the user to decide the range, when to recharge etc and can be implemented in online battery management systems. A historic data-based SoH classification algorithm can help in identifying the current SoH of a battery and can help the user to make decisions on the operating conditions of the battery. Additionally, the dynamic variation of SoC depends on the battery's SoH. Hence, the information derived from the SoH classifier can help in better prediction of SoC in a battery.

1.1. Battery State of Health

Battery state of health is largely identified as a measure of the battery aging process due to capacity degradation or power fade. Estimating battery SoH can help in predicting when the battery should be removed from operation. Since battery health is a difficult measure to quantify, battery SoH is defined in literature as the ratio of the maximum available capacity to the battery nominal capacity. The capacity of the battery could degrade over its operational period due to voltage effects (over-voltage, under-voltage or over-discharge), thermal effects (thermal runaway, reduced power due to low-temperature operation), and mechanical effects (expansion/contraction of electrodes, mechanical stress) (Suresh et al., 2016). When the battery SoH drops below a critical value, the battery is considered to be of poor health and is removed from operation. A newly produced battery would have 100% SoH. The determination of battery SoH is a challenging task as it cannot be directly measured. Several methods are reported in literature for SoH estimation, which includes using battery capacity degradation models developed using first principles models and battery state estimators and data-based lifetime estimation models developed using algorithms like fuzzy logic or neural networks (NN) trained using data collected from accelerated aging experiments (Berecibar et al., 2016; Severson et al., 2019). However, these techniques use various derived features as input and hence, are difficult to implement for industrial applications.

1.2. Battery State of Charge

One of the tasks of the Battery Management System (BMS) is the accurate estimation of the energy stored within the battery for its efficient utilization by avoiding situations of over-charge and over-discharge in a battery. The battery SoC is a physical quantity of the battery that is related to the average Li^+ ion concentration in the negative electrode, which in turn is related to the passage of current during charging or discharging. The battery SoC is a derived quantity and cannot be measured directly. There are various model-based SoC estimation methods where the model could be first principles-based, surrogate first principles-based, data-based or hybrid models (Suresh et al., 2016). Neural network is a powerful and scalable machine learning (ML) tool that can handle multiple inputs and outputs and has good generalization and self-learning abilities. This makes NN suitable for modeling battery characteristics for SoC estimation. Multilayer perceptron architecture trained using the backpropagation algorithm has been used by researchers to model a battery using labeled charge-discharge cycling data available in various

repositories that are obtained through carefully designed experiments. Some of the works cited in the literature (Chemali et al., 2018) use NNs to predict the SoC using battery states such as measured current, voltage, temperature, remaining capacity assuming that the samples are independent of each other. However, the SoC in a battery not only depends on the present values of the independent features like current and voltage, but it also depends on the path taken to reach the present state. In other words, SoC at any time t also depends on the past values of SoC, current, voltage, etc. To account for these causalities, a NN-based NARX (non-linear auto-regressive model with exogenous inputs) model is used in this paper as it has capabilities for time-series modeling, unlike a regular neural net.

2. Battery Data

The data is collected from 48V-3.9kWh battery modules with NCA (Nickel Cobalt Aluminum) electrochemistry used for telecommunication application. The battery module consists of 14 cells, PCBs, and a built-in Battery Management System (BMS). A new battery has 100% SoH, which implies that it has full capacity when it is completely charged. When the battery reaches its end of life, the value of SoH variable is 0%, and this indicates that the battery capacity is 70% of what it had at the beginning of its life. The module starts discharging whenever the network voltage drops below the battery terminal voltage with a lower cut-off voltage of 42Vdc. The module recharges with a constant current/constant voltage (CC/CV) profile when the terminal voltage of the battery drops below the network voltage with an upper cut-off voltage of 56 Vdc. The time-series data for the module charge/discharge current, maximum module temperature, module voltage, SoC and SoH sampled at every 10 minutes is collected from battery modules in four different health condition defined as Site I (100% SoH) , Site II (74% SoH), Site III (49% SoH) and Site IV (24% SoH) over two different time periods named as Dataset 1 (total 24735 data samples) and Dataset 2 (27503 data samples).

3. SoH Classification

The objective is to identify the health status of the battery by building a model using the current, voltage, temperature, and SoC data from the four sites as input features for capturing the underlying behavior and predict the SoH for a new test dataset. An ensemble supervised learning technique of random forest classifier is implemented for classifying the data samples collected from the battery to determine the health of the battery. The random forest classifier is a versatile classification tool that makes an aggregated prediction using a group of decision trees trained using the bootstrap method with extra randomness while growing trees by searching for the best features among a randomly selected feature subset. The input features are first scaled using Min-Max normalization to bring all values into the range $[0, 1]$. 70% of the total data in Dataset 1 (17314 samples) is used for training, and the remaining 30% (7421 samples) is used for testing the predictive ability of the classifier for the same dataset.

Accuracy is the fraction of samples that are correctly identified with respect to the total number of samples tested. Among 7421 test samples, the algorithm correctly identified 6447 samples and hence, the accuracy is 86.87%. The performance of the developed classifier on the test set of Dataset 1 is evaluated using various metrics and is provided in Table 1 along with the confusion matrix. It could be observed that the recall for Site II data is 0.52, which implies that around half the Site II data points were misclassified.

To determine the ability of the developed model in predicting the SoH for a different instant of operation, we tested the developed model on Dataset 2. The input features of Dataset 2 scaled using Min-Max normalization (based on the training data set used for developing the classifier) is used for determining the SoH. The classification using the random forest model yielded an accuracy of 53.19%. Out of the total mislabelled samples, 51% of samples are from Site II. The confusion matrix and various performance metrics for Dataset 2 are presented in Table 2. The precision, recall, and F1 score of Site II are very low compared to other classes indicating that for all the data points that truly belongs to Site II, the percentage of them being labeled as Site II is only 8%, i.e., the number of false negatives is higher for Site II data. It is possible that the true data for Site II is flawed. Site I has very large false positives (Type I error) while site IV has the highest Type II error (False negatives). Site III data has an overall low F1 score.

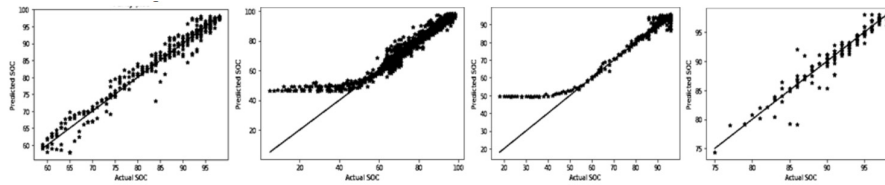


Figure 1: Parity plots of random forest-based SoC prediction for Dataset 2

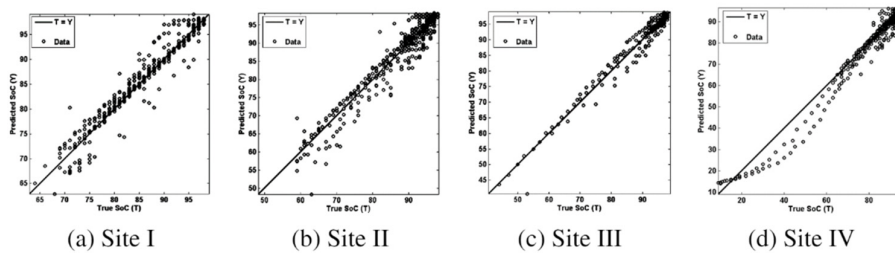


Figure 2: Parity plots of NARX-based SoC prediction for Dataset 1

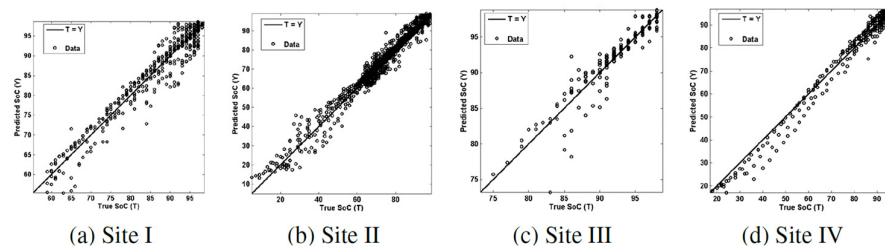


Figure 3: Parity plots of NARX-based SoC prediction for Dataset 2

4. SoC Prediction

A random forest model using time-lagged variables from Site I data of Dataset 1 was implemented to predict the dynamic variation of SoC. The input and output orders were assumed to be 2. When tested on Dataset 2, the model always predicted SoC to be greater than 40%, even though the true SoC was varying between 20% to 100% (ref Figure 1). Whenever the true SoC was below 40, the model predicted it to be close to 40%. The training data obtained from the battery site 1 had battery SOC values in a limited range and did not span the entire 0-100%. Thus, the model was not able to generalize well and

the prediction was of poor quality as shown in the parity plots (Figure 1). NN-based NARX model was used later to improve the quality of SoC prediction.

NARX is a time series modeling method used for modeling non-stationary and non-linear multivariate time series. The NARX model is described by the following equation:

$$y(t) = F(u(t - d_u), \dots, u(t - 1), u(t), y(t - d_y), \dots, y(t - 1)) \quad (1)$$

$u(t)$ and $y(t)$ represent input and output of the model at time t , F is a non-linear function, and d_u and d_y are the input and output order, respectively. The process time delay is assumed to be 0. NARX neural network model is a dynamic recurrent neural network consisting of a feedback connection between its layers. The network prediction is a weighted sum of non-linearly transformed regressor variables. A two-layer feedforward NARX network is used in this work with a hidden layer having a sigmoid transfer function and the output layer having a linear transfer function. The previous values of the input variables and the predicted output variables are stored using tapped delay lines. The output, $y(t)$, being a function of its past values, is sent back to the input layer of the network (through delays). Series-parallel architecture is used to train the network for more accurate predictions using the observed values of battery SoC instead of feeding back the estimated SoC values. This open loop configuration thus becomes similar to feed-forward architecture, where backpropagation could be used for training the network. The model is built using battery current, voltage, temperature, and battery SoH as input features. The SoH information used for SoC prediction is the measured values obtained from battery sites. In case there is no SoH information available, the SoH classifier mentioned in the paper could be used to estimate the SoH, and the accuracy of the SoC prediction will depend upon the accuracy of the classifier. The values of d_u and d_y are assumed to be 2.

Table 1: Confusion matrix and performance metrics of the 4-class SoH classifier on Dataset 1

Dataset 1		Predicted				Dataset 1	Precision	Recall	F1 score
		Site I	Site II	Site III	Site IV				
Actual	Site I	1830	87	114	0	Site I	0.9	0.9	0.9
	Site II	158	738	514	3	Site II	0.84	0.52	0.64
	Site III	40	48	1903	2	Site III	0.75	0.95	0.84
	Site IV	0	6	2	1976	Site IV	1	1	1
						Avg\total	0.88	0.87	0.86

Table 2: Confusion matrix and performance metrics of the 4-class SoH classifier on Dataset 2

Dataset 2		Predicted				Dataset 2	Precision	Recall	F1 score
		Site I	Site II	Site III	Site IV				
Actual	Site I	6813	229	111	5	Site I	0.47	0.95	0.62
	Site II	6270	543	343	3	Site II	0.22	0.08	0.11
	Site III	1073	1528	3445	58	Site III	0.53	0.56	0.55
	Site IV	489	131	2628	3835	Site IV	0.98	0.54	0.74
						Avg\total	0.55	0.53	0.49

70% of the data available for Site I (4642 samples) is used for building the predictive model, and 15% (995 samples) is used for the validation step that ensures network generalization. The remaining 15% of the data from Site I data (995 samples) is used for testing the model. The SoC estimation model is tested on the data from Site II, Site III and Site IV. A single hidden layer is used and the number of neurons in that hidden layer is set to 3. The network is trained using the Levenberg-Marquardt algorithm. The neural net model at epoch 38 was chosen as the best model as it had the least validation error. The training, validation and testing results on Site 1 data of Dataset 1 show very good performance with small RMSE values and R^2 values close to 1 (refer Table 3). SoC prediction using the developed model on Sites II, III and IV of Dataset 1 also show good performance with R^2 values above 0.95 as shown in Table 3. The linear model that best

fits true SoC and predicted SoC for the test set almost coincides with the 45-degree line indicating the goodness of the model as shown in the parity plots presented in Figure 2. Using the NN model built with Site I data of Dataset 1, SoC values are predicted for all the four sites' data in Dataset 2. The results are presented in Figure 3 and Table 4. It could be observed that the model predicts the SoC values for all four classes reasonably well with R^2 values greater than 0.94 in all cases.

Table 3: Performance metrics of NARX neural network on Dataset 1

Site I: Dataset	Mean of residual	Variance of residual	RMSE	R^2	Dataset 1	Mean of residual	Variance of residual	RMSE	R^2
Traihing	3.6×10^{-5}	7.3×10^{-5}	0.008	0.97	Site II	-0.054	0.793	0.892	0.957
Validation	-2.4×10^{-3}	1.4×10^{-4}	0.012	0.99	Site III	0.069	0.160	0.407	0.974
Testing	-0.105	1.93	1.39	0.97	Site IV	0.089	0.985	0.997	0.970

Table 4: Performance metrics of NARX neural network on Dataset 2

Dataset 2	Mean of residual	Variance of residual	RMSE	R^2
Site I	-0.198	2.122	1.47	0.987
Site II	-0.082	0.576	0.763	0.960
Site III	0.072	0.106	0.334	0.942
Site IV	-0.007	0.533	0.730	0.983

5. Conclusion

This paper has two major contributions to battery health monitoring. Firstly, a data-driven model is built using a random forest algorithm on BMS data and is able to determine the health condition of the battery and identify whether it is in good, bad, worse, or worst health. This is critical as battery SoH degradation affects the accurate estimation of battery SoC. Secondly, a NN-based NARX model is developed for SoC estimation taking into consideration the non-linearity, dynamic nature and SoH of the system. The accuracy of predictions shows that the machine learning approaches trained using actual battery data could provide prognostic solutions to battery health management without the need for a complex physics-based model. The future work aims to develop a predictive model for estimating the remaining useful life of battery based on a hybrid model built using battery data and first-principles model for capacity degradation.

6. Acknowledgement

We thank Reliance Industries Ltd. and Reliance Jio Infocomm Ltd. for providing the datasets used in this article. We also thank Raghuvver Vadari from RIL for helpful discussions.

References

- Berecibar, M., Gandiaga, I., Villarreal, I., Omar, N., Van Mierlo, J., & Van Den Bossche, P. (2016). Critical review of state of health estimation methods of Li-ion batteries for real applications. In *Renewable and Sustainable Energy Reviews* (Vol. 56, pp. 572–587). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2015.11.042>
- Chemali, E., Kollmeyer, P. J., Preindl, M., & Emadi, A. (2018). State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach. *Journal of Power Sources*, 400(August), 242–255. <https://doi.org/10.1016/j.jpowsour.2018.06.104>
- Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M. H., Aykol, M., Herring, P. K., Fraggadakis, D., Bazant, M. Z., Harris, S. J., Chueh, W. C., & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5), 383–391. <https://doi.org/10.1038/s41560-019-0356-8>
- Suresh, R., Kumar Tanneru, H., & Rengaswamy, R. (2016). Modeling of rechargeable batteries. *Current Opinion in Chemical Engineering*, 13, 63–74. <https://doi.org/10.1016/j.coche.2016.08.005>

The Adaptive Radial Basis Method (ARBM): An application to the hydrodynamics of liquid-liquid dispersions

Samer Alzyod

CEA, DES, ISEC, DMRC, Univ. Montpellier, Marcoule, France

samer.alzyod@cea.fr, samer.alzyood@yahoo.com

Abstract

In industrial unit operations involving liquid-liquid dispersions, switching to model based prediction is essential for equipment design, scale-up and control purposes. In this regard, an Adaptive Radial Basis Method (ARBM) is presented to model the hydrodynamics behaviour of liquid-liquid dispersions, based on the Population Balance Equation (PBE). In this meshfree method, the number density function is expanded by a series of Gaussians (GA) basis at a finite number of particles along the droplet diameter. The supporting particles are allowed to move to change their locations, based on the distribution properties, to accurately track the dynamic evolution of the number density function. Moreover, the Leave-One-Out Cross Validation (LOOCV) algorithm is adapted to optimize the GA shape parameter in a statistical sense. Accordingly, the ARBM neither inverts badly scaled matrices nor requires large number of collocation particles to reconstruct the distribution shape. The ARBM predictions are validated for stirred tank reactors at different levels. The results revealed that, the method is conservative and able to handle sharp and wide distributions, resulting from breakage and coalescence dominant processes respectively. Based on this, the ARBM is convenient and suitable to model liquid-liquid dispersions.

Keywords: ARBM, PBE, breakage and coalescence, Liquid-liquid dispersions

1. Introduction

Liquid-liquid dispersions are very common in different unit operations ranging from batch stirred tank reactors to solvent extraction columns. In such unit operations, the droplet size distribution is the key parameter, which determines the overall efficiency and the range of the operating conditions. Therefore, an accurate prediction of the size distribution is of prime importance. Due to the growing computational power, the computer-based design approach and process digitalization have gained more attention recently. Consequently, there is a growing demand toward rigorous models, which are able to take all of the possible hydrodynamics interactions into account, to describe the prevailing physical behaviour. Unlike the lumped models, which ignore the discreet nature of the dispersion, the population balance model is based on more fundamental basis to account for the local droplet-droplet interactions, which include the breakage and coalescence phenomena. The PBE has a complex integro-differential form and hence analytical solutions are limited to simplified special cases. As a result of this, intensive research efforts have been carried out to develop efficient numerical methods, in terms of

accuracy and computational cost, to solve the PBE. These methods can be categorized into three main groups namely: stochastic Monte Carlo, moments-based and sectional methods. For more information about these methods, the reader can refer to Kumar and Ramkrishna (1996). In a recent work, a novel Radial Basis Method (RBM) (Alzyod and Charton, 2020) to solve the PBE was presented. The RBM avoids the typical numerical diffusion problems, encountered when applying the sectional methods, and conserves the number density function, which is lost when using the moments based methods. In spite of the RBM accuracy, more collocation particles may be required when dealing with sharp or wide moving distributions, for which additional computational efforts may be required. This inherent problem is also reported in the radial basis functions literature, where a set of adaptive particles can be used to avoid it (Driscoll and Heyudono, 2007). Therefore, the main concern of this work is to derive an adaptive version of the RBM to overcome this fundamental problem.

2. Solution of the PBE using the Adaptive Radial Basis Method (ARBM)

In a homogenous space, the dynamic evolution of the number density function, $n(d, t)$, accounting for breakage and coalescence is given by (Kumar and Ramkrishna, 2000):

$$\begin{aligned} \frac{\partial n(d, t)}{\partial t} = S\{n(d, t)\} = & -\Gamma(d)n(d, t) + \int_d^\infty \Gamma(d)\beta(d/d')n(d', t)\delta d \\ & - n(d, t) \int_0^\infty \omega(d, d')n(d', t)\delta d + \frac{d^2}{2} \int_0^d \eta^{-2/3}\omega(\eta, d')n(\eta, t)n(d', t)\delta d \end{aligned} \quad (1)$$

Where d is the droplet diameter, $\eta(d, d') = \sqrt[3]{d^3 - d'^3}$, $\Gamma(d)$ is the breakage kernel, $\beta(d/d')$ is the daughter distribution function, and $\omega(d, d')$ is the coalescence kernel respectively. The main idea behind the ARBM is to replace the unknown nonnegative number density function appearing in Eq.(1) by a series of converging GA radial basis kernels:

$$\hat{n}(d(t), t) = \left(\sum_{j=1}^{N_c} \lambda_j(t) e^{-\left(\varepsilon_p r_j(d(t))\right)^2} \right)^2 \quad (2)$$

Where $\hat{n}(d(t), t)$ is the reconstructed number density function, $\lambda_j(t)$ is the weighting coefficient, and ε_p is the shape parameter. The norm $r_j(d(t)) = \|d(t) - d_j(t)\|$ represents the Euclidean distance between d and the centre particles d_j . Here, the centres are allowed to change their locations, to track the moving density function as follows:

$$d_j(t) = d_1(t) + \left(\frac{d_{N_c}(t) - d_1(t)}{N_c - 1} \right) (j - 1) \quad j = 1, \dots, N_c \quad (3)$$

Where $d_j(t) \in [d_1(t), d_{N_c}(t)]$. The values of $d_1(t)$ and $d_{N_c}(t)$ are calculated using the first three low-order moments as follows:

$$d_{1, N_c}(t) = \frac{\mu_1}{\mu_0} + \left(1 \mp 6\sqrt{\mu_1\mu_2/\mu_1^2 - 1}, \varepsilon_m \right) \quad (4)$$

Where $\mu_k(t)$ is the k^{th} moment, ε_m is the machine epsilon and the operator $\langle \cdot \rangle$ is used to maintain the maximum positive value. Note that, the adapted centers are distributed around the arithmetic mean (μ_1/μ_0) . Therefore, in case of very sharp distributions (approaching zero variance) the ARBM is still able to reconstruct the distribution located at the mean particle location. To reconstruct the number density function, the required unknown weighting coefficients $(\lambda_j(t))$ appearing in Eq.(2) can be calculated by enforcing the condition: $\sqrt{\hat{n}(d_j(t),t)} = \sqrt{n(d_j(t),t)} = s_j(t)$ for all $j = 1, \dots, N_c$ at the adaptive moving particles, which results in following linear system of equations:

$$\begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_{N_c}(t) \end{bmatrix} = \begin{bmatrix} 1 & e^{-(\varepsilon_p \|d_1(t)-d_2(t)\|)^2} & \dots & e^{-(\varepsilon_p \|d_1(t)-d_{N_c}(t)\|)^2} \\ e^{-(\varepsilon_p \|d_2(t)-d_1(t)\|)^2} & 1 & \dots & e^{-(\varepsilon_p \|d_2(t)-d_{N_c}(t)\|)^2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-(\varepsilon_p \|d_{N_c}(t)-d_1(t)\|)^2} & e^{-(\varepsilon_p \|d_{N_c}(t)-d_2(t)\|)^2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \lambda_1(t) \\ \lambda_2(t) \\ \vdots \\ \lambda_{N_c}(t) \end{bmatrix} \quad (5)$$

This system of equations can be expressed in a compact matrix form as: $S = \phi\lambda$, where ϕ is the radial basis matrix. Note that, ϕ is a real symmetric and positive definite square matrix and hence the resulting linear system of equations has always a unique solution. Indeed, solving this system of equations requires inverting the radial basis matrix, which may lead to inaccurate results if it is nearly singular. To avoid this problem, a proper shape parameter value can be used to minimize the condition number of ϕ . To achieve this, the LOOCV algorithm of Rippa (1999) is adapted and utilized to find the optimal shape parameter value. In this algorithm, the optimal shape parameter is obtained by minimizing the following objective function:

$$e(\varepsilon_p) = \sum_{j=1}^{N_c} |\hat{n}(d_j, t) - n(d_j, t)|^2 = \sum_{j=1}^{N_c} |\lambda_j / \phi_{j,j}^{-1}|^2 \quad (6)$$

Here, $e(\varepsilon_p)$ is the error between the reconstructed and targeted distributions respectively, $\phi_{j,j}^{-1}$ is j^{th} diagonal element of the inverse of radial basis matrix ϕ^{-1} . Note that, using Eq.(6) allows us to minimize the objective function without any prior information about the targeted distribution. Based on this, the resulting linear system of equations given by Eq.(5) can be solved using Cholesky decomposition based on the obtained optimal shape parameter value. Finally, by applying the collocation method at the adapted moving particles locations, Eq.(1) is transformed into the following system of ordinary differential equations:

$$\frac{\partial \hat{n}(d_j(t), t)}{\partial t} = S \{ \hat{n}(d_j(t), t) \} \quad j = 1, \dots, N_c \quad (7)$$

3. Model validation and discussion

This section is devoted to test the ARBM performance. In this regard, the explicit forward first-order Euler's method is implemented to solve the ARBM with a constant time step of $\Delta t = 0.01$ s. Moreover, the integral source terms are approximated using Gauss-Legendre quadrature. To obtain the optimal shape parameter value, a constrained LOOCV-optimization problem is solved using the MATLAB *fmincon* solver. All of the numerical experiments were carried out using a Laptop with 4 GB RAM and 2.4 GHz speed performing a single task.

3.1. Case one: Analytical validation

The first case study is devoted to validate the ARBM predictions analytically for combined breakage and coalescence processes in a uniform stirred tank reactor. In this regard, the analytical solution of McCoy and Madras (2003) is used as a reference solution. The initial locations of the adapted particles are calculated from the initial condition ($3d^2 \exp(-d^3)$). Moreover, a constant coalescence kernel ($\omega = k_c$) is considered, while a power function is assumed for the breakage kernel ($\Gamma(d) = k_b d^3$). The simulation is performed using 25 adaptive particles and the final simulation time is set to 15 s. Figure (1) shows a comparison between the numerical the analytical solutions for two breakage and coalescence dominate cases at three selected time steps: 0, 5 and 15 s respectively.

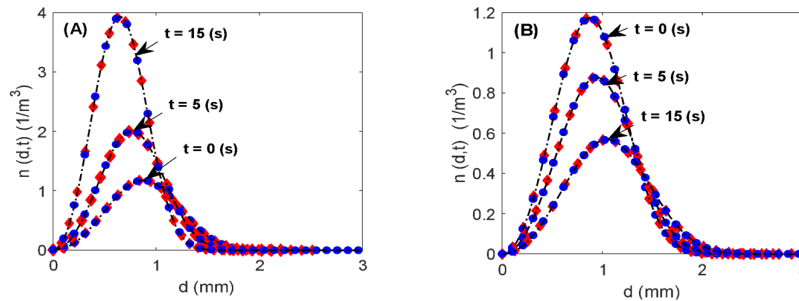


Figure (1): Comparison between the ARBM predictions at the moving particles (diamonds), and at selected stationary particles (circles) and the analytical solution (McCoy and Madras, 2003) (dashed line): (A) Breakage dominant process ($k_b = 0.1$, $k_c = 0.001$). (B) Coalescence dominant process ($k_b = 0.001$, $k_c = 0.1$).

It can be observed that, the adapted particles are able to track the moving distribution for both cases. Therefore, the ARBM reconstructs accurately the shape of distribution at the selected stationary points as compared with the analytical solution. To check the ARBM consistency, the first four low-order moments of the number density are computed and compared with the corresponding analytical moments as depicted in Figure (2). It is clear that, the total number of droplets (μ_0) increases with time for the breakage dominant case, while it decreases for the coalescence dominant case. Moreover, the total volume (μ_3) remains constant as expected for both cases since the total volume of droplets is conserved. Therefore, it can be concluded that the ARBM is a conservative method. An excellent agreement between the numerical and the analytical moments is evidenced.

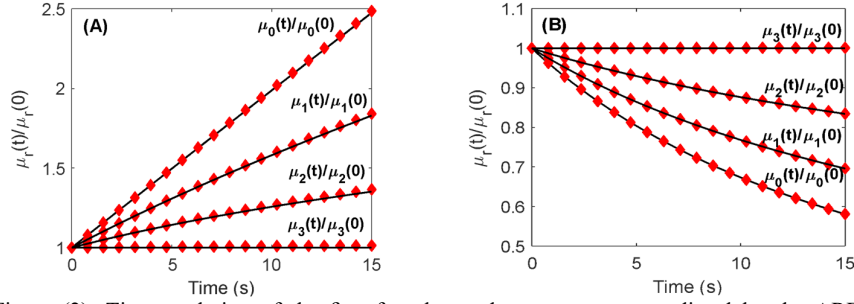


Figure (2): Time evolution of the first four low-order moments as predicted by the ARBM (diamonds), and comparison with the analytical solution (line): (A) Breakage dominant process ($k_b = 0.1$, $k_c = 0.001$). (B) Coalescence dominant process ($k_b = 0.001$, $k_c = 0.1$).

3.2. Case two: Numerical and experimental investigations

In the second case study, the ARBM is numerically validated using the RBM (Alzyod and Charton, 2020) and the Fixed Pivot Technique (FPT) (Kumar and Ramkrishna, 1996). In this regard, the reported experimental data by Qi et al., (2015), in a stirred tank, is used as a reference solution. The chemical test system is heptane in chloroaluminate. Here, the rotational speed is fixed at 500 rpm, while the holdup (α_d) is varied between 0.1 and 0.5 respectively. A normal distribution is used as an initial condition $(a \cdot \exp(-0.5((d - \bar{d})/\sigma)))$.

Concerning the hydrodynamics kernels, the breakage and coalescence kernels are given by (Coualoglou and Tavlarides, 1977):

$$\Gamma(d) = c_1 \frac{\varepsilon^{1/3}}{d^{2/3}(1 + \alpha_d)} \exp\left(-\frac{c_2 \sigma_d (1 + \alpha_d)^2}{\rho_c \varepsilon^{2/3} d^{5/3}}\right) \quad (8)$$

$$\omega(d, d') = c_3 \frac{\varepsilon^{1/3}}{1 + \alpha_d} (d + d')^2 (d^{2/3} + d'^{2/3})^{1/3} \exp\left(-\frac{c_4 \eta_c \rho_c \varepsilon}{\sigma_d^2 (1 + \alpha_d)^3} \left(\frac{dd'}{d + d'}\right)\right)^4 \quad (9)$$

Here c_1 , c_2 , c_3 and c_4 are fitting parameters. These parameters, as well as the initial distribution parameters, are estimated by solving an inverse PBE problem based on the ARBM and kept the same for the RBM and the FPT. The optimization is carried out, using MATLAB *fmincon* solver, to minimize the objective function: $F = \sum_i |n_i^{\text{exp}} - n_i^{\text{sim}}|^2$.

Based on this, the optimized parameters are shown in Table (1).

Table (1): Hydrodynamics optimized parameters

Holdup	a	\bar{d} (mm)	σ (mm)	C_1	C_2	C_3	C_4
0.1	0.1679	0.0892	0.0162	0.5	500.1971	0.3369	482.6460
0.5	0.2267	0.1052	0.0176	0.5	500.1971	0.6461	482.6461

Figure (3) depicts a comparison between the simulated steady state number density function, for two dispersed phase volume fractions, using the three numerical methods and the reported experimental data. A very good agreement is obtained for both cases using the same optimized parameters for all of the numerical methods.

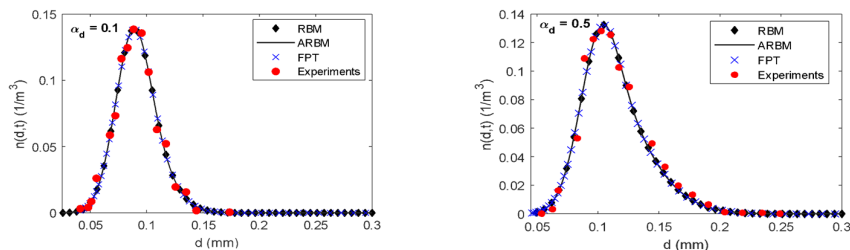


Figure (3): Comparison between the simulated number density and the experimental data at two different dispersed phase holdup values.

4. Summary and Conclusions

The ARBM is introduced to solve the PBE for breakage and coalescence. The method is validated analytically and numerically at different conditions. The ARBM is found to be conservative and able to track the moving number density thanks to the adaptive particles. To assess the ARBM performance, an experimental case study is presented to simulate the hydrodynamics of a stirred vessel. The required kernels parameters are estimated by solving an inverse PBE problem, where a positive validation is obtained.

Acknowledgment

The author would like to thank Dr. Sophie Charton (The French Alternative Energies and Atomic Energy Commission, CEA) for the fruitful discussion, critical reading and helpful comments during the preparation of the present chapter.

References

- S. Alzyod, S. Charton, 2020, A meshless Radial Basis Method (RBM) for solving the detailed population balance equation, *Chem. Eng. Sci.*, 228, 1-11.
- C. Coualoglou, L. Tavlarides, 1977, Description of interaction processes in agitated liquid-liquid dispersions, *Chem. Eng. Sci.*, 32(11), 1289-1297.
- T. Driscoll, A., Heryudono, 2007, Adaptive residual subsampling methods for radial basis function interpolation and collocation problems, *Comput. Math. Appl.* 53, 927-939.
- S. Kumar, D., Ramkrishna, 1996, On the solution of population balance equations by discretization- I. A fixed pivot technique, *Chem. Eng. Sci.*, 51, 1311-1332.
- B. McCoy, G. Madras, 2003, Analytical solution for a population balance equation with aggregation and fragmentation, *Chem. Eng. Sci.* 58, 3049-3051.
- L. Qi, X. Meng, R. Zhang, H. Liu, C. Xu, Z. Liu, P. Klusener, 2015, Droplet size distribution and droplet size correlation of chloroaluminate ionic liquid-heptane dispersion in a stirred vessel, *Chem. Eng. J.*, 268, 116-124.
- S. Rippa, 1999, An algorithm for selecting a good value for the parameter c in radial basis function interpolation, *Adv. Comput. Math.*, 11, 193-210.

Process modeling and multi-criteria optimization of an industrial phosphoric acid wet-process

Ilias Bouchkira^{a,b,c}, Abderrazak M. Latifi^{a,b,*}, Lhachmi Khamar^{b,d} and Saad Benjelloun^b

^a*LRGP, CNRS-ENSIC, Université de Lorraine, Nancy, France*

^b*Université Mohammed VI polytechnique, Benguerir, Maroc*

^c*LCPCE, Université Hassan II, Casablanca, Maroc*

^d*LIPIM, Université Sultan Moulay Slimane, Beni-Mellal, Maroc*
abderrazak.latifi@univ-lorraine.fr

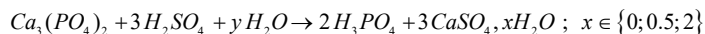
Abstract

This paper deals with modeling and multi-criteria optimization of an industrial phosphoric acid process. More specifically, the digestion tank where the phosphate ore is attacked by a concentrated solution of sulfuric acid is considered. A process model is first developed based on mass and heat balances along with Pitzer's thermodynamic equation. It is then identified by means of a set of data made up of experimental measurements carried out and measurements taken from the literature. The resulting model is then used within a multi-criteria optimization problem to determine the optimal operating conditions that minimize the phosphate losses and increase the productivity of the phosphoric acid while meeting the constraints on the temperature and on the excess of sulfuric acid. The optimization problem is solved by means of the epsilon-constraint method and the results are consistent with the current operating conditions of the digestion tank. They also show that the implementation of optimal results on an industrial scale would significantly improve the performance of the tank and under maximum safety conditions.

Keywords: Modeling, Multi-criteria optimization, Wet-process, Measurements, Phosphoric acid, Phosphate ore.

1. Introduction

Phosphoric acid is mainly produced by digesting a phosphate-containing mineral (generally apatite) using a concentrated sulfuric acid solution. The digestion is carried out in the wet process which is by far the most common route to produce phosphoric acid. The main reaction that takes place in the digestion tank is highly exothermic and given by Becker (1989) as:



During the digestion, sulfuric acid H_2SO_4 dissociates into H^+ and SO_4^{2-} ions. The H^+ ions react with the phosphate compounds containing phosphorus to produce phosphoric acid H_3PO_4 , and simultaneously the SO_4^{2-} ions are captured by the calcium ions Ca^{2+} to produce solid calcium sulfate $CaSO_4 \cdot xH_2O$. The two reaction products are then separated in a filtration unit downstream of the digestion tank. Phosphoric acid is recovered and mainly used in the production of phosphate fertilizers, while the calcium sulfate is removed from the process as an unwanted by-product.

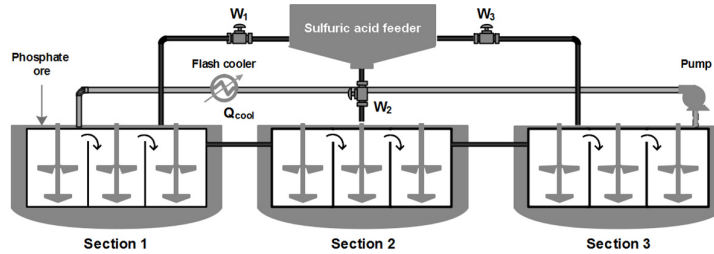


Figure 1: Schematic representation of an industrial phosphate digestion tank

On an industrial scale, the reaction is carried out in a cylindrical tank made up of nine continuous reactors in series, of the same volume and uniformly distributed. The tank is divided into three sections of three reactors each. Sulfuric acid feed flow rate is divided into three streams each feeding a section of the tank. The phosphate ore is fed into the first reactor of the first section and the product leaves the tank at the last reactor of the third section (Fig.1).

Depending on the operating temperature, the calcium sulfate can be gypsum $CaSO_4 \cdot 2H_2O$ or bassanite $CaSO_4 \cdot 0.5H_2O$. It is therefore very important to control the temperature in order to avoid the production of the bassanite since it increases the viscosity of the reaction mixture in the tank and consequently decreases the filtration performances. Moreover, sulfuric acid feed flow rate and its distribution ratios ($w_1; w_2; w_3$) over the three sections are relevant for optimal operation of the production unit. Indeed, a sulfuric acid deficit during the digestion step causes a decrease in the concentration of SO_4^{2-} ions and consequently the Ca^{2+} free ions tend to capture the phosphate-based ions of the same polarity. HPO_4^{2-} ions are more specifically captured to produce the solid brushite $CaHPO_4 \cdot 2H_2O$ thus decreasing the chemical yield of the process since all the solid phase is removed as an unwanted product. These losses are referred to as syncrystallized losses. Conversely, an excess of sulfuric acid causes on the one hand a local increase in temperature due to the heat of dilution, and on the other hand leads to an excessive production of calcium sulfate which forms a solid layer around the phosphate particles. This is referred to as the coating phenomenon which prevents the particles from further attack by sulfuric acid thus decreasing also the chemical yield. These resulting losses are termed unattacked losses.

In this paper, the objective is to determine the operating conditions of the digestion tank that minimize the phosphate ore losses (syncrystallized and unattacked) while meeting the constraints on the excess of sulfuric acid above the stoichiometric amount and on the reaction temperature. Beforehand, Pitzer's thermodynamic model, needed in optimization, is calibrated using experimental measurements.

2. Optimization problem formulation

The optimal operation of the process described above aims to determine the trade-offs between the following objectives: (i) limit the syncrystallized losses by minimizing the production of the brushite, (ii) improve the process productivity by maximizing the production of gypsum, (iii) minimize the unattacked losses by keeping the excess of sulfuric acid above the stoichiometric amount between two limit values set to ensure optimal operation of the process, (iv) control the reaction temperature in order to limit the production of the bassanite and consequently improve the filtration process downstream.

These objectives are taken into account in the formulation of a constrained multi-criteria optimization problem. The first two objectives are defined as optimization criteria while the last two objectives are taken as constraints. The objective functions, the decision variables, the equations of the process model and the constraints of the optimization problem are detailed in the next section.

2.1. Objective functions

In this optimization problem, the two (conflicting) optimization criteria used are defined by the saturation index of the brushite to be minimized all over the three sections of the digestion tank to limit its production, and the saturation index of the gypsum to be maximized to improve the productivity. The two indices are expressed as:

$$f_1 = \sum_j \log \left(a_{Ca^{2+}} a_{HPO_4^{2-}} k_{s_1}^{-1} \right)^j, \quad f_2 = \sum_j \log \left(a_{Ca^{2+}} a_{SO_4^{2-}} k_{s_2}^{-1} \right)^j \quad (1)$$

where a_j are the activities of the components in the reaction mixture, k_{s_1} and k_{s_2} are the solubility of brushite and gypsum respectively and, j refers to the number of the section.

2.2. Decision variables

The decision variables consist of sulfuric acid excess S_c over the stoichiometric amount required by the reactions, its distribution ratios over the three sections, i.e. ($w_1; w_2; w_3$), and the cooling heat Q_{cool} to be removed from the process to control the temperature.

2.3. Equality and inequality constraints

The constraints involved in the optimization problem include (i) the operating constraints of the process, (ii) and the model equations. Two operating constraints are more specifically involved in the problem. The first one sets the values of the upper and lower limits of the excess of sulfuric acid above the stoichiometric amount required by the reactions in order to minimize the unattached losses. These limits are set at 1 % and 3 % (Becker, 1989). The second one concerns the reaction temperature that should remain below 355 K to avoid the production of the bassanite:

$$1 \% \leq S_c \leq 3 \% \quad \text{and} \quad T \leq 355 K \quad (2)$$

The process model equations consist of material balance Eq. (3), charge balance Eq. (4), equations of equilibrium constants Eqs. (5-6), and heat balance Eq. (7).

$$(M)^{tot} = \sum_{i=1}^{NC} \delta_{M,i} m_i, \quad M \in \{S, P, F, Si, Ca\} \quad (3)$$

$$\sum_{i=1}^{NC} z_i \cdot m_i = 0 \quad (4)$$

$$K_j = \prod_{i=1}^{NC} (m_i \cdot \gamma_i)^{\alpha_{ij}}, \quad j = 1, \dots, NR \quad (5)$$

$$\log(K_j) = \log(K_{j0}) + \frac{\Delta H_j}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right) \quad (6)$$

$$Q_H + Q_{AG} + Q_{R+D} = Q_{cool} + Q_{out} \quad (7)$$

where NC and NR are the number of components and reactions respectively, $(M)^{tot}$ is the given total inlet concentration of element M (i.e. sulfur (S), phosphorus (P), Fluorine (F), silica (Si), calcium (Ca)), m_i is the unknown molality of the component i involved in the reactions listed in Table 1, $\delta_{M,i}$ is equal to the number of element M in component i . For example, the mass balance on (F) develops as: $(F)^{tot} = 6 m_{H_2SiF_6} + m_{F^-} + 6 m_{SiF_6^{2-}} + m_{HF} + 2 m_{HF_2^-}$. z_i are the electrical charges, γ_i is the activity coefficient of component i . $\alpha_{i,j}$ is the stoichiometric coefficient of component i involved in reaction j . K_{j0} and ΔH_j refer to the equilibrium constant and the enthalpy of the reaction j at $T_0 = 298 K$ respectively, their values are reported in Table 1. Q_H , Q_{AG} , Q_{R+D} , Q_{out} and Q_{cool} are the enthalpy of the inlet products, the agitation heat, the sulfuric acid dilution heat, the enthalpy of the slurry leaving the reactor and the heat to be removed from the process to control the operating temperature respectively.

In addition to the above equations, Pitzer's thermodynamic model (Pitzer, 2018) is used to compute the activity coefficients as:

$$\ln(\gamma_i) = \frac{z_i^2}{2} f + 2 \sum_j m_j \lambda_{i,j} + z_i^2 \sum_{j,k} m_j m_k \lambda'_{i,j} + 3 \sum_{j,k} m_j m_k \psi_{i,j,k} \quad (8)$$

where i, j, k refer to different components and f is the Debye-Huckel function. $\lambda_{i,j}$, $\lambda'_{i,j}$ and $\psi_{i,j,k}$ are the unknown Pitzer model parameters, they should be determined from the equilibrium measurements carried out in this work and presented in the next section. Furthermore, a suitable estimability analysis method is used to determine the most estimable parameters from the available experiments.

3. Experimental data

Different series of experimental measurements of sulfuric and phosphoric acid solutions have been carried out to measure the molalities of the ionic and molecular species involved in the reaction mixture. For all experiments, 98 w% sulfuric acid and 65 w% phosphoric acid are used to prepare several samples. The temperature and concentration of samples are varied respectively from 298 K to 353 K and from infinite dilution to 4 moles/kgw for sulfuric acid and 12 moles/kgw for phosphoric acid. Moreover, calcium sulfate solubility data are collected from the literature (Shen et al., 2020) they consist of gypsum and anhydrite solubilities for a large range of temperatures and concentrations.

4. Results and discussion

The estimability of the unknown parameters of Pitzer's model is carried out using our recently developed method (Bouchkira et al., 2021) and the most estimable parameters are identified by minimizing the mean-square error between the predictions of the model and the available experiments. As example of results, Figs. 2A-C show the sulfuric and phosphoric acid speciation and the calcium sulfate solubility at 298 K. The high values of the Pearson product-moment coefficient r show the good accuracy of the developed model. The multi-criteria optimization problem is then solved within *GAMS* environment using Baron solver based on a branch-and-reduce algorithm to achieve global optimality. The ε -constraint method (Deb, 2014) was used to transform the multi-criteria optimization problem into a single-objective optimization problem which is solved many times to determine the Pareto front. The *CPU* time needed to perform the simulations is about two hours using a Dell Precision T7810 Bi-Xeon 12 x Core 64GB workstation.

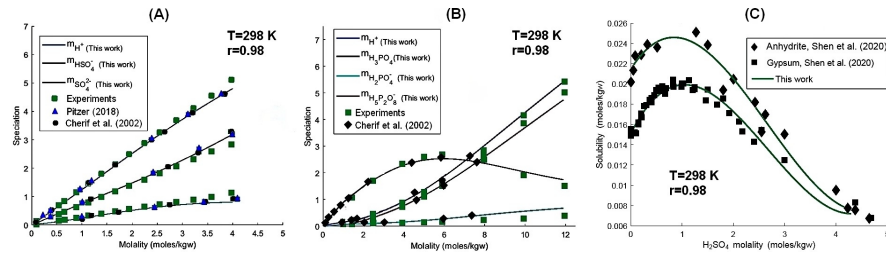


Figure 2: (A, B): Sulfuric and Phosphoric acids speciation, (C): Gypsum and Anhydrite solubility

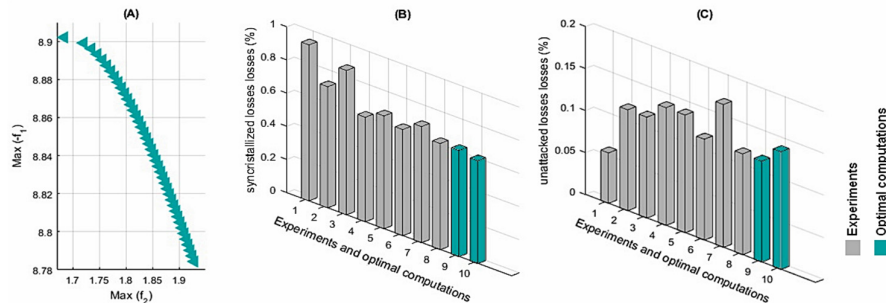


Figure 3: (A): Pareto Front, (B): Syncrystallized losses, (C): Unattacked losses

The Pareto front of the problem (Fig. 3A) shows that: (i) the brushite saturation index is negative for all the optimal solutions meaning that the syncrystallized losses are minimized, (ii) and the gypsum saturation index is positive meaning that the productivity of the industrial units is maximized. To better visualize the optimal solutions, four points of the Pareto front (Fig.3A) are shown in Table 2 with their corresponding decision variables, constraints, cooling heat and phosphate losses. It can be seen that for each point, as the conversion rate of digestion increases from the first to the last section, the demand for sulfuric acid varies in the opposite direction and decreases from the first to the third section for all the computed solutions (i.e., $w_1 \geq w_2 \geq w_3$). Furthermore, sulfuric acid excess lies between 2 % and 3 % which is consistent with the optimal conditions of gypsum crystallization (Becker, 1989). Finally, the heat quantities to be removed via the flash cooler (Fig.1) enable to meet the reaction temperature constraint in order to limit the bassanite production, and to improve the quality of the downstream filtration unit.

Figs. 3B-C show a comparison between industrial losses taken from (El Bouzidi, 2016; Bouchkira and El Fariq, 2017) and the results of two computed optimal solutions chosen from the Pareto front using the Multi-Attribute Utility Theory (Dyer, 2005). The computed solutions are very close to the industrial measurements which are obtained by performing several run-to-run on the plant by changing the values of the decision variables from one run to the next to improve the productivity and minimize the phosphate losses. Moreover, the values of the decision variables used to carry out industrial experiments are sub-optimal and must be adjusted depending on the composition of the phosphate ore which further increases the number of runs. It is easy to understand under these conditions that obtaining these measurements is very costly since it involves significant labor cost, consumes time and huge quantities of reagents, decreases the productivity of the plant and poses some safety issues. Modeling and multi-criteria optimization have proven to be effective since they provide optimal solutions that minimize the number, duration, cost of runs, and security risks.

Table 1: Equilibria involved in the digestion reaction (Aqueous phase)

Equilibria	K_{j0}	ΔH_j (J/mol)	References
1 $\text{H}_2\text{SO}_4 \rightarrow \text{HSO}_4^- + \text{H}^+$	-	-	-
2 $\text{HSO}_4^- \rightleftharpoons \text{SO}_4^{2-} + \text{H}^+$	0.0103	$-1.274 \cdot 10^4$	(Gustafsson, 2011; Pitzer, 2018)
3 $\text{H}_3\text{PO}_4 \rightleftharpoons \text{H}_2\text{PO}_4^- + \text{H}^+$	0.0071	$-4.271 \cdot 10^3$	(Gustafsson, 2011; Ball and Nordstrom, 1991)
4 $\text{H}_2\text{PO}_4^- \rightleftharpoons \text{HPO}_4^{2-} + \text{H}^+$	$4.2 \cdot 10^{-12}$	$-9.266 \cdot 10^3$	(Gustafsson, 2011; Ball and Nordstrom, 1991)
5 $\text{H}_3\text{PO}_4 + \text{H}_2\text{PO}_4^- \rightleftharpoons \text{H}_5\text{P}_2\text{O}_8^-$	0.2550	$-1.329 \cdot 10^4$	(Gustafsson, 2011)
6 $\text{HF} \rightleftharpoons \text{F}^- + \text{H}^+$	$7.2 \cdot 10^{-4}$	$-1.329 \cdot 10^4$	(Gustafsson, 2011)
7 $\text{HF} + \text{F}^- \rightleftharpoons \text{HF}_2^-$	$0.3 \cdot 10^2$	$-4.799 \cdot 10^4$	(Gustafsson, 2011)
8 $\text{H}_2\text{SiF}_6 \rightleftharpoons \text{SiF}_6^{2-} + 2\text{H}^+$	$0.3 \cdot 10^2$	$-4.799 \cdot 10^4$	(Gustafsson, 2011)
9 $\text{CaSO}_4 \rightleftharpoons \text{Ca}^{2+} + \text{SO}_4^{2-}$	0.6394	$-7.200 \cdot 10^3$	(Pitzer, 2018; Shen et al., 2020)

Table 2: Four solutions from the Pareto front (Syn: syncrystallized; Una: Unattacked)

	w_1	w_2	w_3	T^{opt} (K)	S_c (%)	Q_{cool} (Kj/h)	Syn. losses (%)	Una. losses (%)
1	0.73	0.27	0	349.45	2.12	$-1.333 \cdot 10^5$	0.62	0.12
2	0.70	0.30	0	350.25	2.18	$-1.351 \cdot 10^5$	0.64	0.14
3	0.67	0.33	0	350.75	2.23	$-1.355 \cdot 10^5$	0.63	0.16
4	0.63	0.37	0	351.35	2.54	$-1.358 \cdot 10^5$	0.64	0.15

5. Conclusions

Optimal operating conditions of the digestion tank of an industrial phosphoric acid production process are determined by means of a multi-criteria optimization method. The objectives used (i.e. saturation indices) have proved to be effective in minimizing the phosphate losses and increasing the acid productivity. Furthermore, the two inequality constraints on the temperature and on the excess of sulfuric acid above the stoichiometric amount required by the reactions were relevant to ensure an optimal operation of the digestion tank. The optimization results are consistent with the measurements carried out on an industrial installation, thus showing that optimization is a powerful tool for reducing the cost and duration of industrial experiments under maximum safety conditions. However, it is noteworthy that some side reactions can occur due to the impurities present in the phosphate ore which are not considered in this work. Their consideration in future works would undoubtedly improve the performance of the digestion tank.

References

- J. W. Ball, D. K. Nordstrom, 1991. User's Manual for wateq4f. U.S. Geological Survey.
- P. Becker, 1989. Phosphate and phosphoric acid, vol. 6, M. Decker press.
- I. Bouchkira, Y. El Fariq, 2017. Improvement of the 28 % phosphoric acid production units. MSc thesis, ENSA, Khouribga, Morocco.
- I. Bouchkira, A. M. Latifi, L. Khamar, S. Benjelloun, 2021. Global sensitivity based estimability analysis for the parameter identification of Pitzer's thermodynamic model. Reliability Engineering & System Safety, 107263.
- M. Cherif, A. Mgaidi, M. N. Ammar, M. Abderrabba, W. Fürst, 2002. Representation of VLE and liquid phase composition with an electrolyte model: application to $\text{H}_3\text{PO}_4\text{-H}_2\text{O}$ and $\text{H}_2\text{SO}_4\text{-H}_2\text{O}$. Fluid Phase Equilibria 194, 729–738.
- K. Deb, 2014. Multi-objective optimization. In: Search methodologies. Springer, pp. 403–449.
- J. S. Dyer, 2005. Maut multi-attribute utility theory. In: Multiple criteria decision analysis: state of the artsurveys. Springer, pp. 265–292.
- S. El Bouzidi, 2016. Improvement of two phosphoric acid units. MSc thesis, ENSA, Khouribga, Morocco.
- J. P. Gustafsson, 2011. Visual minteq 3.0 user guide. KTH university, Stockholm, Sweden.
- K. S. Pitzer, 2018. Activity coefficients in electrolyte solutions. CRC press.
- L. Shen, H. Sippola, X. Li, D. Lindberg, P. Taskinen, 2020. Thermodynamic modeling of calcium sulfate hydrates in a $\text{CaSO}_4\text{-H}_2\text{SO}_4\text{-H}_2\text{O}$ system from 273.15 to 473.15 k up to 5 m sulfuric acid. Journal of Chemical & Engineering Data 65 (5), 2310–2324.

Towards Automated HAZOPs

Calum Robinson^a, Solomon Brown^a, Joan Cordiner^{a*}

^a *Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom.*

j.cordiner@sheffield.ac.uk

Abstract

An increasing challenge faced in the process industries is the significant reduction of experience levels in chemical plants due to retirements of the baby boomers and more mobile younger generations. Additionally, many plants have ageing equipment which are deteriorating which changes risk levels with time and impacts repair timing decisions that need to be well informed, balancing budgets versus risk. HAZOP teams have less experience to identify and risk rank hazardous scenarios. This work presents a tool that uses criteria to identify appropriate scenarios and their rating, as well as recommend potential appropriate controls and mitigations. The tool allows dynamic HAZOPs that can be updated with the changing condition of safety critical equipment and suggest temporary mitigation measures, whilst also informing leaders of the risk of delaying repair.

Keywords: hazan, automated risk rating, automated hazop

1. Introduction

In development and in manufacture personnel are required to make many decisions. With changing demographics, for example baby boomers retiring and more mobile younger generations, there is a rapid and significant reduction in expertise and experience in all safety critical roles from leadership, engineers, maintenance, and operations technicians. As a result, tools for development through to manufacturing such as expert informed automated hazard analysis and risk assessment are needed. These will inform design and guide engineers in critical areas and guide critical experiments and data needed.

HAZAN (Hazard Analysis) at early-stage development and HAZOP (Hazard and Operability Studies) for detailed design of chemical plants are used very extensively and are dependent on the knowledge and experience that the team carrying out the study have. “Shortcomings could be due to lack of expertise or experience within the HAZOP team”. (Paul Baybutt, 2015). Importantly, the assessments of the severity and frequency can be inaccurate. “When you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.” (Thomson, 1883). Inaccurate risk ranking can lead to under controlling risks and have the potential to lead to catastrophic consequences. Alternatively, we can overspend resources for low level risks assigned to high a ranking.

Using published guidance from verified sources the HAZOP Study Tool was developed that improves upon the current flaws in the traditional method of identifying and risk ranking scenarios.

2. Methodology

The HAZOP Study Tool provides guidance for participants of a study in the areas of hazard identification and risk ranking. The tool consists of three main elements. The first helps the identification of the hazardous properties that the process chemicals being used may have. Comparing these hazardous properties to the operating conditions and guidance from several different sources, e.g the Global Harmonized System of Classification and Labelling of Chemicals (GHS), a set of classification criteria that was developed. These determine the nature and relative severity of the hazard of a chemical substance, allows a severity ranking to be assigned to each property, based upon these criteria as such as shown in Table 1.

Severity Ranking				
Insignificant [1]	Low [2]	Medium [3]	High [4]	Serious [5]
FP > OT by 150°C or more	FP > OT by a range of 149°C - 100°C	FP > OT by a range of 99° - 50°C	FP = OT or FP > by 49°C - 1°C	OT > FP
Not classified as a flammable gas	Not classified as a flammable gas.	Category 2	Category 1B	Category 1A
0.5 atm abs > VP > 0.1 atm abs	1 atm abs > VP > 0.5 atm abs	2 atm abs > VP ≥ 1 atm abs	3 atm abs > VP ≥ 2 atm abs	VP ≥ 3 atm abs
AI Temp > OT by 120°C or more.	AI Temp > OT by 90°C	AI Temp > OT by 60°C	Ai Temp = OT ± 15°C	AI Temp < OT

Table 1 - The severity ranking criteria for specific hazardous chemicals (Element 1 of HAZOP Study Tool).

The second element of the tool aids the identification of the failure rate of different types of equipment. This element addresses a common uncertainty or misjudgment that occurs within HAZOP studies. Lack of understanding regarding the failure types and rate of the equipment makes it very difficult to accurately risk rank this hazard. Element two of the study tool aids the ranking of likelihood of equipment failure based upon quantitative criteria as shown in the sample in Table 2. The tool allows a HAZOP team to identify these failure rates for the process equipment in question and guides the user to a likelihood ranking for this hazard.

Sampled Equipment: Ambient Temperature and Pressure Vessel Large Vessel	Likelihood Ranking				
	Extremely Unlikely [1]	Very Unlikely [2]	Low [3]	Likely [4]	Very Likely [5]
	(Per vessel per year)				

Severity/Type of Release	$\leq 1E^{-06}$	$>1E^{-06}$, $\leq 1E^{-05}$	$>1E^{-05}$, $\leq 1E^{-04}$	$>1E^{-04}$, $\leq 1E^{-03}$	$>1E^{-03}$
Catastrophic	$5.00E^{-06}$				
Major				$1.00E^{-04}$	
Minor				$2.50E^{-03}$	
Roof					$2.00E^{-03}$

Table 2 - The likelihood ranking criteria for equipment failure (Element 2 of HAZOP Study Tool)

The third element of the tool provides information regarding the underlying causes and contributors to both equipment failures and equipment-based accidents to aid scenario identification and ranking. Often there is more than one contributor to an incident. The statistics regarding the number of contributors that occur that result in an equipment-based accident are stated and the main contributor and sub-contributors for each type of equipment failure are outlined.

This element of the tool is laid out in several tables, each table shows a variation of ways in which the contributors to an equipment-based accident can be broken down into more specific events that could cause an incident. Each table contains figures and statistics to show the proportion in which each cause has the possibility to have contributed to an incident.

The first table in the third element of the tool which expresses the breakdown of equipment-based accidents into the percentages that shows the proportion in which type of equipment are involved. For example, this draws attention to items' such as piping as data shows that piping is the type of equipment that most frequently causes equipment-based accidents.

The function of the third table is also to aid the identification of the less frequent accident contributors that still carry a high impact. These contributors are like 'black swan' events. To tool aims to predict what otherwise would be unforeseeable or beforehand seem to be beyond what is expected.

3. Applying the Tool

3.1. Case Study based on Williams Geismar Olefins Plant Incident

The plant in question is the Williams Geismar Olefins Plant in Louisiana, USA. This plant produces ethylene and propylene for the petrochemical industry.

The raw materials of the olefin production are cracked before being cooled and separated. This case study will focus on one small section of this process, the depropaniser. This section of the process uses high-pressure distillation to separate propane and propylene. Assumptions that have been made for the purpose of the case study are as follows, the fractionator operates at 55°C and 19 bar.

3.2. Identifying the Hazards

Using the HAZOP Study Tool embedded within excel, a formal process was carried out to identify the potential hazardous scenarios present within the node in question. The

hazardous scenarios identified are as follows, loss of containment (LOC) of propane and/or propylene from heat exchanger/reboiler, from propylene fractionator and/or from pipework.

3.3. Comparing Traditional Risk Ranking to HAZOP Study Tool Assisted Risk Ranking

To make a valid comparison, the same scenario is compared between the traditional risk ranking method of relying upon the knowledge, opinions, and experience of engineers within the HAZOP team against the risk ranking procedure utilizing the HAZOP Study Tool. Ten BEng level chemical engineering graduates were asked to rank one of the hazardous scenarios that was identified, 'LOC of propane and propylene from propylene fractionator that's operating at 55°C and 19 bar', in terms of severity and likelihood. The results are shown in the figures 1 and 2 below.

3.3.1. Chemical Engineering Graduates Rankings

The results as shown in table 3 that 4 engineering graduates ranked the scenario a ranking of 'High (4)' and 6 a ranking of 'Serious (5)'. Although there is an agreement that the severity is significant, almost 50% of engineers rank the severity differently.

Graduate	Severity Ranking				
	Insignificant [1]	Low [2]	Medium [3]	High [4]	Serious [5]
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					

Table 3 - Chemical engineering graduates severity ranking for hazardous scenario.

The likelihood rankings of the hazardous scenario are significantly widespread as shown in table 4. Three engineers rank the likelihood of the loss of containment of a propylene fractionator, 'Very Unlikely (2) which lies between $1E^{-06}$ and $1E^{-05}$. Whereas on the other side of the spectrum, two engineers believe that the likelihood is 'Likely (4)' which lies between $1E^{-04}$ and $1E^{-03}$.

Graduate No.	Likelihood Ranking				
	Extremely Unlikely [1]	Very Unlikely [2]	Low [3]	Likely [4]	Very Likely [5]
1					
2					
3					
4					
5					
6					
7					

8					
9					
10					

Table 4 - Chemical engineering graduates likelihood ranking for hazardous scenario.

3.3.2. HAZOP Study Tool Assisted Rankings

The use of the HAZOP study tool and each of the elements within the tool allowed a conclusion to be made that both propane and propylene presented a ‘Serious (5)’ level of severity in relation to being a fire and explosion hazard and propane presented a ‘Medium (4)’ level of severity in relation to toxicity to aquatic life and environment.

With focus on the vessel in question, or the propylene fractionator, the HAZOP study tool allows the evaluation of the likelihood of a variety of different types of release, such as, ‘Catastrophic release’, ‘25 – 50mm diameter hole’, ‘13mm diameter hole’ and ‘6mm diameter hole’. In this specific case, each of the types of release as listed were concluded to be ‘Very Unlikely (2)’, except for the ‘6mm diameter hole’ that has a failure rate of $4E^{-05}$ per vessel per year and is ranked as a ‘Low (3)’ level of likelihood.

3.4. Discussion

When comparing the traditional, experience and knowledge-based approach to risk ranking and the HAZOP Study Tool assisted method there is a lot of differences that can be observed:

- When the rankings are collated on a risk matrix the traditional method arrived at five different risk rankings, each of which is coming from a verified source, a qualified engineer within the field of chemical engineering. The study tool assisted rankings are based upon semi-quantitative information gathered from verified industry sources which introduces an element of consistency across all processes, regardless of the team carrying out the HAZOP.
- Each person has varying levels of industry experience and knowledge which have an effect on the rankings that are assigned. The HAZOP study tool allows the user to take use the information on the process and compare it to the criteria that has been created based upon varying sources to allow them to arrive at an accurate ranking. The tool provides guidance for personnel of all levels of experience and enables unexperienced personnel to operate competently within a HAZOP team and avoid misleading a study because of their lack of experience.
- It has been shown that assigning likelihood and severity is dependent on experience. Whereas the tool assisted method focuses on breaking down the hazardous scenario in question and providing semi quantitative information that directs the scenario ranking. The tool assisted method breaks a process down and inspects the failure rate of every piece of equipment providing more of a comprehensive overview of how likely the realization of a hazardous scenario really is.

4. Conclusions

The HAZOP study tool is valuable guidance and expert advisor for teams who increasingly lack the depth of experience and expertise due to demographic changes and

younger generation's increased mobility. Methods of passing down knowledge and rapid access in a way that can be used by inexperienced engineers must be developed.

The quality of the judgement of a team or an individual during a risk assessment is highly dependent on the expertise and experience of the team. The outputs from a risk assessment should be considered to be 'high risk' due to the potential of catastrophic consequences if it is done incorrectly. The addition of semi-quantitative guidance adds expertise that can be used by teams to improve scenario identification, risk ranking and identifying mitigations. The HAZAN and HAZOP study tool is an expert advisory tool and a step in the direction of automated risk assessment for a wide variety of chemical processes.

References

- CSB, 2016. William Geismar Olefins Plant Explosion and Fire-Case Study. [online] Available at: <<https://www.csb.gov/williams-olefins-plant-explosion-and-fire/>> [Accessed 07 January 2021].
- P. Baybutt., 2015. A critique of the Hazard and Operability (HAZOP) study. *Journal of Loss Prevention in the Process Industries*, 33, pp.52 – 58.
- P. Baybutt., 2018. Guidelines for designing risk matrices. *Process Safety Progress*, 37(1), pp.49-55.
- M. Brenner., 2019. Retiring Machinists Will Impact Plant Safety. *Industrial Safety and Hygiene News*, [online] Available at: < <https://www.ishn.com/articles/111930-retiring-machinists-will-impact-plant-safety>> [Accessed 17 August 2020].
- K. Cherry, 2020. What is Cognitive Bias? Verywell, [online] Available at: <<https://www.verywellmind.com/what-is-a-cognitive-bias-2794963>> [Accessed 22 August 2020].
- D. Gadd, D. Keeley, and D. Balmforth, 2003. Good practice and pitfalls in risk assessment. HSE - Research Report 151, [online] Available at: <<https://www.hse.gov.uk/research/rrpdf/rr151.pdf>> [Accessed 24 August 2020].
- GHS, 2011. Globally Harmonized Systems of Classification and Labelling of Chemicals. [online] Fourth Revised Edition. Available at: <https://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev04/English/ST-SG-AC10-30-Rev4e.pdf> [Accessed 21 August 2020].
- HSE, 2017. Failure Rate and Event Data for Use Within Risk Assessments. [online] Available at: < <https://www.hse.gov.uk/landuseplanning/failure-rates.pdf>> [Accessed 27 August 2020].
- K. Kidam and M. Hurme., 2013. Analysis of equipment failures as contributors to chemical process accidents. *Process Safety and Environmental Protection*, [online] 91, pp.61-78.
- J. Kingsley, 2017. 20 Cognitive Biases That Affect Risk Decision Making. [online] Available at: <<https://safetyrisk.net/20-cognitive-biasies-that-affect-risk-decision-making/>> [Accessed 22 August 2020].
- P. Koks, 2015. Six Challenges of Qualitative Data Analysis. [online] OnlineMetrics Available at <<https://online-metrics.com/qualitative-data/>> [Accessed 27 July 2020].
- P. Marhavilas, M. Filippidis, G. Koulinas and D. Koulouriotis, 2020. A HAZOP with MCDM Based Risk – Assessment Approach: Focusing on the Deviations with Economic/Health/Environmental Impacts in a Process Industry.
- Marsh. (2016). *The 100 largest losses 1974-2015* (24th ed.). London: Marsh & McLennan
- M. Othman, R. Idris, M. Hassim, and W. Ibrahim, 2016. Prioritizing HAZOP analysis using analytic hierarchy process (AHP). [online] Available at: <<https://link.springer.com/content/pdf/10.1007/s10098-016-1104-4.pdf>> [Accessed 20 August 2020].
- W. Schulz, 2008. As Baby Boomers Retire *Chemical & Engineering News*, 86(35), p.36
- S. Shin, S. Byeon, J. Sohn, and K. Moon, 2019. Development and Verification of a Risk Index for Evaluating the Chemical Accident Risk of Korean Chemical Enterprises. *International Journal of Environmental Research and Public Health*, 16.

Multi-objective optimization of co-processing of algae hydrothermal liquefaction oil and vacuum gas oil: a survey of algal oil co-processing ratio and gasoline selling price

Shuai Zhang, Qingyu Lei, Le Wu*, Yuqi Wang, Lan Zheng

School of Chemical Engineering, Northwest University, Xi'an 710069, China

lewu@nwu.edu.cn

Abstract

The co-processing of algae hydrothermal liquefaction oil (AHTLO) and vacuum gas oil (VGO) in an existing refinery has been proposed to decrease the production cost of algal bio-fuels by utilizing existing infrastructures of the refinery. Techno-economic analysis of the co-processing of AHTLO and VGO shown that the larger AHTLO co-processing ratio, the higher minimum selling price. In this work, an investigation between the AHTLO co-processing ratio and minimum selling price is conducted by multi-objective optimization (MOO). The proposed MOO model is solved to obtain the Pareto fronts by ϵ -constraint method. The trade-off solution for the co-processing technology is obtained using function evaluation method. The effects of the weight factor on the trade-off solution are also carried out.

Keywords: Algae, HTL, VGO, Co-processing, Multi-objective optimization.

1. Introduction

The development of bio-energy is one of the ways to deal with the problems of resource shortage and environmental pollution. To further reduce the production cost of bio-fuels, the co-processing of bio-oil and vacuum gas oil has been proposed by using the existing refinery infrastructures. The third-generation bio-energy, algae was also used as the raw material to obtain the algae hydrothermal liquefaction oil (AHTLO) which was then co-processed with vacuum gas oil (VGO) in an existing fluid catalytic cracking unit. The co-processing of AHTLO and VGO shown high technical feasibilities (Santillan-Jimenez et al. 2019). A techno-economic analysis shown that the co-processing technique can reduce the production cost of the algal fuels. Furthermore, the minimum selling price (MSP) of gasoline was increased with the increase of AHTLO co-processing ratio. (Wu et al. 2020).

In this work, a multi-objective optimization model is established for the co-processing of AHTLO and VGO in an existing refinery to discuss the relationship between the AHTLO co-processing ratio and minimum selling price (MSP) of gasoline. As there are two kinds of algae production methods, algal turf scrubbers (ATS) and open raceway ponds (ORP), two co-processing scenarios (ATS scenario and ORP scenario) are proposed to illustrate the differences between the two scenarios. The Pareto fronts and the trade-off solution are calculated by ϵ -constraint method and evaluation function method, respectively. The effects of weight factor on the trade-off solution are also discussed.

2. Materials and methods

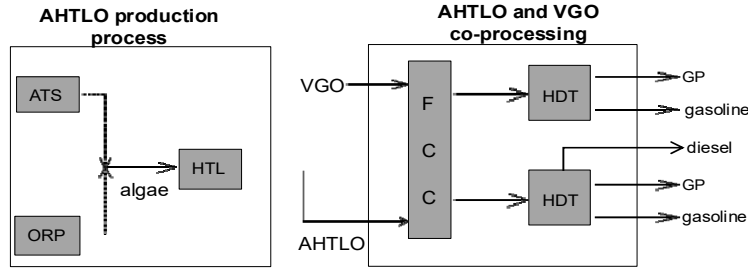


Figure 1: The co-processing diagram

Figure 1 has been shown the basic flowsheet of the co-processing process. This figure can be divided into two parts, the one part is production process of AHTLO, the other part is co-processing of AHTLO and VGO. Algae produced by ATS or ORP can be directly hydrothermally liquefied to obtain AHTLO. Then, the AHTLO and VGO are co-fed into the FCC to produce the co-processing products, gasoline and diesel with bio-carbon. Then, the FCC gasoline and diesel are hydrotreated to obtain the gasoline and diesel fuels.

2.1. Objective functions

In this work, the relation between the co-processing ratio of AHTLO in FCC feed oil and the MSP of gasoline is discussed by using a multi-objective model. It is easy to conduct that more bio-carbon would exist in co-processing product if a larger AHTLO co-processing ratio was adopted. Therefore, the objective functions of the multi-objective model are set as the minimization of the gasoline MSP and the maximization of the AHTLO co-processing ratio. MSP can be calculated by net present value (NPV) and internal interest of rate (IRR) according to the mass balance, energy balance, operating cost, investment cost, profit margin and tax rate. More details can be found in Wu et al. (2020).

$$\max R_{\text{AHTLO}} \quad (1)$$

$$\min \text{MSP} \quad (2)$$

2.2. Constraints

The constraints are main the mass balances of the co-processing process.

$$m_j^{\text{out}} = \sum_i \sum_l m_i^{\text{in}} y_j \quad (3)$$

where m denotes the flow rate, in t/h; superscript *in*, *out* indicate charge and discharge; subscripts i, j are the i^{th} charging port and j^{th} discharging port respectively.

2.3. Trade-off between MSP and co-processing ratio

In this work, the evaluation function is used to obtain the trade-off solutions.

$$u_o^s = (O_o^{\max} - O_o^s) / (O_o^{\max} - O_o^{\min}) \quad (4)$$

$$u^s = \sum_o w_o u_o^s \quad (5)$$

where O indicates the objective function; u is the evolution function; w denotes weight factor; Subscript O indicates O^{th} objective function; Superscript *min*, *max* and *s* are the minimum solution, maximum solution and s^{th} solution of u .

3. Case study

3.1. Basic data

The co-processing of AHTLO and VGO in the FCC unit, the annual operation time is 8400 h and the average annual processing capacity is 1.2 million tonnes. The yield of AHTLO and gas are 44% and 19% in the HTL process. In the FCC unit, the yields of gaseous products, coke, gasoline and diesel are 18%, 6.6%, 99.5% and 91.2%, respectively.

Table 1 presents the basic operating cost of the two algal production method, ATS and ORP, which includes power, nutrients, fresh water, fuel or flocculants, labor and maintenance or insurance.

Table 1: Basic inputs for the ATS and ORP processes (Hoffman et al, 2017)

Operating cost /(\$·t ⁻¹ a)	ATS	ORP
Power	56.12	65.26
Nutrients	/	104.25
Fresh water	/	0.988
Fuel or Flocculants	19.14	56.36
Labor	13.30	14.69
Maintenance/Insurance	29.26	25.53

Note: a) the cost unit is dollar per ash free dry tonne

3.2. Optimal results of ATS scenario

3.2.1. Pareto front in ATS scenario

Figure 2 presents the relations between the MSP of gasoline and the ratio of AHTLO. It can be observed that the MSP increases with the increase of co-processing ratio of AHTLO. When the AHTLO co-processing ratio is 1%, the MSP is 2.04 \$·gal⁻¹. When the AHTLO co-processing ratio is 20%, the MSP is 3.80 \$·gal⁻¹. Considering the relative low MSP of gasoline, the co-processing of VGO and AHTLO to produce gasoline and diesel may be a potential way to partially replace petroleum derived gasoline.

From the breakdown of gasoline price, in ATS scenario. It can be obtained that when the co-processing ratio is 1% and 20%, the proportion of VGO cost in the MSP is the largest proportion, which is 80% and 45%, respectively. Then, as the tax is controlled at 39%, it ranks second with 45% and 42%, respectively. It can be seen from the figure that the cost and tax share of VGO are reduced due to the addition of AHTLO. Since, the scenario does not consume nutrients and fresh water, the proportion of nutrients is 0.

According to Figure 3, the mass balances flow chart of ATS scenario under the condition of AHTLO ratio of 1% and 20%. Since the consumption of energy, nutrition and water can be ignored, in ATS scenario, the input amount is 0 t/h. Algae are produced in ATS,

and AHTLO is obtained by HTL. When the co-processing ratio is 1%, the AHTLO is 1.43 t/h, and the VGO is 141.43 t/h co-processing into FCC. And the obtained gasoline, diesel and GP are 70.91 t/h, 30.38 t/h and 0.73 t/h respectively. When the co-processing ratio of AHTLO is 20%, the AHTLO is 28.58 t/h, and the VGO is 114.28 t/h co-processing into FCC. Products of gasoline, diesel and GP are 70.91 t/h, 30.38 t/h and 0.73 t/h respectively.

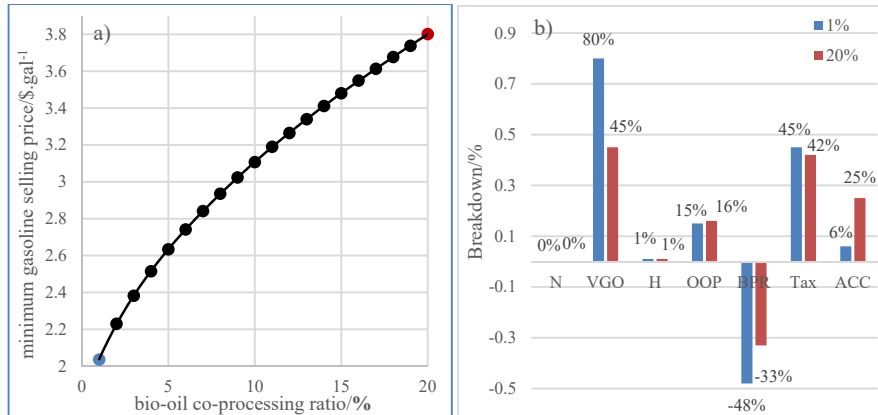


Figure 2: a) The Pareto fronts and the breakdowns of MSPs, b) 1% and 20% AHTLO co-processed (N=Nutrient, H=Hydrogen, U=Utilities, OOC=other operating cost, BPR=by product revenue, ACC=annual capital cost)



Figure 3: Mass balances of ATS scenario (GP=Gas products, fuel gas and LPG)

3.2.2. Effect of the weight factor on the trade-off solution

Effect of the weight factor on the trade-off solution of ATS scenario is shown in Table 2. According to Table 2, when the weight factor is 0.6, the optimal AHTLO ratio is 3% and the corresponding MSP is 2.38 \$·gal⁻¹. The optimal AHTLO ratio would be 8% and MSP would be 2.94 \$·gal⁻¹ if 0.5 is chosen as the weight value. When the weight ratio is less than 0.4, the AHTLO ratio is 20%, and the MSP is 3.80 \$·gal⁻¹; if the weight ratio is more than 0.7, the AHTLO ratio is 1%, and the corresponding MSP is 2.04 \$·gal⁻¹.

Table 2: Effect of weight factor on the trade-off solution of ATS scenario

Weight factor	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Bio-oil ratio/%	20	20	20	20	20	8	3	1	1	1	1
MSP/\$·gal ⁻¹	3.8	3.8	3.8	3.8	3.8	2.9	2.3	2.0	2.0	2.0	2.0
	0	0	0	0	0	4	8	4	4	4	4

3.3. Optimal results of ORP scenario

3.3.1. Pareto front in ORP scenario

As indicated in Figure 4a), the relationship between MSP and co-processing ratio, and in Figure 4b), the Pareto frontier diagram. The results are similar to those under ATS scenario. MSP of co-processing production oil increases with the increase of co-processing ratio of AHTLO. When AHTLO ratio is 1%, MSP is 1.62 \$·gal⁻¹; AHTLO ratio is 20%, MSP is 3.70 \$·gal⁻¹. Whether the ratio of AHTLO is 1% or 20%, the effect is likely to replace part of gasoline and diesel refined that from petroleum. Whether the proportion of AHTLO is 1% or 20%, the co-processing effect can reduce the production cost of gasoline and bio-diesel, and it may also replace part of gasoline and diesel oil refined from petroleum.

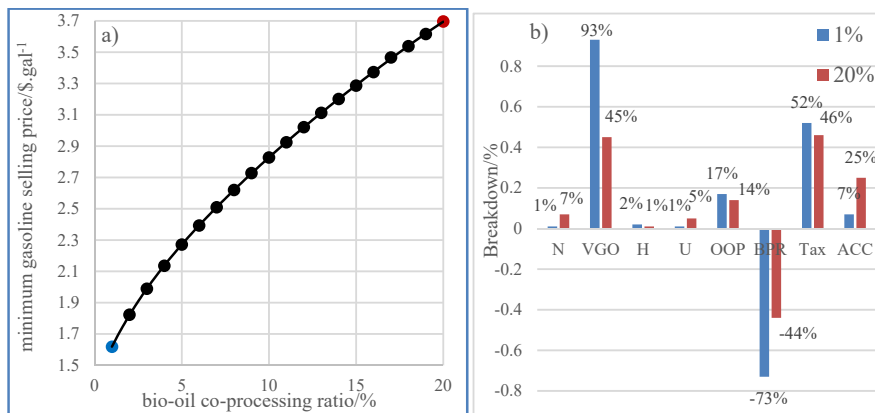


Figure 4: a) The Pareto fronts and the breakdowns of MSPs, b) 1% and 20% AHTLO co-processed (N=Nutrient, H=Hydrogen, U=Utilities, OOP=other operating cost, BPR=by product revenue, ACC=annual capital cost).

The detailed chart which breakdown of gasoline price under the ORP scenario with AHTLO ratio of 1% and 20%. It is similar to ATS scenario. In the MSP, the cost of VOG accounts for 93% and 45% respectively; the second is tax, 52% and 46%, which is also due to the tax control of 39%. Different from ATS, nutrients accounted for 1% and 7%, which was 0% in the ATS.



Figure 5: Mass balances of ORP scenario (GP=Gas products, fuel gas and LPG)

As demonstrated in Figure 5, the mass balance flow chart of ORP scenario under the condition of AHTLO ratio of 1% and 20% is shown. It is similar to ATS scenario, but there are differences in energy, nutrition and water consumption, which are 65.39 t/h and 493.51 t/h respectively; the others are the same as ATS scenario.

3.3.2. Effect of the weight factor on the trade-off solution

Table 3 shows the influence of weight factor on the trade-off solution, which is the same as ATS scenario. When the weight factor is 0.6, the optimal AHTLO ratio is 3% and the corresponding MSP is 1.99 \$·gal⁻¹. The optimal AHTLO ratio would be 8% and MSP would be 2.62 \$·gal⁻¹ if 0.5 is chosen as the weight value. When the weight ratio is less than 0.4, the AHTLO ratio is 20%, and the MSP is 3.70 \$·gal⁻¹; if the weight ratio is more than 0.7, the AHTLO ratio is 1%, and the corresponding MSP is 1.62 \$·gal⁻¹.

Table 3: Effect of weight factor on the trade-off solution of ORP scenario

Weight factor	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Bio-oil ratio/%	20	20	20	20	20	8	3	1	1	1	1
MSP/\$·gal ⁻¹	3.7	3.7	3.7	3.7	3.7	2.6	1.9	1.6	1.6	1.6	1.6
	0	0	0	0	0	2	9	2	2	2	2

4. Conclusion

In this work, a multi-objective optimization of the co-processing of AHTLO and VGO in the existing catalytic cracking pyrolysis is carried out to investigate the relation between the MSP and co-processing ratio. As there are two algae growth technologies, ATS and ORP, two scenarios were proposed to give a comparison between the two technologies when the algae were adopted as the raw materials for the HTL process. This study proved that co-processing of AHTLO and VGO can reduce bio-fuels cost. Indicating that AHTLO and VGO may be a potential way to partially replace petroleum derived gasoline in both cases.

Acknowledgement

The authors gratefully acknowledge funding by the project (No. 21808183) sponsored by Natural Science Foundation of China (NSFC), the project (No. 2020JQ-577) sponsored by Natural Science Foundation of Shaanxi Province and the project (No. 20190602) sponsored by Young Talent Fund of University Association for Science and Technology in Shaanxi, China.

References

- E. Santillan-Jimenez, R. Pace, T. Morgan, C. Behnke, D. Sajkowski, A. Lappas, M. Crocker, 2019, Co-processing of Hydrothermal Liquefaction Algal Bio-oil and Petroleum Feedstock to Fuel-like Hydrocarbons Via Fluid Catalytic Cracking, *Fuel Processing Technology*, 188, 164–171.
- J. Hoffman, R. C. Pate, T. Drennen, J. C. Quinn, 2017. Techno-economic assessment of open microalgae production systems. *Algal Research* 23: 51-57.
- L. Wu, L. Li, Y. Wang, L. Zheng, 2020, Techno-economic analysis of co-processing of vacuum gas oil and algae hydrothermal liquefaction oil in an existing refinery. *Energy Conversion and Management* 224: 113376.

Modeling of a software for sizing a wastewater treatment plant

Anass Qoutbane^a, Btissam Khnajer^b, Essediya Cherkaoui^c, Mohamed Khamar^d and Abderrahman Nounah

Civil Engineering and Environment Laboratory (LGCE); Materials, Water and Environment Team; Higher School of Technology of Salé; Mohammed V University in Rabat, Morocco

Abstract

The new wastewater treatment plants in Morocco are increasingly equipped with intensive biological treatment processes (bacterial beds and activated sludge) to the detriment of lagooning. Activated sludge treatment is a very interesting choice given its satisfactory efficiency.

The objective of this work is to list the formulas already published concerning the activated sludge and lagooning technique after having studied its operation, to draw up the sizing algorithms for the most reliable methods and then to program them in order to develop an easy-to-use software for sizing activated sludge type wastewater treatment plants with a low load saving. This software will contribute to reduce the duration of the study as well as of the implementation of the WWTP project and will constitute a tool for decision making.

This program will help designers to design and verify the purification performance. This tool also allows considerable time savings compared to conventional calculation methods while offering easy data entry.

Keywords : wastewater, treatment, program.

1. Introduction

The dimensioning of a wastewater treatment plant follows several steps, first of all by collecting basic data and essential parameters describing the quality and quantity of effluents, in order to be able to estimate the dimensions of each structure constituting the treatment plant. This estimation is often very expensive and generally requires a lot of time, concentration and energy to the WWTP designers, especially when it comes to working on several variants. The objective of this project is to develop a software called EMASN which allows to determine the dimensions of a wastewater treatment plant by lagooning plus de facto a cost estimate from some basic data. In order to achieve this, we use Excel and the programming language 'Visual basic'.

2. Material and method

2.1. Calculation formulas

2.1.1. Lagooning

Wastewater treatment by lagooning is done by various processes based on physical, chemical and biological phenomena.

The decontamination of wastewater requires a succession of steps involving physical, physico-chemical and biological treatments. Apart from the largest waste present in the wastewater, the treatment must allow, at the very least, to eliminate most of the carbonaceous pollution.

Anaerobic basin :

Basin surface = gross BOD5/ (volume load (Cv)Data in meters

Basin volume = Area * Depth

Residence time =Volume / flow rate

Outgoing pollutant load = Input pollutant load - (Efficiency*Input pollutant load)

Optional basins :

Pond surface = BOD5 load at the outlet of the anaerobic ponds / surface load Cs

Basin V = Basin surface * Depth (1.5 m)

Residence time =Volume / flow rate

Outgoing pollutant load = Input pollutant load - (Efficiency*Input pollutant load)

Curing tanks :

Volume of the tank = flow rate * Residence time (5 d)

S(M) = Basin volume / Depth (1 m)

2.2. Language and tool used

Visual Basic (VB) is a third-generation event-driven programming language and integrated development environment created by Microsoft for its COM1 programming model. Visual Basic is directly derived from BASIC and allows rapid application development, creation of graphical user interfaces, access to databases using CAD, ADO and RDO technologies, and the creation of ActiveX controls or objects. Scripting languages such as Visual Basic for Applications and VBScript are syntactically similar to Visual Basic, but are used and behave significantly differently.

3. Results and discussion

3.1. Client

The client (Fig. 1), allows you to orientate yourself towards your directions concerning the software (dimensioning lagoon, activated sludge, bacterial bed, cost estimate).



Figure 1 : Interface du logiciel

3.2. Calcul des débits

The first sizing step is to calculate the incoming wastewater flow (Fig. 2).

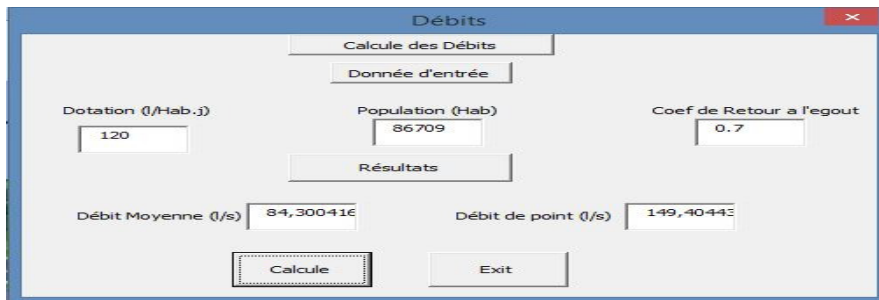


Figure 2: Calculation of Wastewater Inlet Flow Rate

3.3. Calcul des charges polluantes

Through the application of the calculation forum, the user can calculate the pollutant load at the inlet of the WWTP according to the population (Fig. 3).

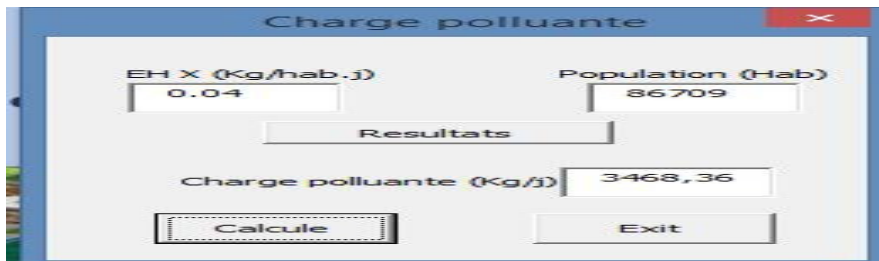


Figure 3: Calculation of the pollutant load at the inlet of the WWTP

3.4. Lagooning

The third step consists in calculating the dimensioning of the treatment basins. (Fig.4).



Figure 4: Basin and Grid Computation Interface

3.4.1. Screening

Before starting the calculation of the treatment basins, it is essential to dimension the elimination grid for large elements.

The screenshot shows the 'Dégrillage' software interface with the following data:

Donnée d'entrée			
Debit de point (l/s)	Espace entre les Barreaux	Epaisseur Des Barreaux (cm)	Vitesse (m/s)
86	0.015	0.002	1.2
Coefficient de colmatage	Tirant d'eau (m)	Facteur de Forme**	Angle d'inclinaison
0.4	0.25	1.8	40

facteur de forme*	
rectangulaire	2,42
circulaire	1,67 à 1,83

Résultats de dimensionnement			
coefficient de vide	0,88235294111	surface de la grille (m ²)	0,20305555551
Longueur (m)	0,38909509711	largeur (m)	0,52186613771
perte de charge	1,01865515711	Nombre des Barreaux	30,6980081031
Refus de la grille (l/h.j)	900		

Buttons: Calculé, Exit

Figure 5: Calculation of grid dimensions

3.4.2 Lagoon dimensions

After the user finishes the calculation of the first step, which is the pre-treatment, he can start the step of calculating the basin volumes. (Fig.6) Based on the calculation formulas and results obtained in the first step, the software calculates

the dimensions of the three treatment basins respecting the residence conditions for each basin, Aerobic, Optional and Maturation.

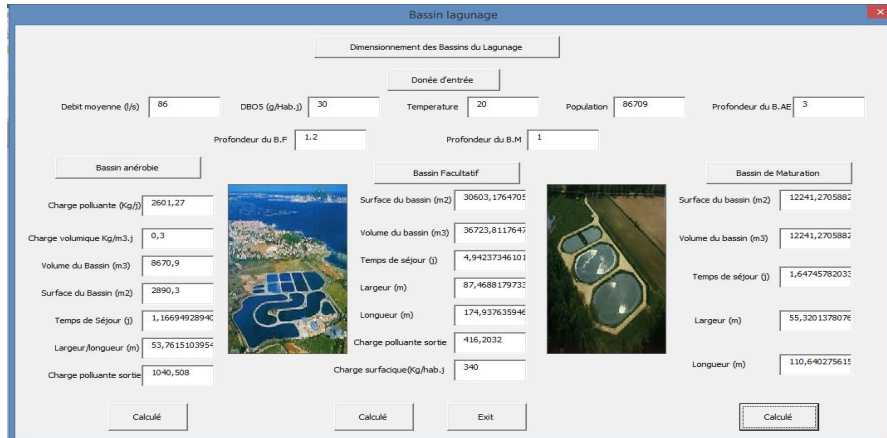


Figure 6: Calculation of basin volumes

3.4.2. Cost Estimate

After determining the technical elements of the project and sizing the volumes of the treatment basins, the user has the power to estimate the project cost. The cost estimate is made in proportion to the number of equivalent inhabitants according to the manual proposed by ONEE-BE. The user must add the field cost according to the study site in order to properly estimate the project and compare the treatment cost with other processes and choose the most appropriate process.

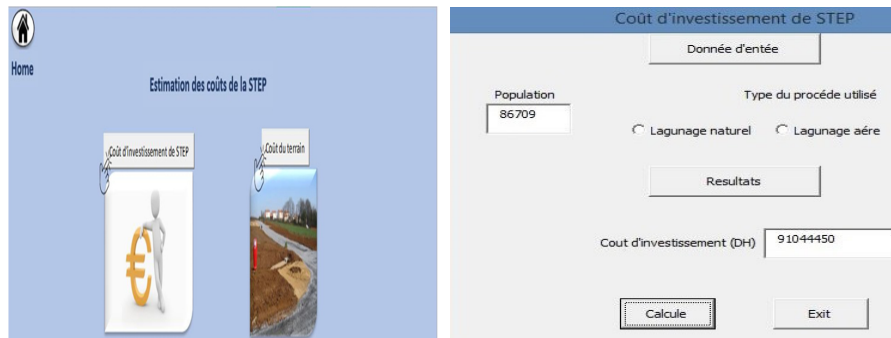


Figure 7: Project Cost Calculation

Conclusion

Today the treatment of liquid effluents, especially wastewater from cities, has become a high priority for the state so that it can preserve its natural resources and protect aquatic ecosystems from pollution caused by the physico-chemical and biological interactions of pollutants with biotopes.

This program makes it easier for designers to design and thus verify the purification performance of natural lagoon-type treatment plants. In addition, this modelled calculation tool offers a significant time saving compared to conventional calculation methods, while providing easy data entry. And it compares the cost of treatment compared to other wastewater treatment processes.

References

- Endamana D., Kenghneim I., Gockowski J., Nya J., Wanji D., Nyemekn D., International Symposium on Water,
- Poverty and Productive uses of Water at the Household Level, Mulders drift, Afrique du Sud (2003)
- P.T Shannon et autres, Epaissement par lots et continu : Prédiction du comportement de sédimentation des lots à partir des données de taux initial avec résultats pour les sphères rigides Ind. Eng. Chem. Fundamentals 3 (1963) 250-260.
- Shannon P., Tory R., Ind. Ind. Chem. Fundamentals 235 (1966) 357-382.
- Majdy, E. Cherkaoui, A. Nounah, M. Khamar Le traitement physico-chimique par coagulation-floculation des rejets d'eaux usées de la ville de Sale. J Mater Environ. Sci. 6 (3) (2015) 834-839ISSN : 2028-2508.
- EL Halouani H Réutilisation des eaux usées dans l'agriculture et son impact sur l'environnement : cas de la ville d'Oujda Thèse de doctorat d'État, Faculté. Sciences Université d'Oujda Mohammed Premier ministre, (1995).
- Qoutbane, E. Cherkaoui, M. Khamar et A. Nounah : caractérisation et méthodes de traitement des eaux usées de la ville de tiflet revue internationale de recherche avancée en ingénierie et technologie (ijaret) volume 11, numéro 6, juin 2020, pp. 80-94, article id : ijaret_11_06_008
- Cygler C. (2008) - Traitement des eaux usées. Prêts pour le recyclage, Hydroplus, 184, p. 36-42.
- Dahir n° 1-95-154 portant promulgation de la loi n° 10-95 sur l'eau.

Joint Chance Constraint Approach based on Data-Driven Models for Optimization Under Uncertainty applied to the Williams-Otto Process

Byungjun You^{a,*}, Erik Esche^a, Joris Weigert^a and Jens-Uwe Repke^a

^aTechnische Universität Berlin, Process Dynamics and Operations Group, Sekr. KWT 9, Straße des 17. Juni 135, Berlin 10623, Germany

**byungjun.you@tu-berlin.de*

Abstract

Uncertainty in parameters is increasingly considered in the optimization of design and operation of chemical process systems. A possible approach solving such optimization problems under uncertainty is the application of the chance constraint method. However, this approach requires high computational effort for calculating the probability of constraint satisfaction. To quickly solve chance-constrained optimization problems, an approach using data-driven models has been developed and proposed for single chance-constraints. In this contribution, the developed framework for data-driven chance constrained optimization is extended to joint chance constraints and analysed on a basic mathematical nonlinear model for easier evaluation of the method and on a simplified Williams-Otto process model with complex and nonlinear behaviour.

Keywords: Optimization under uncertainty, Joint chance-constrained optimization, Data-driven modelling

1. Introduction

Deterministic optimization approaches are well developed in the process industry and are used for offline and online process optimization (Li et al., 2006). However, many uncertainties within and outside of a process, such as feed stream, demanded product amount and kinetic parameters are often not considered, which can lead to serious deficiencies in the system or to economic loss (Giunta, 2004). To tackle this, approaches for process optimization under uncertainty can be applied on the design and operation of process systems in chemical industry. A possible approach solving such optimization problems under uncertainty is the application of chance-constrained (CC) optimization. However, this approach requires high computational effort for calculating the probability of constraint satisfaction for multidimensional distributions in order to achieve the desired probability level (Ostrovsky et al., 2013). To enable CC optimization in real-time applications even for complex processes, which is not possible using conventional approaches, a data-driven CC optimization framework has been proposed (Weigert et al., 2019). The framework has been successfully tested for single chance-constraints, which are satisfied individually with a prescribed confidence level. To increase the flexibility and robustness of the optimization problem, joint chance-constrained programming can be used which is able to guarantee a user-defined probability level of meeting all restrictions simultaneously. This is generally solved through two approaches of the joint chance constrained problem: sampling based and analytical approximation (Yuan et al.,

2017). In this contribution, we extended the existing framework to joint chance constraints as a new method and analysed its behaviour.

2. Optimization under uncertainty

In general, the basic idea of optimization under uncertainty is to integrate the available stochastic information into the optimization problem formulation. To systematically solve the optimization problems under uncertainty, three different methods can be distinguished: (1) Scenario-based optimization approach, (2) Stochastic approach and (3) Parametric approach (Wets et al., 1994). A possible approach solving such optimization problems using the second method is the application of single chance-constrained optimization which can be formulated as follows:

$$\min_u E(\Phi(x, u, \xi)) \quad (1a)$$

$$\text{s.t. } g(x, u, \xi) = 0 \quad (1b)$$

$$\Pr_i\{h_i(x, u, \xi) \geq 0\} \geq \alpha_i \quad \forall i \in \{0, 1, \dots, n_{h_i}\} \quad (1c)$$

$$u^L \leq u \leq u^U \quad (1d)$$

$$u \in \mathbb{R}^{n_u}, \quad x \in \mathbb{R}^{n_x}, \quad \xi \sim \mathcal{N}(\mu, Cov)$$

with expectation operator E , original objective function Φ , decision variables u , state variables x , uncertain parameters ξ , equality constraints g , and inequality constraints h . In most implementations ξ is assumed to be described by a multivariate normal distribution with the mean value μ and covariance matrix Cov . \Pr represents the probability of complying with the inner constraint being feasible to be greater than a user-predefined confidence level α ($0 \leq \alpha \leq 1$). The uncertain input ξ makes the corresponding output variable $h_i(x, u, \xi)$ also uncertain, so further processing of the inequalities must be carried out in order to quantify possible violations and it can be formulated as Eq. (1c). Another way of formulating the equation is joint chance constraint and can be written as (Arellano-Garcia, 2009):

$$\Pr_j\{h_j(x, u, \xi) \geq 0, \quad \forall j \in \{0, 1, \dots, n_{h_j}\}\} \geq \alpha_j \quad (2)$$

In this case, the joint chance constraint requires the satisfaction of all inner constraints to be satisfied simultaneously with a predefined confidence level. In addition, this approach increases the flexibility of the optimization problem in the inequality variable space resulting in decreased values of the objective function while still retaining the overall probability level.

3. Data-driven input-output and uncertainty modelling

A data-driven CC optimization framework (Weigert et al., 2019) has been developed to reduce the computational cost of the existing chance constrained optimization framework and tested successfully for single chance-constraints. The data-driven framework consists of two submodels: a process model (PM) and an uncertainty model (UM). The PM represents the correlation between the decision variables and the expected values of the objective function and the inequality constraints, while the uncertainty model (UM) determines the probability distribution parameters needed to compute the probability of the inequality constraint as a function of all the decision variables.

3.1. Data preparation and data-driven model generation

To generate the data-driven model (here: artificial neural network, ANN), training and testing data have to be prepared based on the formulation of the chance-constrained optimization problem. Data sampling is performed in a space of predefined decision

variables (uniform distribution) and the uncertain parameter (normal distribution) using the framework, which solves the rigorous model in AMPL. Prior to model training, the following details have to be specified: the output uncertainty at each available input point, the expected values of the objective function and the inequality constraints. In particular, a model for describing the probability distribution is selected and its characteristic parameters are specified as the output variables of the UM. Normal and beta distributions are used in this work. Their defining parameters are the variances for the normal and two positive shape parameters (α and β) for the beta distribution, which can be determined by model training. With prepared data sets and the essential parameters, the process and uncertainty model are trained using the machine learning toolbox Scikit-learn (Pedregosa et. al., 2011).

3.2. Data-Driven Model for process optimization under uncertainty

By applying the PM and the UM to the chance-constrained optimization problem of Eq. (1), only decision variables are required to solve this problem, since the influence of x and ξ on the optimization problem is already included in the submodels. The optimization problem can be now simplified to:

$$\min_u E(\Phi(u)) \tag{3a}$$

$$\Pr_k\{h_k(u) \geq 0\} \geq \alpha_k \quad \forall k \in \{0, 1, \dots, n_{h_k}\} \tag{3b}$$

$$u^L \leq u \leq u^U \quad u \in \mathbb{R}^{n_u} \tag{3c}$$

The framework can solve the optimization problem by obtaining expected objective function values from the PM and the probability information of satisfying the chance constraint from the UM. To calculate the probability in chance constrained optimization using the framework, just the cumulative distribution function (CDF) and the expected values of the inequality constraints from the PM are used. These can be derived directly from the outputs of the PM and the UM which depend on the decision variables.

3.3. Computing the probability of a joint chance constraint

One uncertainty model is created for each individual chance constraint defined by the user. In each model, all probability distribution parameters are determined only as a function of all decision variables to calculate the probability constraint. Therefore, the UMs do not depend on each other and can be used in calculating the probability of joint chance constraint method using the following formula:

$$\Pr_{Joint} = \prod_{j=0}^{n_{h_j}} \Pr_j\{h_j(u) \geq 0\} \quad \forall j \in \{0, 1, \dots, n_{h_j}\} \tag{4}$$

In order to carry out and test the approach for joint chance constraints, the proposed method is implemented in the developed framework.

4. Case study

The testing of the extended framework for joint CC is carried out with two different examples: a simple mathematical equation system, and the simplified Williams-Otto process. In both case studies, training and testing datasets (500,000 sample points) are generated according to the procedure described earlier, and data-driven models for the case study are created based on these datasets. The data generation times for the first and second case are 12400 s and 13240 s respectively, whereas their model training times (both the PM and the UM) are 1231 s and 1501 s.

4.1. Case Study I: Simple mathematical model

In order to evaluate the performance of the proposed approach for joint chance constraints, an arbitrary nonlinear equation system with three active constraints is chosen and can be described as follows:

$$\min_{u_1, u_2} (u_1 - 2.1)^2 + \left(\frac{u_2 - 3.9}{40}\right)^2 \quad (5a)$$

$$\Pr_{\text{Joint}} \left\{ \begin{array}{l} 10 \cdot (\xi_3 + u_1 + \xi_2) - u_2 - 75 \geq 0 \\ \exp(-\xi_2 - u_1 + 5.5) - u_2 \leq 0 \\ -10 \cdot (\xi_1 + u_1)^2 - u_2 + 100 \geq 0 \end{array} \right\} \geq \alpha_{\text{Joint}} \quad (5b)$$

with mean values $\xi_1 = 1$, $\xi_2 = 2$, $\xi_3 = 5$ and decision variables u_1 in the interval [1.2, 2.2], u_2 in the interval [2.0, 16.0]. The joint chance-constrained optimization problem is solved for varying α_{Joint} and uncertainty variances σ . To verify the probability of the optimization result (\Pr_{Solver}), Monte Carlo (MC) Simulation is applied based on the uncertainty parameters in the system. To estimate the probability using this method (\Pr_{MC}), numerous simulations with randomly chosen samples from the uncertain parameter distributions are carried out using the original model with fixed decision variables from the optimization result. Then the number of points in the simulation result is counted which simultaneously satisfy all inequality constraints.

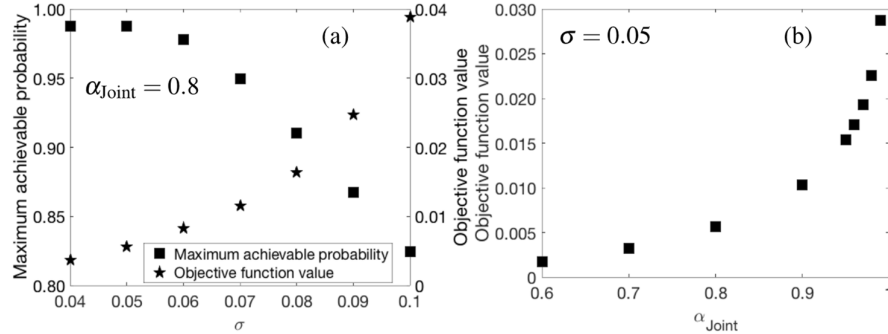


Figure 1: Optimization results: (a) Influence of σ on the objective function value and max. achievable probability, (b) Influence of the probability level α_{Joint} on the objective function value.

Fig. 1 (a) shows the fact that the maximum achievable joint constraint probability decreases with the increasing variance of the uncertainty parameter. This means that all three inner constraints in Eq. (5b) are active simultaneously, affecting the feasible area in which all uncertainty-affected inequality constraints are jointly satisfied. As expected, it can be seen that the objective function value increases with increasing σ value. The objective function values also rise with increasing probability level α_{Joint} (see Fig. 1 (b)), since the optimal solution moves away from the global minimum in order to increase the probability of the distribution being fulfilled. This is also the expected behaviour. According to the optimization results, the probability of the optimization result (\Pr_{Solver}) and the probability obtained by MC simulations (\Pr_{MC}) show a high match rate, and the difference between the two probabilities ($\frac{|\Pr_{\text{MC}} - \Pr_{\text{Solver}}|}{\Pr_{\text{Solver}}}$) is less than 1%. Furthermore, compared to the sampling-based method (Monte Carlo with 10,000 sample points, taking approx. 348 s per iteration), the new approach only takes less than a second to calculate the probability of meeting the joint chance constraint for each optimization iteration. Thus, solving this joint CC optimization problem takes only a few seconds in total.

4.2. Case Study II: Williams-Otto Process

The Williams-Otto process (Williams and Otto, 1960) consists of a continuously stirred tank reactor (CSTR), a decanter, a distillation column and a recycling part of the column into the reactor. Two raw components A and B are fed separately into the CSTR, in which three exothermic reactions take place producing product P, intermediate components C, E and byproduct G. To describe each reaction, an Arrhenius-type kinetic is used. The byproduct G is an expensive waste and hence causes additional costs, which is thus limited to mass fraction 4% at the outlet of the reactor. In addition, it is essential to keep the reactor temperature below 110°C to avoid degradation process. The original model is simplified and therefore the equation system consists of 22 equations. The following optimization problem has been specified for the joint chance constraint:

$$\min_{F_A, F_B} (0.079 - w_P)^2 + (0.045 - w_G)^2 \tag{6a}$$

$$\Pr_{\text{Joint}} \left\{ \begin{array}{l} w_P(F_A, F_B) \geq 0.082 \\ w_G(F_A, F_B) \leq 0.04 \end{array} \right\} \geq \alpha_{\text{Joint}} \tag{6b}$$

The optimization problem has two decision variables, which are the two feed streams for components A and B (F_A and F_B). As uncertain parameters, the three preexponential coefficients of the Arrhenius reaction rates with a standard deviation of 3% of the specified value are used. The objective function is formulated in such a way that the unconstrained global optimum lies outside the feasible area, which enables the constraints (6b) to be activated. The optimization problem for joint chance constraint is formulated based on the description of this section and solved using the data-driven model.

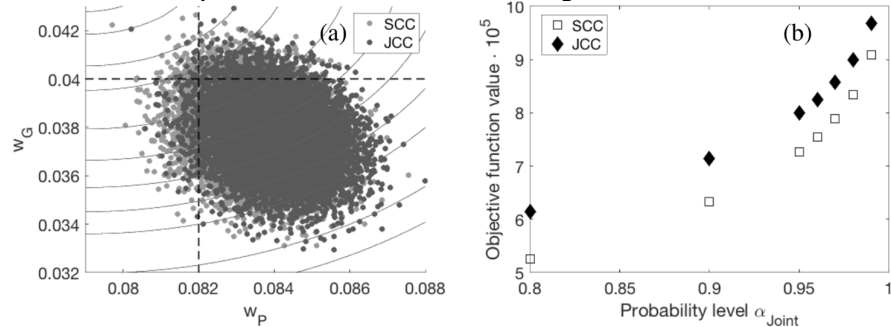


Figure 2: Optimization results: (a) the probability distribution of the concentrations w_P and w_G of the individual sampling point for single and joint chance constraint methods with $\alpha_{\text{Joint}} = 0.95$, (b) Influence of the probability level α_{Joint} on the objective function value.

Figure 2 (a) shows the MC Simulation Results of the concentration w_P and w_G (with 10,000 sample points), taking into account the above mentioned 3 uncertainty parameters at $\alpha_{\text{Joint}} = 0.95$. The single CC method solves the formulated optimization problem by satisfying the two inequality constraints separately ($\Pr_1 \{w_P(F_A, F_B) \geq 0.082\} = 94.8\%$ and $\Pr_2 \{w_G(F_A, F_B) \leq 0.04\} = 94.9\%$, but $\Pr_1 \cap \Pr_2 = 90.4\%$). Since the joint CC approach considers both inequality constraints simultaneously, it is more severe than the single CC with same probability level values. The simulation result distribution keeps a distance from both constraints to jointly satisfy the required probability ($\Pr_{\text{Joint}} = 95.0\%$). In comparison with the single CC, the objective function values from joint CC are shifted to a more conservative solution to simultaneously meet the required concentrations of products and by-products under uncertainty (see Fig. 2 (b)). As already shown in the first case study, the optimization results here also show that the objective function values

increase as expected with increasing α_{Joint} (see Fig. 2 (b)). In this case study, the obtained optimization results also show a good agreement between the probability from the optimization result and probability obtained by Monte Carlo sampling, resulting in a max. deviation of less than 1.5 %. The proposed method also takes less than a second to calculate the probability for each iteration of joint CC optimization compared to the MC sampling technique (approx. 378 s per iteration, with 10,000 sample points). Hence, only a few seconds are required for solving this joint CC optimization problem. Furthermore, the max. operating cooling power limit ($\dot{Q}_{\text{cooling}} \leq 200$ kJ/s), designed not to exceed the max. allowable temperature of the reactor, is integrated into the joint chance constraint (Eq. (6b)) to test the approach for a higher multidimensional distribution. The joint CC optimization problem is formulated and solved successfully at an α_{Joint} of 80 % to 90 %. The maximum possible probability is 90.2 %. The reason for this is that more inequality constraints are jointly considered and therefore the desired probability cannot be reached within a feasible area. The calculation time of the probability for joint CC with these three constraints is not significantly different from the case study II with two constraints.

5. Conclusion and Outlook

In this contribution, an approach for joint CC is proposed to improve the robustness of the single chance constrained optimization solution using a data-driven model. The method is implemented in a developed framework and tested using two case studies. The joint CC method can solve the process optimization problem under uncertainty and provides results while meeting the desired minimum probability level varied from 80% to 99%, which is satisfying all uncertainty-affected inequality constraints jointly. Thus, when constraints are related to safety consideration of a process operation, the joint CC approach using the data-driven model can be applied. In addition, it takes less than a second to calculate the probability of meeting the joint chance constraint. In future work, the extended framework will be tested with measurement data from different chemical processes and will be applied in model-based frameworks for dynamic processes (Weigert et al., 2020). Furthermore, a possible application of the approach for real-time optimization can be investigated.

References

- A. A. Giunta, 2004. Perspectives on Optimization under Uncertainty Algorithms and Applications. 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (1), 1–10.
- P. Li, H. Arellano-Garcia, G. Wozny, 2006. Chance Constrained Programming Approach to Process Optimization under Uncertainty. Elsevier B.V. 32, 25–45.
- J. Weigert, E. Esche, C. Hoffmann, J.-U. Repke, 2019. Generation of data-driven models for chance-constrained optimization. Proceedings of the 9 International Conference on FOCAPD.
- J. Weigert, C. Hoffmann, E. Esche, J.-U. Repke, 2020. Enabling Dynamic Real-Time Optimization under Uncertainty using Data-Driven Chance Constraints. ESCAPE30.
- T. J. Williams, R. E. Otto, 1960. A Generalized Chemical Processing Model for the Investigation of Computer Control. Communication and Electronics. (79), 1–99, 458–473.
- G.M. Ostrovsky, N.N. Ziyatdinov, T.V. Lapteva, 2013, Optimal design of chemical processes with chance constraints, Computers and Chemical Engineering. (59), 74–88
- Harvey Arellano-Garcia, G. Wozny, 2009. Chance Constrained Optimization of Process Systems under Uncertainty, Computers and Chemical Engineering. (33), 1568–1583
- Wets, R. J.-B., 1994, Challenges in Stochastic Programming, WP 94-032. Laxenburg: IIASA
- Yuan Yuan, Zukui Li and Biao Huang, 2017, Robust optimization approximation for joint chance constrained optimization problem, Springer Science+Business Media New York 2016
- Pedregosa, F et. al., 2011. Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825–2830.

CFD Simulations of High Efficiency Gas Cyclones: An Influence of Dustbin Geometry

Pitiwat Pechmanee, Apinan Namkanisorn*, Santi Wattananusorn,
Eakarach Bumrunthaichaichan**

Department of Chemical Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand

* apinan.na@kmitl.ac.th

** b_eakarach@hotmail.com and eakarach.bu@kmitl.ac.th

Abstract

The unsteady state simulation of gas-solid cyclone separator was carried out to investigate the performances of 0.29 m diameter (D_B) cyclones with five different dustbin geometries, including dustbin without dipleg (cylindrical bin) and dustbins with $0.5D_B$, $1.0D_B$, $1.5D_B$, and $2.0D_B$ (divergent conical bin) height divergent conical diplegs. The diameter and total height of five dustbins were $1D_B$ and $2D_B$, respectively. The gas flow and turbulence fields inside the cyclones with the Reynolds number of 280,000 were simulated by Reynolds averaged Navier-Stokes equations (RANS) with Reynolds stress model (RSM). The collection efficiencies were investigated by using discrete phase model (DPM). For model validation, the simulated velocity profiles of the cyclone with cylindrical dustbin have been compared to the previous experimental data available in literature and were in good agreement with the previous results. Further, the simulated results revealed that the Stairmand cyclone with divergent conical and simple cylindrical dustbins respectively represented the highest and lowest collection efficiencies indicated by 50% cut-off diameter, which corresponded to the diameters of 1.692 and 1.744 microns, respectively.

Keywords: CFD, cyclone, dipleg, dustbin, and separation

1. Introduction

Gas cyclones are one of the important gas-solid separators. They are commonly employed in various chemical industries because of their advantages, e.g., simple geometry, no moving parts, low operating cost, etc. They use centrifugal force to remove the solid particles from gas streams. Due to the cyclone advantages, many researchers attempted to enhance their performances by modifying the cyclone parts.

Cyclone inlet, cyclone barrel, cyclone cone, cyclone vortex finder, dipleg, and dustbin were previously studied to improve the cyclone performances by using experiments and computational fluid dynamics (CFD) simulations. The dust outlet geometry was one of the cyclone parts which widely studied by previous works.

Kaya and Karagoz (2009) employed CFD to investigate the gas cyclones prolonged with diplegs. They concluded that the dipleg length considerably affected the collection

efficiency rather than the pressure drop, especially for lower inlet velocities. Elsayed and Lacor (2012) numerically studied the influence of dustbin geometry. They reported that the different pressure drops and collection efficiencies were observed for different dustbin geometries. Moreover, the dust outlet geometry significantly affected the axial velocity profiles. Shapiro and Ben-Shmuel (2012) also used CFD to study the effects of cylindrical and divergent diplegs on pressure drop, cut size, and separation efficiency. Later, the cylindrical dustbins and diplegs (Elsayed et al., 2020) and cylindrical dustbins with various dipleg designs (Parvaz et al., 2020) were numerically investigated.

Although the different dustbin geometries were studied as mentioned earlier, however, the effect of angle for conical dipleg is not investigated. Hence, in this paper, this influence was investigated with the help of CFD. The five different dustbin geometries, including dustbin without dipleg (cylindrical bin) and dustbins with $0.5D_B$, $1.0D_B$, $1.5D_B$, and $2.0D_B$ (divergent conical bin) height divergent conical diplegs, were studied.

2. CFD simulations

2.1. Configurations of gas-solid cyclones

In this paper, the 0.29 m diameter Stairmand cyclone as reported by Hoekstra (2000) was preliminary considered. Further, the Stairmand cyclones with different dust outlet geometry designs were studied. The dimensions of these cyclones can be summarized as depicted in Table 1.

Table 1: Dimensions of the studied gas cyclones

Dimension	Length (mm)				
	$0.0D_B$	$0.5D_B$	$1.0D_B$	$1.5D_B$	$2.0D_B$
Inlet height, a	145	145	145	145	145
Inlet width, b	58	58	58	58	58
Inlet length, L_i	246.5	246.5	246.5	246.5	246.5
Barrel Diameter, D_B	290	290	290	290	290
Cone tip diameter, D_C	108	108	108	108	108
Dustbin diameter, D_D	290	290	290	290	290
Vortex finder diameter, D_V	145	145	145	145	145
Cyclone height, h_T	1,160	1,160	1,160	1,160	1,160
Barrel height, h_B	435	435	435	435	435
Inner vortex finder height, S	145	145	145	145	145
Outer vortex finder height, h_V	435	435	435	435	435
Dipleg height, h_L	-	145	290	435	580
Dustbin height, h_D	580	435	290	145	-

2.2. Grid generation

The hexahedral grids were generated inside gas cyclones with the help of domain decomposition technique (Bumrunghthaichaichan and Wattanusorn, 2019) by using GAMBIT 2.4.6 as demonstrated in Figure 1.

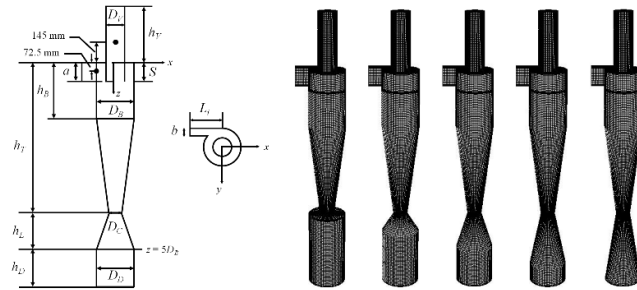


Figure 1: Schematic diagram of cyclone and surface grids for the five tested gas cyclones

2.3. Governing equations

The flow and turbulence fields inside gas cyclones were obtained by solving Reynolds averaged Navier-Stokes equations (RANS) with Reynolds stress model (RSM). Further, the particle motion was simulated by discrete phase model (DPM). The description of these governing equations were clearly described by Elsayed and Lacor (2012).

2.4. Material properties and boundary conditions

The density (ρ) and viscosity (μ) of air were $1.096 \text{ kg}\cdot\text{m}^{-3}$ and $1.81623 \times 10^{-5} \text{ kg}\cdot\text{m}^{-1}\cdot\text{s}^{-1}$, respectively. The solid density were $2,740 \text{ kg}\cdot\text{m}^{-3}$ (Derksen, 2003). At inlet, the velocity-inlet boundary condition type was adopted. The uniform inlet velocity (U_{in}) of $16 \text{ m}\cdot\text{s}^{-1}$ was specified, which corresponded to the Reynolds number based on barrel diameter of 280,000. The turbulence kinetic energy (k) and turbulence kinetic energy dissipation rate (ε) were respectively computed by $k = 1.5(U_{in}l)^2$ and $\varepsilon = C_\mu^{3/4}k^{3/2}/l$. From isotropic turbulence assumption, the normal stresses were $2k/3$. Moreover, the shear stresses were zero. The outflow boundary condition type was assumed at outlet. Further, the no-slip boundary condition and scalable wall functions were employed at the cyclone wall.

2.5. Numerical methods and solution strategy

The ANSYS FLUENT finite volume CFD code was adopted to simulate gas cyclones. The pressure-velocity coupling scheme was SIMPLEC. The spatial discretization schemes of pressure, momentum, turbulence quantities, and Reynold stresses were PRESTO!, QUICK, second order upwind, and first order upwind, respectively (Elsayed and Lacor, 2012). The temporal discretization scheme was second order implicit (Shukla et al., 2011). Further, the time step size of 0.0001 s and maximum iterations per time step of 40 were used. The scaled residuals of 10^{-5} for all quantities were adopted as the convergence criteria.

2.6. Particle motion simulation and collection efficiency prediction

The particle motion was simulated by DPM because of the low particle loading. In order to obtain grade efficiency curves for the five cyclones, the 10,660 uniform particles with the different diameters, including 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, and 10 microns, were released at the cyclone inlet section. The inlet particle and gas velocities were assumed to be identical. Moreover, the collection efficiencies (η) for different particle sizes were evaluated by the number of escaped particles (n_{out}) and the total number of injected particles (n_{total}) and as shown in Eq. (1).

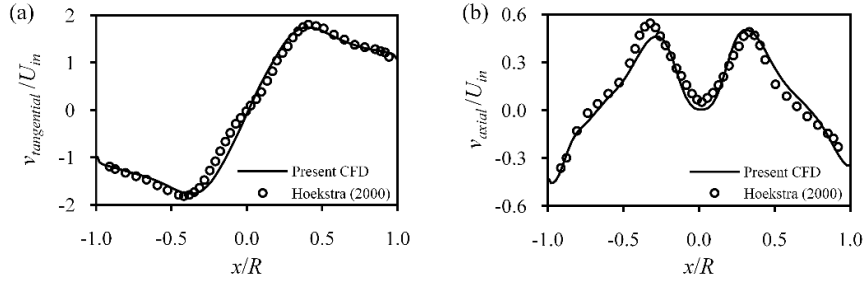


Figure 2: Comparison of (a) tangential velocity profile and (b) axial velocity profile at $z = 0.75D_B$ between the present CFD and the experimental data of Hoekstra (2000)

$$\eta = \left(1 - \frac{n_{\text{out}}}{n_{\text{total}}} \right) \times 100 (\%) \quad (1)$$

2.7. Grid independence study

Regarding to the previous work of Pechmanee (2019), he compared the predicted velocity profiles inside 0.29 m diameter Stairmand cyclone without dustbin of three different grid resolutions, including 90,436, 199,440, and 450,498 cells. The results indicated that the grid independency was obtained by medium grid level. However, in order to eliminate any uncertainties, he employed the fine grid level to study the gas cyclones. So, in the present work, the same grid qualities, e.g., near-wall grid size, growth ratio, etc., were generated inside cyclone barrel and cone. Moreover, the similar grid qualities were also applied inside five different dustbin geometries. The total number of grids for five tested cyclones were 756,788 cells.

3. Validation

The predicted tangential and axial velocity profiles at z of $0.75D_b$ below the cyclone roof were validated by comparing with the experimental data of Hoekstra (2000) as demonstrated in Figure 2. In Figure 2, the predicted profile and maximum value of the tangential velocity are in good agreement with the experiment. Whereas, the comparison shows the underprediction of the axial velocity at the inner vortex region. Due to the complexity of flow inside cyclone, it can be considered that the present CFD model was reasonably employed to study the effect of angle for conical dipleg.

4. Results and discussion

4.1. Cyclone static pressure and pressure drop

Figure 3 represents the contours of static pressure at plane $y = 0$ for five different cyclones. These contours exhibit that the static pressure is found to increase radially from the core vortex at centreline to the cyclone wall. Further, the pressure drop of these cyclones are slightly different. The percentage difference of pressure drop, which can be defined as a ratio of the difference between the maximum and minimum pressure drops to the

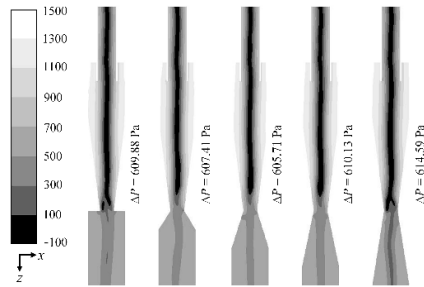


Figure 3: Contours of static pressure at plane $y = 0$ for the five tested gas cyclones (unit in Pa)

maximum value, was lower than 1.5%. It can be implied that the dustbin geometries insignificantly affected the cyclone pressure drop.

4.2. Tangential velocity and axial velocity

The tangential velocity and axial velocity contours of five different cyclones are shown in Figure 4. The results reveal that the tangential and axial velocities inside cyclone barrel and cone are slightly different. At the region near cone tip and dustbin, the velocity fields are significantly different, especially for tangential velocity. It can be summarized that the dustbin shapes influenced on the flow patterns.

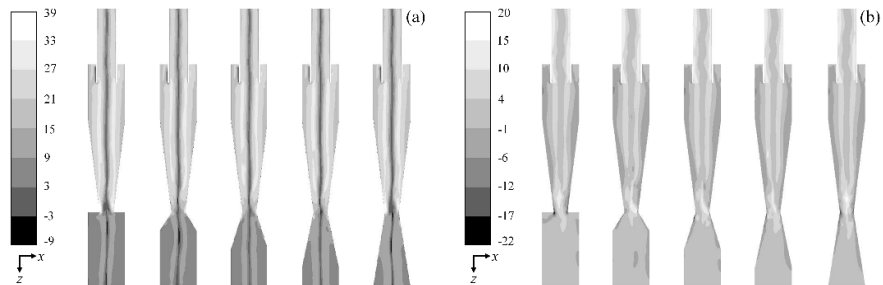


Figure 4: Contours of (a) tangential velocity and (b) axial velocity at plane $y = 0$ for the five tested gas cyclones (unit in $m \cdot s^{-1}$)

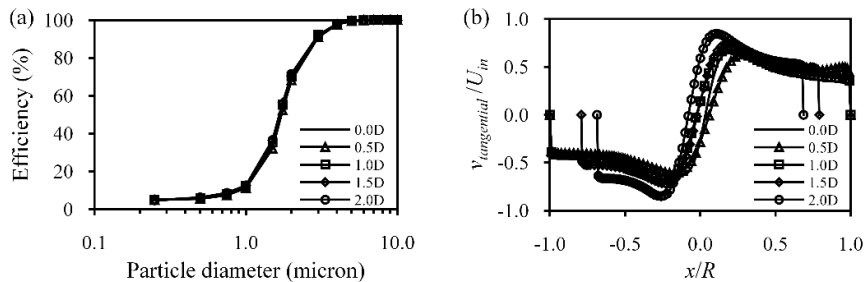


Figure 5: Comparison of (a) grade efficiency curves and (b) tangential velocity profiles at $z = 5D_B$ for five tested cyclones

4.3. Collection efficiency

The grade efficiency curves of five different tested cyclones are plotted in Figure 5(a). Figure 5(a) shows that the cyclone efficiencies are found to increase with increasing the particle sizes and these efficiency curves are similar. However, the 50% cut-off diameters were determined to investigate the collection efficiencies of these cyclones. The cut-off diameters of $0.0D_B$, $0.5D_B$, $1.0D_B$, $1.5D_B$, and $2.0D_B$ cyclones were 1.744, 1.749, 1.711, 1.699, and 1.692 microns, respectively. Meaning that, the cyclone with divergent conical bin ($2.0D_B$ cyclone) provided the best separation performance as comparing with the others because of the highest magnitude of tangential velocity as shown in Figure 5(b).

5. Conclusions

In this paper, the effect of angle for cyclone conical dipleg was investigated by using RANS with RSM turbulence model and DPM. The validation confirmed that the present CFD model was reasonably adopted to study the conical dipleg effect. The simulated results indicated that the dustbin shape affected on gas flow patterns, especially for tangential velocity. Moreover, the 50% cut-off diameter of $2.0D_B$ cyclone was about 3% better than cyclone with cylindrical dustbin ($0.0D_B$ cyclone). Meaning that, the change in angle for cyclone conical dipleg can improve the separation performance.

Acknowledgement

The authors would like to sincerely thank to College of Advanced Manufacturing Innovation, King Mongkut's Institute of Technology Ladkrabang, for supplying the ANSYS FLUENT software.

References

- Bumrunthaichaichan, E., Wattananusorn, S., 2019. CFD modelling of pump-around jet mixing tanks: a reliable model for overall mixing time prediction. *J. Chinese Inst. Eng.* 42, 428–437.
- Derksen, J.J., 2003. Separation performance predictions of a stairmand high-efficiency cyclone. *AIChE J.* 49, 1359–1371.
- Elsayed, K., Lacor, C., 2012. The effect of the dust outlet geometry on the performance and hydrodynamics of gas cyclones. *Comput. Fluids* 68, 134–147.
- Elsayed, K., Parvaz, F., Hosseini, S.H., Ahmadi, G., 2020. Influence of the dipleg and dustbin dimensions on performance of gas cyclones: An optimization study. *Sep. Purif. Technol.* 239, 116553.
- Hoekstra, A.J., 2000. Gas Flow Field and Collection Efficiency of Cyclone Separators. Technische Universiteit Delft.
- Kaya, F., Karagoz, I., 2009. Numerical investigation of performance characteristics of a cyclone prolonged with a dipleg. *Chem. Eng. J.* 151, 39–45.
- Parvaz, F., Hosseini, S.H., Elsayed, K., Ahmadi, G., 2020. Influence of the dipleg shape on the performance of gas cyclones. *Sep. Purif. Technol.* 233, 116000.
- Pechmanee, P., 2019. Simulation of Fluid Flow Field Inside Gas Cyclone Separator by Computational Fluid Dynamics. King Mongkut's Institute of Technology Ladkrabang.
- Shapiro, M., Ben-Shmuel, Y., 2012. Modeling performance of a tangential flow cyclone: Effects of secondary outlet geometry and boundary conditions. *Part. Sci. Technol.* 30, 431–450.
- Shukla, S.K., Shukla, P., Ghosh, P., 2011. Evaluation of numerical schemes using different simulation methods for the continuous phase modeling of cyclone separators. *Adv. Powder Technol.* 22, 209–219.

InvQMOM: A Simple Inversion Method that Works

Menwer Attarakih,^{a,b,*}, Armin Fricke^{b,c}, Hans-Jörg Bart^{c,b}

^a*The University of Jordan, Chemical Engineering Department, Amman, 11942, Jordan*

^b*Population Balance Modeling GmbH, Poststr 12, 10178 Berlin, Germany*

^c*CGC Capital-Gain Consultants GmbH, Poststrasse 12, 10178 Berlin, Germany*

^d*TU Kaiserslautern, 67653 Kaiserslautern, Germany*

m.attarakih@ju.edu.jo

Abstract

We present a universal inversion method to reconstruct the underlying probability density function (pdf) and its cumulative distribution (cdf) from the embedded information in the quadrature nodes and weights of QMOM. These nodes and weights are evolved according to the kinetics governing the interactions among the particles of the dispersed phase such as nucleation, condensation, evaporation, splitting and aggregation. This pdf is represented by a dynamic histogram supported with moving bin boundaries. The inverted pdf and cdf are general and free of any prior assumption about the type of the distribution. We called this method “InvQMOM” which is supposed to converge to the exact pdf and cdf as the number of quadrature nodes increases. The order of convergence of the inverted pdf is $O(h^2)$ and $O(h^3)$ for cdf, where h is the spacing between the two quadrature nodes that encompass the node from which the point functional value is reconstructed.

Keywords: InvQMOM, probability density, cumulative distribution, Inverse.

Introduction

Population balances is an established transport framework which is used to model the evolution of dispersed phase number density concentration. These systems are of core industrial, biological and medical importance where numerical solutions are called for in general (Bart et al., 2020, Durr and Buck, 2020). The Quadrature Method of Moments (QMOM) is one of the most elegant reduction methods used to solve such complex transport equations (McGraw, 1997). Since its introduction by McGraw (1997) it has found considerable theoretical and applied research during the last two decades (Bart et al., 2020). These applications cover wide areas which include aerosols microphysics, nanoparticle formation, turbulent mixing, dispersed phase flows, crystal morphology, cell growth and differentiation and uncertainty propagation through dynamical systems, to name but a few (Attarakih et al., 2019, 2020, Bart et al., 2020, Durr and Buck, 2020). Since QMOM is an integral averaging method, there is no simple direct and general inversion algorithm to reconstruct the unknown weight function which describes the distribution of the particulate phase from the available average information embedded in its moments. To fill this gap, a universal inversion method is proposed to reconstruct the probability density function and its cumulative distribution from the adaptive quadrature nodes and weights of QMOM. The weights of QMOM are local information propagated along the characteristics of the internal particle space according to the kinetics governing the interactions among the particles of the dispersed phase which include nucleation, condensation, evaporation, splitting and aggregation. This pdf can be represented by a

dynamic histogram supported by moving bin boundaries. The inverted pdf and cdf are general and free of any prior assumption about the type of the sought distribution. We called this method Inverse the QMOM: “InvQMOM” to emphasize the inversion of the pdf from the characteristic points and weights (probability per unit system volume) of the QMOM. The order of convergence of the inverted pdf is $O(h^2)$ and $O(h^3)$ for cdf, where h is the spacing between the two quadrature nodes along the particle property space that encompass the node from which the point functional value is reconstructed.

InvQMOM

The integral form of the Population Balance Equation in one particle property space (particle diameter x) over an average number concentration function $f(x,r,t)$ is given by:

$$\begin{aligned} \frac{\partial m_r}{\partial t} + \nabla \cdot \langle \mathbf{u} \rangle m_r + \int_{x_{\min}}^{x_{\max}} \left[\frac{\partial(Gx^r f)}{\partial x} \right] dx' = \int_{x_{\min}}^{x_{\max}} f \Gamma x^r [\pi_r - 1] dx' \\ + \int_{x_{\min}}^{x_{\max}} \int_{x_{\min}}^{x_{\max}} [\omega f(x) f(x') [x^3 + x'^3]^{r/3} - x^r - x'^r] dx dx', r = 0, 1, \dots \end{aligned} \quad (1)$$

The nonlinear integral source term takes into account particle breakage and aggregation while the integral particle growth term appears as an additional positive source term. The mean velocity vector of the particulate phase is $\langle \mathbf{u} \rangle$, the r^{th} moment of the number concentration function $f(x, \cdot)$ is m_r and π_r is the r^{th} moment of the dimensionless daughter particle size with respect to a given daughter particle distribution where $r = 0, 1, \dots$. Due to the general form of the respective particle growth, breakage and aggregation functions G , Γ and ω , the integral terms in Eq.(1) cannot be written in terms of m_r and hence present a challenging closure problem. These integrals remain without general closure algorithm until the introduction of the QMOM by McGraw (1997) where an adaptive and boundary free Gauss-Christoffel quadrature is introduced by transforming the source term in terms of a finite set of low order moments m_r , $r = 0, 1, \dots, N_q$. This transformation leads to the loss of the local information in the particle number concentration function $f(x, \cdot)$ which becomes latent in the low order moments m_r . Decoding of this information during the evolution of these moments in space and time is a nontrivial problem (Attarakih et al., 2020, 2020, Durr and Buck, 2020). In theoretical physics, this problem is known as the classical moment problem (CMP) and in particular the Stieltjes or Hausdorff moment problem (Gzyl & Tagliani, 2010). The CMP tries to recover a unique and positive number concentration function $f(x, \cdot)$ from its low order moments. The methods of inversion and the method of moments were reviewed recently by Attarakih et al. (2013, 2019, 2020) and Durr and Bruck (2020) respectively. Motivated by the QMOM, one can physically interpret the adaptive quadrature node as characteristic points (discrete curves) in the t - ξ space for homogeneous flow systems where ξ is the particle property space and t is time.

The QMOM

The QMOM can be mathematically presented as an adaptive Gauss-Christoffel quadrature for $0 \leq \xi < \infty$ as follows:

$$(\zeta_j, w_j) = Al(m_r), r = 0, 1, \dots, 2N_q - 1, j = 0, 1, \dots, N_q - 1 \quad (2)$$

where (w) is quadrature weight vector, (ζ) is quadrature node vector and (Al) is a suitable inversion algorithm which is required to solve the following system of nonlinear equations and function of m_r only:

$$m_r = \sum_{j=0}^{N_q-1} w_j \zeta_j^r, \quad r = 0, 1, \dots, 2N_q - 1 \quad (3)$$

Two famous algorithms (A1) are used to solve Eq.(3) which are the PD-Algorithm as adapted by McGraw (1997) and the Chebyshev-QMOM algorithm proposed and evaluated by Upadhyay (2012). Upadhyay (2012) showed that the Chebyshev-QMOM is superior to the PD-Algorithm. The Chebyshev-QMOM algorithm is found to be more robust than the PD-algorithm and can be used for a wider class of problems when high number of nodes is required as in our present work.

InvQMOM with local integral limits

Note that the quadrature weights (w) given by Eq.(2) are probabilities of f(x,.) per unit system volume sampled carefully from precomputed points (nodes) (ζ). By definition, these probabilities are given by:

$$w_j = \int_{\zeta_{j-1/2}}^{\zeta_{j+1/2}} f(x,.) dx, \quad j = 0, 1, \dots, N_q - 1 \quad (4)$$

where $\zeta_{j\pm 1/2}$ are unknown grid boundaries (histogram bins); but they satisfy $\zeta_{j-1/2} < \zeta_j < \zeta_{j+1/2}$. If f(x,.) is normalized, then the vector (w) is composed of true probabilities that are satisfying the natural constraint ($\sum w_j = 1$). The problem now can be posed as follows: Given an accurate value of the true probability (w_j) at a characteristic point ζ_j , then invert Eq.(4) to find an estimate of f(ζ_j ;:). This is an inverse problem where a unique solution is required with sufficient accuracy to approximate f(ζ_j ;:) from the available information. The integral in Eq.(4) can be approximated using the one-point Gauss-Legendre quadrature or the midpoint rule which results in the following normalized discrete probability density approximation of f(x,.) at each characteristic point ζ_j :

$$f^*(\zeta_j;:) = \frac{1}{\Delta\zeta_j} \frac{w_j}{\sum_j w_j} + O(\Delta\zeta_j^2), \quad j = 0, 1, \dots, N_q - 1 \quad (5)$$

where $\Delta\zeta_j = \zeta_{j+1} - \zeta_{j-1}$ with the constraint $\zeta_{j-1} < \zeta_{j-1/2} < \zeta_j$ and $\zeta_j < \zeta_{j+1/2} < \zeta_{j+1}$. Eq.(5) provides a complete theoretical inversion of the latent probability density function $f^*(x,;) = f(x,;)/m_0$. The order of convergence is proportional to the square of the distance between the (j-1) and (j+1) characteristic curves of the particle size.

$$F^*(\zeta_{j+1/2};:) = \frac{1}{m_0} \sum_{j=0}^j w_j + O(\Delta\zeta_j^3), \quad j = 0, 1, \dots, N_q - 1 \quad (6)$$

Moreover, the cumulative distribution of f(x;:) can be easily derived from Eq.(5) and is presented in Eq.(6), where m_0 is the zero moment of f(x,.) and $\zeta_{j+1/2} = 0.5[\zeta_j + \zeta_{j+1}]$. It is clear by comparing the order of errors in Eq.(5) and (6) that the inversion of cumulative distribution function from the available QMOM information (ζ_j, w_j) is more accurate than the recovery of the pointwise values of f(x;:). The two normalized pointwise and cumulative densities given by Eqs.(5) and (6) are formally the two inversion formulas for the QMOM and are named the “InvQMOM”.

Results and discussion

In this section we present sample results of using the proposed InvQMOM. The method is validated intensively including analytical cases where exact solution is available, cases of fluid hydrodynamics due to droplet breakage in continuous industrial stirred tank, uncertainty propagation in smoke height prediction in fire system design and uncertainty of the reproduction number of the SIR (Susceptible, Infectious and Recovered) model used in the prediction of COVID-19 dynamics. Due to the limited space, two practical and intensive case studies are presented below.

Droplet breakage in industrial scale vessel

As an example of our inversion results, Figure (1-Left) shows a sample from 26 low-order moments as function of time for the population balance equation describing the dynamics of a continuous industrial scale vessel with 10 m³ volume (Alopaeus et al., 1999).

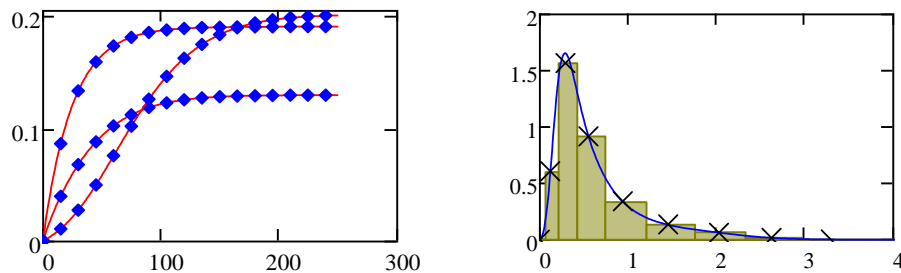


Figure (1): **Left:** Comparison between the first three low-order moments (m_0 , m_2 , m_3 starting from origin) using Chebyshev-QMOM and the OPOSPM-Weibull corrected method for liquid droplet breakage in a continuous industrial stirred tank (x-axis is time). **Right:** Reconstructed pdf using InvQMOM with 13 quadrature nodes at final time of 250 s (x-axis is dimensionless droplet diameter) as compared to OPOSPM-Weibull method (solid line).

The continuous red lines are calculated by the Chebyshev-QMOM while the blue diamonds are based on the pointwise evolution of the probability density function using the OPOSPM-Weibull corrected method (Attarakih and Bart, 2020). The accuracies of the two methods are extremely high; however, the Chebyshev-QMOM method requires CPU time around 50 percent less than that of the OPOSPM-Weibull corrected method at the expense of lost distribution. Now having a set of accurate moments at hand, the pdf is reconstructed from the available nodes and weights using the InvQMOM method (Eq.(5)) as shown in Figure (1-Right). The accuracy of the inverted pdf as compared to the continuous one as predicted by the OPOSPM-Weibull corrected method (solid line) is remarkable (Attarakih & Bart (2020)). Note that the InvQMOM is a post processing step where the pdf and cdf can be reconstructed. The only limitation that might arise is the difficulty to calculate a number of nodes more than 10 by the Chebyshev-QMOM. However, it is found that this limitation is not a problem for pure particle growth as described by the population balance equation. Since InvQMOM is a post processing step, it has no profound effect on the CPU time of QMOM where the inversion step has negligible CPU time and does not depends on the original QMOM complexity.

Uncertainty propagation of smoke layer height

The model that we used here is the Available Safe Egress Time (ASET) model (Upadhyay & Ezekoye, 2008). This is a deterministic fire model with input parameters (area of enclosure and ventilation, height and location of events) that are assumed to be known

with no random components. However, the heat release rate is considered to be uncertain variable and is sampled from the historical data of known distribution. Here the beta distribution is used with two shape parameters and a scaling factor to model the heat release rate.

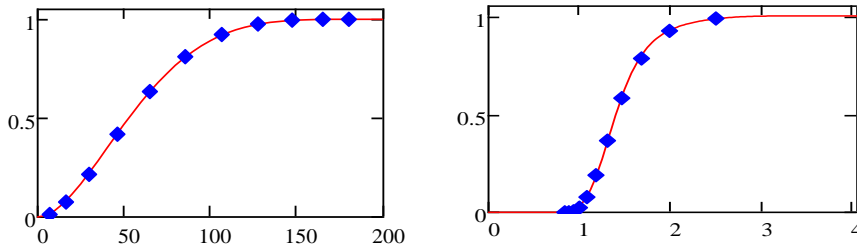


Figure (2): **Left:** Comparison between beta cdf used to model uncertainty in the input heat release rate (solid line) and the inverted cdf using InvQMOM (filled diamonds) with Chebyshev-QMOM ($N_q = 12$, x-axis is R). **Right:** Comparison between inverted smoke layer height cdf using InvQMOM (filled diamonds) at $t_c = 120$ s (x-axis is smoke layer height) with the reconstructed cdf (solid line) using the Differential Maximum Entropy Method (Attarakih & Bart, 2014).

Mathematically, let the stochastic heat release rate be (R) with a pdf $\beta(R/b, s_1, s_2)$ where $s_1 = 1$ and $s_2 = 4$ are shape factors and $b = 200$ is scaling parameter (case I in Upadhyay & Ezekoye (2008)). This input stochastic uncertainty is propagated as an output of the ASET model with pdf $f(z, t)$. For a given critical time (t_c) at which $z(t_c)$ is the smoke layer height from the ground, the input uncertainty is propagated by the equivalence principle between the two pdf's: $\beta(R, t_0)dR = f(z, t_c)dz$ where the propagated information is conserved between $t = t_0$ to $t = t_c$. The smoke layer height $z(t)$ is obtained by the dynamic solution of ASET model with p parameters (Upadhyay & Ezekoye, 2008):

$$\frac{dz(R, t)}{dt} = G(p, R), z(0) = z_0 \tag{7}$$

$$m_r = \int_0^\infty Z(t_c)^r f(z) dz = \int_0^\infty Z(t_c)^r m(R) dR = \sum_{j=0}^{N_q-1} w_j Z_j(t_c)^r \tag{8}$$

The classical way to propagate the uncertainty in the heat release rate is to use the Monte Carlo (MC) simulation by sampling the random variable (R) from its known distribution ($\beta(R)$). Then the fire model is integrated using a given random sample (R) up to $t = t_c$. Repeating this procedure sufficient number of times will produce the required cdf of the propagated uncertainty of $z(R, t_c)$. If the model is sufficiently complex, then the MC method will be computationally expensive. To circumvent this, we follow the same idea presented by Upadhyay & Ezekoye (2008) where the characteristic values (ζ_j) are used as samples of the random variable (R). Keeping in mind that the propagated information is conserved, then the low order moments of $f(z, t_c)$ are calculated from those of the pdf of the input uncertainty as given by Eq.(8). To this end, the output uncertainty cdf is required to provide all the statistical information required for risk assessments. With the QMOM nodes and weights at hand, the InvQMOM is used to reconstruct the cdf of smoke layer height at given time (t_c) where the results of this case study are shown in Figure (2). In Figure (2-Right), the cdf of the propagated input uncertainty (Left) is extracted from the moments of Eq.(8) using InvQMOM (Eq.(6) with 12 nodes) which is compared to the

inverted cdf (solid line) using the DMaxEntM (Attarakih & Bart (2014)). This cdf is useful for risk assessments since the probability of the layer being below any given value is immediately available from the inverted cdf as shown in Figure (2-Right). Note that 12 quadrature nodes were used to sample the input uncertainty from the cdf of beta distribution as shown in Figure (2-Left) (case I in Upadhyay and Ezekoye (2008)).

Conclusions

The InvQMOM is a post processing method which can be used in complex computational environment such as CFD and online real time simulations. As a main conclusion, the InvQMOM is a universal distribution reconstruction method when the QMOM and its variants are used. It needs only a sufficient number of nodes and weights to work with negligible computational cost. Using different analytical and practical case studies, it is found that five to ten nodes were sufficient to reconstruct the pdf and cdf with an accuracy of $O(h^2)$ and $O(h^3)$ respectively.

References

- V. Alopaeus, J. Koskinen & K. I. Keskinen, 1999, Simulation of population balances for liquid-liquid systems in a nonideal stirred tank. Part I description and qualitative validation of model. *Chemical Engineering Science*, 54, 5887-5899.
- M. Attarakih & H.-J. Bart, 2020, Beyond OPOSPM: A Corrected Maximum Entropy Weibull Distribution for Solving Population Balances. *Computer-Aided Chemical Engineering*, 48, 691-696.
- M. Attarakih, H.-J. Bart & M. Abu-Khader, 2019, On the solution of the population balance equation: From global to local constrained maximum entropy method. *Chemical Engineering Science*, 209, 115168.
- M. Attarakih & H.-J. Bart, 2014, Solution of the Population Balance Equation using the Differential Maximum Entropy Method (DMaxEntM): An application to liquid extraction columns. *Chemical Engineering Science*, 108, 123-133.
- M. Attarakih, 2013, Integral formulation of the population balance equation: Application to particulate systems with particle growth, *Computers & Chemical Engineering*, 48, 1-13.
- H.-J. Bart, H. Jildeh & M. Attarakih, 2020, Population Balances for Extraction Column Simulations-An Overview. *Solvent Extraction and Ion Exchange*, 38, 14-65.
- R. Durr & A. Buck, 2020, Approximate Moment Methods for Population Balance Equations in Particulate and Bioengineering Processes. *Processes*, 8, 1-18.
- H. Gzyl & A. Tagliani, 2010, Stieltjes moment problem and fractional moments. *Applied Mathematics and Computation*, 216, 3307-3318.
- R. McGraw, 1997, Description of aerosol dynamics by the quadrature method of moments. *Aerosol Science & Technology*, 27, 255-265.
- R. R. Upadhyay & O. A. Ezekoye, 2008, Treatment of design fire uncertainty using Quadrature Method of Moments. *Fire Safety Journal*, 43, 127-139.
- R. R. Upadhyay, 2012, Evaluation of the use of the Chebyshev algorithm with the quadrature method of moments for simulating aerosol dynamics. *Journal of Aerosol Science*, 44, 11-23.

A Dual-Layer Attention-Based LSTM Network for Fed-batch Fermentation Process Modelling

Kai Liu, Jie Zhang

School of Engineering, Merz Court, Newcastle University, Newcastle Upon Tyne NE1 7RU, UK

jie.zhang@newcastle.ac.uk

Abstract

Recurrent neural network (RNN) is a dynamic neural network where the current network output is related to the previous outputs. Long short-term memory network (LSTM) has emerged as a high-performance RNN. However, the original LSTM does not consider variable and sample relevance for process modelling. To overcome this problem, the paper proposes a Dual-layer Attention-based LSTM (DA-LSTM) network to model a fed-batch fermentation process. In the proposed DA-LSTM, LSTM is used to extract features of the input data and multiple time series results of the hidden layer, an encoder input attention mechanism is to select relevant driving series in the input data sequence, and a temporal decoder attention mechanism is used to measure the importance of encoder hidden states. The model with this deep architecture for high-level representations can learn very complex dynamic systems. To demonstrate the effectiveness of the proposed method, a comparative study with the original LSTM, signal attention-based LSTM is carried out. It is shown that the proposed method gives better modelling performance than others.

Keywords: LSTM, attention mechanism, machine learning, fed-batch processes, encoder-decoder

1. Introduction

Batch processes are suitable for the responsive manufacturing of high value-added products such as pharmaceuticals, polymers and semiconductors. In recent years, due to the increasing demand for high value-added products and the high variety of such products, the proportion of batch processes in modern industry has increased rapidly. In batch processes, maximising the yield while satisfying the quality and operation constraints is the main objective, but it is hard to achieve because of the process characteristics including non-steady-state condition, batch-to-batch variations and strong nonlinearity. Process optimisation based on detailed mechanistic models for responsive manufacturing using batch processes is infeasible or very difficult to implement in practice. To overcome this difficulty, data-driven modelling and optimisation using statistical and machine learning techniques should be capitalised for batch processes.

In order to analyze and mine the information from monitoring data, several data-driven modelling methods can be utilized for modelling complex batch processes, such as principal component analysis, principal component regression, and partial least squares.

Compared with these data-driven modelling methods, artificial neural networks which also known as machine learning have attracted an increased attention in data-driven nonlinear modelling. However, due to complex physical and chemical reaction mechanisms, industrial processes, especially batch processes, have inherently complex nonlinear dynamic behavior. To model highly nonlinear processes, recurrent neural networks (RNN) are widely used. In contrast with feedforward neural networks, some neuron outputs are feedback to the neurons themselves or other neurons in the previous layer, so the basic characteristic of RNN is to exhibit dynamic temporal behavior and process arbitrary sequences of inputs by its internal memory. The traditional RNNs suffer gradient vanishing and/or exploding problems when computing the gradients of network outputs with respect to network weights (Pascanu et al., 2013).

To overcome gradient vanishing or exploding, the gated recurrent neural networks have been established, including long short-term memory (LSTM) and gated recurrent units (GRU) (Schmidhuber et al., 1997). LSTM and GRU networks allow creating gradient paths that neither disappear nor explode along time. Although LSTM is very useful in capturing long-term dependencies, it cannot focus on different variables in different time steps. To solve this problem, a well-constructed autoencoder network which is also named as encoder-decoder network is developed (Cho et al., 2014). Based on the gated RNN units, the sequence-to-sequence problems can be solved by encoder-decoder network. Through the encoder part, the input sequence is converted into a fixed-length vector, and then the decoder converts the generated fixed-length vector into an output sequence. However, as the length of the input sequence increases, the performance of the encoder-decoder network will rapidly decline. Moreover, due to the quality variables are usually predicted based on long lagged input sequences, this problem is more serious when applied to product quality prediction in batch processes. Attention-based encoder-decoder architecture attempts to distinguish hidden states with different attention weights among all time steps of a prediction window and could be a solution to this problem.

To address the above issues, a new multiple-stage attention-based LSTM network is proposed in this paper. It is inspired by the theory of human attention that the human attention is modelled by multiple-stage attention mechanism. In the first attention stage, attention mechanism is used to adaptively extract the relevant information in the input series at each time step by referring to previous encoder hidden state. In the second attention stage, a temporal attention mechanism is proposed to select high correlation encoder output. Both attention mechanisms are integrated with double layers long short-term memory networks and can be trained by back-propagation.

2. Methodology

2.1. Long Short-Term Memory Network

Long short-term memory network is an advanced recurrent neural network (Hochreiter and Schmidhuber, 1997) and provides a well-constructed structure by establishing “gates” in its basic unit which is named as “cell”. These gates can capture both the long-term memory and short-term memory along the time steps and avoid gradient exploding and/or vanishing in standard RNNs. The gates are named as “forget gate”, “input gate”, and “output gate”. The detail structure of LSTM is presented in Figure 1.

The operation of an LSTM unit can be summarized as:

$$\mathbf{i}_t = \sigma_g(\mathbf{W}_i \cdot \mathbf{x}_t + \mathbf{U}_i \cdot \mathbf{h}_{t-1} + \mathbf{b}_i) \quad (1)$$

$$\mathbf{f}_t = \sigma_g(\mathbf{W}_f \cdot \mathbf{x}_t + \mathbf{U}_f \cdot \mathbf{h}_{t-1} + \mathbf{b}_f) \quad (2)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \cdot \mathbf{x}_t + \mathbf{U}_c \cdot \mathbf{h}_{t-1} + \mathbf{b}_c) \quad (3)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \quad (4)$$

$$\mathbf{o}_t = \sigma_g(\mathbf{W}_o \cdot \mathbf{x}_t + \mathbf{U}_o \cdot \mathbf{h}_{t-1} + \mathbf{b}_o) \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (6)$$

where the activation functions are $\sigma_g(x) = \frac{1}{(1+e^{-x})}$ and $\sigma_g(x) = \frac{1}{(1+e^{-x})}$ and $\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, W_i W and U_i U are weight parameters and \mathbf{b} b_i is bias, $e \odot$ denotes the pointwise multiplication.

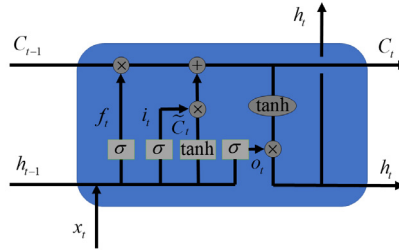


Figure 1. Graphical illustration of the long short-term memory network

2.2. Attention-based Encoder-Decoder Networks

An encoder-decoder is a popular framework in deep learning, especially in sequence-to-sequence modelling. The key idea is to encode the input sequences into the hidden state sequence and then the hidden state sequence is converted into a fix-length vector. After that, the decoder generates a target output with the pervious fix-length vector. However, there is an obvious drawback of encoder-decoder framework that, with the increase in the length of the input sequence, the context vector will lose long-term information from input sequence and this will lead to rapid reduction in prediction performance. Due to this issue, the attention mechanism is a novel addition of encoder-decoder networks. In attention mechanism, an approved context vector \mathbf{c} , \mathbf{c}_t is better than a fixed context vector \mathbf{c} , an individual attention weight w_i , W_i of each encoder hidden state \mathbf{h}_i , h_i is calculated, and the approved context vector \mathbf{c}_t , \mathbf{c}_t is defined as (Bahdanau et al., 2015)

$$\mathbf{c}_t = \sum_{i=1}^T \mathbf{w}_i \mathbf{h}_i \quad (7)$$

In order to optimize the attention-based encoder-decoder network performance in industrial application and inspired by some theories of human attention (Hübner et al., 2010), a new Dual-layer Attention-based LSTM network (DA-LSTM) is proposed. In the first attention stage, attention mechanism is used to adaptively extract the relevant information in input series at each time step by referring to previous encoder hidden state. In the second attention stage, a temporal attention mechanism is proposed to select high correlation encoder output.

The input attention mechanism can be achieved as follow:

$$e_t^i = \mathbf{V}_e \tanh(\mathbf{W}_e \cdot \mathbf{h}_{t-1} + \mathbf{U}_e \cdot \mathbf{s}_{t-1} + \mathbf{Z}_e \cdot \mathbf{x}_t^i + \mathbf{b}_e) \quad (8)$$

and

$$\alpha_t^i = \frac{\exp(e_t^i)}{\sum_{i=1}^n \exp(e_t^i)} \quad (9)$$

where \mathbf{h}_t is hidden state and \mathbf{s}_t is cell state in the encoder LSTM unit.

When the input attention weights are obtained, the input sequence can be converted to

$$\tilde{\mathbf{x}}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \dots, \alpha_t^n x_t^n)^T \quad (10)$$

The decoder with temporal attention can be calculated by:

$$r_t^i = \mathbf{V}_d \tanh(\mathbf{W}_d \cdot \mathbf{d}_{t-1} + \mathbf{U}_d \cdot \mathbf{s}_{t-1}^d + \mathbf{Z}_d \cdot \mathbf{h}_t + \mathbf{b}_d) \quad 1 < i < T \quad (11)$$

and

$$\beta_t^i = \frac{\exp(r_t^i)}{\sum_{j=1}^T \exp(r_t^j)} \quad (12)$$

where \mathbf{d}_t is the decoder hidden state and \mathbf{s}_t is the cell state in the decoder LSTM unit.

When the temporal weights are obtained, the approved context vector $\tilde{\mathbf{c}}_t$ can be achieved by

$$\tilde{\mathbf{c}}_t = \sum_{i=1}^T \beta_t^i \mathbf{h}_i \quad (13)$$

2.3. Proposed modelling strategy

The model is trained by a mini-batch stochastic gradient descent optimization called Adam optimizer (Kingma and Ba, 2015), which is a widely used optimization algorithm

for training neural networks and has advantages such as fast convergence and good adaptation to local minima issue. Moreover, in the training procedure, two training methods, which are teacher forcing method and non-teacher forcing method, are used together. Figure 2 shows the flowchart of the proposed model.

The mean squared error (MSE) is used as the loss function in this paper which can be defined as:

$$MSE = E((y - y_p)^2) \tag{14}$$

Where y and y_p are the target value and model prediction respectively.

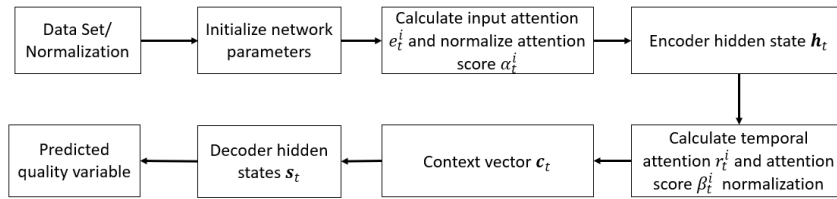


Figure 2. Flowchart for proposed model framework

3. Results and discussions

The benchmark industrial penicillin fermentation simulator, IndPenSim (Goldrick et al., 2015), is used to produce simulated process operation data. Ten benches of data generated by IndPensim are used in model development. Eight batches are used as training data, one batch is used as validation data, and the remaining batch is used as the unseen testing data.

To investigate the effectiveness of DA-LSTM, it is compared with other two baseline methods which are attention-based encoder-decoder network (Encoder-Decoder) and long short-term memory network. The comparison between DA-LSTM and encoder-decoder is to evaluate the input attention mechanism effectiveness on complex bio-processes. With all these LSTM-based methods, 30 controllable and monitoring variables are taken as input variables, and the penicillin concentration is taken as the target output. In all these three models, the number of hidden units is 50, the maximum training epochs are 500, the initial learning rate is 0.005, and the dropout factor is 0.2.

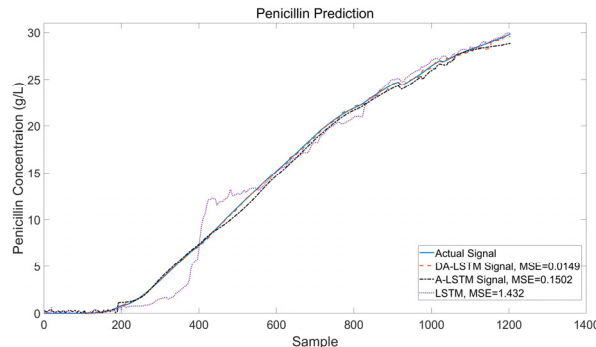


Figure 3. Predictions of penicillin concentration

For visual comparison, the penicillin concentration prediction result is shown in Figure 3 with the MSE of each method shown in the legend. Figure 4 shows the corresponding prediction error. From Figures 3 and 4, it can be seen that comparing to the normal LSTM prediction result, the prediction of DA-LSTM and encoder-decoder are more accurate, especially when the slope starts to get steep. From Figure 3, it can also be seen that the prediction error increases when penicillin starts to ferment. At this period, the input variables are changed rapidly to control the fermentation process. This shows the encoder-decoder framework are more robust against the noise in input data. On the other hand, the prediction accuracy of DA-LSTM outperforms Encoder-Decoder because adaptively extracting driving series by input attention mechanism can provide more reliable input features to make accurate predictions. Overall, the DA-LSTM integrating the input attention mechanism as well as temporal attention mechanism achieves the lowest MSE among all three methods.

4. Conclusions

In this paper, a new dual-layer attention-based LSTM network is proposed and it integrates an input attention mechanism and a temporal hidden state attention mechanism. The encoder input attention mechanism is for selecting relevant driving series in the input data sequence, and a temporal decoder attention mechanism is used to measure the importance of encoder hidden states. Based on the preliminary results, the DA-LSTM can improve the performance of LSTM on modelling complex bioprocesses, especially with multiple input variables. The developed model will be used in process optimisation and control in the future.

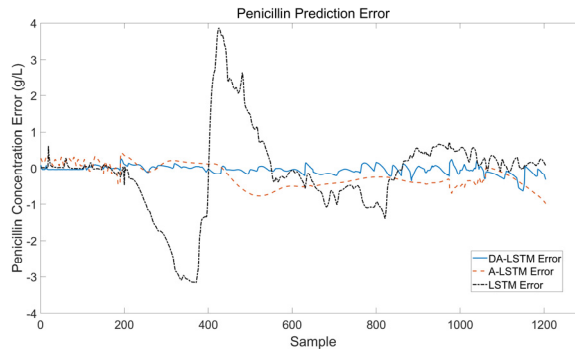


Figure 4. Errors in penicillin concentration prediction

References

- Bahdanau, D., Cho, K., & Bengio, Y., 2015. Neural machine translation by jointly learning to align and translate. 3rd International Conference on Learning Representations (ICLR 2015), San Diego, United States; 7-9 May 2015, arXiv preprint arXiv:1409.0473.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), Doha, Qatar, 25-29 October 2014, 1724-1734.

- Goldrick, S., Stefan, A., Lovett, D., Montague, G., Lennox, B., 2015. The development of an industrial-scale fed-batch fermentation simulation. *Journal of Biotechnology*, 193, 70–82.
- Hochreiter, S., & Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Hübner, R., Steinhauser, M., & Lehle, C., 2010. A dual-stage two-phase model of selective attention. *Psychological Review*, 117(3), 759.
- Kingma, D. P., & Ba, J., 2015. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations (ICLR 2015), San Diego, United States, 7-9 May 2015.
- Pascanu, R., Mikolov, T., & Bengio, Y., 2013. On the difficulty of training recurrent neural networks. 30th International Conference on Machine Learning (ICML 2013), Atlanta, United States, 16-21 June 2013, 2347-2355.

Resilience and Survivability-aware Optimal Design and Operation of Interconnected Supply Chains

Spyridon D. Tsolas, M. M. Faruque Hasan*

Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843-3122, USA.

**hasan@tamu.edu*

Abstract

Pandemic-induced lockdowns, restrictions on commercial activities, or natural disasters can disrupt a supply chain for prolonged time periods. These disruptions significantly impact network connectivity and consumer demands which in turn affect the capacity and profitability of a supply network. In this work, we address the resilience and economic survivability of interconnected networks during connectivity disruptions. Resilience is the ability to defend against and recover from potential disruptions. Economic survivability (ES) is the ability to maintain positive cash flows. We introduce optimization-based metrics to quantify the resilience and ES in supply chain design and operation. We also develop multi-scenario and multi-period models and systematic approaches which lead to supply chains with maximum profitability for desired resilience and ES.

Keywords: Network Resilience, Supply Chain Survivability, Interconnected Supply Chains

1. Introduction

Disruption events can impair the economic performance of supply chains, and negatively impact the social welfare since demands cannot always be met. Graph-theoretic resilience quantification (Latora and Marchiori, 2001; Yazdani et al. 2011), connectivity reinforcing, installing additional capacity and emergency supplies (Turnquist and Vugrin, 2013; Zhang et al., 2020) and other approaches have been proposed to tackle network disruptions. However, pandemic-induced lockdowns (e.g., COVID-19), geo-political conflicts, sanctions and restrictions on commercial activities, and sudden natural disasters give rise to new challenges (Ivanov and Dolgui, 2020; El-Halwagi et al., 2020). Some of these lead to remote employment, productivity reduction, loss of connectivity links and sudden changes in product demands. These disruptions often continue for prolonged time periods, which may eventually lead to partially or completely shutting down a supply chain.

In this work, we first address the resilience of interconnected networks against connectivity disruptions. We use resilience quantification metrics to analyze the system performance against single-connectivity failures. We also propose a mixed-integer linear program (MILP) to design resilient systems with minimum overdesign for interconnected supply chains (Chapter 2). Next, we introduce the concept of economical survivability (ES) to express the sensitivity of a supply chain network to demand decreases (Chapter 3). A case study is presented to illustrate the concepts in Chapter 4.

2. Resilience Analysis of Interconnected Networks

We address the resilience analysis of interconnected supply chains with multiple products. A representative example is the Water-Energy Nexus (WEN) (Tsolas et al., 2018), where the energy and water networks act upon shared energy and water sources, acting as water and energy sinks respectively (Fig. 1). First, we analyze WEN resilience by quantifying the performance of interconnected supply chains after single-edge failures. Specifically, we measure the impact of removing one designed connection, and repeat it for the rest of a given infrastructure. The impact is measured using the following metrics: (i) network structural robustness ρ , which is the fraction of single-connectivity removals, after which the rest of the network is still feasible; (ii) operational resilience θ , which is the average fraction of demand satisfaction across all removals; and (iii) excess nexus capacity ζ , defined as the excess consumer demands that can be satisfied in nominal operation without disruptions.

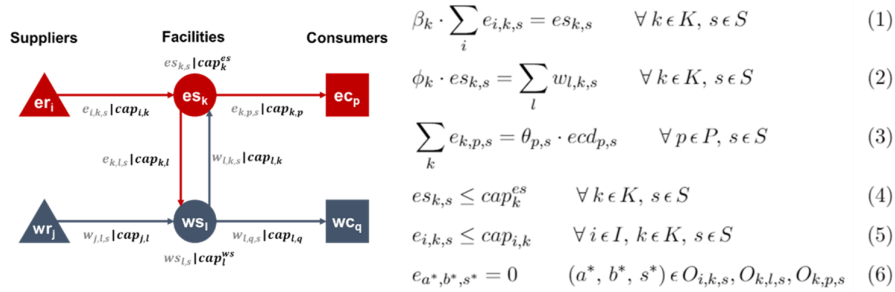


Figure 1: Water-energy nexus design and operation model

Over-designing improves the ability to operate after connections fail, but also increases the cost. One can avoid over-designing by paying a penalty if demand violation occurs. To capture this trade-off, we introduce the concept of minimum cost of resilience (*MCOR*), which is defined as $MCOR = \frac{\text{Resilient Design Cost} - \text{Nominal Design Cost}}{\text{Nominal Design Cost}}$. *MCOR* is the normalized cost difference between an overdesigned WEN (improved resilience) and the nominal design network (no disruptions).

Consider a WEN with the I, J energy and water resources, P, Q energy and water consumers, and K, L potential energy and water sources. To account for supply chain resilience in the design phase and ensure minimum overdesign (*MCOR*) we introduce disruption scenarios, equal to the number of potential connections to be designed. In each scenario, one connecting flow is imposed to be zero. The operating variables (generation, connecting supplies) and all operating balances are expanded for every scenario $s \in S$. Representative operating scenario-dependent balances are shown in Eqs. (1)-(6). Eq. (1) calculates the generation of energy from the energy sources k at every scenario s , based on the total input and a conversion factor β_k . Eq. (2) captures the dependence of energy generation from the water network via the intensity factor ϕ_k . Eq. (3) ensures that the demands of energy consumer p are satisfied for every scenario s . Also, operational resilience θ is a decision variable, as unsatisfied demands are penalized as cost. Eqs. (4) and (5) connect the scenario-dependent operating variables with the maximum design capacities. Finally, Equation (6) fix the operating flow rates of the missing links to zero, for each scenario-connection combination O_{a^*,b^*,s^*} . These combined with the operating equations ensure that balances hold across all disruption scenarios. Finally, the optimizer

will determine the cost-minimum common maximum capacities ($cap_k^{gen}, cap_{i,j}, \dots$) that will bound the scenario-dependent operating variables. The objective is to minimize the total WEN infrastructure cost, which is comprised of the fixed investment, capital, and scenario-dependent operating expenditures.

3. Economic Survivability of Supply Chains

In this section, we consider geographically interdependent supply chains with the same product. The interdependence arises from the raw material and product flows between supply chains from different regions. We consider the multi-period design and operation of a 5-echelon supply chain with raw material suppliers I , warehouse storage facilities J , processing facilities K , distribution centers L , and consumer centers M . In the design problem, given are consumer center demands that need to be satisfied in multiple regions. The objective is to determine the optimal connectivity and combination of storage and processing facilities to be designed, which will lead to minimum total network cost.

The supply chain design gives rise to an MINLP formulation, governed by operating balances to determine the time-dependent generation $gen_{k,t}$, flow rates $f_{i,j,t}$, and inventory levels of storage facilities $stor_{j,t}^{wh}$, to satisfy the consumer demands $d_{m,t}$. Representative operational constraints are depicted below. Eq. (7) ensures the availability of the raw material suppliers $a_{i,t}$ is respected. Eq. (8) and (11) calculate the inventory levels of warehouses J, L across the periods t . Eq. (9) expresses the product generation balance at processing facilities K , given by the total input from warehouses J and a conversion factor β_k . Finally, equation (12) ensures that the total flows reaching the demand centers from the distributors L will satisfy the consumer demands.

$$\sum_j f_{i,j,t} \leq a_{i,t} \quad \forall i \in I, t \in T \quad (7)$$

$$stor_{j,t}^{wh} = stor_{j,t-1}^{wh} + \sum_i f_{i,j,t} - \sum_k f_{j,k,t} \quad \forall j \in J, t \in T \quad (8)$$

$$\beta_k \cdot \left(\sum_j f_{j,k,t} \right) = gen_{k,t} \quad \forall k \in K, t \in T \quad (9)$$

$$gen_{k,t} = \sum_l f_{k,l,t} \quad \forall k \in K, t \in T \quad (10)$$

$$stor_{l,t}^{dc} = stor_{l,t-1}^{dc} + \sum_k f_{k,l,t} - \sum_m f_{l,m,t} \quad \forall l \in L, t \in T \quad (11)$$

$$\sum_l f_{l,m,t} = d_{m,t} \quad \forall m \in M, t \in T \quad (12)$$

We now address the survivability-aware design of supply chains in the event of lockdown-induced demand fluctuations. Our objective is to ensure that multi-regional supply chains maintain positive cash flows during disruption events. We express this by introducing economic survivability (ES). Disruption-induced demand decreases create negative cash flows since existing systems are designed for nominal higher demands. We define ES as the threshold of demand fraction for which the total profit of the supply chain remains non-negative. We define λ_t as this demand decrease fraction with lower values corresponding to higher survivability. The total profit $P(D_0)$ is given by the total sales revenue $REV(D_0)$ minus the capital cost of investment $FC(D_0)$ and the total operating cost $OC(D_0)$: $P(D_0) = REV(D_0) - FC(D_0) - OC(D_0)$.

The expression is calculated so that the network is designed to satisfy nominal demands

D_0 . If there is a demand reduction $D_t = \lambda_t \cdot D_0$, the capital cost of investment is not affected since it is a function of the capacity variables, and the original network was constructed for nominal demands $FC(D_0)$. However, the total revenue $REV(D_t)$ and operating cost $OC(D_t)$ are reduced by λ_t , hence $REV(D_t) = \lambda_t \cdot REV(D_0)$ and $OC(D_t) = \lambda_t \cdot OC(D_0)$. So, we can calculate or impose the economic survivability threshold of a supply chain using the following expression:

$$P(D_t = \lambda_t \cdot D_0) = \lambda_t \cdot REV(D_0) - FC(D_0) - \lambda_t \cdot OC(D_0) \geq 0 \quad (13)$$

$$\lambda_t \geq \frac{FC(D_0)}{REV(D_0) - OC(D_0)} \quad (14)$$

Our objective is to minimize this demand threshold to make supply chains more survivable.

4. Results

4.1. Resilience-aware Design

In the first part of the results we demonstrate the resilience-aware design of a WEN. We utilize the MILP multi-scenario formulation to obtain the resilient grass-root design for a given initial superstructure. For parametric values of external grid purchasing cost, which is the λ parameter that penalizes the unsatisfied demands, we derive the resulting resilience indices and the *MCOR*.

The distribution of the performance metrics is demonstrated in **Fig. 2a**. For low penalties λ , a network is not required to satisfy the demands, since the nexus relies solely on external supply purchasing. The metrics in that region are equal to zero. As λ increases, the network's size and contribution to the demands (Operational resilience θ) increase as well. Interestingly, for very high penalty λ , the network is 100% self-sufficient in terms of demand satisfaction and 100% resilient against single-connectivity disruptions ($\theta = \rho = 100\%$). Finally, the excess demand satisfaction ζ is almost 50% in that region. This entails that the nexus can satisfy 50% increased demands in the absence of disruptions.

In **Fig. 2b** the *MCOR* distribution is presented. In the low penalty region, the multi-scenario formulation provides the same solution as the non-resilient economical optimization of the same initial superstructure. Also, since there is no network to satisfy the demands, no connectivity disruptions can be observed. For higher penalties, the network is created to contribute to the demand satisfaction, increasing the *MCOR* until its maximum value. The last region corresponds to the maximum investment increase to fortify the supply chain with provisions to satisfy all demands against all disruption

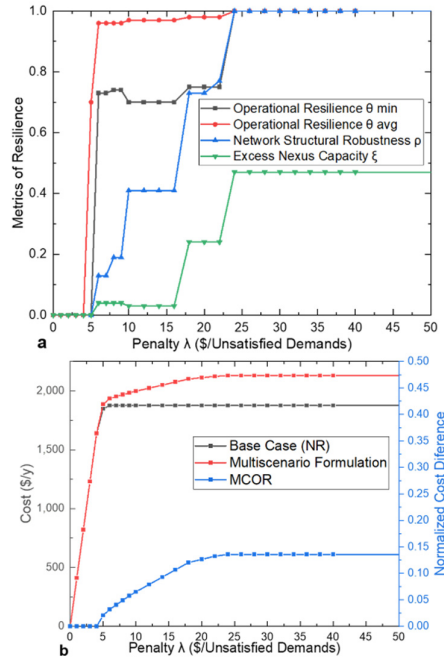


Figure 2: Resulting resilience metrics (a) and *MCOR* (b) for parametric penalty λ .

scenarios. However, this investment cost increase ($MCOR$) corresponds to the minimum possible that can achieve 100% resilience. That is compared to heuristic-based resilience improvements and over-design strategies.

4.2. Survivability-aware Design

Finally, we explore the effect of economic survivability in the design of multi-regional supply chains. We consider a supply chain spanning across two regions $R=2$, with suppliers $I=2$, warehouses $J=3$, facilities $K=4$, distribution $L=3$, and consumer centers $M=2$. The distribution of the initial nodes across the corresponding regions is demonstrated in **Fig. 3**.

For the first part of the results, we maximize the global economic survivability ES_{tot} , by minimizing the threshold λ_{tot} , for the whole supply chain. The resulting network (**Fig. 3b**), leaves region 2 without any warehouses allocated and receives the required flows from region 1. The total profit is 219, and the total survivability threshold is $\lambda_{tot} = 0.652$. However, the calculated individual thresholds are $\lambda_{r=1} = 0.929, \lambda_{r=2} = 0.426$. This entails that the first region can handle smaller demand decreases and still maintain positive cash flows. This can be explained by the additional investments of the warehouse that operates in region 1 and supplies to region 2.

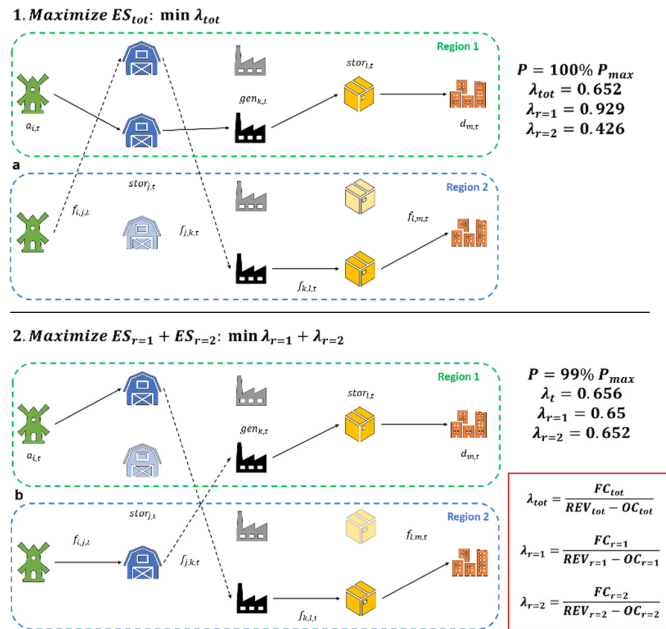


Figure 3: Comparison between optimizing a two-region supply chain for (a) global and (b) local survivability.

In the second part, we minimize the summation of the individual regional thresholds $\lambda_{r=1}, \lambda_{r=2}$. The resulting network (**Fig. 3b**) has total profit of 217 units, total survivability threshold $\lambda_{tot} = 0.656$, and $\lambda_{r=1} = 0.65, \lambda_{r=2} = 0.652$. In this case, the optimizer still allows inter-regional supplies, but now allocates each region with its required facilities. We have assumed that the capital investments and operating expenditures of inter-

regional supplies are compensated in the source region's total cost. As a result, the profit of the whole supply chain decreases and obtains 99% of the maximum potential profit. The total survivability is almost equal to the first case, but the individual thresholds are more balanced and closer to the average. The fixed investment of the facilities is also shared across the different regions more evenly.

We can conclude with the following remarks: (i) There is a strong positive correlation between return on investment (*ROI*) and economic survivability (*ES*). (ii) Maximizing global economic survivability (*ES*) leads to higher profitability of the whole supply chain. (iii) Maximizing local economic survivability $ES_{r=1} + ES_{r=1} + \dots$ ensures that investments on storage and generation facilities is more evenly distributed across multiple regions, in the expense of lower profit.

5. Conclusions

We have introduced two key concepts for the resilience- and survivability-aware analysis and design of interconnected supply chain networks. The first concept is the minimum cost of resilience (*MCOR*) and allows us to analyze and compare different network configurations for safe-guarding interconnected networks in the events of potential transportation link failures. The second key concept is the economic survivability condition, which ensures that a supply chain is immune to sudden changes in demands due to prolonged lockdowns or restrictions on inter-regional flows. We found that there is a strong correlation between economic survivability and return-on-investment, both at the global and local supply chain levels. Therefore, it is important to evenly allocate infrastructure to ensure the survivability of individual local regions that constitute a global chain. While this may reduce the profit, it ensures the sustainability of the overall supply chain in the long run.

References

- Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical review letters*, 87(19), 198701.
- Yazdani, A., Otoo, R. A., & Jeffrey, P. (2011). Resilience enhancing expansion strategies for water distribution systems: A network theory approach. *Environmental Modelling & Software*, 26(12), 1574-1582.
- Turnquist, M., & Vugrin, E. (2013). Design for resilience in infrastructure distribution networks. *Environment Systems & Decisions*, 33(1), 104-120.
- Zhang, G., Zhang, F., Zhang, X., Wu, Q., & Meng, K. (2020). A multi-disaster-scenario distributionally robust planning model for enhancing the resilience of distribution systems. *International Journal of Electrical Power & Energy Systems*, 122, 106161.
- Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. *International Journal of Production Research*, 58(10), 2904-2915.
- El-Halwagi, M. M., Sengupta, D., Pistikopoulos, E. N., Sammons, J., Eljack, F., & Kazi, M. K. (2020). Disaster-Resilient Design of Manufacturing Facilities Through Process Integration: Principal Strategies, Perspectives, and Research Challenges. *Frontiers in Sustainability*, 1, 8.
- Tsolas, S. D., Karim, M. N., & Hasan, M. F. (2018). Optimization of water-energy nexus: A network representation-based graphical approach. *Applied energy*, 224, 230-250.

A Hybrid Modeling Approach for Liquid Entrainment Fraction Predictions

Yushi Deng,^a Selen Cremaschi,^{a*} Mario R. Eden,^a Carlos Avila,^b Haijing Gao,^b Ivan Mantilla^b

^a*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*

^b*Chevron Energy Technology Company, Houston, TX 77002, USA*

Abstract

Liquid entrainment fraction is a key parameter for designing and optimizing downstream oil and gas production equipment and needs to be reliably estimated. In this study, a hybrid modeling (HM) approach, which combines existing mechanistic and semi-mechanistic (white-box) models and data-driven (black-box) models, is introduced for estimating entrainment fraction and its uncertainty. The hybrid model incorporates six white-box models and discrepancy models developed using Gaussian Process Modeling (GPM) trained with 1,562 experimental data. The average root-mean-square-error (RMSE) of the hybrid model is 0.06, which is a 61% reduction compared to the RMSE of the white box models (0.17).

Keywords: Hybrid Modeling, Entrainment Fraction, Gaussian Process Modeling

1. Introduction

In two-phase annular flow, the liquid forms a thin film around the pipe while the gas flows in the core of the pipe. Some portion of the liquid is entrained from the liquid film to the gas core. The ratio of the entrained liquid droplets mass flow rate to the total liquid mass flow rate in the pipeline is defined as the liquid entrainment fraction (Cioncolini and Thome, 2010). The liquid entrainment fraction is an essential parameter for the estimation of pressure drop, flow rate, liquid holdup, dry-out in annular flow, as well as for designing and optimizing separation facilities (Al-Sarkhi et al., 2012). Although mechanistic and semi-mechanistic models have been developed for predicting the liquid entrainment fraction, this is a complicated process, and none of the models capture all relevant phenomena, nor can they be applied over a wide range of operating conditions.

Hybrid modeling (HM), where part of a model can be formulated using first principles (white-box), and part of the model has to be inferred from data because of a lack of understanding of the mechanistic details (black-box) (Marquardt, 2002), is a good approach for estimating liquid entrainment fraction in two-phase flows. There are numerous hybrid modeling applications in chemical engineering, e.g., Zahedi et al. (2011) developed an ethylene oxide (EO) fixed bed reactor model with reaction kinetics estimated by an artificial neural network (ANN). In a more recent example, Bangi and Kwon (2020) integrated first principles models with a deep neural network (DNN), and applied it to a hydraulic fracturing process.

In this paper, a hybrid model is developed to estimate the liquid entrainment fraction. A database composed of entrainment fraction measurements and six semi-mechanistic models for predicting liquid entrainment fraction (white-box models) is constructed. The black-box models were then built for predicting the model discrepancy, which is defined

as the difference between the experimental and predicted entrainment fraction. The black-box models were trained using the Gaussian Process Modeling (GPM) (Williams and Rasmussen, 2006). The GPM outputs are the mean and variance of the model discrepancy, which were used to estimate confidence intervals. To obtain a more robust and accurate prediction, the bootstrap aggregating (bagging) technique is applied.

A brief overview of the six white-box models and the database used to build the black-box model is given in Section 2. Section 3 gives the details of the hybrid modeling approach. The result and discussion are given in Section 4. Finally, conclusions are summarized in Section 5.

2. Liquid Entrainment Fraction Database

We selected three widely-used models for predicting liquid entrainment fraction in different flow orientations, namely horizontal and vertical pipelines, as flow orientation significantly changes the underlying entrainment phenomena. The white-box models selected are Oliemans et al. (1986), Zhang et al. (2003), and Cioncolini & Thome (2010) for vertical pipeline orientation; and Pan & Hanratty (2002b), Mantilla (2008), and Nakazatomi & Sekoguchi (1996) for horizontal orientation.

The database contains 1,562 liquid entrainment fraction measurements at different experimental conditions, which are collected from open sources. Classified by the pipeline orientation, the number of measurements in vertical and horizontal orientations are 1,083 and 479, respectively. Nine experimental conditions (pipe diameter, inclination angle, gas density, liquid density, gas viscosity, liquid viscosity, gas-liquid surface tension, superficial gas velocity, and superficial liquid velocity) are the input variables that are relevant for entrainment fraction.

3. Hybrid Model Development Methodology

The liquid entrainment fraction measurement (y^e) at an operating condition can be decomposed and expressed by Eq. (1) (Jiang et al. 2013).

$$y^e = y^m + \delta \quad (1)$$

In Eq. (1), y^m is the entrainment fraction prediction of the white-box model, and δ is the model discrepancy, which is the difference between the entrainment fraction measurement and white-box model prediction. Here, we develop GPM-based black-box models (Jiang et al., 2013) to estimate the model discrepancy.

The training and validation data for developing the black-box models are obtained using the liquid entrainment measurements and the corresponding white-box model predictions for each pipeline orientation. For example, Oliemans et al. (1986) model is employed to calculate entrainment fractions and the model discrepancies for all 1,083 experimental conditions corresponding to the vertical pipeline orientation. These conditions are the inputs (X), and the calculated model discrepancy is the output (δ) for training the discrepancy model for Oliemans et al. (1986) model. This process resulted in six training and validation data sets.

3.1. Gaussian Process Modeling (GPM) Development with Bagging

The Gaussian Process (GP) algorithm is a supervised learning method with a theoretical basis in statistics. A GP is characterized by its mean function and covariance function or kernel function (Williams and Rasmussen, 2006). A constant mean function and square exponential kernel function with different length scales for each input dimension is utilized. Then, the GP prior is represented in Eq. (2):

$$f \sim \mathcal{GP}(m, k)$$

$$m(x) = a, \text{ and } k(x_p, x_q) = \sigma_f^2 \exp\left(-\sum_{h=1}^d \frac{(x_p - x_q)^2}{2l_h^2}\right) + \sigma_n^2 \Delta_{pq} \quad (2)$$

In Eq. (2), f represents the underlying function value at the input x , $m(x)$ is the mean function of the GP, and $k(x_p, x_q)$ is the covariance function representing the spatial covariance between any two points (x_p and x_q) at the process. In the covariance function, d is the dimension of input x , l_h represents the characteristic length scale corresponding to the h^{th} dimension of input x , σ_n^2 is the output variance, and the parameter σ_f^2 is output-scale amplitude. Δ_{pq} is Kronecker delta, which is one if $p = q$ and zero otherwise. The hyperparameters for the GP in Eq. (2) are $\theta = \{a, \sigma_f, \sigma_n, l_1, l_2, \dots, l_d\}$, and they are estimated by the Maximum Likelihood Estimation (MLE) (Rasmussen, 2003) using a Python-based package, GPy (GPy, 2012).

As shown in Eq. (3), the value of the mean function and its variance for the unseen test location X^* are inferred by calculating the posterior distribution using the training data set (X, δ) via Eq. (3).

$$f_* | X, \delta, X_* \sim \mathcal{N}(\bar{f}_*, \text{cov}(f_*))$$

$$\bar{f}_* = K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} \delta \quad (3)$$

$$\text{cov}(f_*) = K(X_*, X_*) - K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*)$$

where f_* is the outputs of the test data given the inputs X and outputs δ of the training data and the inputs of the test data X^* . K is the covariance matrix. I is the identity matrix. Eq. (4) is employed to estimate model discrepancy $\hat{\delta}$ and its variance $\sigma_{\hat{\delta}}^2$ using the GPM at an unseen experimental condition.

$$\hat{\delta} = \bar{f}_*$$

$$\sigma_{\hat{\delta}}^2 = \text{cov}(f_*) \quad (4)$$

In this study, bootstrap aggregating (bagging) is utilized to obtain more robust and accurate predictions. In bootstrap aggregating, a set of models are developed to characterize the same relationship between input and output variables. One hundred training data sets are generated with replacement and used to train one hundred models for predicting model discrepancy that corresponds to each white-box model. To test the model's generalizability, for each model, only data that is not included in the training data set is used as testing data. The averaged predictive mean and variance from each model developed using bagging is calculated based on the property of the Gaussian mixture model shown in Eq. (5) and Eq. (6) (Chen and Ren, 2009):

$$E(\hat{\delta}) = \frac{1}{C} \sum_{c=1}^C \hat{\delta}(c) \quad (5)$$

$$\text{Var}(\hat{\delta}) = \frac{1}{C} \sum_{c=1}^C \sigma_{\hat{\delta}}^2(c) + \frac{1}{C} \sum_{c=1}^C (\hat{\delta}(c) - E(\hat{\delta}))^2 \quad (6)$$

where $E(\hat{\delta})$ is the averaged model discrepancy, $\text{Var}(\hat{\delta})$ is the averaged variance from one hundred GP models. Variable C represents the number of single GP models, which is

equal to 100 in this work. The final form of the predicted entrainment fraction from bagging GPM is shown in Eq. (7).

$$\hat{y}^e = y^m + E(\hat{\delta}) \quad (7)$$

$$95\% \text{ Confidence Interval} = \hat{y}^e \pm 1.96\sqrt{\text{Var}(\hat{\delta})}$$

3.2. Metrics to Evaluate the Hybrid Model Performance

To assess the prediction performance of the hybrid model, three quantitative metrics, Root-Mean-Square-Error (RMSE), Area Metric (AM), and Calibration Score (CS), are adopted. The RMSE is a widely used metric to measure the overall difference between prediction and measurement. The definition of RMSE is shown in Figure 1, where J is the number of predicted points. A lower RMSE value indicates a smaller difference between prediction and actual value, which represents a better model accuracy.

The AM compares both the mean and the variance of measurement and prediction (Ferson and Oberkampf, 2009). In Figure 1, the AM is the shaded area between the cumulative probability distributions of y^e and \hat{y}^e . A lower AM value represents a more accurate prediction. An AM value higher than the RMSE indicates that the experimental measurement is covered by the predicted confidence interval.

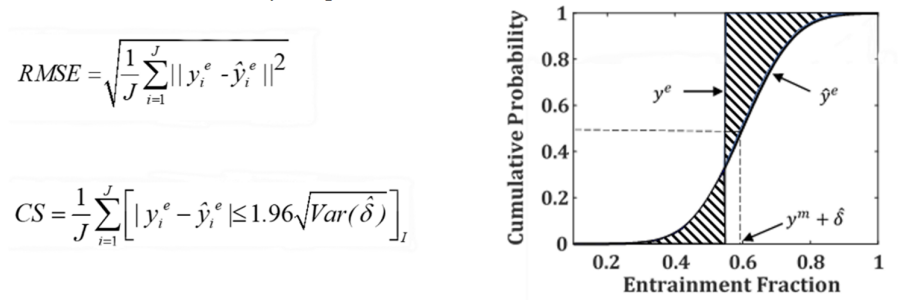


Figure 1. Definition of RMSE, AM and CS

The CS is the fraction of experimental measurements that are covered within a selected confidence interval (Richardson et al., 2019). In Figure 1, $[]_I$ represents the Iverson bracket. In our study, the 95% confidence interval is used. This metric is used as a minimum requirement check.

4. Results and Discussion

The performance metrics for the hybrid models are compiled in Figure 2. The RMSE and AM results are shown in Fig. 2(a), while the CS result is shown in Fig. 2(b). The AM for the white-box model is equal to the RMSE as there is no variance information. As can be seen from Fig. 2(s), there is a 60% reduction in both the average RMSE and AM with the introduction of hybrid models. If the results for vertical (Fig. 2(a) left of the dotted line) and horizontal pipeline (Fig. 2(a) right of the dotted line) orientations are compared, the RMSE and AM of models for the vertical orientation is lower (better) compared with the models for the horizontal orientation. Figure 2(a) reveals that the bagging technique reduced the RMSE and AM for all models compared to using a single GP model and improves the hybrid model prediction accuracy. The lower AM values and CS being closer to 0.95 for the bagging GP models indicate that bagging provides a more moderate and reliable confidence interval.

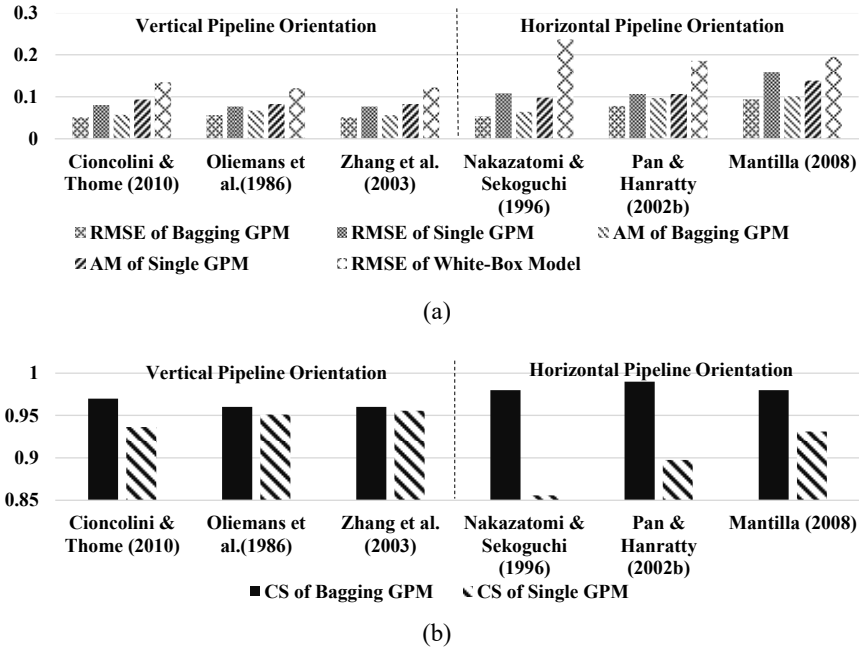


Figure 2. Entrainment Fraction Prediction Result

5. Conclusions

In this study, a hybrid modeling approach, which combines white-box models and GPM-based black-box models, is developed to estimate the liquid entrainment fraction in two-phase flow. Six mechanistic and semi-mechanistic models explaining the underlying physical phenomena are employed as white-box models to predict entrainment fraction, while the GPM is used as a black-box model to estimate the difference between experimental measurements and white-box model predictions. The bagging technique is added to obtain a more robust and accurate black-box model. The results reveal that the hybrid modeling approach yields models that have lower RMSE than the white-box models. Furthermore, the GPM-based hybrid model provides an estimate of the prediction uncertainty, which is not possible to obtain with the white-box models. Future work will investigate approaches to increase the hybrid model prediction accuracy and shrink the confidence interval, such as data clustering and input and output transformations prior to hybrid model training.

Acknowledgments

The financial support for this project is provided by Chevron Energy Technology Company.

References

Al-Sarkhi, Abdelsalam, Cem Sarica, and Bilal Qureshi. 2012. "Modeling of Droplet Entrainment in Co-Current Annular Two-Phase Flow: A New Approach." *International Journal of Multiphase Flow* 39: 21–28.

- Bangi, Mohammed Saad Faizan, and Joseph Sang Il Kwon. 2020. "Deep Hybrid Modeling of Chemical Process: Application to Hydraulic Fracturing." *Computers and Chemical Engineering* 134.
- Chen, Tao, and Jianghong Ren. 2009. "Bagging for Gaussian Process Regression." *Neurocomputing* 72(7–9): 1605–10.
- Cioncolini, Andrea, and John R. Thome. 2010. "Prediction of the Entrained Liquid Fraction in Vertical Annular Gas-Liquid Two-Phase Flow." *International Journal of Multiphase Flow* 36(4): 293–302.
- Ferson, Scott, and William L Oberkampf. 2009. "Validation of Imprecise Probability Models." *International Journal of Reliability and Safety* 3(1–3): 3–22.
- GPy. 2012. "GPy: A Gaussian Process Framework in Python."
- Jiang, Zhen, Wei Chen, Yan Fu, and Ren Jye Yang. 2013. "Reliability-Based Design Optimization with Model Bias and Data Uncertainty." *SAE International Journal of Materials and Manufacturing* 6(3).
- Marquardt, Wolfgang. 2002. "Adaptivity in Process Systems Modeling." In *European Symposium on Computer Aided Process Engineering-12*, eds. Johan Grievink and Jan B T - Computer Aided Chemical Engineering van Schijndel. Elsevier, 42–56.
- Rasmussen, Carl Edward. 2003. "Gaussian Processes in Machine Learning." In *Summer School on Machine Learning*, Springer, 63–71.
- Richardson, Robert R., Christoph R. Birkel, Michael A. Osborne, and David A. Howey. 2019. "Gaussian Process Regression for in Situ Capacity Estimation of Lithium-Ion Batteries." *IEEE Transactions on Industrial Informatics* 15(1): 127–38.
- Williams, Christopher K I, and Carl Edward Rasmussen. 2006. *2 Gaussian Processes for Machine Learning*. MIT press Cambridge, MA.
- Zahedi, G, A Lohi, and K A Mahdi. 2011. "Hybrid Modeling of Ethylene to Ethylene Oxide Heterogeneous Reactor." *Fuel Processing Technology* 92(9): 1725–32.

RETHINKING BALANCE MODELS APPLIED TO PFRS AND EXTRACTING NEW INSIGHTS

João Manzi

Federal University of Campina Grande, Department of Chemical Engineering

Rua Aprigio Veloso 882, Campina Grande - PB - CEP: 58428-830 Brazil

joao.manzi@ufcg.edu.br

Abstract

This paper makes a detailed investigation of the construction of the balances of mass, energy and entropy, aiming at a comprehensive analysis of the behavior of tubular reactors (Plug Flow Reactors - PFRs). The correlation between the physical base (space-time) and thermodynamic concepts was also explored with a view to making the relationship between the physical and thermodynamic quantities clearer. While doing so, new insights were obtained which led to showing that these balances can be used to improve the performance of tubular reactors in a simple and inexpensive way. A typical case study is discussed in order to demonstrate that the contribution of this paper is to put forward a model that can also be applied to PFRs and improve their performance.

Keywords: balances, entropy production, energy saving, cost.

1. Introduction

From statements made at conferences or even from comments in reviews of papers, it seems clear to us that there is a gap in understanding the main concepts of thermodynamics and in how best to apply them in the context of chemical processes. This has generated a wide array of misunderstandings, particularly when applied to models for plug flow reactors (PFRs).

When our thoughts turn to process analysis, simulations, balances, building models, and some other techniques, the excellent textbooks by Thomas Edgar and David Himmelblau (Edgar et al., 2001; Himmelblau and Bischoff, 1968), though first published more than 20 years ago, still come to mind today. However, some details are not focused on in much depth and it can therefore make an exploration of the concepts that underpin them difficult, especially for beginners.

The classical case is concerned with saving energy for which the First Law of Thermodynamics is the major reference to establish the procedures. However, using only this Law does not appropriately solve such issue, because it does not deal with the nature constraints imposed on the processes. A quick analysis, carried out when the Gibbs function $\Delta G = \Delta H - T\Delta S$ is taken into account, reveals that when the process occurs (ΔG negative) and since ΔS is always increasing, then on minimizing energy (ΔH), ΔG becomes more negative and this can seem to be good for the process, because it becomes more spontaneous. However, this occurs at the expense of an increase in the level of entropy, thus indicating a loss of the level of organization of the reactive system, which may indicate a higher production of by-products has taken place. Hence, the Second Law of Thermodynamics should also be used to deal appropriately the system, so as to generate

a more realistic scenario, taking into account a trade-off between energy and entropy (Andresen,2011; Bispo et al.2013). For these reasons, this paper tackles and explores the constructive details of these balances, as a result of which new insights were obtained.

2. Theoretical basis

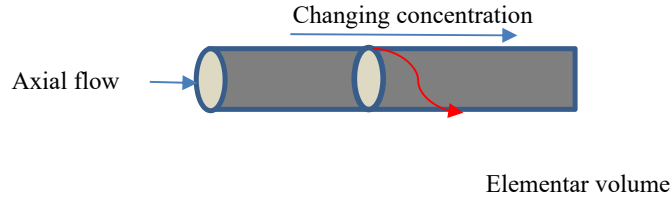


Figure 1: Outline representation of a PFR

In dealing with mass balance, the following equation can be derived as a function of time (t), length (l), angle (θ) and radius (ϕ).

$$\frac{\partial m_i}{\partial t} \frac{dt}{dt} + \frac{\partial m_i}{\partial l} \frac{dl}{dt} + \frac{\partial m_i}{\partial \theta} \frac{d\theta}{dt} + \frac{\partial m_i}{\partial \phi} \frac{d\phi}{dt} = [\text{generation or consumption of } i] + \text{mass. dif} \quad (1)$$

Given that $\frac{dC_i}{dt} = \frac{1}{V} \frac{dn_i}{dt} = \frac{1}{VM_i} \frac{dm_i}{dt}$ results in $\frac{dm_i}{dt} = \mp r_i VM_i$, then Equation 1 can be rewritten as:

$$\frac{\partial m_i}{\partial t} + \frac{\partial m_i}{\partial l} \frac{dl}{dt} + \frac{\partial m_i}{\partial \theta} \frac{d\theta}{dt} + \frac{\partial m_i}{\partial \phi} \frac{d\phi}{dt} = \mp r_i VM_i + [\text{mass diffusion}] \quad (2)$$

r_i is the reaction rate of generation or consumption of the component i while M_i is its molecular mass. Since $m_i = C_i VM_i$ and the transport by mass diffusion can be given by $J_m = D_{ij} \left[\frac{\partial^2 C_i}{\partial l^2} + \frac{1}{\phi^2} \frac{\partial^2 C_i}{\partial \theta^2} + \frac{1}{\phi} \frac{\partial}{\partial \phi} \left(\phi \frac{\partial C_i}{\partial \phi} \right) \right]$, where D_{ij} denotes the diffusion coefficient, Eq. 2 does not seem to meet the needs of analysis, because the working variables, i.e., those normally used, have not been adequately inserted. Consequently, the working variable as concentration should be introduced. Therefore, Equation 2 can be rewritten as:

$$\frac{\partial C_i}{\partial t} + \frac{\partial C_i}{\partial l} \frac{dl}{dt} + \frac{\partial C_i}{\partial \theta} \frac{d\theta}{dt} + \frac{\partial C_i}{\partial \phi} \frac{d\phi}{dt} = \mp r_i + D_{ij} \left[\frac{\partial^2 C_i}{\partial l^2} + \frac{1}{\phi^2} \frac{\partial^2 C_i}{\partial \theta^2} + \frac{1}{\phi} \frac{\partial}{\partial \phi} \left(\phi \frac{\partial C_i}{\partial \phi} \right) \right] \quad (3)$$

Having in mind the general formulation applied to the energy balance, the following Equation can be derived in which energy expresses E (total) with regard to the kinetic (E_k), potential (E_p) and internal energy (U). Given that $H = U + PV$ which can be rewritten as $H = U + E_p$, it is easy to note that total energy can be replaced by H , enthalpy, since kinetic energy can be neglected. Thus, the energy balance can be given by the following Equation:

$$\frac{dH(t,l,\theta,\phi)}{dt} = \frac{\partial H}{\partial t} + \frac{\partial H}{\partial l} \frac{dl}{dt} + \frac{\partial H}{\partial \theta} \frac{d\theta}{dt} + \frac{\partial H}{\partial \phi} \frac{d\phi}{dt} = \pm \dot{Q} \pm \dot{W} + \text{ther. diffusion} \quad (4)$$

It should be pointed out that Eq. 4 represents the state variable H as a function of space and time. However, in order to explore the thermodynamic behavior of the system, what is needed is to make this Equation appropriate, by introducing the state variable H as a function of $H(T, P, n_i)$, where T , P and n_i denote temperature, pressure and moles of component i , given by:

$$dH = \frac{\partial H}{\partial T} dT + \frac{\partial H}{\partial P} dP + \sum_i \bar{H}_i dn_i \quad (5)$$

where \bar{H}_i represents the partial molar enthalpy. By placing Eq. 5 into Eq. 4, Eq. 6 can be obtained:

$$\begin{aligned} & \left[\left(\frac{\partial H}{\partial T} \right)_P \frac{dT}{dt} + \left(\frac{\partial H}{\partial P} \right)_T \frac{dP}{dt} \right] + \left[\left(\frac{\partial H}{\partial T} \right)_P \frac{dT}{dt} + \left(\frac{\partial H}{\partial P} \right)_T \frac{dP}{dt} \right] \frac{dl}{dt} + \left[\left(\frac{\partial H}{\partial T} \right)_P \frac{dT}{d\theta} + \left(\frac{\partial H}{\partial P} \right)_T \frac{dP}{d\theta} \right] \frac{d\theta}{dt} + \\ & \left[\left(\frac{\partial H}{\partial T} \right)_P \frac{dT}{d\phi} + \left(\frac{\partial H}{\partial P} \right)_T \frac{dP}{d\phi} \right] \frac{d\phi}{dt} = \pm \dot{Q} \pm \dot{W} + [\text{heat generated by the reactive system}] + \\ & [\text{thermal diffusion}] \end{aligned} \quad (6)$$

or rewritten as:

$$\begin{aligned} & \left[\left(\frac{\partial H}{\partial T} \right)_P \frac{dT}{dt} + \left(\frac{\partial H}{\partial P} \right)_T \frac{dP}{dt} \right] + \left[\left(\frac{\partial H}{\partial T} \right)_P \frac{dT}{dt} + \left(\frac{\partial H}{\partial P} \right)_T \frac{dP}{dt} \right] \frac{dl}{dt} + \left[\left(\frac{\partial H}{\partial T} \right)_P \frac{dT}{d\theta} + \left(\frac{\partial H}{\partial P} \right)_T \frac{dP}{d\theta} \right] \frac{d\theta}{dt} + \\ & \left[\left(\frac{\partial H}{\partial T} \right)_P \frac{dT}{d\phi} + \left(\frac{\partial H}{\partial P} \right)_T \frac{dP}{d\phi} \right] \frac{d\phi}{dt} = \pm \dot{Q} \pm \dot{W} + \Delta H_R(-r.V) + K \left[\frac{\partial^2 T}{\partial l^2} + \frac{1}{\phi^2} \frac{\partial^2 T}{\partial \theta^2} + \frac{1}{\phi} \frac{\partial}{\partial \phi} \left(\phi \frac{\partial T}{\partial \phi} \right) \right] \end{aligned} \quad (7)$$

It should also be noted that Eq. 4 does not contain such a heat generated by the reactive system. It is important to emphasize that the heat generated by the reactive system arose in Eq. 6 due to the difference between the partial molar enthalpy between reagents and products. The dependence relationships between H and T or P are well known and they can be easily found by means of the Maxwell equations.

Since the heat generated by the chemical reaction is given by $\Delta H_R(-r.V)$ and considering the mass balance besides the thermal diffusion given by (J_Q) , then Eq. 6 can be rewritten as shown in Eq.7.

To make the in-depth analysis of the process, which was mentioned in the introduction, the entropy balance must be introduced to meet analytical needs. Therefore, the Second Law of Thermodynamics should be used.

$$\frac{dS(t, l, \theta, \phi)}{dt} = \frac{\partial S}{\partial t} + \frac{\partial S}{\partial l} \frac{dl}{dt} + \frac{\partial S}{\partial \theta} \frac{d\theta}{dt} + \frac{\partial S}{\partial \phi} \frac{d\phi}{dt} = \frac{\delta \dot{Q}}{T} + \delta \dot{\sigma} + \text{entro. diffusion} \quad (8)$$

where \dot{Q} denotes the heat flow through the surface and $\dot{\sigma}$ represents the entropy production rate while Equation 8 expresses entropy change as a function of space and time. In the same way that was followed for previous balances, the entropy change must be expressed as a function of T , P and n_i to give thermodynamic sense to the analysis. The entropy change can be given by:

$$dS = \frac{\partial S}{\partial T} dT + \frac{\partial S}{\partial P} dP + \sum_i \bar{S}_i dn_i \quad (9)$$

By introducing Eq.9 into Eq. 8, the following equation can be obtained:

$$\begin{aligned} & \left[\left(\frac{\partial S}{\partial T} \right)_P \frac{dT}{dt} + \left(\frac{\partial S}{\partial P} \right)_T \frac{dP}{dt} \right] + \left[\left(\frac{\partial S}{\partial T} \right)_P \frac{dT}{dt} + \left(\frac{\partial S}{\partial P} \right)_T \frac{dP}{dt} \right] \frac{dl}{dt} \\ & + \left[\left(\frac{\partial S}{\partial T} \right)_P \frac{dT}{d\theta} + \left(\frac{\partial S}{\partial P} \right)_T \frac{dP}{d\theta} \right] \frac{d\theta}{dt} + \left[\left(\frac{\partial S}{\partial T} \right)_P \frac{dT}{d\phi} + \left(\frac{\partial S}{\partial P} \right)_T \frac{dP}{d\phi} \right] \frac{d\phi}{dt} \\ & = \frac{\delta \dot{Q}}{T} + \delta \dot{\sigma} - \Delta S_R(-rV) + [\text{entropy diffusion}] \end{aligned} \quad (10)$$

where $\dot{\sigma}$ denotes the entropy production rate. Entropy diffusion (J_s) is given by $J_s = \frac{J_Q}{T}$ (Umpierre, 2015).

REDUCING THE DIMENSIONALITY OF THE SYSTEM

In practice, it is always important to make some simplifications. Thus, it is not difficult to verify for most practical cases that the following conditions can be assumed:

- There is no shaft work, the fluid properties are constant, the velocity of the fluid is sufficiently high to neglect the effects of diffusions, the external temperature is maintained constant, only the effects and interactions along the length are considered and the drop in pressure is negligible.
- In addition, another substantially significant one can be assumed, namely: the steady state.

As the objective is to bring the system as close as possible to reversibility, which corresponds to a particular steady state and since the above conditions are easy to verify in practice, considering also that $v_l = \frac{dl}{dt}$ and $\Delta S = \frac{\Delta H}{T}$, because $\Delta G=0$ when the system goes towards reversibility, then Eqs. 3, 7 and 10 can be reduced to:

$$\begin{aligned} \frac{\partial c_i}{\partial t} v_l &= \mp r_i ; \quad \left[\left(\frac{\partial H}{\partial T} \right)_p \frac{dT}{dt} \right] v_l = \pm \dot{Q} + \Delta H_R (-r \cdot V) ; \\ \left[\left(\frac{\partial S}{\partial T} \right)_p \frac{dT}{dt} \right] v_l &= \frac{\delta \dot{Q}}{T} + \delta \dot{\sigma} - \Delta S_R (-rV) \end{aligned} \quad (11)$$

Thus, by rearranging Equations in (11), it is easy to conclude the following:

$$\dot{\sigma} = 2 \frac{\Delta H_R}{T} (-rV) \quad (12)$$

Eq. 12 expresses the rate of entropy production for a PFR, which satisfies the restrictions aforementioned. For more complex cases and with reduced restrictions, the complete structure previously formulated should be used. The resulting set of couple differential Equations can be easily solved by means of one of several numerical methods widely cited in the literature (Fogler, 2016).

3. Results and Discussion

A simple but significant case has been used as a case study, for which the reduction of the dimensionality of the system has been taken into account. The classical tubular reactor presented by Bilous (Bilous and Amundson, 1956) also addressed by Himmelblau (Himmelblau and Bischoff, 1968) considers a simple reaction $A \rightarrow B$, irreversible, first order, for which the data sheet, extracted from Bilous and Amundson (1956). By using such data, jointly with the balance model established, the behavior of the concentration of the reagents and the system temperature are shown below. Additionally, since the 2nd Law of Thermodynamics plays a pivotal role in the optimization of results, then the entropy production rate and the entropy production for two wall temperatures are also depicted below (Jaynes, 1980). It should be emphasized that the model with the aforementioned restrictions fully reproduces the results obtained by Bilous and Amundson.

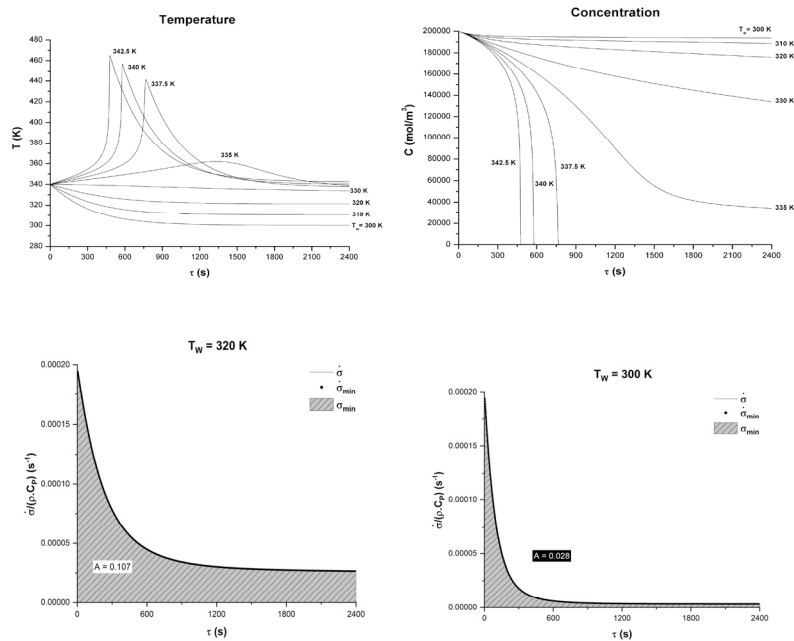


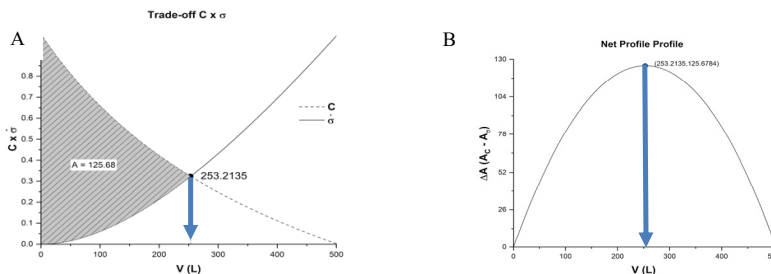
Figure 2: Concentration profile, temperature and the entropy production. The rate of entropy production and its maximum value are also shown for two wall temperatures (T_w). From graphical results presented above it is noted an increase in the conversion as too in the production of entropy, with probably the generation of by-products, which can energetically overload the separation section. Thus, the global energy saving may not meet the initial expected savings.

Additional insights derived from entropy analysis

The idea to develop a relation between concentration and entropy comes from cost analysis that it seems to be direct and very relevant, with a view to determining the optimum length of the reactor. This can be done by establishing a trade-off between conversion and the entropy production. The graph below shows the quantities, concentration and the entropy production, with data normalized on a scale from 0 (zero) to 1 (one) and plotted on the Cartesian plane.

The area under the concentration curve in Figure 3 A shown below can be viewed as income because it generates the product of economic interest while the area under the entropy production curve in the same Figure can be regarded as the entropic cost (variable cost). Thus, the shaded region can be considered as an indicator of net profit because it is the result of the difference between the revenue and the cost. Note also from the intersection point of the two curves, that the area between the concentration and the entropy production may indicate a monetary loss because the operating costs denoted by the production of entropy is higher than the revenue represented by the concentration.

Figure 3 B shows a difference ΔA between the area (A_C) of concentration (profit) and the area (A_σ) of entropy production (cost), as a measure of the net profit or loss. Another relevant feature of the graph shown in Figure 3 B is the maximum point of the curve that represents the difference ΔA between the areas of concentration less entropy production that is, the maximum profit. If $\Delta A < \Delta A_{max}$, this would generate a smaller profit.



Figures. 3 A: Trade-off between concentration and the entropy production. 3 B: Difference ΔA between the area (A_c) of concentration (profit) and the area (A_σ) of entropy production (cost), as a measure of the net profit or loss.

Concluding Remarks

This paper showed the development in detail of the balances of mass, energy and additionally the entropy balance applied to PFR with a view to bridging the conceptual gaps that the literature has discussed over the last 20 years.

The results also showed the behavior of the system in relation to conversion, temperature and the entropy production rate, in addition to the entropy production. It should be noted that the result for conversion and temperature of the system reveal the same behavior obtained by Bilous (Bilous and Amundson, 1956). It was also shown that from the economic point of view, the intersection of the curves, conversion and the entropy production, establishes the optimal point, which denotes economic efficiency. Taking into account the optimal point and since there is a strong relationship between it and the length of the reactor, then it is possible directly to determine the optimal length of the reactor for producing the best economic results, mainly regarding the saving of energy due to reduced overloading of by-products for separation section.

Finally, the methodology has proved to be a feasible strategy, easy to apply and of low operating cost.

References

- Andresen, B., 2011. Current Trends in Finite-Time Thermodynamics. *Angewandte Chemie International Edition* 50, 2690–2704.
- Bilous, O., Amundson, N.R., 1956. Chemical reactor stability and sensitivity: II. Effect of parameters on sensitivity of empty tubular reactors. *AIChE Journal* 2, 117–126.
- Bispo, H., Silva, N., Brito, R., Manzi, J., 2013. On the equivalence between the minimum entropy generation rate and the maximum conversion rate for a reactive system. *Energy Conversion and Management* 76, 26–31.
- Edgar, T., Himmelblau, D., Lasdon, L., 2001. *Optimization of Chemical Processes*, 2nd ed. McGraw-Hill Science, New York.
- Fogler, H.S., 2016. *Elements of Chemical Reaction Engineering*, 5th ed. Pearson.
- Himmelblau, D.M., Bischoff, K.B., 1968. *Process analysis and simulation - deterministic systems*, 1st ed. John Wiley & Sons, Inc.
- Umpierre, Alexandre, 2015. *Introdução à Termodinâmica do Não-Equilíbrio Clássica*, 1st ed. GEN LTC. (in Portuguese).

Digital Twin Model Development for Chemical Plants Using Multiple Time-Steps Prediction Data-Driven Model and Rolling Training

Jia-Lin Kang,^{a*} Somayeh Mirzaei,^a Yao-Chen Lee,^b Yao-Cheng Chuang,^b
Marvin Frias,^b Cheng-Huang Chou,^b San-Jang Wang,^b David Shan Hill Wong,^{b*}
Shi-Shang Jang^b

^a*Department of Chemical and Materials Engineering, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan ROC*

^b*Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, ROC*

jlkang@yuntech.edu.tw; dshwong@gmail.com

Abstract

Data-driven operation monitoring and optimization of chemical plants can be performed using Digital Twin, along with intelligent algorithms. This study proposed a sequence-to-sequence rolling training algorithm to overcome the challenge of rolling predictions. The data were generated through dynamic simulation of the vapor-recompression C3 process using Aspen Plus. Studies showed that StS with rolling training could better fit the real data than the StS model. Moreover, StS with rolling training was able to present efficient long-term predictions as Digital Twin.

Keywords: Digital Twin, sequence-to-sequence model, rolling training, long-term prediction.

1. Introduction

Digital twin (DT) defines the application of holistic simulations to virtually mirror the physical systems. Adopting digital twins can assist engineers to monitor productions, evaluate the benefits and safety of the control in an isolated virtual environment, and operation strategies by interacting with the twin model before going online (Rosen et al. 2015; Eckhart and Ekelhart, 2018; He et al. 2019).

A digital twin can be developed through the first principle models (FPMs) and data-driven models (DDMs) (Chou et al. 2019). Commercial DT solutions for chemical plants are generally based on data-driven models improved purely from the measured data of the targeted industrial plants (Brenner and Hummel, 2017; Magargle et al. 2017). DTs based on data-driven methods depend on the black-box models created to attain correlations between the inputs and outputs of the plant (Pantelides and Renfro, 2013). These systems are applied to find specific production anomalies or to achieve production forecasts. However, they cannot be used to predict abnormal plant operation situations not covered by the available collected data because they are based only on measurement data attained from the plant. Moreover, they need proficient interpretation and are thus complicated to scale up (Eckhart and Ekelhart, 2018). Data-driven DT application depends on the automation and the monitoring systems to provide information about the plant state (Pantelides and Renfro, 2013). In contrast, the advantage of the first principle model can be ensuring that a physical process response

is correct. The drawbacks are including a lot of domain knowledge is required, a lot of information may not be readily available, the model development life cycle will be a very long-time, the accuracy of the model may be limited and restricted by the use of commercial software (Boschert and Rosen, 2016; Martínez 2018). Some researchers have studied the application of DT in engineering optimization (Rosen et al. 2015; Gabor et al. 2016). There is no guarantee that the data-driven method can extrapolate beyond existing data because very little domain information about the process is used. Therefore, there are always doubts about whether such models can serve as true DTs. Sequence-to-Sequence (StS) model is the closest approach as a DT, but the prediction is not long enough (Chou et al. 2019).

This study proposes a novel training algorithm to ensure that the StS model is able to generate infinite long-term rolling predictions as a DT. The paper combines the research within the dynamic simulation and rolling prediction. DT is developed and used for the quality and process control of a chemical plant. The procedure not only minimizes the output errors but also errors in the prediction of the hidden state as a rolling prediction without new sensor input is performed.

2. Methodology

2.1. Process Description

Aspen Plus was employed for dynamic process simulation of the vapor-recompression C3 (propane-propylene) distillation column, and about 10^6 data were generated with random disturbances of the inlet compositions and flow rate. The vapor-recompression C3 flowsheet is shown in Figure 1.

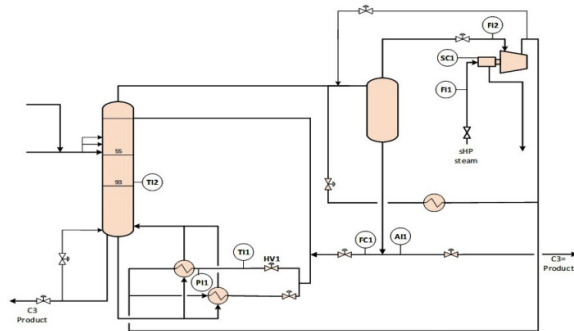


Figure 1. VRC C3 process schematic.

Data of three manipulated variables (u_t), five sensor variables (sv_t) such as flows, pressures, temperatures, and one quality variable (qv) were available every minute, as Table 1 presents. 10^6 data of variables were collected by averaging every 10 minutes, in 2~3 years. The manipulated variables were adjusted in the process, but the sensor variables were measured. The StS model was trained for the rolling prediction. Then, the results of StS model and StS with rolling training for quality and sensor variables were compared for one day and eight days of rolling predictions. The coefficient of determination (R^2), a statistical measure of the goodness of fit with respect to variations, of test data predictions was used as an indicator of the prediction accuracy of the model. The higher coefficients are indicators of better goodness of fit for the observations.

Figure 2 shows StS rolling prediction of the quality variable for one day. The model could only predict well for the first nine hours while the error increased along with the

rolling. R^2 was calculated 0.217 for this case and it confirmed that horizontal prediction is not equal to the rolling prediction. Therefore, the StS model was trained for the rolling based on the algorithm mentioned in Figure 3.

Table 1. Variables of the process

Variables		
u_1	SC1	Compressor speed control
u_2	FC1	Reflux rate control
u_3	HV1	Valve control
qv_1	AI1	Bottom composition measurement
sv_1	FI1	Steam rate measurement
sv_2	FI2	Flow of top vessel measurement
sv_3	TI1	Temperature of major reflux measurement
sv_4	TI2	Temperature of distillation measurement
sv_5	PI1	Pressure of major reflux measurement

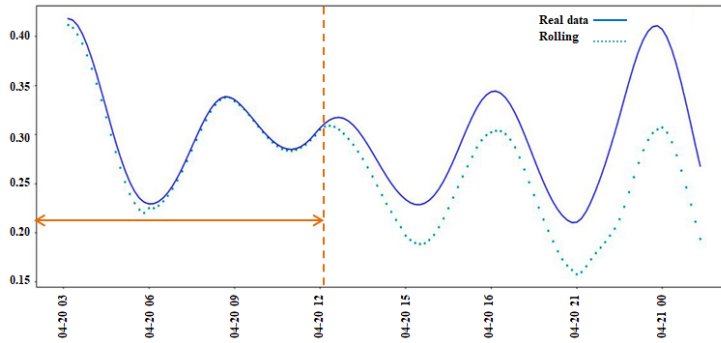


Figure 2. StS Rolling prediction of one day for the quality variable.

2.2. Rolling training algorithm

The new procedure designs rolling training for the StS model in each epoch (Figure 3). The StS model is composed of two sections comprising an encoder (En) and a decoder (De), each of which is a chain of cells corresponding to different time steps. The encoder is an observer, and the decoder is a predictor. Based on Figure 3, when

$$\begin{aligned} \text{Rolling} = 1 & \quad h_{En,r,t} = h_{De,r0,t} & r = 0, \dots, W \\ \text{Rolling} = 2 & \quad h_{En,r,t+1} = h_{De,r0,t+1} & r = 0, \dots, W \\ \text{Rolling} = W-1 & \quad h_{En,r,t+(W-1)} = h_{De,r0,t+(W-1)} & r = 0, \dots, W-1 \end{aligned}$$

$r = 0$ is defined as horizontal prediction.

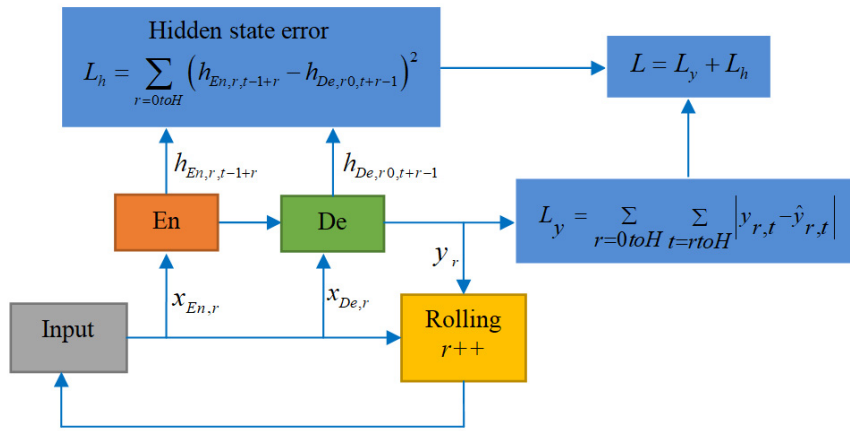


Figure 3. Architecture of StS network with rolling training

3. Result and discussion

3.1. Rolling prediction of quality variable

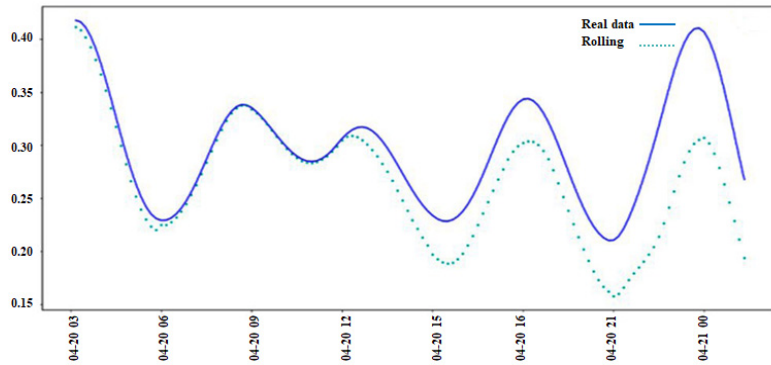


Figure 4. Rolling prediction for one day using the StS model.

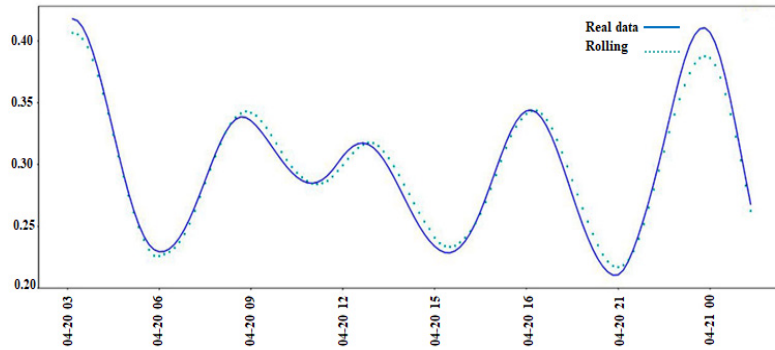


Figure 5. Rolling prediction for one day using StS with rolling training.

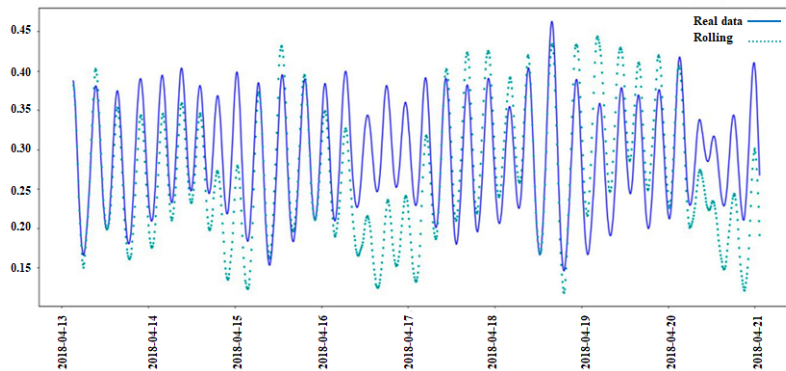


Figure 6. Rolling prediction for eight days using the StS model.

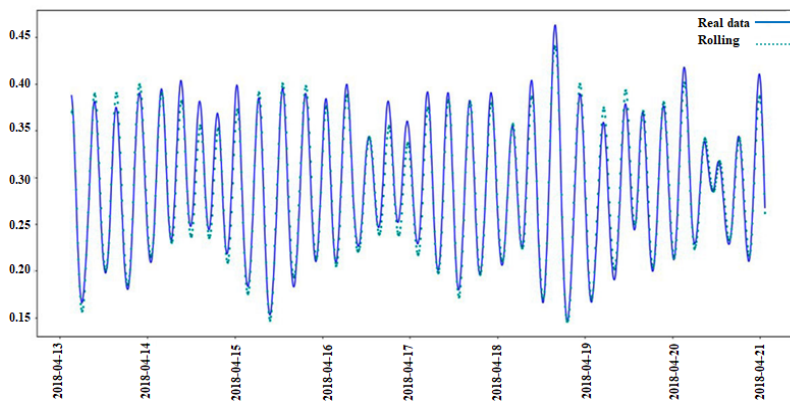


Figure 7. Rolling prediction for eight days using StS with rolling training.

Table 2. Comparison of R^2 for variables predictions

	R^2			
	StS (1 day)	StS with rolling training (1 day)	StS (8 days)	StS with rolling training (8 days)
qv ₁	0.217	0.975	0.170	0.972
sv ₁	0.433	0.950	0.370	0.975
sv ₂	0.906	0.990	0.956	0.996
sv ₃	-6.703	0.734	-2.088	0.895
sv ₄	0.911	0.985	0.877	0.994
sv ₅	0.906	0.985	0.868	0.993

A comparison study of the StS model and StS with rolling training is shown in Figures 4–7 and Table 2. As shown in Figure 5, the rolling prediction of the quality variable for one day was able to closely follow the real data, and R^2 improved by 0.76 compared with the StS model (Figure 4). According to Figures 6 and 7, rolling prediction of the quality variable for eight days did not match the real data for the StS model but, for long-term rolling prediction, StS with rolling training could show good performance and R^2 improved by 0.80. Table 2 confirms the superiority of StS with rolling training

compared with the StS model. For all variables, the errors for StS with rolling training are less than the StS model, especially for qv_1 , sv_1 , and sv_3 that the difference is considerable. Rolling training on the StS model could significantly improve sv_3 because R^2 increased by 7.43 and 2.98 for one day and eight days predictions, respectively. Moreover, the determination coefficients of StS with rolling training for eight days could better fit the real data compared with one-day rolling training. Besides, sv_3 could present an improvement of 21.93% in the long-term prediction.

4. Conclusions

In this study, the rolling training algorithm could enhance the performance of long-term StS rolling predictions in the process of vapor-recompression C3. The model could predict dynamic responses over a near infinite horizon using various manipulation sequences that were not included in the training data, with only a short window of observed data. StS model with long-term rolling predictions could better fit the real data than the StS model. Therefore, the StS model with rolling training was concluded as a true data-driven digital twin of a physical simulator in this study.

References

- S. Boschert, R. Rosen, 2016, Digital twin-the simulation aspect, Mechatronic futures, Springer, Cham, 59-74.
- B. Brenner, V. Hummel, 2017, Digital twin as enabler for an innovative digital shop floor management system in the ESB Logistics Learning Factory at Reutlingen-University, Procedia Manufacturing, 9, 198-205.
- C.-H. Chou, H. Wu, J.-L. Kang, D. S.-H. Wong, Y. Yao, Y.-C. Chuang, S.-S. Jang, J. D.-Y. Ou, 2019, Physically Consistent Soft-Sensor Development Using Sequence-to-Sequence Neural Networks, IEEE Transactions on Industrial Informatics, 16, 4, 2829-2838.
- M. Eckhart, A. Ekelhart, 2018, Towards security-aware virtual environments for digital twins, Proceedings of the 4th ACM workshop on cyber-physical system security, 61-72.
- T. Gabor, L. Belzner, M. Kiermeier, M. T. Beck, A. Neitz, 2016, A simulation-based architecture for smart cyber-physical systems, IEEE international conference on autonomic computing (ICAC), 374-379.
- R. He, G. Chen, C. Dong, S. Sun, X. Shen, 2019, Data-driven digital twin technology for optimized control in process systems, ISA transactions, 95, 221-234.
- R. Magargle, L. Johnson, P. Mandloi, P. Davoudabadi, O. Kesarkar, S. Krishnaswamy, J. Batteh, A. Pitschikani, 2017, A simulation-based digital twin for model-driven health monitoring and predictive maintenance of an automotive braking system, Proceedings of the 12th International Modelica Conference, Prague, Czech Republic, 132, 35-46.
- G. S. Martínez, S. Sierla, T. Karhela, V. Vyatkin, 2018, Automatic generation of a simulation-based digital twin of an industrial process plant, 44th Annual Conference of the IEEE Industrial Electronics Society, 3084-3089.
- C. C. Pantelides, J. Renfro, 2013, The online use of first-principles models in process operations: Review, current status and future needs, Computers & chemical engineering, 51, 136-148.
- R. Rosen, G. Von Wichert, G. Lo, K. D. Bettenhausen, 2015, About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine 48, 3, 567-572.

Parameter estimation and sensitivity analysis for a diesel hydro-processing model

Esin Iplik^a, Panagiotis Tsirikoglou^b, Ioanna Aslanidou^a, Konstantinos Kyprianidis^a

^a *School of Business, Society and Engineering, Mälardalen University, 721 23 Västerås, Sweden*

^b *Limmat Scientific AG, 6300, Zurich, Switzerland*

Abstract

Model-based approaches are essential for the operation, optimization, and control of applications in the process industry. Different structures are often investigated to build representative and robust models, and a set of parameters with the same attributes are required to utilize them effectively. Parameter estimation gets arduous with the increasing complexity of the process, the model, and the size of the parameter space. In this work, a parameter-estimation problem based on a steady-state model of diesel hydrodesulfurization is investigated using gradient-based and gradient-free optimizers. The optimal parameter sets obtained are then assessed in terms of performance and computational time for the different optimizers. Furthermore, the sensitivity of the various parameters is also investigated. Due to the catalytic reactions in this process, some parameters have to be updated depending on the catalyst activity. In addition to the initial estimation, the updated parameters are also studied, and instead of a time-based one, a tolerance-based recalculation schedule is suggested. Finally, the robustness of the final model is analyzed by giving different operating conditions and feed characteristics. The adaptive parameter approach proved better data fitting capabilities by improving the coefficient of determination for temperature predictions.

Keywords: parameter estimation, optimization, sensitivity analysis.

1. Introduction

Environmental regulations to decrease the land and sea vehicles' SO_x emissions have urged refiners to produce low sulfur fuels. For the production of ultra-low-sulfur diesel, ultra-low-sulfur gasoline, and low-sulfur marine grade fuels, to meet the most recent Emission Control Area (also known as global sulfur cap) regulations (IMO, 2016), hydrodesulfurization (HDS) systems are used at various locations in a refinery, i.e., pretreatment of naphtha isomerization, reforming, and heavy-oil cracking feedstock, post-treatment of fluid catalytic cracking product. HDS reactors generally consist of a series of fixed catalyst beds at 300 – 400 °C and 30 – 130 bar (Jiménez et al., 2007).

Tasks of design, simulation, optimization, and control of HDS are aided by mathematical models of the system akin to other refinery operations, which differ from most chemical process models by the number of compounds and chemical reactions involved. There are different approaches to model such systems used for these tasks, some more convenient than the others reflecting the behavior of high numbers of compounds and reactions. To describe the kinetics of the HDS reactions, power-law (Dorneles de Mello et al., 2018) and Langmuir - Hinshelwood (Borgna et al., 2004; Jarullah et al., 2011) models are widely used in the available literature. While the power-law models are simpler with fewer parameters involved, Langmuir – Hinshelwood models are more informative on

the surface-catalyzed reactions. Furthermore, a lumping methodology is commonly used for HDS models. Discrete lumping (Yang et al., 2017), continuous lumping (Aydin et al., 2015; Elizalde & Ancheyta, 2012), and structure-oriented lumping (Ghosh et al., 2009) offer varying levels of fidelity, and these methods are previously discussed according to the required input data and computational effort (Iplik et al., 2020).

Discrete lumping and continuous lumping models are often used for simulation and optimization studies due to their simplicity compared to structure-oriented lumping. During the commercial process with fixed bed reactors, catalyst loses its activity due to coke and metal deposition in the porous structure (Maity et al., 2013). Although this phenomenon is studied and modeled (Takatsuka et al., 1996), and it causes an increase in the operational temperature, it is rarely included in the simple power-law models. If the deactivation is not included in the model, an optimization with a final sulfur concentration constraint will fail to estimate optimal operating conditions as the catalyst activity decreases.

A continuous lumping model for a hydrocracker, by Elizalde and Ancheyta (2014), is an example considering catalyst deactivation with an additional time-dependent model equation for activity decay that has a constant parameter dependent on the operating conditions and feed. As useful as it is for single crude processing refineries, most refineries operate with varying feedstock; therefore, it is hard to assume a constant decay. In this work, we suggest a reaction rate constant update schedule based on the difference between the model predictions and the process measurements.

2. Methods

Continuous lumping models are based on the assumption of the liquid mixture being a continuum with respect to a selected attribute (e.g., true boiling point, carbon number, reactivity) (Becker et al., 2015). This type of model is found useful for petrochemical applications due to its ability to represent the complexity of the mixture with a few parameters. The model used for this work consists of 1D differential mass and energy balances (Eq. 1 and 2) and a correlation of true boiling point to reactivity derived by Sau et al. (1997) (Eq. 3). True boiling point (TBP) is normalized by equation 4, and the distribution of the sulfur species is shown by equation 5. The term N represents the number of species in the mixture, and as $D(k)$ is their distribution function, equation 6 ensures the conservation of species (Chou & Ho, 1988). First-order kinetics was considered for all the reactions. Details of the model can be found in Elizalde and Ancheyta's work (2012) that used the same mass balance equation.

$$\frac{dc(k, z)}{dz} = -\frac{1}{\vartheta} k e^{-Ea/R(1/T-1/T_0)} c(k, z) \quad (1)$$

$$\frac{dT}{dz} = \frac{1}{\vartheta \cdot c_p} \int_0^{k_{max}} k e^{-Ea/R(1/T-1/T_0)} c(k, z) D(k) (-\Delta H(k)) dk \quad (2)$$

$$k = k_{min} - k_{max} \ln(e^{-1} - 1) \theta^{1/\beta} \quad (3)$$

$$\theta = \frac{TBP - \min(TBP)}{\max(TBP) - \min(TBP)} \quad (4)$$

$$D(k) = \frac{di}{d\theta} \frac{d\theta}{dk} = N \frac{d\theta}{dk} \quad (5)$$

$$\frac{1}{N} \int_0^{k_{\max}} D(k) dk = 1 \quad (6)$$

The model with the given equations has five parameters, k_{\min} , k_{\max} , β , E_a , and ΔH . While k_{\min} and k_{\max} represent the reactivities of the highest and the lowest boiling point compounds that consist of a sulfur atom, respectively, β is the exponential correlation coefficient. E_a and ΔH are the activation energy and the heat of reaction; however, rather than belonging to a single reaction, these values represent all the reactions of the continuous mixture.

The initialization of the concentration distribution on the TBP curve was formulated as a constrained optimization (Govindhakannan & Riggs, 2007) and solved with a sequential quadratic programming algorithm. The sulfur compounds are normally distributed on the same TBP curve. Differential material and energy balance equations were solved with a 5th order Runge – Kutta method. The parameter estimation problem was structured as a sum of temperature and sulfur residuals given in equation 7.

$$\min_p \left(\sum_{n=1}^3 ((T_{data,n} - T_{pred,n}(p))^2) + \sum (s_{data} - s_{pred}(p))^2 \right) \quad (7)$$

Two methods were tested to solve the problem, gradient-based optimization and gradient-free optimization. Both solvers were given constraints to prevent impractical results such as negative activation energy. Multiple initial points were randomly distributed by Latin hypercube sampling, and in total, 190 initialization sets were employed for the gradient-based solution. Matlab fmincon solver with an interior point algorithm was used as the gradient-based method, and simulated annealing was used as the gradient-free method. Unlike the gradient-based solution, simulated annealing was initialized only once due to its ability to avoid local minima.

The data was collected from a real HDS plant with three catalyst beds; therefore, all the parameters were estimated for each bed. The resulting parameter sets were evaluated on test data. A sensitivity analysis was applied to the best fitting parameter set by perturbation of $\pm 10\%$ on each parameter. Finally, an update schedule for the reaction rate constant was tested, recalculating the parameters within $\pm 10\%$ constraints of their latest value, when any temperature prediction has an error over 2 K. While updating the deviating bed's parameters, activation energy and specific heat of reaction were not recalculated. The cost function of the update included only the relevant bed's temperature residual and the final sulfur residual.

3. Results and Discussion

Using the temperature and the sulfur measurements of the process, the parameter estimation problem was solved with two different algorithms. Within the analysis of the

190 resulting parameter sets of the gradient-based algorithm, 62% of them were found in local minima, compensating the temperature residual with the sulfur residual or vice versa. The feasible parameter sets, together with the one obtained from the simulated annealing algorithm, were tested by using a data set that was not included in the solution of the parameter estimation problem. Consecutive data sets were selected for the parameter estimation and the test case to eliminate the effect of catalyst deactivation on the reaction constants. The best-fitting parameter set was found by the gradient-based optimization, given in Table 1.

Table 1: Estimated best-fitting parameters.

1 st bed		2 nd bed		3 rd bed	
$k_{\min 1}$ [h^{-1}]	0.89	$k_{\min 2}$ [h^{-1}]	1.49	$k_{\min 3}$ [h^{-1}]	1.49
$k_{\max 1}$ [h^{-1}]	2.45	$k_{\max 2}$ [h^{-1}]	9.95	$k_{\max 3}$ [h^{-1}]	1.51
β_1 [-]	9.96	β_2 [-]	1.00	β_3 [-]	5.79
E_{a1} [kJ/kmol]	65500	E_{a2} [kJ/kmol]	66500	E_{a3} [kJ/kmol]	51800
ΔH_1 [$\text{kJ/kg}_{\text{sulfur}}$]	-7555	ΔH_2 [$\text{kJ/kg}_{\text{sulfur}}$]	-5230	ΔH_3 [$\text{kJ/kg}_{\text{sulfur}}$]	-8278

One point to keep in mind is the time needed to estimate these parameters. The gradient-free algorithm took roughly 48 hours to calculate a single set of parameters, which was not the best fit. The gradient-based algorithm estimated a set of parameters, on average, every 15 minutes that resulted in 190 parameter sets taking the same time. With the increasing number of dimensions and the nonlinearity of the models, parameter estimation requires high computational time, and the fast gradient-based algorithms might fail to calculate a feasible set. However, with a simple set up, the design space can be investigated thoroughly, and a useful set of parameters can be estimated.

Perturbation of the parameters given in Table 1 by $\pm 10\%$ showed that the changes in E_a and ΔH do not have a considerable effect on the temperature or sulfur concentration predictions of the model. On the other hand, k_{\min} , k_{\max} , and β changes affected the model outputs significantly. The extent of the changes is given in Figure 1 for some parameters of the first bed. Since the sulfur measurement is only available for the final product, the right-hand side axes of all the graphs show the effect of each parameter on the final value, whereas the left-hand side axes demonstrate the impact on the given bed temperature.

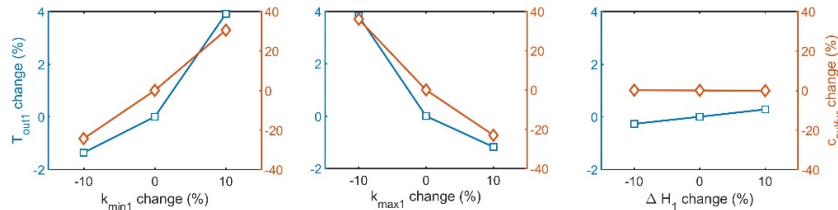


Figure 1: Sensitivity analysis results of some parameters. (\square - Temperature, \diamond - Sulfur)

Based on the sensitivity of the parameters, k_{\min} , k_{\max} and β were selected for the update procedure to explain the catalyst deactivation, therefore, to predict the relevant values better. Temperature measurements were chosen to set the update threshold, and each catalyst bed was evaluated separately. When the bed temperature prediction failed to stay in the 2 K error threshold region, k_{\min} , k_{\max} , and β of the related bed were recalculated by using the last measurements as explained in the previous section.

As expected, the first bed parameters were updated more often due to faster deactivation. None of the updates were periodic; the time passed between the updates varied between

two to eight weeks. Lacking a regular trend shows the benefit of the error threshold based parameter updates. The difference in the feed characteristics causes a difference in the deactivation pattern.

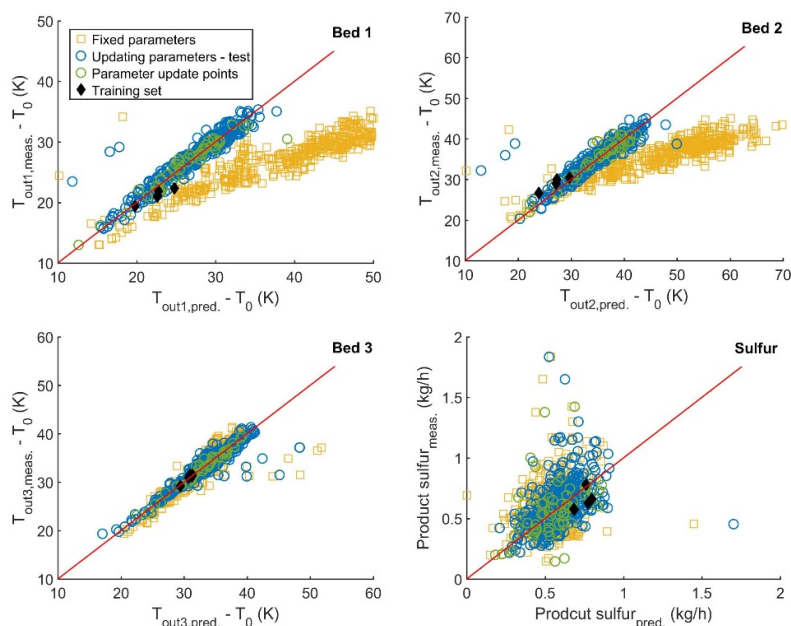


Figure 2: Model outputs with and without the parameter update schedule

The difference between the temperature measurements and the model predictions are given for bed temperatures and sulfur content in Figure 2. The improvements in the model prediction capabilities can be seen from the graphs. Especially in the first two catalyst beds, the fixed parameter set fails to give reasonable estimates as the temperature increases. The update procedure prevents this significant deviation that indeed occurs due to lower catalyst activity. The final sulfur content prediction has a higher error than the other model outputs and, its coefficient of determination increased to 0.65 from 0.45 after the updates.

4. Conclusion

A steady-state continuous lumping model was used to explain real data of an HDS reactor with three catalyst beds. The parameter estimation problem was solved with a gradient-based and a gradient-free algorithm. The best fit was chosen, and the sensitivity of the parameters was checked in the $\pm 10\%$ region. The parameters related to the reaction rate constants were found to be more sensitive than the activation energy and the heat of reaction values. The sensitive parameters were selected to be recalculated when the model failed to predict the temperature with a greater error than 2 K, assuming that the reason was the deactivation of the catalyst over time. The model predicted the bed temperatures better with the update schedule. The deactivation is compensated with a temperature increase, and a high temperature causes a premature deactivation. Therefore, it is crucial to update the models to predict the desired process values correctly; an outdated model cannot be used for optimization or control. The refineries using changing feed should

consider a check of prediction deviations as the deactivation will differ according to the processed mixture.

References

- Aydin, E., Celebi, A. D., Sildir, H., Arkun, Y., Canan, U., Is, G., & Erdogan, M. (2015). Dynamic modeling of an industrial diesel hydroprocessing plant by the method of continuous lumping. *Computers and Chemical Engineering*. <https://doi.org/10.1016/j.compchemeng.2015.06.005>
- Becker, P. J., Celse, B., Guillaume, D., Dulot, H., & Costa, V. (2015). Hydrotreatment modeling for a variety of VGO feedstocks: A continuous lumping approach. *Fuel*. <https://doi.org/10.1016/j.fuel.2014.08.032>
- Borgna, A., Hensen, E. J. M., Van Veen, J. A. R., & Niemantsverdriet, J. W. (2004). Intrinsic kinetics of thiophene hydrodesulfurization on a sulfided NiMo/SiO₂ planar model catalyst. *Journal of Catalysis*. <https://doi.org/10.1016/j.jcat.2003.09.018>
- Chou, M. Y., & Ho, T. C. (1988). Continuum theory for lumping nonlinear reactions. *AIChE Journal*. <https://doi.org/10.1002/aic.690340914>
- Dorneles de Mello, M., de Almeida Braggio, F., da Costa Magalhães, B., Zotin, J. L., & da Silva, M. A. P. (2018). Kinetic modeling of deep hydrodesulfurization of dibenzothiophenes on NiMo/alumina catalysts modified by phosphorus. *Fuel Processing Technology*. <https://doi.org/10.1016/j.fuproc.2018.04.010>
- Elizalde, I., & Ancheyta, J. (2012). Modeling the simultaneous hydrodesulfurization and hydrocracking of heavy residue oil by using the continuous kinetic lumping approach. *Energy and Fuels*. <https://doi.org/10.1021/ef201916s>
- Elizalde, I., & Ancheyta, J. (2014). Modeling catalyst deactivation during hydrocracking of atmospheric residue by using the continuous kinetic lumping model. *Fuel Processing Technology*. <https://doi.org/10.1016/j.fuproc.2014.02.006>
- Ghosh, P., Andrews, A. T., Quann, R. J., & Halbert, T. R. (2009). Detailed kinetic model for the hydro-desulfurization of FCC naphtha. *Energy and Fuels*. <https://doi.org/10.1021/ef900632v>
- Govindhakannan, J., & Riggs, J. B. (2007). On the construction of a continuous concentration-reactivity function for the continuum lumping approach. *Industrial and Engineering Chemistry Research*. <https://doi.org/10.1021/ie0607191>
- IMO, October 2016. Marine environment protection committee. 70th Session.
- Iplik, E., Aslanidou, I., & Kyprianidis, K. (2020). Hydrocracking: A perspective towards digitalization. *Sustainability (Switzerland)*. <https://doi.org/10.3390/su12177058>
- Jarullah, A. T., Mujtaba, I. M., & Wood, A. S. (2011). Kinetic parameter estimation and simulation of trickle-bed reactor for hydrodesulfurization of crude oil. *Chemical Engineering Science*, 66(5), 859–871. <https://doi.org/10.1016/j.ces.2010.11.016>
- Jiménez, F., Kafarov, V., & Nuñez, M. (2007). Modeling of industrial reactor for hydrotreating of vacuum gas oils. Simultaneous hydrodesulfurization, hydrodenitrogenation and hydrodearomatization reactions. *Chemical Engineering Journal*. <https://doi.org/10.1016/j.cej.2007.03.080>
- Maity, S. K., Ancheyta, J., Alonso, F., & Rayo, P. (2013). Hydrodesulfurization activity of used hydrotreating catalysts. *Fuel Processing Technology*. <https://doi.org/10.1016/j.fuproc.2012.09.012>
- Sau, M., Narasimhan, C. S. L., & Verma, R. P. (1997). A kinetic model for hydrodesulfurization. *Studies in Surface Science and Catalysis*. [https://doi.org/10.1016/s0167-2991\(97\)80040-7](https://doi.org/10.1016/s0167-2991(97)80040-7)
- Takatsuka, T., Wada, Y., & Inoue, S. I. (1996). A Catalyst Deactivation Model for Residual Oil Hydrodesulfurization and Application to Deep Hydrodesulfurization of Diesel Fuel. *ACS Symposium Series*. <https://doi.org/10.1021/bk-1996-0634.ch030>
- Yang, Y., Dai, F., Li, C., Xiang, S., Yaseen, M., & Zhang, S. (2017). Kinetic Evaluation of Hydrodesulfurization and Hydrodenitrogenation Reactions via a Lumped Model. *Energy and Fuels*. <https://doi.org/10.1021/acs.energyfuels.7b00496>

Model Development of Amine Regeneration Process with Electrodialysis Reclamation Unit

Jonghun Lim,^{a,b} Jesung Lee,^{a,c} Hyungtae Cho,^a Junghwan Kim^{a,*}

^a*Green Materials and Processes R&D Group, Korea Institute of Industrial Technology, 55, Janga-ro, Ulsan 44413, Korea*

^b*Chemical and Biomolecular Engineering, Yonsei University, 50, Yonsei-ro, Seoul 03722, Korea*

^c*Chemical and Biochemical Engineering, Dongguk University, 1, Pildong-ro, Seoul 04620, Korea*

kjh31@kitech.re.kr.82

Abstract

In this study, we developed a model of amine regeneration process with electrodialysis reclamation unit based on the removal of heat stable amine salt (HSAS) and conducted techno-economic analysis (TEA) by applying the developed model. The HSAS produced in the amine regeneration process acted as a corrosive factor and reduced the efficiency of heat exchanger and the amount of amine available. During development, the model was divided into the amine gas sweetening process and amine regeneration unit construction. The thermodynamic equation considered electrolyte phase equilibrium using the electrolyte nonrandom two liquid model. A RadFrac model was used for regeneration, the removal process of H₂S, and the HSAS removal unit considered the reaction with the sodium hydroxide solution by utilizing the RCSTR model. According to TEA, the total capital investment increased by \$1,823,530, but the total plant cost decreased by \$485,994; therefore, the payback period was calculated at 3.75 y. We believe that the model could provide a guideline for process application if implemented for the optimization of other variables, such as the consumption of raw materials.

Keywords: heat stable amine salt, amine regeneration process, electrodialysis reclamation unit

1. Introduction

Recently, sulfur compounds, such as H₂S, which occur in the refinement of crude oil, are considered the main cause of air pollution. Unlike other air pollutants, they account for about 80% of the emissions generated during the combustion of plants and factories. In addition, they have a fatal effect on the human body and act as a corrosion factor in the process. Hence, many factories focus on the removal of these compounds, with most of them following the amine gas sweetening process for this purpose. This process involves the absorption and removal of H₂S generated during the combustion of crude oil using amine. After the absorption of H₂S, the amine is regenerated from the stripper and reused; this process is called amine regeneration. The absorption and regeneration mechanism of the amine is due to the pK_a difference from H₂S. In the absorber of the amine gas sweetening process, the pK_a difference is greater than one, and thus, the amine is

absorbed from H_2S because it operates at relatively low temperature and high pressure. Subsequently, the regeneration of amine occurring in the pKa difference is smaller than one by creating relatively high and low pressure conditions in the stripper of the amine regeneration process. However, if the amine reacts with relatively strong acids, such as formate, acetate in the acid gas, and even strippers under high and low pressure conditions, heat stable amine salt (HSAS) is produced because the pKa difference is greater than one. Because the HSAS remains in the process and acts as a corrosion factor, reducing the efficiency of the heat exchanger, it should be removed through additional processes. Among the various methods, electro dialysis reclamation is a method of the selective elimination of HSAS through a negative ion and positive ion permeability membrane via a reaction with a basic solution, such as NaOH, which has high electricity consumption and device cost. Therefore, when practically implementing the electro dialysis reclamation units, we must compare them with conventional processes through techno-economic analysis. Here, we developed a model of the amine regeneration process with electro dialysis reclamation unit and validated the model with the real data of the commercial plant. We carried out the economic assessment by applying the developed model, calculated the net present value (NPV) through the annual cash flow considering depreciation and inflation rate, and estimated the payback period (PBP) for annual total plant cost (TPC) reduction in comparison with the increased total capital investment (TCI).

2. Process description

Figure 1 depicts a simple workflow for the amine gas sweetening process. First, to remove the H_2S generated during the refinement of crude oil from sour gas, the gas enters the sour gas absorber. The sour gas absorber typically uses a tray column, and the bubble cap on each plate is deployed to facilitate the contact between gas and liquid. The H_2S is afterward absorbed and removed through an exothermic reaction under relatively low-temperature and high-pressure conditions by contacting with amine (Eq. (1)).

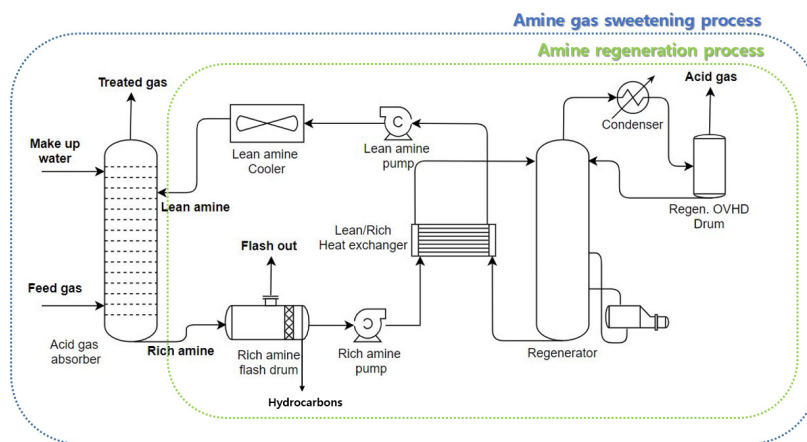


Figure 1 - Schematic diagram of amine gas sweetening process.

The amine that absorbs H₂S, called rich amine, is discharged into the lower part of the absorber and undergoes the amine regeneration process.



The hydrocarbons in rich amine are removed by weight difference in rich amine flash drum, which operates at low pressure to maintain the amine flow to the regenerator. To reduce the heat load of the reboiler in the regenerator and remove H₂S, the rich amine is preheated through the Lean/Rich heat exchanger before entering the regenerator. Subsequently, it enters the regenerator for regeneration. The regenerator utilizes the tray column like sour gas absorber. The steam produced in the reboiler applies sufficient heat to the amine solution to aid the removal reaction. After the regeneration, acid gas is cooled through the condenser and released via the treatment process. Thereafter, the regenerated amine circulates in the lower part for reuse. The amine that removed the H₂S can be reused through the removal process and is called lean amine. The regenerated lean amine is reused by reducing the temperature using the Lean/Rich heat exchanger, thereby entering the absorber through Lean amine cooler and filter. In addition to H₂S, the sour gas produced during the combustion of crude oil contains relatively acidic substances, such as acetic acid and formic acid, which generate HSAS when reacted with amine. The typical HSAS contains acetate and formate, and the salt-producing reaction formula is as follows (Eq. (2), (3)).



3. Model development

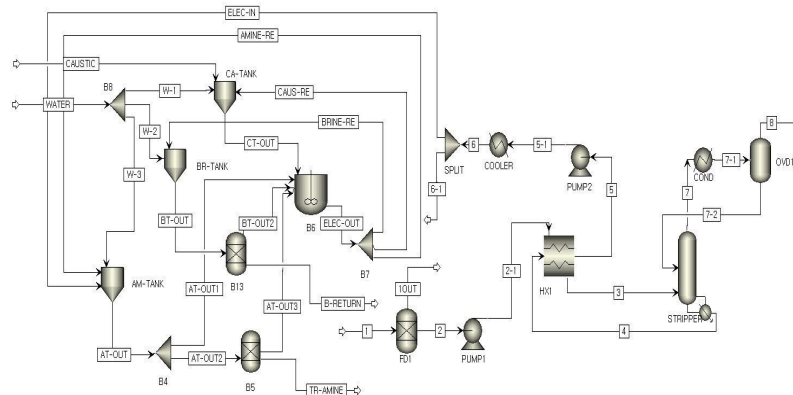


Figure 2 - Schematic diagram of amine gas sweetening process.

In this study, Aspen plus V10 was used to simulate the amine regeneration process. Figure 2 presents a schematic diagram of the amine gas sweetening process. The electrolyte nonrandom two-liquid model was used as the thermodynamic model for considering the gas phase and the chemical reaction mechanism of the absorption and regeneration of H₂S in the amine regeneration process using the Redlich–Kwon equation. The equilibrium constant of absorption, regeneration, and HSAS removal reactions was obtained by minimizing the Gibbs free energy. In process modeling, SEP1 was simulated using the Separator model as rich amine flash drum, which input the split fraction as much as the emission flow from the actual process. Afterward, the rich amine entered the Lean/Rich Heat Exchanger, which was HX1, and heat exchange occurred. HX1 used MHeatX model and entered the outlet temperature of the actual condition. The regenerator was simulated using the RadFrac model and referenced the operating condition in the actual process, such as feed stream to stage, reboiler type. It is also reflected in the model considering the reaction of the regeneration of H₂S. Finally, EL-STACK was simulated using the RCSTR model as the electro dialysis stack. Operation conditions were set to 40 °C and 5 atm, assuming equal temperature and pressure, and the reactor volume was entered referring to design basis. The design conditions are based on those of a commercially operated electro dialysis reclaiming unit with a capacity of 0.5 kg-mol/day.

Table 1 lists the validation results for design. These results indicate that the composition was mostly consistent with the simulation results; however, there was a difference in the amount of HSAS and H₂S removed. This is because the difference of the values was calculated using the equilibrium constant obtained by minimizing the Gibbs energy, which was influenced by other factors in the actual process, such as temperature and pressure. However, factors that affect the equilibrium constant must be derived through experimental study. Therefore, in the future, the model could be modified considering other factors that affect the reaction.

Table 1 - Validation results for design [kg/h].

Category	Stream ELEC-IN		Stream TR-AMINE		Stream B-RETURN	
	Design	Simulation	Design	Simulation	Design	Simulation
H ₂ O	23.0340	22.9770	23.0340	22.9760	256.3300	256.6800
H ₂ S	0.0366	0.0000	0.0366	0.0000	0.0000	0.0000
MDEA	12.3430	12.3430	14.8250	14.6760	0.0000	0.0001
MDEAH ⁺	3.1372	3.1372	0.6336	0.7843	0.0000	0.0000
HCOO ⁻	1.1165	1.1165	0.2021	0.2791	0.0000	0.0000
CH ₃ COO ⁻	0.0770	0.0770	0.0462	0.0193	0.0000	0.0000
HCOONa	0.0000	0.0000	0.0000	0.0000	1.3813	1.2650
CH ₃ COONa	0.0000	0.0000	0.0000	0.0000	0.0428	0.0801
NaOH	0.0000	0.0000	0.0000	0.0000	5.2966	5.3468

4. TEA

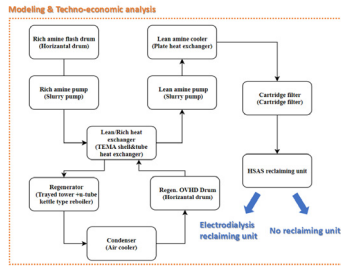


Figure 3 - TEA diagram of the process.

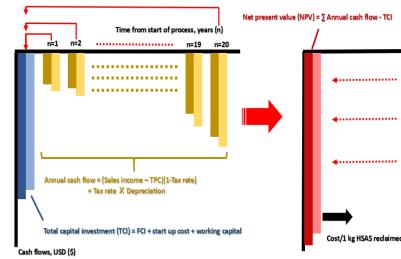


Figure 4 - Cash flow of the process

Techno-economic analysis (TEA) is performed to determine the potential for investment by reviewing economic feasibility in advance. Figure 3 presents the TEA diagram of the process. Here, we conducted TEA with existing processes to determine their economic feasibility with electrolysis reclamation unit using the developed model. The NPV was calculated based on the 20-year recovery period and PBP was estimated using the decreased annual TPC. Table 2 lists the equipment costs and Table 3 the utility costs of the reclamation unit obtained from the Aspen Process Economic Analyzer software.

Table 2 - Equipment cost of the process.

Category	Equipment Cost [\$]
Flash drum	74,400
Rich amine pump	9,044.53
Heat exchanger	1,374,600
Stripper	463,500
Condenser	352,900
Lean amine pump	8,711.29
Cooler	16,100
Filter	50,600
Overhead drum	14,600
Electrodialysis reclaiming unit	1,550,000

Table 3 - Utility cost of the process.

Category	Usage	Unit
Cooling water	134,942.0	\$/y
Steam	4,937,273	\$/y
Electricity	73,937.60	\$/y
Makeup water	22.44294	\$/y

The TCI includes equipment, labor, and land costs necessary for the construction and installation of the initial process. Generally, it includes the fixed capital cost (FCI), startup cost, and working capital. The TPC is the cost incurred by the production or service of a product. Tables 4 and 5 list the TCI and TPC of each process, respectively.

Table 4 - TCI of each process.

Classification	Conventional process	Improved process	Unit
FCI	11,916,857	11,916,857	\$
Startup cost	1,191,685.7	1,191,685.7	\$
Working capital	2,313,272.3	2,586,801.7	\$
TCI	15,421,815	17,245,345	\$

Table 5 - Total plant cost (TPC) of each process.

Classification	Conventional process	Improved process	Unit
Manufacturing cost	10,591,552	10,004,559	\$
General expense	0.000000	0.000000	\$
Reclamation unit cost	-	101,000	\$
TPC	10,591,552	10,105,558	\$

The NPV is the conversion of annual cash flows throughout the investment period into the present value (Eq. (4)). If the NPV has a positive value, it is considered a reasonable investment. However, because the amine regeneration process covered in this study is not a profitable process, the NPV always has a negative value when costs are converted according to current prices. The NPV of the two processes were calculated, and units with lower absolute values were determined to be economical. Table 6 lists the NPV of each process.

Table 6 - NPV of each process.

Category	NPV	Unit
Conventional process	-2,539,231	\$/kg HSAS reclaimed
Improved process	-2,531,225	\$/kg HSAS reclaimed

The NPV calculations revealed that the installation of additional processes increased TCI, but NPV had smaller values owing to low raw material and maintenance costs, thereby making the improved process economical. In addition, PBP was calculated at approximately 3.75 y in comparison with increased TCI.

5. Conclusions

In this study, we developed a model of the amine regeneration process with electro dialysis reclamation unit. The simulation results using the developed model were validated against the data of the actual commercial process. In TEA, TCI increased by approximately \$1,823,530, while the annual TPC decreased by \$485,994. Hence, the cost per kilogram of HSAS reclaimed decreased by approximately \$8,006. We believe that this model could be a guideline for further optimization of real processes

Acknowledgments

This study was conducted with the support of the Korea Institute of Industrial Technology as “Development of On-line Amine Purifier (kitech EE-20-0004)” and “Development of hybrid model and software to optimization of ash removal system in recovery boiler for power generation (JH-20-0014).”

References

- Y. R. Choi, J. H. Kim, I. Moon, 2020, Simulation and economic assessment of using H₂O₂ solution in wet scrubber for large marine vessels, *Energy*, 194, 1, 116907.
- J. H. Kim, K. T. Tak, I. Moon, 2012, Optimization of Procurement and Production Planning Model in Refinery Processes Considering Corrosion Effect, *Industrial & Engineering Chemistry Research*, 51, 30, 10191-10200.

CO₂ methanation: Reactor modelling and parametric analysis

Erasmus Mancusi,^{a*} Piero Bareschino,^a Annunziata Forgione,^a
Claudio Tregambi,^a Francesco Pepe^a

^aDipartimento di Ingegneria, Università degli Studi del Sannio, Piazza Roma 21, Benevento, 82100, Italia

erasmo.mancusi@unisannio.it

Abstract

The fundamental process step within the production of SNG from biomass is methane production. In the present work, methanation is considered to be carried out in a series of adiabatic fixed bed reactors with inter-cooling and product recycle. A robust mathematical model for the considered configuration has been developed that would essentially capture the fundamental operations of the methanation process and help to provide optimal guidelines for industrial operations. In detail, several CO/CO₂ feed ratio compositions and recycle ratios were analyzed.

Keywords: Methanation, Adiabatic Packed Bed Reactors, Power-to-Methane, Dynamical Model.

1. Introduction

The reduction of CO₂ emissions and, more in general, the sustainable production of adequate amounts of energy are unescapable challenges that our societies have to tackle. In this context Power-to-Methane (PtM) is a very interesting technology, for the role that it can play in the storage and distribution of electric energy produced by renewable sources, overcoming their intrinsic uncertainty (Bareschino et al. (2020)). Actually, methane is widely used in the industry, energy, and transportation sectors worldwide and its already developed grid distribution network makes its production preferable when compared to other gaseous fuels. Relying on the features of Chemical Looping Combustion (CLC) residual biomasses can be used as source of heat and CO₂ (Bareschino et al. (2020), Diglio et al. (2017a)). Indeed, in this process the combustion of a fuel is carried out in such a way as to reach an inherent separation of CO₂ and water from the combustion flue gas, leading to a gaseous stream that is ready to be fed to the methanation reactor (Adanez et al. (2012), Diglio et al. (2018)). Furthermore, high purity CO₂ streams may also be obtained from integration of Calcium Looping (CaL) with concentrated solar power (Tregambi et al. (2021)). The conversion of CO₂ to SNG needs hydrogen, which in real applications is usually produced from excess renewable energy (solar and/or wind) (Rönsch et al. (2015)). Carbon dioxide methanation is a highly exothermic and reversible reaction favored at low temperature and high pressure. The catalysts are easily deactivated due to the thermal stress and the methane yield is limited because of the exothermic reaction approaching the chemical equilibrium. The design and development of a methanation reactor is therefore complex owing to both the large amount of heat produced at high pressure and mass transfer and equilibrium

limitations. Several reactor concepts have been consequently proposed and investigated in literature to control the temperature rise associated to the highly exothermic Sabatier reaction so to avoid catalyst sintering while approaching the best thermodynamic condition for the process (Kopyscinski et al (2010)), ranging from multiple adiabatic layers fixed beds with inter-cooling and optional product recycle (e.g. TREMP or Lurgi methanation) to fluidized bed reactors. The reactor configuration studied in the present work consists of three adiabatic fixed bed reactors with gas recycle and inter-stage cooling to displace the gas temperature in the direction of higher equilibrium conversion. A recycle loop is adopted to decrease the hot spot temperature by diluting the inlet reactants concentration. In this background, the main goal is to develop a robust mathematical model for the proposed configuration. Such a model would essentially capture the fundamental operations of the methanation process and help to provide optimal guidelines for industrial operations. In particular, by assuming a production scenario, several recycle ratio and feed composition have been studied.

2. Kinetic scheme and mathematical model

The reported methanation process is carried out in a series of adiabatic fixed bed reactors with nickel-based catalysts and inter-cooling between each stage. In Fig. 1 a schematic layout scheme of the reported methanation process is depicted. After entering the reactor, the temperature of the gas mixture rises to values between 300 and 700 °C because of the exothermicity of the methanation reaction. At the exit of each reactor stage, cooling and water condensing occurs so to partially remove water and adjust the dew point of the gas, in this way the equilibrium is shifted towards the products.

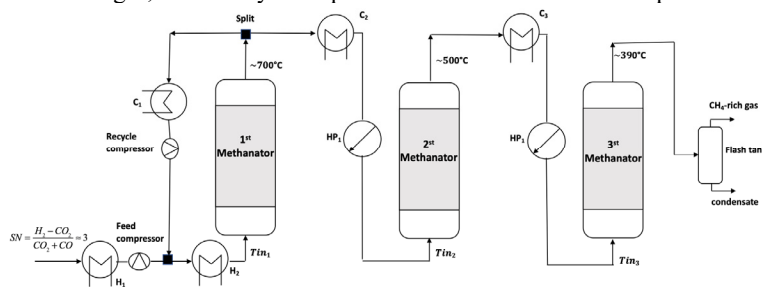


Figure 1 – Adiabatic fixed bed methanation with inter-cooling and recycle; H₁-H₂: heater, C₁-C₄: cooler, HP₁-HP₂ high-pressure boiler and 3 adiabatic methanation reactors, see (Kopyscinski et al., 2010) for more details.

Methanation unit inlet pressure was set to 20 bar, according to Rönisch et al. (2015), so to achieve high conversion degree. The recycle ratio (molar gas flow rate of the recycle stream divided by that of the stream leaving the reactor) was varied between 0.5 and 2.0 both to shift the balance towards the formation of methane and to limit reactor temperature below 700 °C in order to avoid catalyst sintering and carbon deposition (Diglio et al. 2107b).

To describe axial temperature and concentration profiles in each fixed bed methanation reactor, a numerical 1D model has been used. In the model it is assumed that the reactor is adiabatic and pseudo-homogeneous, hence neglecting the heat and mass transfer resistances between the phases and within the catalyst. Therefore, the model complexity is considerably lower in comparison to heterogeneous models (Kopyscinski et al. (2010)). Finally, homogeneous properties are considered, and catalyst surface is

assumed to be completely exposed to the bulk fluid conditions, so fluid-to-particle heat and mass transfer resistances are neglected. An analogous model for CO methanation has been developed and validated by Er-rbib and Bouallou (2014).

The material balance for each gas component ($i=CH_4, CO, CO_2, H_2, H_2O$) was written as:

$$\varepsilon_g \frac{\partial C_i}{\partial t} = -u_{sg} \frac{\partial C_i}{\partial x} - (1 - \varepsilon_g) \rho_c r_i \quad (1)$$

where x is the dimensionless axial position along each reactor belonging in $[0,1]$, ε_g represent the bed porosity, C (kmol.m^{-3}) the gas concentration, ρ_c (kg.m^{-3}) the packed-bed density, and r_i ($\text{kmol.kgcat}^{-1}.\text{s}^{-1}$) the rate of consumption or formation of i -species ($i=CH_4, CO, CO_2, H_2, H_2O$); this latter is determined by summing up the reaction rates of that species in all the reactions R_j (see Table 1) according to the stoichiometric coefficient as follow:

$$r_i = \sum_{j=1}^3 (v_{i,j} R_j) \quad (2)$$

The gas superficial velocity (u_{sg} (m.s^{-1})) has been calculated as follow:

$$u_{sg}(x, t) = \frac{PM_{in}}{PM} T(x, t) u_{sg, in} \quad (3)$$

where PM is the molecular weight (kg.kmol^{-1}) and the subscript in represents the inlet conditions. Finally, the energy balances were written as:

$$(\varepsilon_g \rho_g c_{pg} + (1 - \varepsilon_g) \rho_c c_{pc}) \frac{\partial T}{\partial t} = -(\varepsilon_g \rho_g c_{pg}) \frac{\partial T}{\partial x} - \rho_c \sum_{j=1}^3 (\Delta H_{R_j}) R_j \quad (4)$$

where ρ and c_p ($\text{J.kg}^{-1}.\text{K}^{-1}$) represent the density and heat capacity for gas (g) and solid (c), respectively, while ΔH_{R_j} (kJ.kmol^{-1}) is the reaction enthalpy for R_j reaction. The complete CO/CO₂ methanation reaction scheme is reported in Table 1.

Table 1 – Reactions scheme and associated standard enthalpies of reactions.

Reaction	ΔH_{298} (kJ.kmol^{-1})	
$CO + 3H_2 \Leftrightarrow CH_4 + H_2O$	-206	R1
$CO_2 + H_2 \Leftrightarrow CO + H_2O$	41	R2
$CO_2 + 4H_2 \Leftrightarrow CH_4 + 2H_2O$	-165	R3

The detailed kinetic model for Ni-based catalyst and the reaction rate constants can be found in Xu and Froment (1989).

The mathematical model Eqs. (1)-(4) is completed with the following boundary and initial conditions:

$$C_i(0, t) = C_{i, in}, C_i(x, 0) = 0 \quad (5)$$

$$T(0, t) = T_{in}, T(x, 0) = T_0 \quad (6)$$

The operating conditions, reactor volumes, and catalyst properties used in the simulations are reported in Table 2.

Partial differential equations governing reactor dynamics were reduced to a set of ordinary differential equations by the application of finite differences over a uniform grid of 200 discretization nodes. The reduced set of ordinary differential equations has been numerically solved by making use of the Fortran library DLSODES (e.g. (Altimari et al., 2012)). Both reactions R1 and R3 (Table 1) are exothermic and characterized by a substantial volume contraction.

Table 2 – Parameters used in the simulations.

Parameter	Value	Parameter	Value
P, bar	20.0	V_3, m^3	35
$T_{in1}, ^\circ C$	300	$c_{pc}, J.kg^{-1}.K^{-1}$	1100
$T_0, ^\circ C$	400	$\rho_c, kg.m^{-3}$	2350
V_1, m^3	7	ϵ_g	0.4
V_2, m^3	16	$u_{sg0}, m.s^{-1}$	1

The high reaction enthalpy associated with the above reactions, reported in Table 1, results in large potential adiabatic temperature increase that may cause catalyst sintering and possibly leads to carbon particles formation. To avoid catalyst loss and carbon deposition it is therefore recommended to operate at a temperature below 700 °C (Rönsch et al. (2015)). Accordingly, the reactor can be cooled with the help of product recirculation, this being state-of-the-art for adiabatic methanation reactors during steady-state operation. The recycle ratio is typically chosen in a range between 0.5 and 3.0 for methanation reactors (Matthischke et al. (2018)).

3. Results

Several types of feed stream were considered as feed for the above described reactors network. Namely, a CO₂ pure stream leaving a CLC process and two different syngases coming from PW (poultry waste) and PE (polyethylene) CLC-gasification (Qingyu et al. (2019)) are considered. For the synthesis of methane, the additional hydrogen required for the methanation process comes from an array of high-pressure polymer electrolyte membrane cells (HP-PEM) (Bareschino et al. 2020). According to reactions' stoichiometries (R1 and R3 in Table 1), the stoichiometric number (SN) for the composition of the feed is:

$$SN = \frac{mol_{H_2} - mol_{CO_2}}{mol_{CO} + mol_{CO_2}} \quad (7)$$

The SN value is 3 and according to this value, the hydrogen demand by HP-PEM unit has been defined.

In Fig. 2 are shown methane molar fraction and temperature time series at reactors exit. The methane molar fraction increases moving along the reactor series and at the exit of the last reactor a H₂ and CO₂ conversion degree of 0.97 and 0.96 is respectively observed, corresponding to a $y_{CH_4}=0.96$ on a dry basis. Under the same operating conditions, the time series of the temperature at the exit of each reactor are depicted in

Fig. 3. The adiabatic temperature rise in the first, second and last reactor is 280 °C, 200 °C, and 45 °C, respectively.

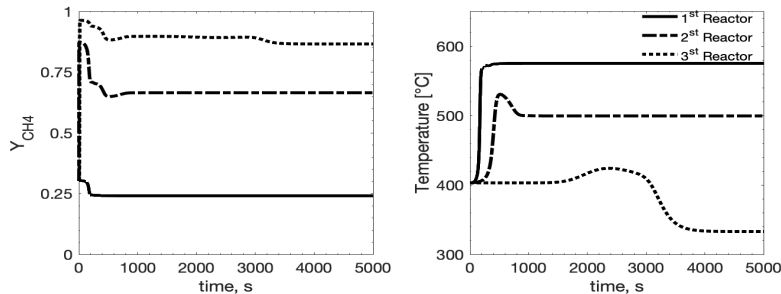


Figure 2 – Methane molar fraction and temperature at the exit of each reactor according to the legend for $y_{CO_2in}=0.2$ and $y_{H_2in}=0.8$, and recycle ratio equal to 1 (other parameters are those reported in Table 2).

To gain insight about the effectiveness of the displacement of gas temperature in the direction of higher equilibrium conversion by intermediate cooling (Fig. 2), the evolution of the H₂ conversion is illustrated in Fig. 3 for 3 methanation reactors as well as the adiabatic temperature effect. Each diagonal line represents an adiabatic reactor (inlet and outlet conditions), and each horizontal line represents an intermediate cooling.

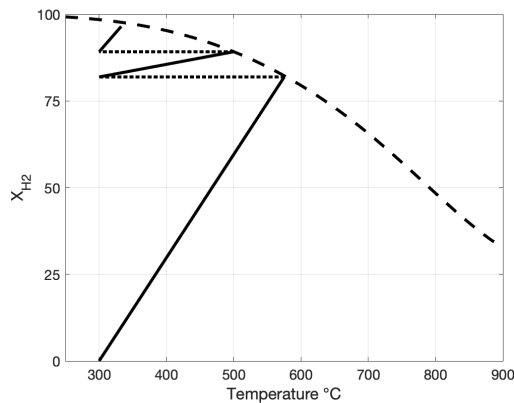


Figure 3 – H₂ conversion vs temperature

When the syngas is produced by CLC-Gasification the following feed composition are considered, $y_{CO}=0.14$, $y_{CO_2}=0.09$ and $y_{H_2}=0.77$ for PW and $y_{CO}=0.21$, $y_{CO_2}=0.04$ and $y_{H_2}=0.77$ for PE gasification. As previously discussed for SN=3 in the fed the required hydrogen is produced by HP-PEM system. Due to the higher exothermicity of R1 respect to R3 (Table 1), increasing the CO/CO₂ ratio in the feed increases the maximum temperature reachable. Therefore, to guarantee that the maximum temperature does not exceed 700 °C, a higher recycle ratio (RR) is required. In particular, for the GCL output by PW a $RR=2$ is used while for GDC output by PE a $RR=2.5$ is considered.

It is important to underline that the concentration of methane on the dry basis obtained in this scheme matches in all the considered cases the level required for its direct supply in the natural gas grid (see Table 3) wherever the minimum required threshold is set by special regulations for the injections of SNG at > 95% methane content.

Table 3 – Methanation unit results on dry basis.

	PW	PE
CH ₄	95.5	95.3
X _{H2}	97.4	98.1

4. Conclusions

The performance of the methanation unit was evaluated by using a series of adiabatic fixed bed reactors filled with Ni supported on alumina, with inter-cooling, water condensation at the exit of each reactor, and product recycle. Several feed compositions and recycle ratio has been investigated to achieve high reactor performances while limiting temperature increase so to avoid catalyst sintering. The results show that in all the investigated cases the methane concentration on the dry basis is high enough to be injected in the methane grid without the necessity of further upgrade processes.

References

- A. Adanez, F. Abad, P. Garcia-Labianno, L. Gayan, L.D. Diego, 2012, Progress in chemical looping combustion and reforming technologies, *Prog. Energ. Combust.*, 38, 2, 215–282.
- P. Altimari, E. Mancusi, L. Russo, S. Crescitelli, 2012, Temperature wave-trains of periodically forced net-works of catalytic reactors, *AIChE J.*, 58, 3, 899–913.
- P. Bareschino, E. Mancusi, M. Urciuolo, A. Paulillo, R. Chirone, F. Pepe, 2020, Life cycle assessment and feasibility analysis of a combined chemical looping combustion and power-to-methane system for CO₂ capture and utilization, *Renew. Sust. Energ. Rev.*, 130, 109962.
- G. Diglio, P. Bareschino, R. Solimene, E. Mancusi, F. Pepe, P. Salatino, 2017a, Numerical simulation of hydrogen production by chemical looping reforming in a dual fluidized bed reactor, *Powder Technol.*, 316, 614–627.
- G. Diglio, P. Bareschino, E. Mancusi, F. Pepe, 2017b, Numerical Assessment of the Effects of Carbon Deposition and Oxidation on Chemical Looping Combustion in a Packed-bed Reactor, *Chem. Eng. Sci.*, 160, 86–96.
- G. Diglio, P. Bareschino, E. Mancusi, F. Pepe, 2018, Techno-Economic Evaluation of a small-scale power generation unit based on a Chemical Looping Combustion Process in Fixed Bed Reactor network, *Ind. Eng. Chem. Res.*, 57, 33, 11299–11311.
- H. Er-rbib, C. Bouallou 2014, Modeling and simulation of CO methanation process for renewable electricity storage, *Energy*, 75, 1, 81–88.
- J. Kopyscinski, T.J. Schildhauer, S.M.A Biollaz, 2010, Production of synthetic natural gas (SNG) from coal and drybiomass – A technology review from 1950 to 2009, *Fuel*, 89, 1763–1783.
- Q. Liu, C. Hu, B. Peng, C. Liu, Z. Li, K. Wu, H. Zhang, R. Xiao, 2019, H₂/CO ratio syngas production from chemical looping co-gasification of biomass and polyethylene with CaO/Fe₂O₃ oxygen carrier, *Energy Convers. Manag.*, 199, 111951.
- S. Matthischke, S. Roensch, R. Güttel, 2018, Start-up Time and Load Range for the Methanation of Carbon Dioxide in a Fixed-Bed Recycle Reactor, *Ind. Eng. Chem. Res.*, 57, 18, 6391–6400.
- S. Rönsch, J. Schneider, S. Matthischke, M. Schlüter, M. Götz, J. Lefebvre, P. Prabhakaran, S. Bajoh, 2015, Review on methanation – From fundamentals to current projects, *Fuel*, 9, 82–102.
- C. Tregambi, P. Bareschino, E. Mancusi, F. Pepe, F. Montagnaro, R. Solimene, P. Salatino, 2021, Modelling of a concentrated solar power – photovoltaics hybrid plant for carbon dioxide capture and utilization via calcium looping and methanation, *Energy Convers. Manag.*, 230, 113792.
- J. Xu, G. Froment, 1989, Methane steam reforming, methanation and water-gas shift: I. Intrinsic kinetics, *AIChE J.*, 35, 85–96.

An uncertainty-aware hybrid modelling approach using probabilistic machine learning

Rasmus Fjordbak Nielsen^a, Nima Nazemzadeh^a, Martin Peter Andersson^b, Krist V. Gernaey^a, Seyed Soheil Mansouri^a

^a*Process and Systems Engineering Centre (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads 228A, 2800 Kgs. Lyngby, Denmark*

^b*Combustion and Harmful Emission Control Centre (CHEC), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads 228A, 2800 Kgs. Lyngby, Denmark*

Abstract

Hybrid modelling has caught renewed attention in many fields of engineering in the last two decades. By combining machine learning with first principles modelling, hybrid modelling is in many cases a more pragmatic modelling approach compared to first principles modelling, and at the same time a more robust alternative to data-driven modelling. However, quantifying uncertainty associated with hybrid models has not been investigated in detail thus far. Thereby, in practice, some models fail to reliably provide information for their performance under uncertainty. In this work, an integrated probabilistic modelling approach is presented for simultaneous modelling and uncertainty quantification using a hybrid model structure. The approach accounts for three types of uncertainty, including training data uncertainty, process stochasticity and model structure uncertainty. To demonstrate the advantages of this approach, the modelling strategy is highlighted through the modelling of a flocculation process. Here, mass and population balance models are combined with a probabilistic machine learning based kinetic model for estimating the particle phenomena kinetics. The model predictions are compared to predictions from a deterministic hybrid model counterpart.

Keywords: Hybrid modelling, Probabilistic modelling, Machine learning.

1. Introduction

With the increasing demand for fast process development alongside the requirement of a transition to sustainable productions, there has never been more focus on generating and utilizing mathematical models for chemical and biochemical processes. Modelling of these processes based on traditional first principles is in many cases not a trivial task and may easily require a significant number of time-consuming experiments. In some cases, the cost of developing a process model is even assessed not to be commensurate with the benefits of having a process model in the first place. This results in process models not being applied to the extent possible, even though they could facilitate a faster process development and secure a more sustainable production.

To reduce the time and efforts spent on modelling a chemical/biochemical process, various machine learning approaches have been proposed. The application of

data-driven modelling has been accelerated in the last two decades by the significant improvements in various sensor techniques, advancements in systems engineering and digitalization of data collection, and the development of highly efficient computational techniques for training data-driven models. Especially the combination of first principles modelling of fundamental phenomena and machine learning, also known as hybrid modelling, has caught renewed attention in the past few years. Application of first principles modelling here ensures that model predictions obey chemical and/or physical laws, which make the machine learning more robust and requires less training data. Hybrid modelling has been applied for modelling of various chemical and biochemical processes showing good predictive capabilities (Glasse and Von Stosch, 2018). More advanced hybrid model structures have also been proposed recently, where one example is the integration of computational chemistry into hybrid modelling, which may facilitate multi-scale modelling (Nazemzadeh, 2020). Recently, Nielsen et al. (2020a) have demonstrated how even complex hybrid models can be trained in real-time by utilizing automatic differentiation and backpropagation.

Many of the hybrid models suggested in literature are lacking a quantification of the prediction uncertainty, which makes it difficult to assess whether the model prediction is supported by the training data or it is a result of an uncertain extrapolation. Knowing about this uncertainty is crucial in the development of new processes where data may be scarce, but also when a hybrid model is used for process control. For instance, it was recently shown by Nielsen et al. (2020b) that a hybrid model based predictive control may suffer from the lack of uncertainty quantification, as it may converge to a solution that is based on a highly uncertain prediction.

With a part of the hybrid model being data-driven, the model accuracy and uncertainty will heavily depend on the data quantity and data quality used for training. This also means that the model uncertainty will change, when hybrid models are used for adaptive learning, where data is continuously acquired from a process and used for training the process model. Therefore, it is necessary to re-estimate this uncertainty continuously based on the current training dataset.

In this work, a hybrid modelling approach is presented, which integrates the quantification of model uncertainty into the model training. The uncertainty quantification includes the training data uncertainty, the uncertainty due to process stochasticity and the model structure uncertainty. The uncertainty of the training data is here estimated based on prior knowledge of the sensor precision, whereas the process stochasticity and the model structure uncertainty are estimated based on the training data. The modelling approach is subsequently applied for modelling of a flocculation process and compared to a deterministic hybrid model.

2. Probabilistic hybrid model

In this work, a hybrid model structure presented by Nielsen et al. (2020a) will be used, where a set of mechanistic models based on first principles are supplemented with a machine learning model that predicts a number of hidden process variables, \bar{y} . The machine learning model estimates the hidden process variables using the current measured process variables, \bar{x} , and control actions, \bar{z} . The mechanistic models are here formulated as a system of differential equations for the measured process variables, \bar{x} . To train the hybrid model, a set of time-series data is required for the measured process

variables, where the hybrid model is trained to predict the process dynamics between pairs of measurements that are distanced with a small time-horizon Δt .

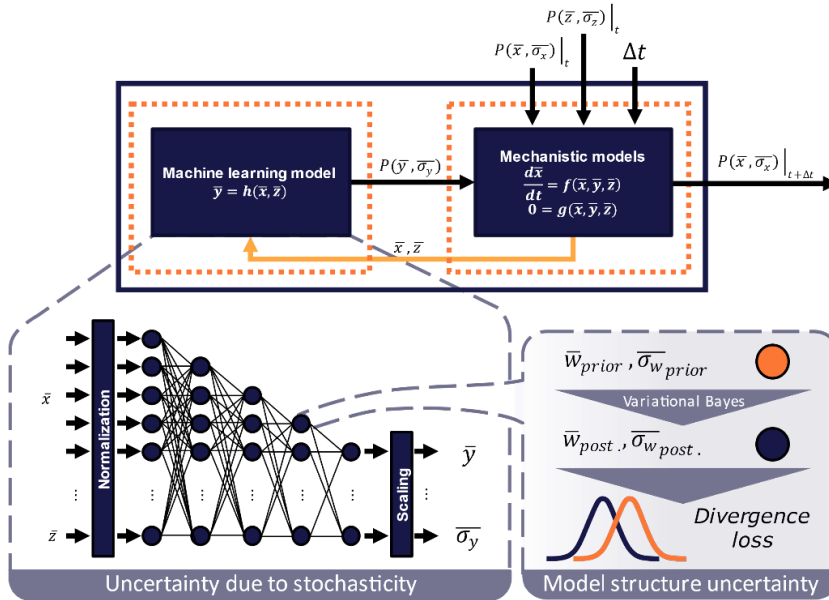


Figure 1: Probabilistic hybrid model structure, based on the hybrid model structure suggested by Nielsen et al. (2020a)

In this work, the sources of model prediction uncertainty in the hybrid model are divided into three, addressing the training data, the process stochasticity, and the model structure uncertainty respectively. The three types of uncertainty will be treated slightly differently, and will be summarized in the following sub-sections, followed by a procedure for training the probabilistic model. A schematic overview of the probabilistic hybrid model structure can be found in Figure 1. To simplify the uncertainty quantification in this work, we assume that structural model uncertainty of the mechanistic models can be neglected, and that the only source of model uncertainty is within the machine learning model.

2.1. Training data uncertainty

To estimate the uncertainty of the training data, prior sensor/sampling knowledge is required. A probability distribution must be specified for each measured process variable to estimate the uncertainty due to sensor precision. Furthermore, the standard deviation for each individual process sensor reading must be estimated. This is done based on prior knowledge of the specific sensor resolution and/or by estimating the uncertainty due to sampling.

2.2. Process stochasticity

To estimate the process stochasticity, the machine learning model is expanded to not only predict the mean value of the process variables, \bar{y} , but also to estimate the standard deviation of these process variables, $\bar{\sigma}_y$, corresponding to their stochasticity. This means that the output size of the machine learning model is doubled compared to a

deterministic hybrid model counterpart. A probability distribution must here be selected for each hidden process variable.

2.3. Model structure uncertainty

To estimate the model structural uncertainty, a probabilistic machine learning model is used as the data-driven model. Compared to a traditional deterministic machine learning model, this type of machine learning model will not only give a prediction, but also quantify the uncertainty of the prediction. This is done by treating the machine learning weights, \bar{w} , as statistical parameters. The mean values of the weights, \bar{w} , and their corresponding standard deviations, $\bar{\sigma}_w$, are here estimated using variational inference, based on the training data. To carry out the variational inference, a prior probability distribution needs to be specified. In this work, the prior probability distribution is set to have a trainable mean value and a fixed standard deviation, corresponding to Empirical Bayes. To carry out the variational inference in this work, the stochastic variational inference method by Kingma and Welling (Kingma, 2014) is used.

After model training, for each model prediction, a new set of weights is sampled from the inferred probability distributions, and thereby estimates the uncertainty of the model. This does not differ significantly from a parameter-based uncertainty analysis of a conventional model. However, as the weights, \bar{w} , in a machine learning model can have much greater impact on the model structure compared to parameters in a conventional first principles model, this uncertainty quantification will not only estimate the parameter uncertainty but also the overall model structure uncertainty.

2.4. Model training

The presented probabilistic hybrid model is trained using a gradient descent method, where the gradient is calculated using automatic differentiation. The loss function that needs to be minimized (see Equation 1 and 2) is the negative log-likelihood, based on the mean and standard deviation of the model predictions of the future measured process variables, \bar{x} , after the various model uncertainties have been propagated through the hybrid model. An additional term in the loss function is furthermore added, consisting of the divergence between the prior probability distribution and the posterior probability distribution of the model weights, which allows for training the prior mean.

$$loss = -\log(\mathcal{L}(\bar{w}, \bar{\sigma}_w | \bar{x}_{data})) + \sum D_f(p_{w,posterior} || p_{w,prior}) \quad (1)$$

$$\min_{\bar{w}, \bar{\sigma}_w} loss \quad (2)$$

When training the probabilistic hybrid model, for each training epoch, a number of samples, n_s , are drawn from the probability distributions for the measured process variables, \bar{x} , with the measured mean and their corresponding estimated standard deviation, $\bar{\sigma}_x$. These samples are then used as inputs to the machine learning model, where the model weights have been sampled from the posterior probability distribution. For each sample, the predicted mean, \bar{y} , and standard deviation, $\bar{\sigma}_y$, are used to sample the stochastic hidden variables. The hidden stochastic variables are then propagated through the set of mechanistic models alongside the sampled initial process variables, \bar{x} ,

resulting in n sample predictions for each training entry, which can be transformed into probability distributions where the likelihood of a given outcome can be calculated. Note that the samples from the various probability distributions are only drawn once for one training entry as it is here assumed that the hidden process variables will remain constant for short time-horizons Δt , which is used during model training.

3. Case study: A lab-scale flocculation of silica

In this section, the presented probabilistic hybrid model will be applied for modelling of a lab-scale flocculation process of silica particles in water, using experimental data that was obtained from a 200 mL stirred glass reactor. The particle size distribution and pH were monitored for in total 11 batch operations, with varying pH. The particle size distribution was here monitored by at-line image analysis using the ParticleTech solution (ParticleTech ApS, Farum, Denmark). pH was monitored using an in-line pH probe. A low pH was observed to promote agglomeration of the silica particles and stirring would result in breakage of particles. This process has previously been examined and modelled using a hybrid machine-learning assisted model by Nielsen et al. (2020a), using the same dataset for training the model, and will here be used for comparison to the predictions from the probabilistic hybrid model. To allow for comparison, the same non-linear discretization scheme has been applied.

As for the previously demonstrated deterministic hybrid model (Nielsen et al., 2020a), the mechanistic models here consist of a discretized population balance model that accounts for the mass balances involved in the agglomeration and breakage phenomena. pH is assumed to be perfectly controlled, meaning that it is possible to model the future pH using the control action z . The machine learning model used in the probabilistic hybrid model is here a deep Bayesian neural network (BNN) with in total 4 layers, where the prior probability distributions are normal distributions for all model weights. The BNN estimates the bin-specific rate of effective agglomerations, the bin-specific rate of breakage and the bin-specific daughter-particle distribution coming from breakage. The stochasticity of the predicted rates is assumed to follow normal distributions.

The estimated standard deviation for the pH measurements was estimated based on the resolution of the sensor, which was 0.01. The estimated standard deviation for the discretized particle size distribution was done based on the sensor resolution and the sampling uncertainty. First, the uncertainty due to the image analysis sensor resolution was introduced. As the image resolution of the ParticleTech solution is 0.5 $\mu\text{m}/\text{pixel}$, this was used as the standard deviation. For each detected particle, gaussian noise with a mean of zero and a standard deviation of 0.5 μm was added to the detected particle diameter. This was done for 100 samples, where each result was discretized based on a non-linear discretization scheme. For each discretized result, the standard error of sampling was estimated, giving a mean and standard deviation of the discretized particle size distribution.

The probabilistic hybrid model structure was trained using TensorFlow and TensorFlow Probability by Abadi et al. (2014), where both the Bayesian neural network and the set of mechanistic models were implemented as custom models. For training the model, 20 samples were used for each training entry, resulting in a roughly 20 times longer training time compared to the corresponding deterministic hybrid model (Nielsen et al., 2020a). End-of-batch predictions from the probabilistic hybrid model and the

deterministic hybrid model counterpart can be seen in Figure 2. The predictions are based on a set of test data coming from a low-pH batch, where the time-horizon of the prediction is one hour.

When comparing the mean prediction of the probabilistic hybrid model and the single-point predictions from the deterministic hybrid model, both models can be seen to succeed in capturing the agglomeration kinetics to an acceptable degree. The probabilistic hybrid model does however offer more information, in the form of reasonable estimations of the prediction uncertainty. It is here evident that the predictions are quite uncertain, especially

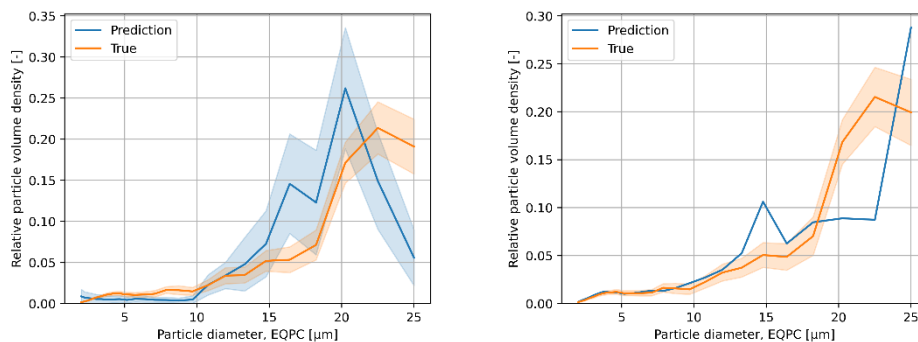


Figure 2: End-of-batch predictions from the probabilistic hybrid model (left) and the deterministic hybrid model counterpart by Nielsen et al. (2020a) (right)

for the agglomeration phenomenon, which cannot be seen easily from the deterministic hybrid model. Further studies are however needed to confirm the reliability of these uncertainty predictions, and the methodology must be tested for varying sets of training data to check its ability to adapt the uncertainty quantification accordingly.

4. Conclusions

In this work, a hybrid model structure and training procedure has been proposed where prediction uncertainty is integrated into the modelling process. This approach could allow for quantifying the model uncertainty for a given set of training data, by addressing uncertainty related to training data, process stochasticity and model structure uncertainty. At the cost of an increased computational time for training, the probabilistic hybrid model has been shown capable of giving a reasonable quantification of the model prediction uncertainties. Additional studies are however needed to confirm these findings.

5. Acknowledgements

This work partly received financial support from the Greater Copenhagen Food Innovation project (CPH-Food), from Novozymes, from EU's regional fund (BIOPRO-SMV project) and from Innovation Fund Denmark through the BIOPRO2 strategic research centre (Grant number 4105-00020B).

References

J. Glassey, M. Von Stosch, 2018, Hybrid Modeling in Process Industries, CRC Press

- R. F. Nielsen, N. Nazemzadeh, L. W. Sillesen, M. P. Andersson, K. V. Gernaey, S. S. Mansouri, 2020a, Hybrid machine learning assisted modelling framework for particle processes, *Comput. Chem. Eng.*, 140, 106916
- R. F. Nielsen, K. V. Gernaey, S. S. Mansouri, 2020b, A Hybrid Model Predictive Control Strategy using Neural Network Based Soft Sensors for Particle Processes, *Comput. Aid. Chem. Eng.*, 48, 1177-1182
- N. Nazemzadeh, L. W. Sillesen, R. F. Nielsen, M. N. Jones, K. V. Gernaey, M. P. Andersson, S. S. Mansouri, 2020, Integration of Computational Chemistry and Artificial Intelligence for Multi-scale Modeling of Bioprocesses, *Comput. Aid. Chem. Eng.*, 48, 295-300
- D. P. Kingma, M. Welling, 2014, Auto-Encoding Variational Bayes, 2nd International Conference on Learning Representations
- M. Abadi et al.. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from [tensorflow.org](https://www.tensorflow.org).

Modelling and Dynamic Optimization of beer fermentation towards optimal flavor and operation

Carina L. Gargalo,^{a*} Alex Matesanz Salvadó,^a Ulrich Krühne,^a Krist V. Gernaey^{a*}

^a *Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 228A, 2800 Kongens Lyngby, Denmark*

kvg@kt.dtu.dk

Abstract

Beer production is an elaborate process, where, from the same critical ingredients, a greatly complex mixture of compounds is obtained depending on the operating conditions. The different combinations of these species, also called off- and on-flavors, are responsible for each beer's distinctive taste.

Improving any step of the beer production process will significantly influence the success of a brewery. Therefore, it is essential to carefully describe the beer production process's different stages to optimize process conditions and beer flavor. To do so, in this work, a comprehensive fermentation model and optimization problem were formulated to determine how an up-to-date industrial beer production should optimally operate. To achieve this, a dynamic optimization problem using a Genetic Algorithm is formulated and implemented to maximize the conversion of substrate to ethanol and minimize fermentation time while setting the final acceptable concentrations of the by-products as strict constraints. Results show that each by-product affects beer flavor as well as process performance in a unique manner.

Therefore, we believe that this work contributes to a comprehensive understanding of the impact of different off- and on-flavors on the optimal beer flavor profile and optimizing process conditions.

Keywords: fermentation, modelling, optimization, beer flavor

1. Introduction

Brewing is one of the oldest processes in which biotechnology has been used to produce alcoholic beverages. The main production steps in beer brewing are: (i) wort production, alcoholic fermentation and maturation, further processing, and stabilization. Fermentation and maturation are the lengthiest steps in the brewing process; fermentation takes three to six days, and the maturation stage takes up to two weeks (Vassilev et al., 2013). Wort, a sugar-rich intermediate, is produced from a starch source such as malted barley (Rodman and Gerogiorgis, 2017). The fermentation occurs through the yeast's enzymatic activity, where sugar is converted to pyruvate, and, in anaerobic conditions, the yeast then converts pyruvate to ethanol and CO₂. Off- and on-flavor compounds, being by-products, are produced during fermentation and maturation (Bosse and Griewank, 2014). Thus, beer flavor is highly dependent on the

yeast's metabolism. There are over 800 compounds present in the final beer, and they are categorized as follows: higher alcohols, esters, fatty acids, vicinal diketones (VDKs), aldehydes, and sulfidic compounds. However, not all of these species have the same impact on the final beer flavor. Higher alcohols, esters, and VDKs are the strongest flavor impacting compounds produced by yeast. Higher alcohols and esters are desirable volatile constituents of a pleasant beer (on-flavor), while VDKs and sulfidic compounds are considered off-flavors. The main sulfidic compounds are H_2S and SO_2 which are highly related to the yeast activity during fermentation. The concentration profile of all the flavor-active species must be under control in order to keep some compounds from dominating and destroying the delicate flavor balance.

Therefore, this work's primary goal is to model and optimize the beer fermentation process to provide the breweries with a competitive edge by improving beer flavor and decreasing costs. To achieve this, the following steps are taken: (i) a comprehensive fermentation model is developed based upon kinetic models, including a yeast growth model, ethanol and by-products production (fusel alcohols, esters, vicinal diketones, acetaldehyde, and sulfidic compounds); (ii) a sensitivity analysis is performed on the initial variables of fermentation model; and, (iii) a dynamic optimization problem is formulated and implemented based upon the genetic algorithm strategy to optimize the flavor profile and operation.

2. Methods

2.1. *Fermentation model scope, development & implementation*

The model developed and implemented in this work is built upon the lager beer fermentation model presented in (Ramirez and Maciejowski, 2007), including the biomass growth model, amino acids uptake, ethanol formation, and flavor model. In this model, there are twelve yield parameters that relate the yeast growth and nutrients uptake to the production of flavor compounds, and these are assumed to be constant (except the growth and amino acids uptake rate). They are obtained by averaging the values given in (Gee and Ramirez, 1994). Thus, in this study, an important modification to the model has been made to account that the enzymatic activity involved in the production mechanisms varies with temperature. Therefore, not only the kinetic rates should change with temperature, but the yield coefficients should also be temperature-dependent. For this reason, instead of assuming that the production yields remain constant during the fermentation, we have introduced temperature dependence on said yields by using a 2nd order polynomial equation. This has shown to be a good option, but the details of such adjustments are not presented here due to the space limitations.

Furthermore, to the best of our knowledge, there is no more freely available data on the production of the remaining flavor compounds. Since the current work aimed not to perform experiments, only ten flavor species have been included in the final model. The model has been implemented in MATLAB and is available upon request.

2.2. *Sensitivity analysis*

A sensitivity analysis is performed on the initial variables to test the effects of upstream processing conditions and different raw materials. To do this, one-factor-at-a-time (OAT) sensitivity analysis is applied. It is one of the most common methods to perform sensitivity analysis since it is simple and easy to implement. The OAT approach

modifies one input variable at-a-time, keeping the remaining parameters at their baseline (Saltelli and Annoni, 2010). Any change observed in the output will unambiguously be due to the changes in the input. However, one disadvantage is that OAT does not account for the effect of simultaneous variation of input variables; thus, it doesn't spot correlations between inputs.

In this work, as presented in Section 3, we assess the impact on the final concentrations of the biomass, ethanol, and by-products when having a 25 % variation in the initial variables (e.g., G_0 , M_0 , X_0).

2.3. *Dynamic optimization*

Several model parameters influence the yeast metabolism and thus final beer flavor profile. Among these parameters, in order to reach optimal flavor and operation, fermentation time and the temperature were selected as the variables to be optimized. The reactor's operating temperature has a significant impact on the resulting beer flavor and aroma by affecting the yeast growth and reaction rates as well as the yield coefficients. To increase the profit margin, breweries continuously try to minimize the fermentation time. However, a minimum time is required to consume all the sugars, produce enough ethanol, and allow the re-assimilation of VDKs to occur successfully (Smart, 2003). Moreover, these parameters are two of the few easily adjustable parameters in an existing fermentation process.

The genetic algorithm (GA) was chosen as the optimization strategy to optimize time and the temperature profile. Genetic algorithms are stochastic search algorithms based on natural selection and natural genetics principles. GA has been chosen since it avoids getting trapped in local optima like other traditional optimization methods. This is due to the fact that being a population-based algorithm, it leads to a better exploration of the search space. For more details on GA and its implementation, please refer to (Potgieter and Engelbrecht, 2007).

3. **Results & Discussion**

3.1. *OAT Sensitivity Analysis*

As previously stated in Section 2.2, a 25 % variation (uncertainty range) is applied to the initial variables to test their impact on the concentrations of substrate (Table 1) during the fermentation and on the biomass, ethanol, and by-products concentration (Table 2).

As observed in Table 1, the initial concentration of glucose and the initial yeast concentration are undoubtedly the initial concentrations with a greater influence on sugars and amino acids' consumption rates. This is due to the compounding effect given by the fact that the presence of sugars boosts yeast growth. In contrast, once the initial concentration of one amino acid is changed, only the corresponding amino acid consumption rate is impacted.

The maximum variations observed regarding the final concentrations of biomass, ethanol, and by-products (flavor species) are presented in Table 2.

Changes in the initial maltose and yeast concentration are highly influenced by the concentration of acetaldehyde, VDKs, ethyl caproate, and ethyl acetate. Furthermore, the initial concentrations of amino acids seem to be the most influential parameters in the production of all fusel alcohols and isoamyl acetate (directly related to the formation of isoamyl alcohol). Furthermore, the relationship between ethyl acetate's final concentration is also affected by changes in the initial concentration of maltose. This is

because ethyl acetate is directly linked to sugar metabolism. Thus, higher initial concentrations of maltose lead to a higher final concentration of ethyl acetate.

Table 1: OAT sensitivity analysis: 25 % variation on initial variables and their impact on the time required to drop to 50% of substrate concentration. Red represents highly sensitive input-output pairs, white stands for zero sensitivity.

unc - 0.25	Time required to reduce 50% concentration					
	Glucose	Maltose	Maltotriose	Leucine	Isoleucine	Valine
Initial variables (wort composition)						
G_0	24.1000	15.8000	13.2000	12.4000	14.8000	13.5000
M_0	0.0000	6.1500	8.1900	0.3190	0.5810	0.7520
N_0	0.0000	0.2050	0.6970	0.3190	0.2910	0.5010
X_0	32.1000	28.5000	27.8000	23.6000	25.3000	26.8000
L_0	0.0000	0.0000	0.0000	8.3100	0.0000	0.0000
I_0	0.0000	0.0000	0.0000	0.0000	10.5000	0.0000
V_0	0.0000	0.0000	0.0000	0.0000	0.0000	8.5200

Table 2: OAT sensitivity analysis: 25 % variation on initial variables and their impact on the final concentration of biomass, ethanol, and by-products. Red represents highly sensitive input-output pairs, white stands for zero sensitivity.

unc - 0.25	Final concentration									
	Biomass	Isobutyl alc.	Isoamyl alc.	2-Meth-1-Butan	Propanol	Ethyl acetate	Ethyl caproate	Isoamyl acetate	VDK	Acetaldehyde
Initial variables										
G_0	3.0200	0.0000	0.0282	0.0000	0.0000	3.9600	7.4100	0.0282	5.4300	9.7400
M_0	4.5400	0.0000	0.0617	0.0003	0.0001	10.9000	11.1000	0.0617	11.8000	12.9000
N_0	1.4600	0.0000	0.0123	0.0000	0.0000	2.1200	3.5600	0.0123	3.7700	2.4200
X_0	26.6000	0.4460	0.4670	0.0190	0.2860	22.6000	29.0000	0.4670	24.6000	12.2000
L_0	0.0000	0.0000	25.0000	0.0000	0.0000	0.0000	0.0000	25.0000	0.0000	0.0000
I_0	0.0000	0.0000	0.0000	25.0000	9.3700	0.0000	0.0000	0.0000	0.0000	0.0000
V_0	0.0000	25.0000	0.0000	0.0000	15.6000	0.0000	0.0000	0.0000	0.0000	0.0000

3.2. Dynamic optimization: time and temperature

As stated in Section 2.3, GA is the optimization strategy chosen to perform the dynamic optimization of the beer fermentation process (time and temperature profile). Four fitness functions (Eqs. 1 to 4) have been designed to test and analyze the following objectives.

To minimize: the concentration of off-flavors (fusel alcohols, VDKs, acetaldehyde) and fermentation time; *To maximize:* concentration of on-flavors (ethanol and esters).

$$FF(O1) = Q_{IB} + Q_{IA} + Q_{MB} + Q_P + Q_{EA} + Q_{EC} + Q_{IAc} + Q_{VDK} + Q_{AAI} + Q_{EtOH} + Q_t \quad (1)$$

$$FF(O2) = Q_{EtOH} + Q_t \quad (2)$$

$$FF(O3) = Q_{EA} + Q_{EC} + Q_{IAc} + Q_{EtOH} + Q_t \quad (3)$$

$$FF(O4) = Q_{IB} + Q_{IA} + Q_{MB} + Q_P + Q_{VDK} + Q_{AAI} + Q_t \quad (4)$$

The fitness function penalizes low concentrations of esters, ethanol and fermentation time as well as high concentrations of fusel alcohols, VDKs, and acetaldehyde. The fitness function will only consider penalties above (off-flavors + time) or below (on-flavors) the extreme values. If the variables are above the minimum population value (off-flavors + time) or below the maximum population value (on-flavors), the penalization is given by the squared difference to the extreme value. Besides, the fermentation time is considered as a penalization factor for all the fitness functions. This is due to the fact that the targeted minimum fermentation time is reached after 120 hours. In Eqs. 1 to 4, all variables have the same importance. Thus all penalties have been normalized for time and all the flavor species. Due to space limitations, the penalization functions, Q_i , are not described here.

The optimal temperature profiles obtained with the four fitness functions are presented in Figure 1. In O2, where the goal is to maximize the ethanol concentration (while minimizing the fermentation time), the optimal profile keeps the temperature at its maximum value (287.65 K). In contrast, when the aim is to minimize the off-flavor species and fermentation time (O4), the optimal profile maintains the temperature at the minimum acceptable value (283.65 K). In O3, the optimal profile is consistently close to the maximum temperature value; however, the temperature declines marginally in the fermentation's early stages. Finally, when all the variables are included in the optimization problem (O1), the optimal profile starts at the maximum temperature (287.65 K) and, after 15 hours, drops to the minimum temperature (283.65 K) and, after 60 hours, the temperature rises and reaches 287.65 K once again.

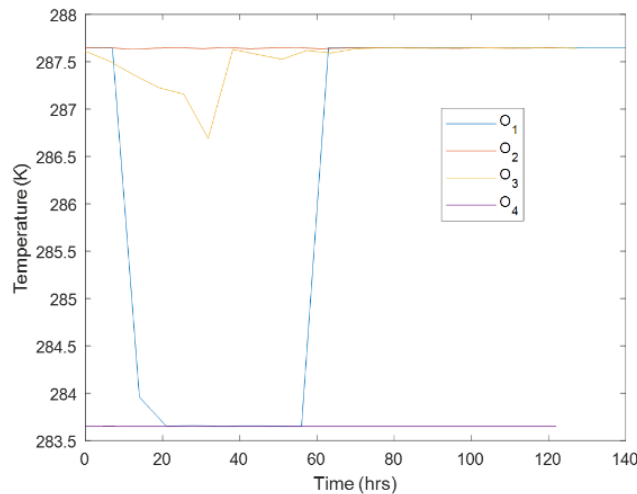


Figure 1: Optimal temperature profiles for optimization scenarios O1 to O4.

In scenarios O2 and O3, the penalization values of isobutanol, isoamyl alcohol, and 2-methyl-1-butanol are relatively high. That means that when the fermentation temperatures are close to the maximum value during the entire fermentation (O2 and O3), the fusel alcohols' production increases significantly, leading to undesired levels. Similar behavior is observed in scenario O4, where the low optimal temperature leads to low values of esters and ethanol production (also undesired). Hence, these low final

concentrations lead to the high penalization values for ethanol and esters in run O4. As mentioned above, due to space constraints, the values of all the penalization factors for the four optimal temperature profiles are not presented.

4. Conclusions

This fermentation model has been carefully examined and improved, as well as tested through a sensitivity analysis. The model is able to assess the trade-off between time-minimization and the final states of the fermentation process, and it has been designed to include all temperature effects. Finally, an optimization problem is solved based on the GA method, which has been applied in order to obtain an optimal flavour profile and within the minimum fermentation time possible. The results obtained from the optimization study show that low temperatures keep the production of the undesired compounds low and high temperatures are required to maximize the ethanol production. Due to the model challenges presented before, it is audacious to claim that the exact optimal profile obtained with the optimization O1 is the best temperature profile for the studied beer fermentation. However, we are confident to affirm that the temperature must be kept low during the first days of the fermentation to avoid over-production of off-flavour species, and afterwards it must be increased to enhance the ethanol and esters production in a minimal time. Future work is two-fold, the authors want to explore the effect of variables like pitching rate and wort composition and also expand the fermentation model to include other flavor compounds and for that we are planning experiments.

Nomenclature

G_0 - Glucose concentration for $t=0$	IB - Isobutanol	IAC - Isoamyl acetate
M_0 - Maltose concentration	IA - Isoamyl alcohol	VDK - Vicinal diketones
N_0 - Maltotriose concentration for $t=0$	MB - 2-methyl-1-butanol	AAI
X_0 - Biomass concentration for $t=0$	P - 2,3-Pentadione	EtOH - ethanol
I_0 - Isoleucine concentration for $t=0$	EA - Ethyl acetate	t - time
V_0 - Valine concentration for $t=0$	EC - Ethyl caproate	

References

- Bosse, T., Griewank, A., 2014. Optimal control of beer fermentation processes with Lipschitz-constraint on the control. *J. Inst. Brew.* 120.
- Gee, D.A., Ramirez, W.F., 1994. A flavour model for beer fermentation. *J. Inst. Brew.* 100, 321–329.
- Potgieter, G., Engelbrecht, A.P., 2007. Genetic algorithms for the structural optimisation of learned polynomial expressions. *Appl. Math. Comput.* 186, 1441–1466.
- Ramirez, W.F., Maciejowski, J., 2007. Optimal beer fermentation. *J. Inst. Brew.* 113, 325–333.
- Rodman, A.D., Gerogiorgis, D.I., 2017. Dynamic optimization of beer fermentation: Sensitivity analysis of attainable performance vs. product flavour constraints. *Comput. Chem. Eng.* 106, 582–595.
- Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. *Environ. Model. Softw.* 25, 1508–1517.
- Smart, K., 2003. *Brewing Yeast Fermentation Performance, Brewing Yeast Fermentation Performance*. Wiley.
- Vassilev, S., Naydenova, V., Badova, M., Iliev, V., Kaneva, M., Kostov, G., Popova, S., 2013. Modeling of alcohol fermentation in brewing-comparative assessment of flavor profile of beers produced with free and immobilized cells, in: *Proceedings - 27th European Conference on Modelling and Simulation, ECMS 2013*. European Council for Modelling and Simulation, pp. 415–421.

On the fitting of bathtub-shaped failure models to lifetime data for selective maintenance optimization

Teemu J. Ikonen,^a Iiro Harjunkoski^{a,b,*}

^a*Aalto University, School of Chemical Engineering, Kemistintie 1, 02150 Espoo, Finland*

^b*Hitachi ABB Power Grids Research, Kallstadter Str. 1, 68309 Mannheim, Germany*

**i.iro.harjunkoski@hitachi-powergrids.com*

Abstract

The maximum likelihood estimate is a method for fitting failure models to lifetime data. In the literature, a commonly used practice is to find a combination of model parameter values where the partial derivatives of the log-likelihood are zero. We show that greater log-likelihood values can be found by using the Nelder-Mead optimization algorithm with adaptive parameters. We demonstrate that the improved fitting has a significant impact on the decision-making on a selective maintenance optimization problem, defined using the failure model by Sarhan and Apaloo (Reliab Eng & Syst Saf, 2013, 112, 137-144).

Keywords: model fitting, reliability, optimization, decision-making.

1. Introduction

The components of an industrial production plant degrade over time, reducing the overall reliability. An unexpected failure in a critical component may cause significant production losses. In order to improve the reliability of the plant, the operators plan maintenance shutdowns, during which degraded components are either replaced or repaired. However, as modern production plants often comprise thousands of replaceable/repairable components, the operators are rarely able to maintain all the components. Identifying the best subset of maintenance actions, subject to time and cost, is referred to as selective maintenance optimization. In the corresponding literature, the component lifetimes are typically assumed to follow either the Weibull or exponential distribution with given parameters without directly considering any lifetime data (Cao et al., 2018). Thus, we recently performed a selective maintenance optimization study that links lifetime data into selective maintenance optimization with the focus on the bathtub-shaped failure rates (Ikonen et al., 2020).

The bathtub-shaped failure rate is a combination of three contemporaneous failure modes: a decreasing infant mortality rate, a constant random failure rate, and an increasing failure rate due to degradation. The literature is fairly established in terms of failure distributions of such models. However, the fitting of model parameters to lifetime data is often performed inadequately. While some authors explicitly maximize the log-likelihood function (e.g. by PROC NLMIXED in SAS or MaxBFGS in the Ox language), a commonly used approach is to find a point in the parameter space where the partial derivatives of the log-likelihood are zero. This approach is, for example, performed by Xie et al. (2002), El-Gohary et al. (2013), and Sarhan and Apaloo (2013) when fitting their failure models to the widely studied lifetime dataset by Aarset (1987). As the log-

likelihood functions of the models are non-convex, the resulting fitting may well correspond to a local optimum, or even a saddle point.

In this paper, we first show that better fits, i.e., greater log-likelihoods, can be found to the lifetime dataset by Aarset (1987) for the failure models proposed in the three abovementioned studies by using the Nelder-Mead algorithm with adaptive parameters (Gao and Han, 2012). Second, we demonstrate how the improved fitting of the failure model by Sarhan and Apaloo (2013) affects the decision-making on a selective maintenance optimization problem.

2. Model fitting

The failure rate $h(t)$ and the reliability function $R(t)$ of the failure model by Sarhan and Apaloo (2013)¹ are defined as

$$\left\{ \begin{array}{l} h(t) = \frac{\lambda\beta\gamma \left(\frac{t}{\alpha}\right)^{\beta-1} e^{(t/\alpha)^\beta + \lambda\alpha(1-e^{(t/\alpha)^\beta})}}{\left[1 - e^{\lambda\alpha(1-e^{(t/\alpha)^\beta})}\right]^{1-\gamma} + e^{\lambda\alpha(1-e^{(t/\alpha)^\beta})} - 1} \\ R(t) = 1 - \left[1 - e^{\lambda\alpha(1-e^{(t/\alpha)^\beta})}\right]^\gamma \end{array} \right. \quad (1)$$

where $\lambda, \alpha, \beta, \gamma > 0$ are the parameters of the model and $t \geq 0$ is time. The likelihood function is defined as

$$\mathcal{L}(\theta) = \prod_{i=1}^n h(t_i)^{d_i} R(t_i), \quad (3)$$

where θ is an array of the model parameters, n is the number of samples in the lifetime dataset, t_i is the service time of component i and d_i indicates whether the component has failed ($d_i = 1$) or is still functioning, i.e., a right-censored sample ($d_i = 0$). In the dataset by Aarset (1987), all components were operated until failure, so let us assume here that $d_i = 1, \forall i \in \{1 \dots n\}$. Substituting Eqs. (1) and (2) into Eq. (3), and taking the natural logarithm from both sides, yields the log-likelihood function

$$\begin{aligned} \ln \mathcal{L}(\theta) = & n[\alpha\lambda + (1 - \beta) \ln\alpha + \ln\beta + \ln\lambda + \ln\gamma] \\ & - \alpha\lambda \sum_{i=1}^n e^{\left(\frac{t_i}{\alpha}\right)^\beta} + \frac{1}{\alpha\beta} \sum_{i=1}^n t_i^\beta + (\beta - 1) \sum_{i=1}^n \ln t_i \\ & + (\gamma - 1) \sum_{i=1}^n \ln \left(1 - \exp\left\{\lambda\alpha \left(1 - e^{(t_i/\alpha)^\beta}\right)\right\}\right). \end{aligned} \quad (4)$$

Sarhan and Apaloo (2013) maximize the log-likelihood function, $\ln \mathcal{L}(\theta)$ (Eq. (4)), by solving numerically a system of four non-linear equations. These equations are defined by setting the partial derivative of $\ln \mathcal{L}(\theta)$ with respect to the four model parameters $\theta = \{\alpha, \beta, \gamma, \lambda\}$ to zero. The same approach is used by Xie et al. (2002) and El-Gohary et al. (2013).

¹ For brevity, we do not include the correspondents of Eqs. (1), (2) and (4) for the models by Xie et al. (2002) and El-Gohary et al. (2013). The equations are available in the papers in question.

In this paper, we maximize $\ln \mathcal{L}(\theta)$ (Eq. 4), by the Nelder-Mead algorithm with adaptive parameters (Gao and Han, 2012), which we implemented using the SciPy.minimize library (Harris et al., 2020). Further, in order to avoid local maxima, we perform a multi-start of 100 optimization procedures with randomized starting points. The function evaluations are implemented using NumPy (Virtanen et al., 2020).

The dataset by Aarset (1987) contains the lifetimes of 50 components (Table 1). We fitted the model by Sarhan and Apaloo (2013), as well as the models by Xie et al. (2002) and El-Gohary et al. (2013), to this dataset using the above described approach. The resulting log-likelihoods $\ln \mathcal{L}(\theta)$ and the model parameters are listed in Table 2. Our approach yields log-likelihoods greater than those reported in the references for all three models.

Table 1: Lifetime dataset by Aarset (1987). The dataset consists of failure times of 50 components, reported here on multiple lines.

0.1	0.2	1	1	1	1	1	2	3	6	7	11	12	18	18	18	18
18	21	32	36	40	45	46	47	50	55	60	63	63	67	67	67	67
72	75	79	82	82	83	84	84	84	85	85	85	85	85	86	86	

Table 2: Log-likelihoods $\ln \mathcal{L}(\theta)$ and parameters of the trained models.

model	method	$\ln \mathcal{L}(\theta)$	parameters
Sarhan and Apaloo (2013)	Nelder-Mead	-203.6972	$\alpha = 79.3949, \beta = 39.0646, \gamma = 1.4040e-2, \lambda = 1.1116e-12$
	reference	-213.858	$\alpha = 49.05, \beta = 3.148, \gamma = 0.145, \lambda = 7.181e-5$
Xie et al. (2002)	Nelder-Mead	-231.6466	$\alpha = 13.7467, \beta = 0.5877, \lambda = 8.7597e-3$
	reference	-236.247	$\alpha = 110.0909, \beta = 0.8408, \lambda = 1.41e-2$
El-Gohary et al. (2013)	Nelder-Mead	-222.2441	$\lambda = 8.9592e-5, c = 8.2786e-2, \theta = 0.2625$
	reference	-224.080	$\lambda = 1.43e-3, c = 4.4e-2, \theta = 0.421$

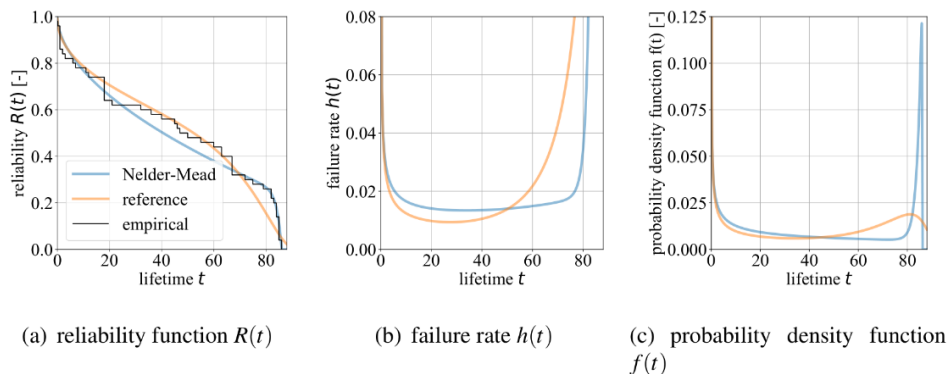


Figure 1: Probability distributions of the model by Sarhan and Apaloo (2013) with the tuned (by the Nelder-Mead algorithm with adaptive parameters) and the reference parameters.

Figure 1 shows the reliability function $R(t)$, failure rate $h(t)$, and probability density function $f(t)$ of the model by Sarhan and Apaloo (2013) with the tuned parameters,

obtained by the Nelder-Mead algorithm, and the reference parameters. The distributions are clearly different. First, the reliability function with the tuned parameters is steeper in the time period of $t = 80 \dots 90$ than that with the reference parameters, capturing the frequent occurrences of failures at the end of time in the dataset (c.f the empirical reliability function). Second, the failure rate $h(t)$ remains nearly constant longer with the tuned parameters (until around $t = 75$) than with the reference parameters (until around $t = 50$). In other words, the ‘bottom of the bathtub’ is wider with the tuned parameters than with the reference parameters, indicating that the failures due to degradation appear later.

3. Selective maintenance optimization

In our earlier work (Ikonen et al., 2020), we derived a variable describing the change in the component-specific reliability ΔR^y , if it is replaced, and presented a contour plot of ΔR^y in the space of component age a and the length of the next operation window t_w . The change in reliability is defined as

$$\Delta R^y = R^y - R^0 = R(t_w|0) - R(a + t_w|a)F \quad (5)$$

where F is a parameter indicating whether the component is functioning before the maintenance break, and R^y and R^0 are the reliabilities if the component is replaced or left untouched, respectively. The conditional reliabilities in Eq. (5) are

$$\begin{cases} R(a + t_w|a) = \frac{R(a + t_w)}{R(a)} \\ R(t_w|0) = R(t_w). \end{cases} \quad (6)$$

$$(7)$$

Figure 2(a) and Figure 2(b) show the contour plots of ΔR^y for the model by Sarhan and Apaloo (2013) with the reference and tuned parameters, respectively. In the earlier work, we pointed out that, if the component is functioning before the maintenance break and the hazard rate is bathtub-shaped, a region exists in the contour plot where the reliability is negative (indicated by the white region in Figure 2(a) and Figure 2(b)). The replacement of a functioning component lying in this region is not sensible.

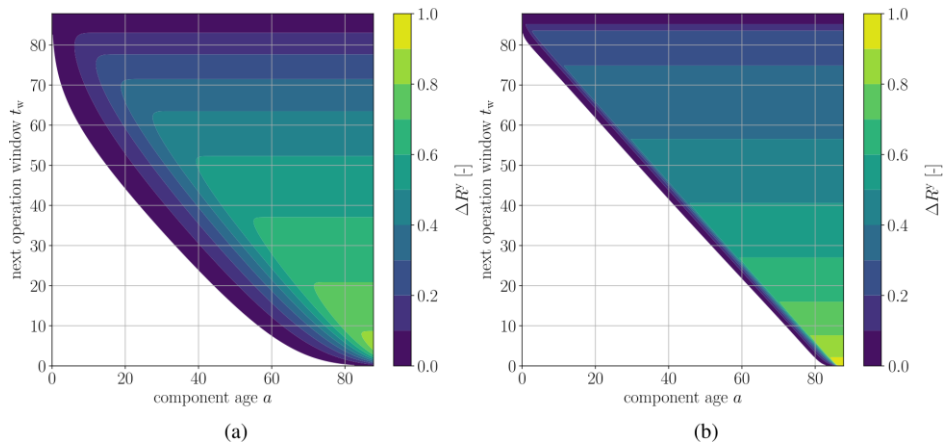


Figure 2: The improvement in reliability, ΔR^y , if a functioning component is replaced, based on the model by Sarhan and Apaloo (2013) with the reference (Subfigure a) and tuned parameters (Subfigure b).

In the earlier work (Ikonen et al., 2020), we studied a small-scale bi-objective selective maintenance optimization problem as an illustrative example. The two conflicting objectives are to maximize the overall reliability of the system, R_{sys} , and to minimize the maintenance cost c_{tot} . The system consists of 18 components with various levels of criticality. The reliabilities of all components in the system are assumed to be equivalent to that used in generating the dataset by Aarset (1987). The decision variables define whether individual components are repaired, replaced, or left untouched, and the number of maintenance personnel, p , invited to perform the maintenance actions. The next planned operation window has a length of $t_w = 10$. For more details of the optimization model and the selective maintenance optimization problem the reader is referred to Sections 3.2 and 4, respectively, of the paper by Ikonen et al. (2020).

Let us now return to this optimization problem and solve its Pareto front using the failure model by Sarhan and Apaloo (2013) with the tuned and reference parameters². We transform the bi-objective optimization problem into a series of single-objective optimization problems by ϵ -constraint method (Haimes et al., 1971), and solve them using the global optimization solver BARON 20.4.14 (Kılınç and Sahinidis, 2018). Figure 3 shows the obtained Pareto fronts, as well as representative solutions. Each grid element in the representative solutions corresponds to a single component, the age of which has been indicated by its color.

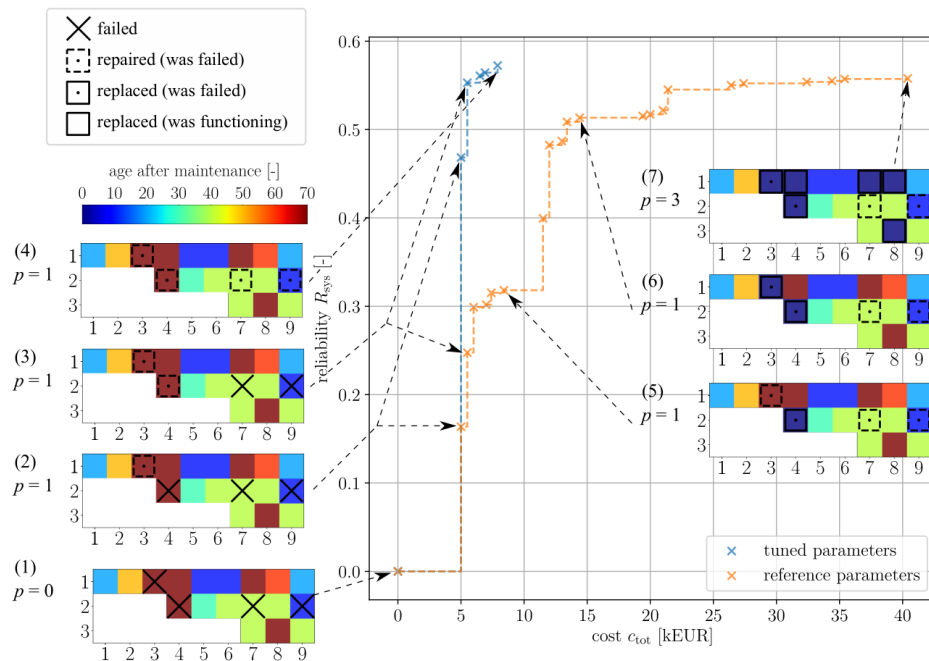


Figure 3: Pareto fronts of the selective maintenance optimization problem, in which the component-specific reliabilities are determined using the model by Sarhan and Apaloo (2013) with the tuned and reference parameters.

² The results obtained with the reference parameters have already been reported in Ikonen et al. (2020).

Starting from the bottom left corner of the Pareto fronts, the first three solutions are the same with both parameters. However, the modeled system reliabilities, as well as the following solutions on the Pareto front are drastically different. The reason is that with the tuned parameters, all components lie in the region of Figure 2(b), where the replacement of a functioning component is not sensible. Thus, even at the top right corner of the corresponding Pareto front, only the failed components are repaired (see Representative solution (4)). The use of reference parameters would result in unnecessary (and costly) component replacements, which would not improve the overall reliability of the system.

4. Conclusions

Fitting a failure model to lifetime data by maximizing the log-likelihood requires solving a non-linear, and often non-convex, optimization problem. We show that performing a multi-start of the Nelder-Mead optimization algorithm with adaptive parameters (Gao and Han, 2012) yields greater log-likelihoods for the models by Xie et al. (2002), El-Gohary et al. (2013) and Sarhan and Apaloo (2013) than the reported reference parameters. We demonstrate that the improved fitting of the model has a significant impact on the decision-making of maintenance actions on an optimization problem, defined using the failure model by Sarhan and Apaloo (2013).

Acknowledgment: Financial support from the Academy of Finland is gratefully acknowledged (project RELOOP, decision number 330388).

References

- M. V. Aarset, 1987. How to identify a bathtub hazard rate. *IEEE Transactions on Reliability* 36 (1), 106–108.
- W. Cao, X. Jia, Q. Hu, J. Zhao, Y. Wu, 2018. A literature review on selective maintenance for multi-unit systems. *Quality and Reliability Engineering International* 34 (5), 824–845.
- A. El-Gohary, A. Alshamrani, A. N. Al-Otaibi, 2013. The generalized gompertz distribution. *Applied Mathematical Modelling* 37 (1-2), 13–24.
- F. Gao, L. Han, 2012. Implementing the nelder-mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications* 51 (1), 259–277.
- Y. V. Haimes, L. S. Lasdon, D. A. Wismer, 1971. On a bicriterion formation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man and Cybernetics* (3), 296–297.
- C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, et al., 2020. Array programming with NumPy. *Nature* 585 (7825), 357–362.
- T. J. Ikonen, H. Mostafaei, Y. Ye, D. E. Bernal, I. E. Grossmann, I. Harjunoski, 2020. Large-scale selective maintenance optimization using bathtub-shaped failure rates. *Computers & Chemical Engineering* 139, 106876.
- M. R. Kilinç, N. V. Sahinidis, 2018. Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with BARON. *Optimization Methods and Software* 33 (3), 540–562.
- A.M.Sarhan, J.Apaloo, 2013. Exponentiated modified Weibull extension distribution. *Reliability Engineering & System Safety* 112, 137–144.
- P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature methods* 17 (3), 261–272.
- M. Xie, Y. Tang, T. N. Goh, 2002. A modified weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety* 76 (3), 279–285.

Sampling Space Reduction for Data-driven Modelling of Batch Distillation - Introducing Expert Process Knowledge through Operation Recipes

Gerardo Brand-Rihm^{a,*}, Erik Esche^a, Jens-Uwe Repke^a

^a*Technische Universität Berlin, Process Dynamics and Operations Group, Sekr. KWT 9, Straße des 17. Juni 135, Berlin 10623, Germany*
g.brandrihm@tu-berlin.de

Abstract

Simulation times of complex chemical processes are often too long for real-time applications, especially when first principles nonlinear dynamic models are used. Data-driven surrogate models of those first principles models trained with simulation data offer a competitive alternative for advanced process control (APC) applications. When sampled with standard methods, processes with a narrow feasible dynamic input space generate datasets with scarce feasible samples. As a representative example, the startup, operation and shutdown of a vacuum batch distillation process separating a binary methanol-water mixture was simulated using a rigorous dynamic pressure-driven model in Aspen Plus dynamics. Expert knowledge was introduced in the form of operation recipes to reduce the vast dynamic input space to a time independent parameter sampling. Using the generated datasets in the form of timeseries, nonlinear ARX and RNN systems were identified. The prediction accuracies of both data-driven models were compared, taking into account their model complexity.

Keywords: data-driven modelling, batch distillation, dynamic simulation, sampling

1. Introduction

In general, it is prohibitive to conduct the minimum amount of experiments necessary for process optimization in real chemical plants. Therefore, process modelling and simulation is used to generate those predictions. Even so, simulation times are often too long for real-time applications, especially when first principles models are used. To counter this, data-driven surrogate models of those mechanistic models are trained with simulation data to reduce computation time during online operation, sacrificing prediction accuracy (Caballero and Grossmann (2008)). One important challenge arises when processes with a narrow feasible input space are sampled with standard methods (such as amplitude-modulated pseudorandom binary sequences (APRBS), among others. See Isermann and Münchhof (2011)), generating datasets with scarce feasible samples.

As a descriptive example, in this contribution a batch distillation process consisting of a vacuum column separating a binary methanol-water mixture is simulated using a rigorous dynamic pressure-driven model in Aspen Plus dynamics (APD) from cold and empty start-up and subsequent shutdown with nitrogen purge. Optimization of the startup of reactive distillation processes has been performed by Reepmeyer et al. (2004); Forner et al. (2006); Lukacs et al. (2006) and Carmona (2011), but using computationally expensive first principles models. Startup strategies for pressure swing distillation were also developed by Klein and Repke (2007), and distillation of three-phase mixtures was also

analysed by Tran et al. (2003), yet based also on first-principles models. Complex dynamic processes like these show a narrow feasible input domain, where standard nonlinear dynamic sampling methods either generate input-output data in a non-representative domain, or more often sample over a hypercubic domain defined by input upper and lower bounds which leads to mostly non-converging simulations. To tackle this, expert knowledge is introduced in the form of operation recipes (Löwe (2001)) to reduce the vast dynamic input space to a time independent parameter sampling.

To test the usability of this sampling method in real-time applications, data-driven models are trained and tested. As a benchmark, the prediction accuracy of a more conventional nonlinear autoregressive model with exogenous variables (NARX) with a tree partition nonlinearity is compared to a recurrent neuronal network (RNN) with two long short-term memory (LSTM) layers with a similar amount of trainable parameters.

2. Methods

2.1. Simulated batch distillation process

A batch distillation process for the separation of a water-methanol mixture shown in Fig. 1 was modelled in APD, with geometric dimensions and specifications shown in Tab. 1. Pressure changers (valves and pumps) had to be placed at every material stream. NRTL was chosen as thermodynamic activity coefficient model with N_2 as Henry component. Phase equilibrium efficiency was chosen as 100%.

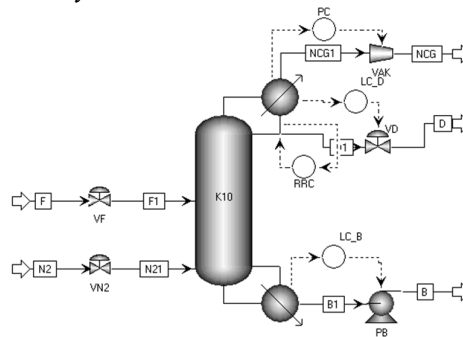


Figure 1: Screenshot of the model implemented in APD, including control loops, inertization stream (N_2) and non-condensable gases (NCG). The distillate tank was simulated as a cumulative distillate flow.

Table 1 Operating conditions and plant specifications for a sieve tray column

parameter	value	units	description
$w_{F,MeOH}$	0.5	kg/kg	feed mass fraction of methanol (rest is water)
T_F	21.0	$^{\circ}C$	feed temperature
h	8.0	m	column height
D	80.0	mm	column inner diameter
V_B	0.08	m^3	max. volume of sump
h_{weir}	0.08	m	weir height
n_{st}	10		number of equilibrium stages
S_F	5		feed stage
UA^{Cond}	1.0	kW/K	overall heat transfer coeff. times area of condenser
$c_{p,Cond}^{med}$	4.187	$kJ/(kg \cdot K)$	cooling medium specific heat capacity
UA^{Reb}	1.635	kW/K	overall heat transfer coeff. times area of reboiler
$c_{p,Reb}^{med}$	1.670	$kJ/(kg \cdot K)$	heating medium specific heat capacity

2.2. Dynamic sampling through recipes

Batch distillation processes are inherently dynamic. Since in this contribution the scope is to generate surrogate models capable of predicting whole operation cycles, this process was simulated starting from a cold and empty state at ambient pressure, with all valves closed, pumps off and heating and cooling mediums at 21°C. Subsequent shutdown with nitrogen inertization was also simulated. The developed first principles model was set up to have 12 inputs (see Tab. 2) and 26 outputs (although an arbitrary amount can be selected from the flowsheet simulator). Since performing classical nonparametric input samplings would lead to mostly non converging simulations (due to nonphysical behaviours with strong nonlinearities like pumping out empty trays), process knowledge is introduced to constrain the dynamic input sampling space using operation recipes. Different recipes are described in Löwe (2001) and Wang et al. (2003), among others. Here the “conventional strategy” (according to Löwe) described in Tab. 3 will be adopted. A time independent parameter sampling on the parameter set $\Theta = \{\theta_1 \dots \theta_{20}\}$ is translated to dynamic input signals. Hammersley sampling over a hypercubic domain of $[-10\%, +10\%]$ from nominal values with normal sampling distribution was chosen.

Table 2 Selected plant inputs

ID	variable	description
1	x_{VN2}	nitrogen valve position
2	PC^{SP}	pressure controller setpoint
3	x_{VF}	feed valve position
4	RRC^{SP}	reflux ratio controller setpoint
5	$T_{Reb}^{med,in}$	reboiler heating medium inlet temperature
6	LC_D^{MODE}	level controller mode ($\in \{0 = auto, 1 = manual\}$)
7	LC_D^{SP}	condenser level controller setpoint
8	LC_D^{OP}	condenser level controller output
9	PC^{MODE}	pressure controller mode ($\in \{0 = auto, 1 = manual\}$)
10	PC^{OP}	pressure controller output
11	LC_B^{MODE}	sump level controller mode ($\in \{0 = auto, 1 = manual\}$)
12	LC_B^{OP}	sump level controller output

Table 3 “Conventional strategy” starting from an empty and cold column

step	description
1	wait for time θ_1 with system cold and empty
2	generate vacuum ramping down PC^{SP} to θ_2 during a period of θ_3
3	start filling sump with feed ramping up x^{VF} to θ_4 during a period of θ_5
4	wait until sump level reaches θ_6 and close feed valve ($x^{VF} = 0$)
5	heat up reboiler ramping up $T_{med,in}^{Reb}$ to θ_7 during a period of θ_8
6	wait until condenser level reaches θ_9
7	start extracting distillate shrinking RRC^{SP} from ∞ to θ_{10}
8	turn on condenser level controller ($LC_D^{MODE} = 0, LC_D^{SP} = \theta_{11}$)
9	rise reboiler temperature (ramp up $T_{med,in}^{Reb}$ to θ_{12} during a period of θ_{13})
10	wait until achieving cumulative distillate composition θ_{14}
11	stop distillate extraction ($LC_D^{MODE} = 1, RRC^{SP} = 0$)
12	stop generating vacuum ($PC^{MODE} = 1, PC^{OP} = 0$)
13	cool down reboiler ramping down $T_{med,in}^{Reb}$ to θ_{15} during a period of θ_{16}
14	wait until reaching a reboiler temperature of θ_{17}
15	start inertization ramping up x_{VN2} to θ_{18} during a period of θ_{19}
16	empty bottoms ($LC_B^{MODE} = 0, LC_B^{OP} = \theta_{20}$)
17	stop inertization ($x_{VN2} = 0$)

2.3. Design and training of data-driven models

In this contribution, the more classical NARX and newer RNN models are analysed in terms of prediction accuracy over a testing dataset. As an extension of linear time-invariant polynomial models to the nonlinear case, NARX extends ARX models by a nonlinear block (Schoukens and Ljung (2019)). Here, a tree partition nonlinearity was chosen. The amount of tree partition units of the NARX model was set to generate a model with a similar complexity as the RNN (in terms of trainable parameters n_p). The architecture of both data-driven models is specified in Tab. 4. If not specified, default settings of the corresponding toolboxes were used.

The prediction accuracies can be analysed comparing the mean squared errors (MSE) between min-max normalized testing (y_{stc}) and predicted (\hat{y}_{stc}) data for each model, according to:

$$MSE = (n_s n_t n_y + 1)^{-1} \sum_{s=1}^{n_s} \sum_{t=1}^{n_t} \sum_{c=1}^{n_y} (y_{stc} - \hat{y}_{stc})^2$$

where s, t, c are sample, timestep and channel indexes with cardinalities n_s, n_t and n_y .

Table 4 Chosen data-driven model architectures and number of trainable parameters n_p

model	NARX	RNN
architecture	1 input regressor with input delay of 1 timestep for each output. Nonlinearity: tree partition with 231 units. focus: prediction method: PEM	sequence of layers: <input type="checkbox"/> LSTM (64 units), returning sequences <input type="checkbox"/> LSTM (64 units), returning sequences off <input type="checkbox"/> dense layer solver: adam loss function: MSE
toolbox	Matlab/System Identification Toolbox version 9.13	tensorflow 2.1.0/Keras
n_p	729,456	730,426

3. Results and discussion

A total of 1000 Hammersley parameter samples were generated. 645 of these led to converging simulations, shown in Fig. 2. Each complete batch cycle consists of $n_t = 401$ timesteps with a sample time of $T_s = 0.01$ h.

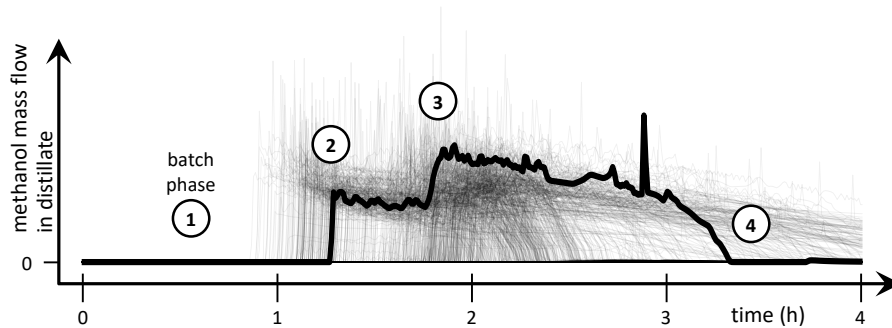


Figure 2: Simulated mass flow of MeOH in the distillate for converged runs in APD. One sample is highlighted in **black**, others are shaded. Visible batch phases: (1) cold phase, (2) first distillate production, (3) further heating and (4) distillate depletion and shutdown.

The input time series \mathbf{X} generated by the recipe in Tab. 3 as well as the output time series \mathbf{Y} from APD were collected in a 3D data structure in the form: $\dim(\mathbf{X}) = (n_s, n_t, n_x)$ with $n_x = 12$ input channels, and $\dim(\mathbf{Y}) = (n_s, n_t, n_y)$ with $n_y = 26$ output channels. A training/testing data ratio of 3/1 was selected to split the converged samples ($n_s = 645$) into sample sizes $n_s^{train} = 483$ and $n_s^{test} = 162$. The resulting training and testing datasets were min-max normalized for each corresponding channel in the training set.

For similar complexity (in terms of trainable parameters), the MSE over all testing samples was $5.24 \cdot 10^{-3}$ for the NARX model and $9.83 \cdot 10^{-3}$ for the RNN model. Fig. 3 shows one testing sample example and the predictions of both data-driven models.

Although higher, the MSE achieved by the RNN model and its lack of unexpected peaks makes this combination of recipe sampling and RNN as surrogate model a promising strategy for APC applications. The peaks observed in the NARX prediction might be consequence of overfitting by forcing its number of parameters n_p to equal those used in the RNN. To tackle this issue, alternative model architectures and their hyperparameters must be studied in more depth.

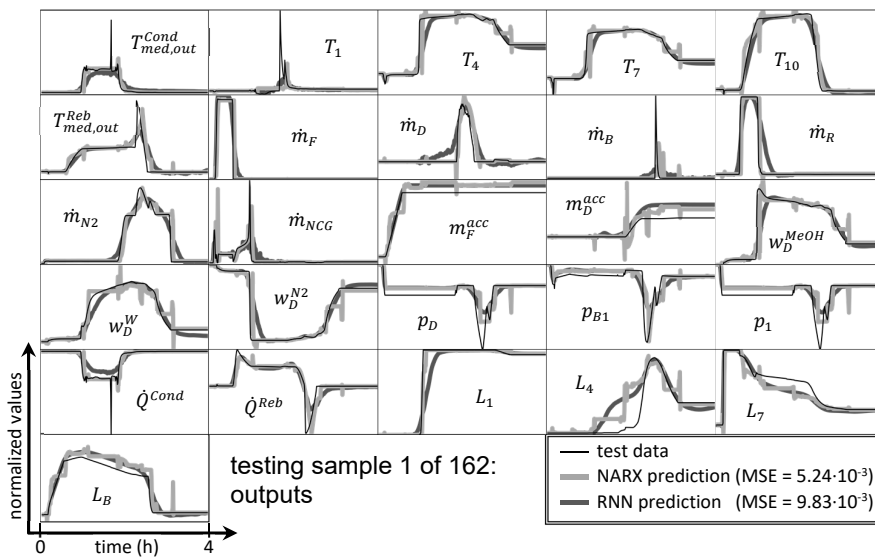


Figure 3: Outputs predicted by NARX and RNN models for one of 162 testing samples

4. Conclusion and outlook

A dataset for the identification of data-driven models was generated from a pressure-driven dynamic batch distillation column simulation in Aspen Plus Dynamics. Even though expert knowledge in the form of operation recipes was introduced to reduce the dynamic sampling space to a time invariant parameter space, 35.6% of the Hammersley samples led to non-converging APD simulations. NARX and RNN models were identified using data from converged simulations, where the NARX model achieved a higher prediction accuracy on a testing dataset than the RNN model in terms of its MSE with a similar number of trainable parameters, although showing unexpected peaks in its

prediction. Therefore, the combination of recipe sampling and RNN as surrogate model shows promising properties for APC applications.

The methodology of dynamic sampling through recipes may be applied to other complex dynamic processes where the feasible sampling domain is highly constrained. Adaptive sampling strategies might decrease the amount of non-converging simulations and enhance prediction accuracies of data-driven models. Since data from non-converged simulations carry information about operation constraints, in future works those trajectories will be included in the training dataset. Furthermore, the architecture of NARX and RNN models may also be studied in more depth to avoid overfitting by varying the amount of regressors and trainable model parameters.

Batch operation may be optimized based on data-driven models, and these models may be adapted for their use in real-time applications.

5. Acknowledgements

This research has been supported by the project "KI-Inkubator-Labore in der Prozessindustrie - KEEN", funded by the Bundesministerium für Wirtschaft und Energie (BMWi) under grant number 01MK20014T.

References

- J. A. Caballero, I. E. Grossmann, 2008. An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal* 54 (10), 2633–2650.
- I. Carmona, 2011. Design and optimaler Betrieb alternativer Prozessvarianten der reaktiven Batch-Rektifikation. Ph.D. thesis, TU Berlin.
- F. Forner, M. Döker, J.-U. Repke, J. Gmehling, G. Wozny, sep 2006. Zeitoptimales Anfahren der Reaktivdestillation in Packungs- und Bodenkolonnen. *Chemie Ingenieur Technik* 78 (9), 1284.
- R. Isermann, M. Münchhof, 01 2011. Identification of Dynamic Systems: An Introduction with Applications. Springer.
- A. Klein, J.-U. Repke, oct 2007. Zweidruck-Batchdestillation zur Trennung homogener azeotroper Gemische. *Chemie Ingenieur Technik* 79 (10), 1639–1647.
- K. Löwe, 2001. Theoretische und experimentelle Untersuchungen über das Anfahren und die Prozeßführung energetisch und stofflich gekoppelter Destillationskolonnen. Ph.D. thesis, TU Berlin.
- T. Lukacs, C. Steger, J.-U. Repke, E. Rev, Z. Lelkes, oct 2006. Design of batch reactive distillation with equilibrium limited consecutive reactions. *Systèmes D'Information, Modélisation, Optimisation et Commande en Génie des Procédés*.
- F. Reepmeyer, J.-U. Repke, G. Wozny, 2004. Time optimal start-up strategies for reactive distillation columns. *Chemical Engineering Science* 59 (20), 4339–4347.
- J. Schoukens, L. Ljung, 2019. Nonlinear system identification: A user-oriented roadmap. CoRR abs/1902.00683. URL <http://arxiv.org/abs/1902.00683>
- T. K. Tran, J.-U. Repke, G. Wozny, mar 2003. Anfahren dreiphasiger Rektifikationskolonnen: Simulation und Experiment. *Chemie Ingenieur Technik* 75 (3), 194–199.
- L. Wang, P. Li, G. Wozny, S. Wang, 2003. A startup model for simulation of batch distillation starting from a cold state. *Computers & Chemical Engineering* 27 (10), 1485–1497.

On the acceleration of global optimization algorithms using a cutting plane decomposition and machine learning

Asimina Marousi, Antonis Kokossis,*

School of Chemical Engineering, National Technical University of Athens, Zografou Campus, 9, Iroon Polytechniou Str. GR-15780, Athens, Greece

akokossis@mail.ntua.gr

Abstract

The paper presents an accelerated approach for global optimization that is applied on quadratic programming problems. The global optimization is based on a decomposition method using cutting planes that are generated, analysed, and screened using advanced analytics. The work capitalizes and builds on innovations by Baltean-Lugojan et al. (2019) who recently presented a generic and effective outer approximation method suitable for semidefinite relaxations. The use of data analytics is applied in populations (P) of cutting planes, experimenting with different metrics and clustering methods. The proposed approach achieves a reduction in the integrality gap by 18-30% with the largest reductions in relation to larger problems (jumbo problems: 100 variables, 75% density).

Keywords: global optimization, outer- approximation, cutting planes, quadratic programming, machine-learning

1. Introduction

Global optimization is a dynamic and thriving area of research with several powerful methods already in place. The effectiveness of algorithms varies with the particular form of the objective function and constraints, the number and type of variables and constraints, and/or the problem sparsity/structure. Even for smooth objective functions and constraints (e.g. polynomials), the optimization problem is generally difficult to solve while providing proper guaranties for the optimum and solution at acceptable times. State-of-the-art (SoA) solvers in global optimization rely on iterative algorithms that establish progress through local optimization steps or, in the case of decomposition methods, in the successive optimization of sub-problems. Cutting plane approximation methods classify among the most promising algorithms of that sort. As a general remark, algorithms generate data that are not fully turned to account. Emerging technologies in machine learning and advanced data analytics can be deployed to exploit such data streams as they are generated and produced internally. Data populations may relate to individual iterations or relate to the entire sequence of iterations. Open challenges include means to translate interim data into exploitable populations, methods to map data analytics with decisions, and techniques to formulate approximations suitable to improve optimization efficiency and speed. In this paper we examine the potential of embedding data analytics in conjunction with SoA optimization methods, specifically advances in quadratic programming featuring outer-approximation methods and cutting planes following the work by Baltean-Lugojan et al. (2019). The new approach leads to a new

class of separation problem and is particularly amenable to the data-driven approach studied in the paper. The purpose of the research has been to particularly examine the correlation of the cutting planes as they are selected by the algorithm in the QP problems and to formulate selection criteria appropriate for each individual iteration. In Section 2, the approximation of the original QP and the decomposition of the cutting planes are presented. Section 3 demonstrates the methodology establishing data metrics and analytics. The emphasis in the research experiments has been to improve the solution quality. Results are presented in Section 4. They consider a wide range of examples and rather encouraging evidence.

2. Problem description and background

The optimization problem is considered a nonconvex quadratic problem that is box and linearly constrained, i.e.

$$z_{qp} = \min_x \{x^T Q x + c^T x \mid Ax \leq b, x \in [0,1]^N\} \quad (1)$$

with an N -variable vector x , $A \in \mathbb{R}^{p \times N}$ and $Q \in \mathbb{R}^{N \times N}$ assumed to be an indefinite matrix. For the problem to be solved a series of reformulations and relaxations are taking place as proposed in Sherali and Fraticelli (2002). Each quadratic term $x_i x_j$ is replaced by a new variable X_{ij} . Let the lifted variables $X_{ij} \forall i, j$ form the symmetric matrix $X = xx^T$ and let $Q \cdot X = \text{Tr}(Q^T X) = \sum_{i,j} Q_{ij} X_{ij}$, representing the Forbenius inner product. Then z_{qp} is lower bounded by,

$$z_{qp}(\mathcal{B}) := \min_{x,X} \{Q \cdot X + c^T x \mid Ax \leq b, x \in [0,1]^N \text{ and } (x, X) \in \mathcal{B}\} \quad (2)$$

parametric on any convex set \mathcal{B} that adds valid constraints to the basic lifted formulation of the quadratic problem.

The relaxation of the nonconvex $X = xx^T$ to $X \succcurlyeq xx^T$, or equivalently $\begin{bmatrix} 1 & x^T \\ x & X \end{bmatrix} \succcurlyeq 0$ results in the semidefinite relaxation (SDP) of the quadratic problem with a positive semidefinite (PSD) restriction Sherali and Fraticelli (2002); Qualizza et al. (2012). The SDP relaxation is augmented by the reformulation-linearization technique (RLT), Anstreicher (2009), for $0 \geq x_i, x_j \leq 1$.

In the initial round of optimization, the PSD constraint is omitted. Let the solution of the initial round be X^*, x^* , to evaluate if the PSD constraint is guaranteed, eigendecomposition is performed on $\begin{bmatrix} 1 & x^{*T} \\ x^* & X^* \end{bmatrix}$, with t the negative eigenvalues of the matrix. If $t = 0$ then PSD is met, if $t \geq 1$ then PSD is violated. For every t the corresponding eigenvector v_k is used to generate the violated cutting planes $v_k^T \begin{bmatrix} 1 & x^T \\ x & X \end{bmatrix} v_k \geq 0, \forall k \in 1, \dots, t$ which are used as PSD constraints in the following round.

Qualizza et al. (2012), observed that the generated cuts are few, one cut per negative eigenvalue, and also very dense, i.e., almost all entries in v_k are nonzero, causing the reoptimization of the linear relaxation to slow down. For that reason, they introduced a heuristic for the sparcification of the PSD cuts. To overcome the

problem of few dense cuts, Baltean-Lugojan et al. (2019), introduced low-dimensional approach leading to lighter linear relaxations.

With \mathcal{P} denoting the power set of the vertex $V: \{1, \dots, N\}$ and $\rho \in \mathcal{P} (\rho \subseteq V)$ any arbitrary index subset, let $x_\rho \in \mathbb{R}^{|\rho|}$ the vector slice of x and $X_\rho \in \mathbb{R}^{|\rho| \times |\rho|}$ the submatrix slice of X . For any subset of \mathcal{P} the following semidefinite relaxation is introduced,

$$(\forall \mathcal{F} \subseteq \mathcal{P}) \mathcal{P}(\mathcal{F}) := \left\{ (x, X) \mid \forall \rho \in \mathcal{F} : \begin{bmatrix} 1 & x_\rho^T \\ x_\rho & X_\rho \end{bmatrix} \geq 0, X_{ii} \leq x_i \forall i \in \rho \right\} \quad (3)$$

A fixed cardinality $n (1 \leq n \leq N)$ is imposed on \mathcal{P} such that:

$$\mathcal{P}_n := \{ \rho \in \mathcal{P} \mid |\rho| = n \}, \text{ with } |\mathcal{P}_n| = \binom{N}{n} \quad (4)$$

A separation problem arises since there are $\binom{N}{n}$ available inequalities to create cutting planes. In Baltean-Lugojan et al. (2019), the separation problem is addressed with one of the following strategies: feasibility, or optimality and combined selection. In feasibility strategy cutting planes are selected based on most negative eigenvalues of the SDP relaxations. Optimality strategy represents the improvement of the objective function resulted by the selection of a specific cut and is calculated by an artificial neural network. Combined strategy selects optimality cuts with negative eigenvalues.

In order to make the Master problem efficient we need to find the best underestimators which will allow faster convergence of the optimization algorithm. We are imposing an extra sorting criterion for the inequality constraints added in the Master problem by Baltean-Lugojan et al. (2019). The cutting planes should not only be highly ranked based on eigenvalues(feasibility) or the improvement of the objective function (combined) but they should also be independent to each other. They rationale here is that if one cutting plane is selected, there must be a guarantee that any additional selected cutting plane added to the list of constraints provides new information for the problem. Both may be informative enough however the information they provide may be overlapping.

3. Proposed acceleration methodology and machine learning experiments

The proposed algorithm consists of 3 main steps that involve:

- Step 1: Solution of $z_{qp}(\mathcal{B})$ from Eq. (2)
- Step 2: Creation and evaluation of ordered sets of cutting plane populations
- Step 3: Screening and selection of cutting planes

The convex set \mathcal{B} initially consists of RLT constraints. Each solution of Eq. (2) is attained by the Branch and Cut algorithm (Step 1). For each on-going solution the decomposition provides for the low-dimensional PSD constraints. Each constraint is examined regarding the selected measure/condition for feasibility (Baltean-Lugojan et al., 2019); all PSD sub-problems satisfying the condition (e.g., $\lambda_\rho < 0$) enter the population (P) used for analysis. (P) consists of N-D vectors with $\frac{n}{N} \cdot 100\%$ density and its population is ordered in descending feasibility. The analysis uses both Euclidean metrics and alternative metrics introduced by means of an *affinity distance*, $d_a(x, y)$, that is formally introduced

thereafter. The complementarity within (P) is determined by clustering techniques (Step 2) such as k-means and agglomerative clustering that are deployed using fixed and/or variable numbers of clusters. Hybrid methods explored sequential implementations of Euclidean and affinity metrics.

The affinity metric $d_a(x, y)$ holds similarities to the Hamming distance and was used to evaluate differences between vectors for each sub-problem population. For $x, y \in \mathbb{R}^N$, the affinity metric is defined by,

$$d_a(x, y) := \sum_i^N [1 - g(x_i, y_i)], \text{ where } g(x_i, y_i) = \begin{cases} 1, & \text{if } x_i = y_i \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

In Step 2 the Euclidean and/or the affinity distance are applied to analyse the population of (P). In the case of the affinity metric defined by Eq (5), the distance between the vector, \tilde{x}_1 , that is corresponding to $|\lambda|_{max}$, is first computed against all other vectors in P. Vectors featuring $d_a(\tilde{x}_1, \tilde{y}) = 2$ are declared similar and subsequently removed from (P) into a cluster. The analysis proceeds with the next vector \tilde{x}_2 ; the process terminates once the full population is clustered. In Step 3, vectors are selected from the created clusters to suggest cutting planes and to update the convex hull \mathcal{B} . Within each cluster, a shortcut criterion (C1) can be used to select the highest-in-rank vector ($max|\lambda_\rho|$) while other criteria may also be applied featuring additional levels of sophistication. A mixed use of metrics has also been applied where clustering is implemented twice in Step 2 and 3: clustering using the Euclidean distance and selecting cutting planes using the affinity norm (Hybrid1) or with the use of the affinity metric first and the Euclidean next (Hybrid2).

4. Results

Machine learning experiments are performed over the test set of BoxQP studied by Baltean-Lugojan et al. (2019). For all computational experiments $n = 3$, namely the dimensionality of the cutting planes. The feasibility is the cut selection measure. The number of rounds and cuts are dictated by literature so that to compare the proposed approach with reference work (Rf1). The number of iterations in the optimization algorithm is 20; the added cuts for each round (e.g. the selection size) are set to 100. Following Baltean-Lugojan et al. (2019), the convergence limit and the final solution is achieved when cut rounds are 40 and cut/round are set to 5% of the available sub-problems. Results are evaluated based on the convergence of the algorithm and the final solution. The computational experiments are carried in python 3.5 using cplex 12.8 python API solver and scikit-learn v0.2 (Pedregosa et al., 2011) package for k-means and agglomerative clustering.

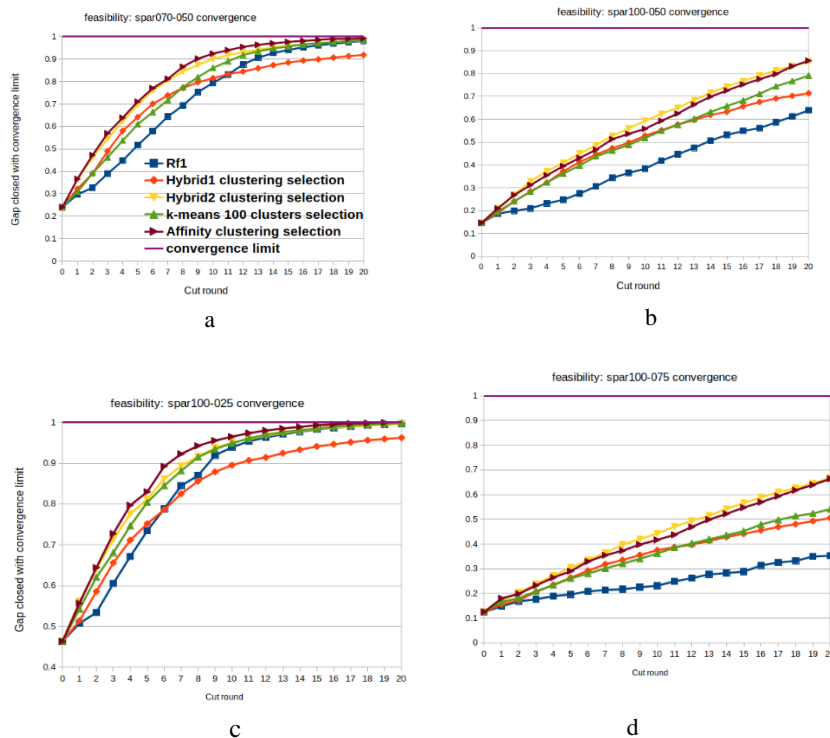


Figure 1 Gap closure between $z_{qp}(\mathcal{B})$ and the convergence limit targeted by the outer-approximation using the feasibility measure to rank the population P. The different clustering methods are embedded in the reference algorithm (Rf1) resulting in an extra sorting of P

As a general remark, the proposed approach achieves significant gap closures that are especially evident in problems of higher complexity as shown in Fig.(1b) (15-21% gap decrease) and Fig.(1d) (18-31% gap decrease). Accordingly, the approach proves particularly useful in cases where conventional methods feature large convergence gaps. To a lesser degree, improvements are achieved in the lower-density and lower-dimensionality problems shown in Fig.(1a) and Fig.(1c). The convergence performance in using the affinity metric in Step 2 performs with similar trends to Hybrid2. The use of affinity in Step 3 performs with similar trends to Hybrid1 and proves an inferior option suggesting that the use of the affinity metric is a much stronger choice when applied in earlier stages of analysis. The shortcut criterion (C1) is generally found to outperform other sophisticated options studied in this work.

Clustering based on $d_a(x, y)$ provides remarkable improvements not only over the reference algorithm but also over the conventional clustering method of k-means. Furthermore, the convergence of the optimization solution follows a consistent trendline in all cases, demonstrating that the $d_a(x, y)$ is a suitable metric for a wide range of QP instances. The emphasis of the paper has been to exclusively reduce the integrality gap. The reduction has subsequently affected CPU times with an increase reported either in the use of the affinity metric or, primarily, the use of clustering algorithms. The affinity metric used for clustering increased CPU times by 5-60 times. The largest delays are observed in small complexity instances where the original algorithm has fast convergence. Over-the-self clustering algorithms delayed the algorithm by 7-25 times with computational time increasing with problem complexity. Given the current emphasis of the paper, there is significant scope to improve such times in future developments.

5. Conclusions

The use of data analytics by cutting plane algorithms has proved an essential drive for improvements in global optimization. The conclusion is demonstrated with a study of a wide range of optimization problems. Results demonstrate significant room to reduce the cutting plane populations generated by the conventional algorithms. The affinity metric introduced in Eq (5) proved a promising approach to handle cutting plane redundancies and to support clustering. Reinforcing the methodology by selection criteria biased on both feasibility and complementarity performance has enabled improvements in the original algorithm. The use of the affinity metric has generally outperformed general-purpose clustering (e.g. k-means, agglomerative clustering) while in hybridized forms the metric led to better results though without impressive improvements. Remarkably, the clustering approach has a greater impact in high-dimensional & medium-to-high density problems indicating a promising and encouraging scope to improve existing algorithms.

6. Acknowledgement

The authors are grateful to generous assistance and collaboration with Dr Ruth Misener and Radu Baltean-Lugojan from Imperial College. The sharing exchanges have been instrumental to test developments for several problems also to compare the proposed algorithms with established benchmarks and the state-of-the-art. None of that would be possible without their assistance. The authors acknowledge financial support by the H2020-EU.1.3.3. Grant RENESING II (H2020-778332).

References

- K. M. Anstreicher, 2009. Semidefinite programming versus the reformulation-linearization technique for nonconvex quadratically constrained quadratic programming. *Journal of Global optimization* 43, 471–484.
- R. Baltean-Lugojan, P. Bonami, R. Misener, A. Tramontani, 2019. Scoring positive semidefinite cutting planes for quadratic optimization via trained neural networks.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. VanderPlas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12.

A. Qualizza, P. Belotti, F. Margot, 2012. Linear programming relaxations of quadratically constrained quadratic programs. *Mixed Integer Nonlinear Programming* 154.

H. Serali, B. Fraticelli, 2002. Enhancing rlt relaxations via a new class of semidefinite cuts. *Journal of Global Optimization* 22, 233–261.

Towards a Digital Twin: A validated Digital Model of a Pilot Scale Bioreactor producing Green Fluorescent Protein (GFP)

Ole Laustsen Lave^a, Carina L.Gargalo^a, Ryan Barton^b, Christoph Bayer^c, Isuru A. Udugama^a and Krist V. Gernaey^{a,*}

^a *Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Lyngby, Denmark*

^b *Biomufacturing Training and Education Center (BTEC), Raleigh, North Carolina State University, USA*

^c *Department of Process Engineering, TH Nuernberg, Nuernberg, Germany*

kvg@kt.dtu.dk

Abstract

Industrial scale bio-manufacturing processes are transitioning towards Industry 4.0, which promises to change the competitive landscape. As part of this (r)evolution, there is an increased focus on developing digital twins of critical operations, such as fermentation processes. A key economic driver for the enthusiastic embrace of digital twins by this industry is the promise of increased process efficiencies and resource utilization, mainly through improved plant operations. However, for digital twins to move from the “current hype” to a beneficial solution in the industry, some challenges must be overcome. This manuscript attempts to identify these challenges and opportunities by developing a Digital Model of a Pilot Scale Bioreactor producing Green Fluorescent Protein (GFP) currently used for Teaching purposes.

Keywords: Digital Twins, Validated models, Bioreactors, Fermentation.

1. Introduction

The digitalization movement in bio-manufacturing promises improved process operations through informed decision-making (Udugama, Gargalo, et al., 2020). One key pillar in the discussion of digitalization is the concept of Digital Twins. Digital Twins, through the creation of virtual plants, accurately represent the behavior, including the “look and feel” of a physical plant (Lukowski et al., 2019).

In Process Systems engineering, the development of mathematical/process models has been a standard practice for decades. There are remarkable examples of benchmark simulations such as the MIT-Novartis Continuous pharmaceutical production process (Benyahia et al., 2012) and the Benchmark Simulation Model (BSM) for wastewater treatment plants (Rosen et al., 2006) being amongst a countless number of mathematical models. In the bio-manufacturing industry, mathematical models have also been used for decades and continue to be a popular alternative to physical experiments when it comes to design and optimization of different processes and unit operations (Mears et al., 2017; Narayanan et al., 2020). Today, these mathematical models are often referred to as Digital Model and make up the core “prediction engine” of a Digital Twin (Lopez et al., 2020).

The Accelerated Innovation in Manufacturing Biologics (AIM-Bio) project is a joint initiative between North Carolina State University and the Technical University of Denmark (DTU) established to create a “world-class program in bioprocess research and development and workforce training that focuses on products and technologies for the future of biopharmaceutical manufacturing”. As a part of this initiative, the (PROSYS) center at DTU collaborates with BTEC at North Carolina State University to develop digital models of their pilot scale bio-manufacturing facility, primarily used for education purposes. The overall goal of this endeavor is to develop Validated Digital models that can be in the form of Operator Trainer Simulators to enhance students' learning.

As in many bio-manufacturing processes, fermentation is a central part of the pilot scale bio-manufacturing operations. BTEC produces GFP as a model protein through an *E. coli* based fermentation process. More specifically, the process employs an induced fermentation batch using *E. coli* (BL21(DE3)) with a plasmid pET17-b to produce GFPuv with a T7 promoter (IPTG) in a Sartorius Stedim 300-L Bioreactor. This bioreactor is equipped with a standard set of sensors such as temperature, pressure, pH, and off-gas measurements. An industrial distributed control system (DCS) is used to operate the reactor, while a process historian is used for logging the available process data. Moreover, advanced analytical equipment is used for off-line monitoring of the batch's progression, similar to standard industrial practice.

This contribution aims to understand the level of detail and the necessary auxiliary elements (e.g., GUI) that a mathematical model of a fermentation process needs to have in order to fulfill the role of being a standalone Virtual Plant. Furthermore, this work will also enable us to better understand the level of detail/fidelity, and additional complexity that a mathematic model must incorporate so as to become a Digital Twin.

2. Digital Model Development

As an initial push towards developing a fully-fledged digital model, a macro-kinetic process model that accounts for overflow metabolism and the associated acetate excretion and substrate consumption inhibition was set up in Matlab Simulink. The ordinary differential equations describing the mass balances were set up in accordance with earlier models reported by Xu et al. (Xu et al., 1999), Anane et al. (Anane et al., 2017) and Pham et al. (Pham et al., 1998).

2.1. Macro kinetic model formulation

The set of ordinary differential equations uses five inputs and a set of auxiliary algebraic equations describing the intracellular kinetics: biomass formation and oxygen consumption, substrate, and acetate consumption. The substrate is primarily glucose and is modded accordingly, while 15 g of ampicillin are added as antibiotics are added at BTEC reactor to control unwanted bacteria growth, this is not modelled in the digital model. There are six output/state variables, which are: volume, biomass concentration, extracellular substrate concentration, acetate concentration, dissolved oxygen, and mass of GFP produced.

The inputs used in the model were chosen due to the fact that they are important and manipulatable physical variables: feed flow rate (L/h), feed glucose concentration (g/L), temperature (C), headspace pressure (atm), volumetric mass transfer coefficient. The Matlab model developed by Anane et al. (2017) was used as the starting point for developing a Digital Model. And the temperature was kept at a constant 30 °C.

2.2. Structural Modifications to the Model

Two main structural changes were carried out in this work. First and foremost, the model was converted into a Matlab Simulink operated model. This allows the model to be run similar to a prediction engine, where outputs and inputs can be altered externally in real time, lending it to be used as a Digital Model. Secondly, along with the process variables, the model was extended to also have other key state variables as output, in case that they can be monitored at the pilot plant in the future.

2.3. Introduction of GFP

The mechanistic model of the GFP production after the induction with IPTG has been developed using standard biochemical yield kinetics (Figure 1). The kinetics are based on how much substrate is available for product formation and given as total mass because GFP is formed intercellularly with minor extracellular concentration as a result of lysis and only being sampled after harvesting the batch. After induction, the cell production is also reduced, assuming that all induced cells only produce GFP and do not reproduce. The switch from the growth phase to the production phase is modeled based on mixing experiments performed at BTEC on the actual reactor. This results in a nearly instantaneous switch from growth to production due to the reactor's short mixing times. IPTG can be toxic to cells when used at high concentrations causing cell stress. However, the concentration used at BTEC is optimized so that this problem is negated.

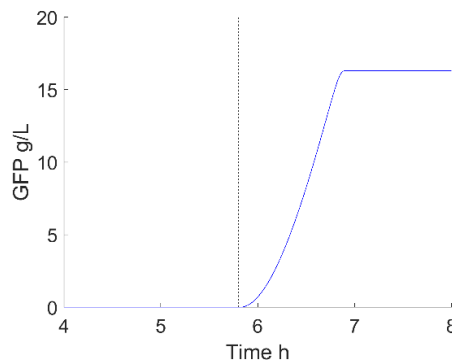


Figure 1. Current modeled GFP increase with IPTG induction at 5.8 hours (dotted line).

3. Challenges and Opportunities

3.1. Modelling the Cell line

A key learning from the initial endeavor was that an accurate digital representation of the process requires us to go beyond the standard mathematical modeling and explore the cell factory. To this end, a digital model of the process seems to require a multi-scale approach. Of particular interest was to understand how IPTG influences the cell behavior so that the cell is induced to produce GFP. From a chemical engineering perspective, this type of process behaviour can be modeled using a kinetic equation activated when IPTG is introduced into the reactor. However, based on BTEC's observations, this concept of a hard production switch is incorrect. Instead, the observations suggest that the cell growth rate is significantly reduced at that induction time point. Even before induction, some "minor" GFP production rate is expected. Moreover, the reduction in growth rate is a

function of the concentration of IPTG introduced into the reactor. Considering all these aspects, efforts are underway to model the cell itself, where parameters such as reduced growth rate and GFP production rate present in the mechanistic model will be determined by solving the cell model at appropriate time intervals during a batch.

3.2. Lack of Process Data

A critical limitation that had to be overcome in the development of the digital model was the lack of process data that was initially available for parameter estimation and model validation purposes. This is another key aspect in transitioning a mathematical process model into a digital model.

The main benefit of adopting the *E. coli* growth model developed by Anane et al. (Anane et al., 2017) is that the process parameters related to biomass growth have been validated against an actual fermentation process, although in a lab-scale setup. To identify if the process parameters suggested in (Anane et al., 2017) yield an accurate digital representation of the fermentation process at BTEC, a Monte-Carlo analysis on the process model was performed, where key process parameters including qS_{max} and Y_{em} were varied. The results of this analysis were plotted together with the single process data set that was available. They are displayed in Figure 2 (a) and (b).

Figure 2 shows that the data points and the process model have a similar but slightly shifted growth curve when only varying a single parameter from values based on lab-scale data (Anane et al., 2017). Note however that a lack of data points during regular periods of operation means that further process data is required to make a definitive conclusion. Compared to biomass growth rate data, only a final value of GFP produced is available for GFP production. This is due to the fact that GFP is stored intracellularly, and by following the current standard operating procedure (SOP), the amount of GFP produced in a bio-reactor is measured by processing the biomass into cell paste using a sample taken at the end of the process.

A clear plan has been put in place to improve the quality of the data to be gathered in the future. Firstly, the volume of data gathered will be improved, as multiple experimental runs are currently planned and protocols in place to collect the process data generated. Specific experiments are also designed to capture the GFP production rate through running trial experiments where batches are not induced. This will allow for substrate consumption ratio for biomass production vs. protein yield to be calculated throughout a batch.

3.3. Validation vs. Complexity

Based on the current trends and interpretation of digital models, which, in short, need to be a virtual representation of the physical process, only a multi-scale model that integrates the “cell factory” and the reaction kinetics would be sufficiently complex and accurate. To achieve this, a digital model of such a process should include this multi-scale phenomenon and even consider fluid dynamic dependencies due to the potential heterogeneity present within a 300 L bioreactor. Such a multi-scale model is more complicated than the current *off-the-shelf* solution that employs a set of Ordinary Differential Equations to describe a bioreactor's time-dependent behavior. If the necessary process data is available, such a multi-scale model would potentially be much more accurate in mimicking an actual bioreactor's behavior.

However, from a practical standpoint, there are significant limitations in gathering process data. For example, other than temperature, pH, and flow information, BTEC currently requires manual operator intervention to collect and analyse off-line measurements to determine dissolved oxygen and glucose concentration. Furthermore, capturing effects such as pH gradients is impractical as only a single pH sensor is present

in the reactor. Another important consideration is the data preprocessing (e.g., GFP concentration measurements), which is tedious and time-consuming. All in all, a decision needs to be made regarding the level of detail to be included in the final digital model. There are two alternatives: (i) go back to the current approach (ODEs representing the reactor); or (ii) gather the necessary data at the cell level, as well as potentially other aspects (e.g., bioreactor fluid dynamics), and integrate them into the model. In pursuing the second approach, concepts such as high-throughput screening can be used to generate data related to the cell model, which then can be used for validation purposes at the cell level. Novel measurement technology such as free following sensors (Gargalo et al., 2020) can potentially be used to gather information on the bioreactor's fluid dynamics. Moreover, fluorescence spectroscopy can be used to track GFP production in real-time, which would enable the validation of both cell and reactor models.

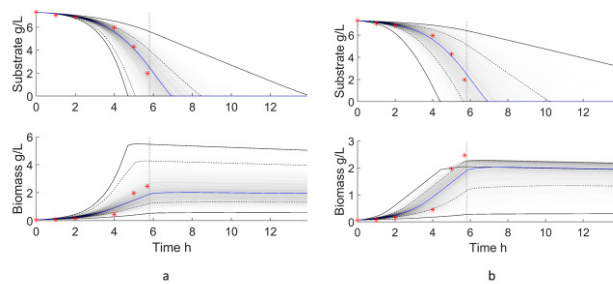


Figure 2. Monte-Carlo simulation: 2000 samples varying (a) Y_{em} (b) qS_{max} with induction time. 95% confidence interval (CI) is shown as dotted lines. Outermost incident (solid line), process data (*).

3.4. The role of a Graphical User Interface (GUI)

A GUI plays a crucial role in transitioning a mathematical process model into a digital model. This is because a GUI makes a mathematical process model have the “look and feel” of a physical process. In practice, many physical processes in the bio-manufacturing industry are often controlled through a controller GUI. To this end, the development of a graphical user interface allows students to explore a mathematical model's behavior in a similar way to operating a real process (Udugama, Germaey, et al., 2020).

4. Conclusions

A Digital Model of the BTEC Pilot Scale Bioreactor producing Green Fluorescent Protein (GFP) has been developed and implemented in Matlab Simulink. Furthermore, by analyzing the GFP production details, it was concluded to develop a multi-scale approach to increase the digital model's fidelity. To this end, data on different scales, e.g., at the cell level and bioreactor fluid dynamics, will be collected and reflected in the model. In summary, this contribution laid the foundation in developing a detailed mathematical model with the necessary auxiliary elements (such as a GUI) to fulfill the role of being a standalone virtual plant both for in education and in industrial scale bio-manufacturing processes.

5. Acknowledgement

The authors wish to acknowledge the financial support provided Novo Nordisk Foundation (AIMBio) project (Grant number NNF19SA0035474).

References

- Anane, E., López C, D. C., Neubauer, P., & Cruz Bournazou, M. N. (2017). Modelling overflow metabolism in *Escherichia coli* by acetate cycling. *Biochemical Engineering Journal*, *125*, 23–30. <https://doi.org/10.1016/j.bej.2017.05.013>
- Benyahia, B., Lakerveld, R., & Barton, P. I. (2012). A Plant-Wide Dynamic Model of a Continuous Pharmaceutical Process. *Industrial & Engineering Chemistry Research*, *51*(47), 15393–15412. <https://doi.org/10.1021/ie3006319>
- Gargalo, C. L., Udugama, I., Pontius, K., Lopez, P. C., Nielsen, R. F., Hasanzadeh, A., Mansouri, S. S., Bayer, C., Junicke, H., & Gernaey, K. V. (2020). Towards smart biomanufacturing: a perspective on recent developments in industrial measurement and monitoring technologies for bio-based production processes. *Journal of Industrial Microbiology & Biotechnology*. <https://doi.org/10.1007/s10295-020-02308-1>
- Lopez, P. C., Udugama, I. A., Thomsen, S. T., Roslander, C., Junicke, H., Mauricio-Iglesias, M., & Gernaey, K. V. (2020). Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation. *Biofuels, Bioproducts and Biorefining*, *bbb.2108*. <https://doi.org/10.1002/bbb.2108>
- Lukowski, G., Rauch, A., & Rosendahl, T. (2019). The Virtual Representation of the World is Emerging. In *Future Telco* (pp. 165–173). https://doi.org/10.1007/978-3-319-77724-5_14
- Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., & Gernaey, K. V. (2017). Mechanistic Fermentation Models for Process Design, Monitoring, and Control. *Trends in Biotechnology*, *35*(10), 914–924. <https://doi.org/10.1016/j.tibtech.2017.07.002>
- Narayanan, H., Luna, M. F., von Stosch, M., Cruz Bournazou, M. N., Polotti, G., Morbidelli, M., Butté, A., & Sokolov, M. (2020). Bioprocessing in the Digital Age: The Role of Process Models. *Biotechnology Journal*, *15*(1), 1–10. <https://doi.org/10.1002/biot.201900172>
- Pham, H. T. B., Larsson, G., & Enfors, S.-O. (1998). Growth and energy metabolism in aerobic fed-batch cultures of *Saccharomyces cerevisiae*: Simulation and model verification. *Biotechnology and Bioengineering*, *60*(4), 474–482. [https://doi.org/10.1002/\(SICI\)1097-0290\(19981120\)60:4<474::AID-BIT9>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0290(19981120)60:4<474::AID-BIT9>3.0.CO;2-J)
- Rosen, C., Vrecko, D., Gernaey, K. V., Pons, M. N., & Jeppsson, U. (2006). Implementing ADM1 for plant-wide benchmark simulations in Matlab/Simulink. *Water Science and Technology*, *54*(4), 11–19. <https://doi.org/10.2166/wst.2006.521>
- Udugama, I. A., Gargalo, C. L., Yamashita, Y., Taube, M. A., Palazoglu, A., Young, B. R., Gernaey, K. V., Kulahci, M., & Bayer, C. (2020). The Role of Big Data in Industrial (Bio)chemical Process Operations. *Industrial & Engineering Chemistry Research*, *59*(34), 15283–15297. <https://doi.org/10.1021/acs.iecr.0c01872>
- Udugama, I. A., Gernaey, K. V., Taube, M. A., & Bayer, C. (2020). A novel use for an old problem: The Tennessee Eastman challenge process as an activating teaching tool. *Education for Chemical Engineers*, *30*, 20–31. <https://doi.org/10.1016/j.ece.2019.09.002>
- Xu, B., Jahic, M., & Enfors, S.-O. (1999). Modeling of Overflow Metabolism in Batch and Fed-Batch Cultures of *Escherichia coli*. *Biotechnology Progress*, *15*(1), 81–90. <https://doi.org/10.1021/bp9801087>

Automated Box-Jenkins Methodology to Forecast the Prices of Crude Oil and Its Derivatives

Ahmet Can Serfidan,^{a,b}, Gurkan Ozkan,^a Metin Türkay,^b

^a*TUPRAS, Address, Petrol Cd. No:25 D:No:25, Turkey*

^b*Koc University, Rumelifeneri, Sariyer Rumeli Feneri Yolu*

Abstract

Developing forecasting models that incorporate that external parameters in addition to past data for crude oil and derivatives are a challenging task since it is highly dependent on economic, geographical, and political issues. However, forecasting the prices is very important for strategic planning and oil refineries' operational decisions. This paper presents an automated tool to predict crude oil prices and their main products by applying Box-Jenkins methodology for the next two months at the beginning of each month in a rolling horizon manner. The resulting forecast is shared with related departments to develop their production plans accordingly. We show that improved accuracy with this forecasting approach is beneficial in any planning and decision-making process and increases profit.

Keywords: price forecasting, crude oil, Cox-Jenkins, SARIMA

1. Introduction

Refineries are continuous-flow manufacturing facilities that have a considerable effect on the global economy. In concise terms, petroleum refineries can be summarized as processing crude oil into a more valuable product by using chemical and physical operations. Production planning and process scheduling are among the most important activities of refineries for maximizing profit. Several software packages optimize the refinery operations; these tools need various data such as feedstock prices and products, demands for all products and byproducts, and supply limits [1]. Since the usual planning period covers the next two months, forecasting the future prices of products is critical.

Linear programming-based planning systems determine the production plans for all units in the refinery. The general approach for the production and scheduling process has two phases: (i) Determining monthly rolling plans for crude selection and conducting refinery operations in line with forecasting demands. (ii) based on the monthly plans, implementing short-term plans for finding operational strategies regarding crude availability, product delivery, operational and logistic constraints, and economic issues. [2]

As well as the price of the products itself, the price bracket between related products is a key parameter for making decisions on production planning. For example, the price difference between the naphtha and gasoline shows the refineries how to add value to the upper distillate; either produce gasoline or sell naphtha and reformat separately. The price range of diesel and aviation fuel has a strong impact on the refinery's profitability. The price bracket between low and high sulfur diesel determines the refineries'

profitability of desulphurization units. Furthermore, the prices of low and high sulfur fuel oils are important for determining the quantities of additives during the blends of different types of marine fuels.

This paper presents an automated approach to generate price data forecasting by applying Box-Jenkins methodology to historical price data. As an industrial implementation, the objective is that this monthly-generated data set becomes an input to planning activities in operations of a refinery operated by Tüpraş. The improved accuracy of forecasted data allows us to improve robustness in scheduling.

2. Methodology

Box Jenkins ARIMA methodology refers to a set of procedures for identifying, fitting, and checking ARIMA models with time-series data [3]. Although there are many works related to Box-Jenkins ARIMA methodology, a completely automated Box-Jenkins SARIMA tool does not exist. The objective is to create an automated and comprehensive forecasting approach that generates price predictions for an oil company.

There are two different approaches to monitoring crude oil and derivatives' price: Ratio and crack margin. The ratio is the division of a product price by the crude oil price, whereas crack margin is the price difference of the product and crude oil. Instead of the ratio, monitoring a product's crack margin is much more meaningful because it does not fluctuate rapidly as the crude oil price changes.

In this study, crack margin data is used for modeling because its variance is lower than the ratio. Also, crack margin is a function of both Brent crude oil price and each product's price.

2.1 Data Preprocessing

We start by identifying whether the data is stationary or not [4]. The most usual approach is to analyze the series plots and ACF plots [5]. Since the aim is to develop a full-automated system, we use various tests based on the definition of stationarity.

The corresponding tests and their stopping criterion conditions are as follows:

- (i) Stationary test: Count the number of ACF lags out of the critical level. If the result is more than $(0.2) \times (\text{total number of lags})$, the series is not stationary,
- (ii) Trend stationary test: We use KPSS (Kwiatkowski, Phillips, Schmidt, and Shin) test for this purpose. If the data is stationary, it should not contain trend. [6]
- (iii) Mean stationary test: We divide the data into two pieces, and by using t-test, we check whether the mean of the series of time-dependent or not.
- (iv) Randomness test: We use the Ljung-Box Q test over the ACF of the series. The aim is to check whether series have autocorrelation for a fixed number of lags.
- (v) Augmented Dickey-Fuller Test: We use the test to check whether time series data are stochastically stationary or not.

The algorithm executes the tests mentioned above automatically, and if there is at least one test that shows data is nonstationary, then we take the first derivative of data, namely order of differencing is increased by one (by default, it is zero). These comprehensive tests can easily catch nonstationary data. Nevertheless, we still want to use graphical analyses to inform the user.

At the end of this section, the algorithm determines whether the data is stationary or not. If the data is nonstationary, then it calculates the order of differencing. Original data, box plot of monthly divided data, historical data around the desired month, monthly average data, ACF and PACF plots of monthly data, and normality plots will be shown to the user to justify the findings.

2.2 Model Identification and Diagnostic Checking

To increase the selection process's robustness, we model all possible combinations. We use the fact that orders of AR and MA are usually less than and equal to two. We also increased the upper limit of the AR limit to 7. However, model results do not become more desirable than increasing model complexity. Thereby, we chose to go over all 972 combinations of SARIMA models (nonseasonal differencing order is fixed in the first step, so 972 is the number of all different combinations).

For all 972 different SARIMA models, values of coefficients are calculated by using maximum likelihood estimation. The estimated parameters for the current model is forced to satisfy the following conditions:

- (i) Parameter estimation should converge
- (ii) MA coefficients should be invertible
- (iii) AR coefficients should be stationary

If one model fails to meet these conditions, we drop this model and do not further investigate or use it in the forecasting.

Having estimated coefficients, we check goodness of the fit for each model. This part is diagnostic checking of all available SARIMA models and particularly vital in the Box-Jenkins method. All available SARIMA models are assessed with the following tests:

- Significance of parameters: All the estimated parameters should be significantly different from zero. We use p values for each parameter of a model for this purpose. If the corresponding model has at least one parameter with a higher p -value than 0.1, we diagnose it as having an insignificant parameter.
- Testing heteroscedasticity: The standard deviation of standardized residuals should be varying at a fixed level. If it is not non-constant, then there is a heteroscedasticity issue. In order to detect this issue, we applied the Goldfeld-Quandt test. We tested whether the sum of squares in the first third of the sample is significantly different from the sum of squares in the last third of the sample.
- Testing normality: We tested for the normality of standardized residuals. For this purpose, we used the Jarque-Bera normality test. Standardized residuals should have an approximately normal distribution. Otherwise, it violates one of the Box-Jenkins assumptions.
- Testing serial correlation: We tested whether there is a statistically significant correlation between standardized residuals. The residuals should pose no autocorrelation. Otherwise, there is room for improvement. We applied Ljung and Box and Durbin-Watson tests to check whether residuals are uncorrelated.

When a model passes all these tests, it goes to the next stage, which is the final stage for selecting the best model.

2.3 The Best Model Selection

This section presents each product's best model and is used for forecasting. For this purpose, we take the first 85% of the data as training data and the rest of it as the test data.

We have approximately 120 data points, so the test data window is around 12 data points, namely one year. The best model should capture the trend of last year with minimum error. There are many statistics to check this model accuracy; we used mean squared error for this purpose.

After splitting the dataset into training and test sets, we train all candidate models with training data, and new models are obtained. Then, each time step of the test set is iterated. Namely, candidate models predict one period ahead, and after that, the actual value of the predicted time is given to the model. While preserving the model parameters, new coefficients are estimated with this new value, and again one period ahead prediction is carried out. "The iterative approach allows a new ARIMA model to be trained each time step. Prediction is made each iteration and stored so that all prediction can be compared to actual values at the end of the test set, and an error score is calculated. In this case, a mean squared error score is calculated". The best model has the least error score among the candidate models.

We implemented all steps of the model in python 3.6. Built-in Python libraries carry out some statistical tests. Anyone can use this comprehensive and complete automated tool for any time series data. However, the users must consider the prediction horizons, i.e., daily, weekly, monthly, etc. After this specification, the tool provides forecasting values for the desired prediction horizon

The accuracy of the forecast is tested between January 2017 and May 2018. The approach in testing is quite similar to selecting the best model. First, only the data between January 2009 and December 2016 is given to the model. It selects the best model as discussed in the previous sections. After it predicts the next month, we calculate the monthly average of that predicted month. In the next iteration, the algorithm has data between January 2009 and January 2017, and the code again carries out all stages. The program in any iteration might identify a completely different SARIMA model. We calculate all prediction errors for all months and all products by using the mean average percent error

3. Results and Discussion

To test the forecasting tool's accuracy, we used the actual oil and its derivatives price data published every day. We gather the data daily; however, we used monthly average data since the aim is to predict prices monthly. Moreover, a daily variation of prices is too high to predict, and the autocorrelation function decays so slowly that even after many differencing, it is not stationary. In this study, seven major products and Brent crude oil prices are forecasted for sixteen months ranging from January 2017 to June 2018. At the beginning of every month, the prices are predicted for the next month and the two months ahead. Table 1 shows the average forecasting accuracy for each product and Brent crude oil. As a general conclusion, the prediction for $n+1^{\text{th}}$ month is better than the one for $n+2^{\text{th}}$ as expected since the forecasting power decreases as the time period increases. The prediction accuracy of gasoline is around 90% for both periods. When middle distillate products of aviation fuel, low and high sulfur diesel, are considered, the accuracy ranges from 91% to 95%. The application has some development areas when the bottom products are considered since the accuracy is between 68% and 86%. Although Brent crude oil price is hard to forecast due to factors that cannot be modeled easily, the tool performs well, ranging from 93% to 96%. The forecast means absolute error for naphtha is 1.2. These calculations are done using mean absolute error (MAE) except naphtha because

mean average percentage error (MAPE) is used for naphtha since the magnitude of error is small.

Moreover, to compare our SARIMA Box Jenkins methodology, we implement Single Exponential Smoothing, Double Exponential Smoothing, and Triple Exponential Smoothing (Holt-Winters Method) for one period ahead prediction. The performance measure is the same. Mean absolute average error and the same iteration method are applied (first the model predicts, then the corresponding data is introduced). The following models' optimal parameters are found by using a similar grid search algorithm.

Table 1. Average Forecasting Accuracy for 16 Months by Stepwise Modeling for SARIMA Box Jenkins

	(n+1)th Month Accuracy	(n+2)th Month Accuracy
Gasoline 10 ppm	91,1%	90,3%
Aviation Fuel	95,0%	92,4%
High Sulphur Diesel	93,8%	93,4%
Diesel 10 ppm	93,6%	91,2%
Fuel Oil 1%	78,7%	68,3%
Fuel Oil 3.5%	85,9%	75,0%
Brent Crude Oil	95,9%	93,1%

Table 2. Average Forecasting Accuracy for 16 Months by Stepwise Modeling with Other Models

	Single Exponential Smoothing	Double Exponential Smoothing	Holt-Winters	ARIMA	SARIMA
Gasoline 10 ppm	89.9%	89.1%	88.1%	90.8%	91.1%
Aviation Fuel	93%	93.2%	92.7%	92.2%	95.0%
High Sulphur Diesel	93.8%	93.5%	91.0%	93.2%	93.8%
Diesel 10 ppm	93.4%	93.4%	93.2%	91.7%	93.6%
Fuel Oil 1%	73.7%	73.9%	78.3%	73.3%	78.7%
Fuel Oil 3.5%	84.8%	84.1%	85.8%	84.7%	85.9%
Brent Crude Oil	94.9%	95.2%	93.5%	95.2%	95.9%

Comparisons of actual and forecasted prices are displayed to the user as shown in the following figure:

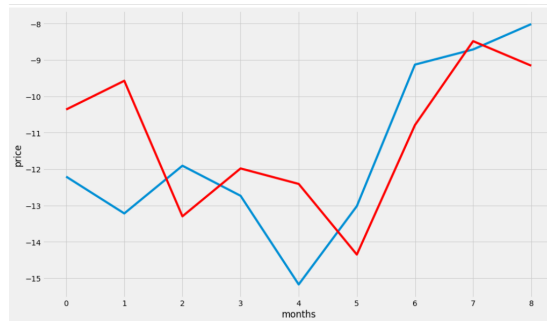


Figure 1 Predicted Price (red) vs Actual Price (blue).

4. Conclusions

This study proposes a new approach to achieve more profitable production planning. We extended the Box-Jenkins method by adding a fully automated and comprehensive grid search algorithm. This approach substitutes human judgment with statistical tests and trying all combinations without giving up any critical stage of Box-Jenkins methodology. This tool provides the numbers and very powerful graphs that justify those numbers. Since it is an interactive tool, the user may accept or reject the proposed results, mainly based on the graphs. We implemented using python 3.6 and made an easy to use interface. The tool only needs the original data, time window, and predicted values. To test model accuracy, we applied it to forecasting petroleum products' price and Brent crude oil price. Generally, our forecasting accuracy is between 90-98%, with one exception, the accuracy of Fuel Oil forecasts is close to 75-85%. When we compare the Box-Jenkins Full Automated SARIMA model to other well-established model types, it is seen that our model has better forecasting accuracy than the other types. This fully automated tool is deployed in Tüpraş Foreign Trading Department. Price prediction dedicated from this tool is one of the critical import inputs they use during the crude selection and planning processes

References

- [1] Göthe-Lundgren, Maud, Jan T. Lundgren, and Jan A. Persson. "An optimization model for refinery production scheduling." *International Journal of Production Economics* 78.3 (2002): 255-270.
- [2] Refinery Short-Term Scheduling with tank farm, inventory and distillation management: An integrated simulation-based approach, by George Chryssoularis, Nikolaous Papakostas, Dimitris Mourtzis.
- [3] Box Jenkins ARIMA methodology refers to a set of procedures for identifying, fitting, and checking ARIMA models with time-series data.
- [4] Peter J. Brockwell, Richard A. Davis, Introduction to Time Series Analysis (3rd ed.) Springer 2016, p. 157-193
- [5] Time Series: Theory and Methods, p 273-329 Peter J. Brockwell Richard A. Davis
- [6] Kwiatkowski D, Phillips PCB, Schdint P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics* 1992; 54(1) 159-178

Systematic Modelling of Transport Processes across Interfaces

Robert Pujan^{a,b} and Heinz A. Preisig^b

^a*DBFZ Deutsches Biomasseforschungszentrum gemeinnützige GmbH, Torgauer Straße 116, 04347 Leipzig, Germany*

^b*NTNU Norwegian University of Science and Technology, Høgskoleringen 5, 7491 Trondheim, Norway*
robert.pujan@dbfz.de

Abstract

This paper discusses the implementation of interfacial transport in the systematic modelling methodology by Preisig (2012) and how to lecture the approach as well as its implications to engineering students. We first elaborate on the theoretical concept of interfaces before abstracting them into topologies that allow for comprehensible as well as precise mathematical assessments of the transport processes involved. Thereby, any interfacial transfer is rooted in its fundamental parts of transport towards, across and away from the interface. These steps can be assessed by only a handful of algebraic equations from an ontology, which covers the entirety of physical concepts, state derivatives, and conservation principles. Due to this, interfacial transport can rapidly be implemented in process models, significantly aiding subsequent model simplification as well as model fitting.

Keywords: process modelling, ontology, topology, dynamics, diffusion

1. Motivation

Many processes like extraction, adsorption and membrane filtration depend on mass, momentum and energy transfer across interfaces. As shown by Preisig (2014b), these processes can conveniently be abstracted into a topology that is a network of capacities; non-overlapping, finite volumes that only interact with each other by the means of directed arcs. Arcs that are flows of extensive quantities (like mass) driven by gradients of intensive properties like concentration, temperature or pressure.

The convenience of topology modelling can further be intensified by providing an expert ontology - a comprehensive collection of fundamental definitions and relations taken from the application-relevant scientific roots, and by implementing all of this in the highly automatised modelling suite *ProMo* [Elve and Preisig (2019); Preisig (2020)]. However, from lecturing engineering students it transpired that further emphasis has to be placed on the concept of interfaces and its realization within the methodology.

Consider the case one is asked to model an adsorption column: Most likely, the initial process topology drawn would be similar to the one depicted in Figure 1. A feed F enters column C , in which the adsorption of component i on adsorbent A proceeds before leaving the unit as the processed product P . The process is defined by the conditions at its boundaries to the environment. Since we assume for now that the environment is infinitely large compared to our process plant, thus not affected by the process, we depict both the feed F and the product tank P as reservoirs (semi-circles)

that are infinite volumes of constant properties. On the other hand, C and A are depicted as finite, distributed capacities (ellipses), indicating that in these volumes intensive properties are dependent on the location of observation. The process of adsorption is implemented by flows of mass (black arcs), heat (dotted red arcs) and volume work (dashed blue arcs) across a boundary (black bar) between C and A . Boundaries are surfaces without any extensive quantity. Usually, they indicate discrete changes between two capacities, like fluid | solid interphases in the sketched adsorption process. Despite its low granularity and abstract nature, this topology appropriately outlines the process in question. However, as we noticed when lecturing students, most people tend to stick with their first, rather rough topology when continuing with the mathematical modelling, not realising that such a low degree of granularity will cause complications in further modelling. In the example of Figure 1, students would often go on to try implementing all transport terms in capacity C , while A is only contributing the adsorption kinetics. Meanwhile, the boundary is just left as a shallow symbol of intent in the topology, not contributing to any mathematical modelling despite being the most-important part of the process; the process-limiting interphase. This approach usually ends in the numerical bottleneck that the system in C will be a partial-differential equation with at least three dependencies:

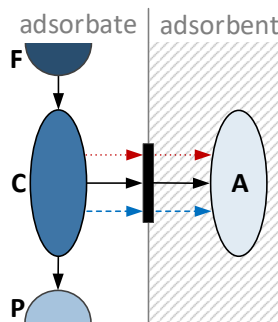


Figure 1: A first topology for any arbitrary adsorption process

The time t , the radial coordinate x and the axial coordinate y . Furthermore, transport processes from C to A are compressed into one misleading term in this approach, neglecting the potentially different nature of connected transports in the direction of x and neither providing nor showing any understanding of the inner workings at hand.

Therefore, we strongly advocate for a systematic approach when modelling processes that exceed interfaces. By expanding on the discussions in Pujan and Preisig (2020), this paper presents the proposed methodology while discussing its implications and contributions to the design and communication of comprehensive but comprehensible process models.

2. Transport processes across all sorts of interfaces

In the following, we refer in our terminology to *interphases* for interfaces between two discrete phases, whereas *intraphases* denominate interfaces without any phase change, like the discrete difference between a flowing fluid in a pipe and its almost-stagnant fluid film at the pipe walls.

When an interphase is introduced to a system, fluids form sub-systems on both sides. These adhere considerably different regimes than the fluid bulks. Those sub-systems are conventionally called films. However, as we will discuss later, the term *film* can be misleading in this context. For now, we define as film the volume that manifests an explicit change of properties on the interphase's normal coordinate, as depicted in the examples of Figure 2. We limit the scope of our remarks on material transport, although the analogy to heat transfer is often self-evident. As the films in Figure 2 indicate, diffusion is in them commonly the predominant mode of transportation, always in a gradient-successive direction. This film diffusion gets slower the further it departs from the interphase. Until it measurably dies out at the level of the bulk fluid. At this point, an intraphase indicates the discrete change between both, the considerably lumped bulk and the distributed film.

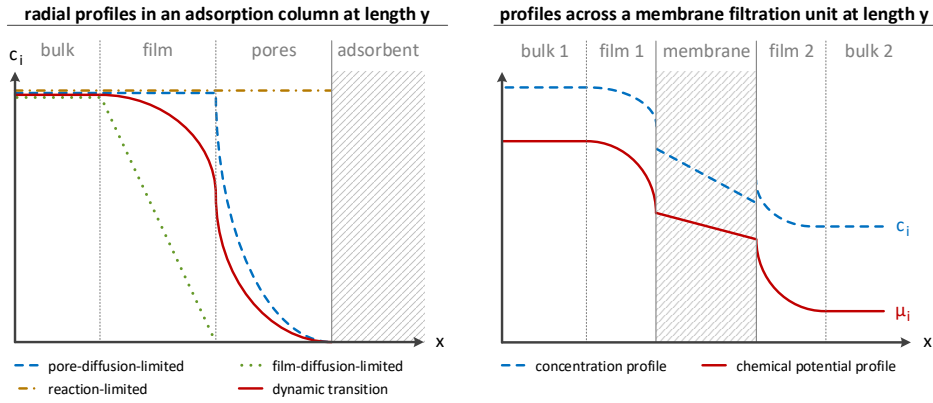


Figure 2: Generic profiles of an adsorption column (left) and a membrane unit (right)

Though it shall be placed emphasis on the fact that the concentration gradient is never actually zero, what would wrongly imply an absence of transport from the bulk to the film. In actuality, the gradient inclines towards zero in infinity and the intraphase is only placed on a x from where on the effective diffusion is conventionally considered to be negligible.

The two examples in Figure 2 splendidly exhibit most kinds of interfaces: *a*) bulk | film intraphases; *b*) fluid | solid interphases; *c*) fluid | pores intraphases, and *d*) fluid | membrane interphases. The first two we discussed already, but further discussion is advisable for the cases of *c* and *d* since, in contrast to *a* and *b*, they strongly rely on the scale of model design. Picture that the plots in Figure 2 would be zoomed out for the adsorption process and zoomed in for the membrane. The film | pores intraphase would turn into a film | adsorbent interphase, whereas the film | membrane interphase would transform to a film | membrane pores intraphase. The implication is clear: Pores may be formed by a solid phase but their contents clearly belong to the fluid phase, thus, a film | pores intraphase is placed if considered to be constructive. A membrane is fundamentally nothing more than an elaborate porous material. Accordingly, defining a membrane as another phase than the fluid may be the conventional as well as constructive approach, but a fluid | membrane "interphase" remains nothing similar to an actual interphase in a fluid | fluid system. The interphase-specific concentration jump may occur in the right plot of Figure 2 just as it would across a water | hexane interphase, however, this effect is not due to phase saturation but sieve stochastics.

Estimating concentrations at interphases is tricky due to discrete jumps. Thus, no approximations can be made on one side of the interface just from the concentration on the other. Fortunately, the chemical potential μ_i does not exhibit jumps so that we can approximate the concentrations with the aid of Equation (1), where μ_i° is the standard, μ_i^{ideal} the ideal and μ_i^{excess} the excess chemical potential. Keep in mind though, even there is no jump in μ_i , functions of intensive properties are *never* continuous across interfaces and we can only approximate interfacial conditions via close-quarter estimations on both sides.

$$\mu_i = \mu_i^\circ + \mu_i^{ideal} + \mu_i^{excess} = \mu_i^\circ + RT \ln(x_i) + \mu_i^{excess} \quad (1)$$

3. Derivation of topology capacities

From the elaborations in the previous section it transpires that interface-exceeding processes have to be modelled with the impact of the interface in mind. To stick with the examples in Figure 2, an adsorption model has to expand a system of bulk - film - pores - adsorbent, while a membrane model should expand bulk 1 - film 1 - membrane - film 2 - bulk 2. Supposedly, these are the lowest granularities for the associated topologies, which we can construct solely from six different capacities: *I*) lumped (circle), *II*) dynamic distributed (ellipse), *III*) steady-state distributed (rectangle), *IV*) boundary (black bar), *V*) point capacity (black dot), and *VI*) reservoirs (semi-circles). According to Preisig (2012), this is enough to define basic operations as elements of a process and to construct complete plant models by generating different combinations of them.

Each capacity inherits the fundamentals of any physics ontology, namely the conservation principles of mass, energy and momentum. Therefore, Equation (2) is part of every capacity, where \dot{n}_i denotes the accumulation of compound i over time and \hat{n}_i the compound flow. In our example, the state variable transformations in Equations (3) and (4) are also amenable to all, with dr_j being the place-in for the coordinate expansion normal to A .

$$\dot{n}_i = \sum \hat{n}_i \quad (2)$$

$$\hat{n}_i = c_i \hat{V} \quad (3)$$

$$V = A dr_j \quad (4)$$

The distinction between capacities results from additional definitions like Equation (5) added to the lumped *I* and *IV-VI* (the distributed capacities are left out since c_i is not constant), and the requirement for some kind of transport mechanism in the distributed systems *II* and *III*. For films and pores, the transport mechanism is the link between molecular flux and flow in Equation (6), and Fick's 1st law in Equation (7). Although we limit our discussion to interface-exceeding processes, note that different transport mechanisms would have to be selected for the membrane, and transport outside of films would have to consider not only diffusional but also convective fluxes. The event-dynamic capacities *III-VI* are set with the constraint $\frac{\partial}{\partial t} = 0$. Additionally, the surface nature of *IV* constrains its amount of extensive quantities to zero while connecting flows only on the boundary's normal coordinate.

$$\dot{n}_i = c_i \dot{V} \quad (5)$$

$$\hat{n}_i = J_{i,r_j} A \quad (6)$$

$$J_{i,r_j} = -D_{i,r_j} \frac{\partial c_i}{\partial r_j} \quad (7)$$

That leaves us with six distinct capacities to build our process on. In the following, we locate vectors normal to membrane or adsorbent on the x coordinate, whereas the perpendicular flows along the process unit are located on y . We further assume processes on other coordinates than x and y , and the distribution of compounds in films, pores and the membrane along y as well as between the adsorbent particles, to be negligible. We know from Figure 2 that at any discrete y , films and pores inhibit diffusion-induced distributions in the direction of x , separated by interfaces. Accordingly, we set dynamic, distributed capacities in the dynamic time-scale of Figure 3, perpendicular to the convective bulk flows and connected via boundaries. Preisig (2014a) explains in detail how topologies are

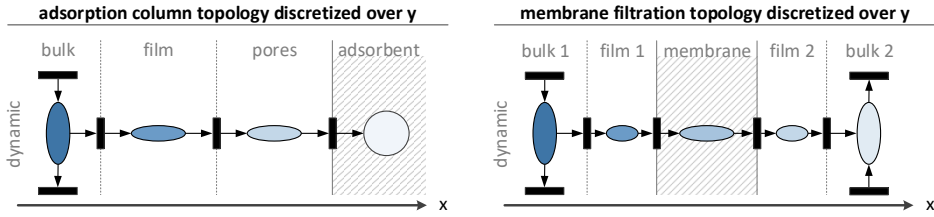


Figure 3: Dynamic topology for any discretionary stage in an adsorption column (left) and a membrane unit (right)

easily translated into an incidence matrix in order to mathematically assess the model. The topologies in Figure 3 depict one theoretical stage for each process, thus a discretization over y . Given an appropriate number of stages - so that $dy \rightarrow 0$ applies - the bulk flow can be reduced to a linear function of the boundary conditions at the respective stage. Furthermore, the input conditions for processes in x -direction do not need to be formulated as a function of y but adhere to the homogeneity assumption. Effectively relieving the numerics from y as one of the model's dependent variables.

$$\int_V \dot{n}_i \, dv = \int_A (\hat{n}_{i,x_0} - \hat{n}_{i,(x_0+dx)}) \, da \quad \Rightarrow \dots \Rightarrow \quad \dot{n}_i \, dx = \hat{n}_{i,x_0} - \hat{n}_{i,(x_0+dx)} \quad (8)$$

$$\hat{n}_{i,(x_0+dx)} = T(x; x_0) = f(x_0) + f'(x_0) \frac{(x-x_0)}{1!} + f''(x_0) \frac{(x-x_0)^2}{2!} + \dots \quad (9)$$

$$\dot{n}_i \, dx = \hat{n}_{i,x_0} - \hat{n}_{i,x_0} - \frac{\partial \hat{n}_i}{\partial x} \, dx - \frac{\partial^2 \hat{n}_i}{\partial x^2} \frac{dx^2}{2} + \dots \quad \Rightarrow \dots \Rightarrow \quad \dot{n}_i = - \frac{\partial \hat{n}_i}{\partial x} \quad (10)$$

$$\dot{c}_i = \frac{D_{i,x}}{dx} \frac{\partial^2 c_i}{\partial x^2} = k_{i,I} \frac{\partial^2 c_i}{\partial x^2} \quad (11)$$

Every diffusional flow in Figure 3 has one input at an arbitrary x_0 and one output at $x_0 + dx$. By recalling that the accumulation in a capacity \dot{n}_i is estimated over the capacity's volume V while connected flows \hat{n}_i are estimated over the capacity's surface A , we derive Equation (8) from Equations (2) and (4). Substituting Equation (5) into (8) yields the balance of a generic lumped capacity. For the distributed capacities in our example however, we have to expand the unknown $\hat{n}_{i,(x_0+dx)}$ in Equation (9) over x_0 by means of a Taylor series $T(x; x_0)$. This, together with $dx \rightarrow 0$, formulates the balance of an arbitrary distributed capacity in Equation (10). The substitution of (10) by the selected transport mechanism and the common assumption that A and the diffusion coefficient $D_{i,x}$ are independent of x , closes with Equation (11) - Fick's 2nd law in disguise.

4. Concluding remarks of advisory

We reason that lecturing process modelling with the presented methodology in mind not only significantly aids comprehensibility but also trains engineering students in compartmentalizing complex processes into their fundamental parts and to design appropriately comprehensive models that always satisfy the conservation principles. As was shown, limiting the topology design to the six introduced capacities together with their respective parts of a fundamental equation ontology enables rapid model design while offering easy mathematical accessibility as well as pacing options in terms of discretization. This easing of numerics also significantly reduces computing times, thus empowering the crafted

models to be utilized in online process control applications.

Since we arranged flows on different coordinates in separate capacities, solving the resulting equations is no longer a problem of numerics but of parameters only. If no $D_{i,x}$ are handily available, multiple interfacial transport concepts like the two-film, penetration or surface-renewal theory are applicable to solve Equation (11) via the mass transfer coefficient $k_{i,l}$ at each side of the interface. Furthermore, different kinds of transport limitations, as depicted in Figure 2, can be realized by means of time-scale based model simplifications in the respective capacities. However, the material diffusion coefficient $D_{i,x}$ must not be confused with the effective diffusion coefficient applied by Nernst (1904). As sketched in Figure 4, the effective film thickness δ is derived from the intensive gradient's tangent at the interface, whereas the film thickness dx in Equation (11) - and therefore the dedicated capacity - denotes the range of measurable distribution.

In this paper, we only discussed the dynamic process depiction. The reason is that dynamic and steady-state processes are two clearly distinct problems in process simulation. In a dynamic process, the knowledge of the conditions on one side of an interface suffices the estimations on the other side. In contrast, steady-state estimations are in their nature boundary-condition problems - Equation (11) with the constraint $\frac{\partial}{\partial t} = 0$ results in a gradient of its boundary conditions. Thus, steady-state interfacial processes have to be assessed from both sides. That raises one particular question, not to be answered within the scope of this paper but still worth considering: Is it actually constructive to lecture process modelling with focus on steady-state as the reputedly easy starting point?

Acknowledgements

As part of the research centre *Bio4Fuels*, this study has received funding from the *Research Council of Norway (RCN)* under the support code 257622 and is hosted by the *Norwegian Centre for Environment-friendly Energy Research (FME)*.

References

- A. T. Elve, H. A. Preisig, 2019. From ontology to executable program code. *Comput. Chem. Eng.* 122, 383–394.
- W. Nernst, 1904. Theorie der reaktionsgeschwindigkeit in heterogenen systemen. *Zeitschrift für physikalische Chemie* 47, 52–55.
- H. A. Preisig, 2012. Thinking ontologies. *Comput. Aided Chem. Eng* 31, 1682–1686.
- H. A. Preisig, 2014a. A graph approach to representing the pressure distribution in complex plants. *Comput. Aided Chem. Eng* 33, 865–870.
- H. A. Preisig, 2014b. Visual modelling. *Comput. Aided Chem. Eng* 34, 729–734.
- H. A. Preisig, 2020. Promo - a multi-disciplinary process modelling suite. *Comput. Aided Chem. Eng* 48, 571–576.
- R. Pujan, H. A. Preisig, 2020. Systematic modelling of flow and pressure distribution in a complex tank. *Comput. Aided Chem. Eng* 48, 1945–1950.

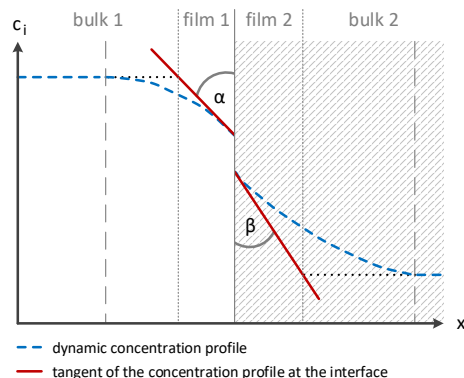


Figure 4: Schematic representation of Nernst's effective diffusion layer thickness

An identification approach to a reaction network for an ABE catalytic upgrade

Elham Ketabchi,^a Tomas Ramirez Reina^{a,*}, Bogdan Dorneanu,^b Harvey Arellano-Garcia^b

^a*Department of Chemical and Process Engineering, University of Surrey, Stag Hill, Guildford GU2 7XH, UK*

^b*LS Prozess- und Anlagentechnik, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus D-03044, Germany*

t.ramirezreina@surrey.ac.uk

Abstract

This contribution presents a kinetic study for the identification of the complex reaction mechanism occurring during the ABE upgrading, and the development of a kinetic model. Employing graph theory analysis, a directed bipartite graph is constructed to reduce the complexity of the reaction network, and the reaction rate constants and reaction orders are calculated using the initial rate method, followed by the calculation of the activation energy and frequency factor for an Arrhenius-type law. Subsequently, using general mass balancing a proposed mathematical model is produced to determine the apparent reaction rates, which are successfully in line with the experimental results.

Keywords: reaction mechanism, kinetic modelling, ABE upgrading.

1. Introduction

The increase in global energy consumption, correlated with the increase in greenhouse gas emissions has prompted interest in the “green” production of fuels and chemicals. Biomass conversion, such as sugar fermentation using the bacteria genus *Clostridium* (Qureshi et al., 2010) to produce an acetone, butanol and ethanol (ABE) mixture is a well-recognised “drop-in” fuel option, and one of the first commercial biofuel production methods. Following upgrading and conversion, the products from the ABE mixture are compatible with fuel or chemicals sourced from other means, making the process an attractive option for the transportation or chemical industry. This opens an avenue for the integration of a bio-refinery with an oil refinery, with the aim of moving towards sustainable fuels and chemicals production.

An important aspect of ABE upgrading requiring understanding is the determination of the very complex reaction network, which could facilitate the conceptualisation of a realistic model of the process. Although various studies regarding the production of long-chain hydrocarbons from ABE are available (Onyestyák et al., 2015; Sreekumar et al., 2015), there is little work concerning the kinetic modelling of the involved reactions.

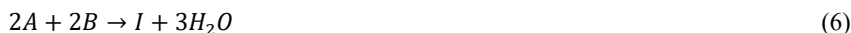
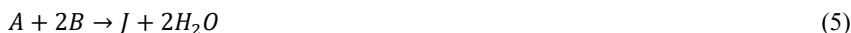
The following sections present the identification of a reaction network for the ABE upgrading based on literature and experimental data (Ketabchi et al., 2020a; Ketabchi

et al., 2020b). This network is used as a basis for calculating the kinetic parameters and a mathematical model of the process. Finally, the obtained model is compared to experimental data to investigate its accuracy.

2. Acetone, butanol and ethanol catalytic upgrade towards long-chain hydrocarbons

Following upgrading and conversion, the ABE mixture's products are attractive fuels or chemicals for transportation and chemical industry. Despite the complexity of the reaction network, it is necessary to identify a starting point so that the kinetics of this reaction can be calculated to produce a preliminary model. Even though various studies regarding the production of long-chain hydrocarbons from ABE are available (Onyestyák et al., 2015; Sreekumar et al., 2015), there is little work concerning the kinetic study and modelling of the reactions.

Starting from these studies, a network of the main reactions taking place during the ABE upgrading is proposed in Equations (1) – (11).



Due to the complexity, with 16 components involved, (A = acetone, B = butanol, E = ethanol, D = isopropyl alcohol, F = 2-pentanone, G = 2-heptanone, I = 4-nonanone, J = 6-undecanone, K = 6-undecanol, L = 4-heptanol, M = 1-butanone, N = mesityl oxide, P = isophorone, plus O₂, H₂O and H₂) including equilibrium, series, reversible, and other type of reactions occurring at once, this network requires further simplification for the scope of this work.

3. Kinetic modelling of the ABE upgrading

Using the reactions in equations (1) – (11), as well as their interrelations, a mathematical modelling theory is applied using three main algebraic parameters. These parameters are: the set of chemical reactions R_i , with i = the number of finite sets of reactions occurring; the set of chemical species S_j , with j = the corresponding species involved in reaction; and lastly, the nonnegative integer stoichiometric coefficient of each component β_{ij} . Altogether, the model is represented as:

$$R_i = \sum_{j=1}^n \beta_{ij} S_j \quad i = 1, \dots, m \quad j = 1, \dots, n \quad (12)$$

Given that these parameters are defined, the kinetic rate equation can be defined as a nonlinear function of the concentration of chemical species S_j , $f_i (C_s(\text{mol/L}))$:

$$r_i = k_i f_i(C_s) \quad i = 1, \dots, m \quad (13)$$

Where r_i is the reaction rate, and k_i is the reaction rate constant.

The reactions are considered to take place in a batch reactor with constant volume. Applying a mass balance on this batch system, a lumped kinetic model in the form of ordinary differential equations (ODEs) is obtained. These equations demonstrate the change of concentration against time in the following way:

$$\frac{dC_i}{dt} = r_i \cdot \beta_{ij} \quad (14)$$

To simplify the kinetic reaction network, the concept of graph theory analysis (Gupta et al., 2016) can be applied. This concept is based on the evaluation of unnecessary reactions and components that are eliminated as a result of the interpolation of complete conversion or quasi-equilibrium reactions.

3.1. Identification of chemical reaction network

In order to identify the connection between the species in the proposed reaction network in Equations (1) – (11), a directed bipartite graph is defined (Figure 1), which is dependent on the components, S , the reactions, R , and a set of direct links.

The application of this approach decreases the numbers of reactions in the network from 14 (due to the reversible reactions (9) – (11) considered each as two separate reactions) to only 7 (or 9, as R4X and R5 are reversible):

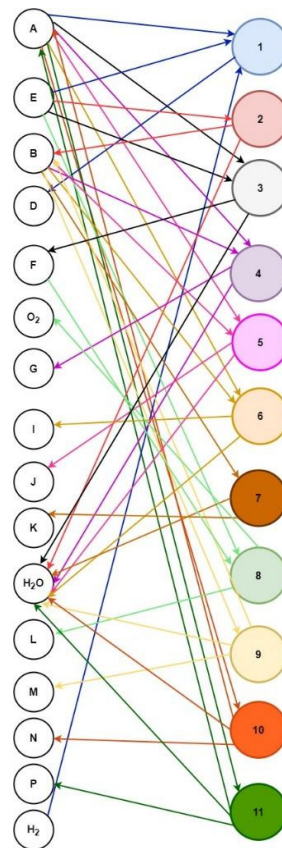
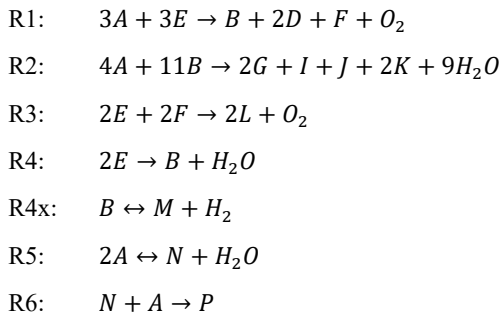


Figure 1. Bipartite graph for the ABE upgrading reaction network

For simplicity in future calculations, the reverse reactions of R4x and R5 are denoted as R-4x and R-5, respectively.

3.2. Kinetic calculations of the chemical reaction network

Considering the chemical reaction network identified in the previous section, the mass balance of the system represented by Equation (14) can be written. The following kinetic law, from which the order (x, y) and the reaction rate (k_i) are determined is used:

$$r_i = k_i c_a^x c_b^y \quad (15)$$

These are calculated based on the initial rate method, in which the rate of reaction at the instant that the reactants are first mixed is considered. This initial rate is assumed to be equal to the average rate of reaction.

Subsequently, the different activation energies and pre-exponential factors associated with each reaction are determined using an Arrhenius-type equation, by fitting experimental data at various temperatures.

Due to the assumption of equating the initial rate to the average rate, the values obtained will be, to some extent, inaccurate. On the other hand, some species will appear halfway through the reaction, making the calculated rates underestimated. Therefore, to reduce these errors, a parameter optimisation is performed using MATLAB R2019a, in an iterative procedure, using the calculated kinetic constants as first guess and optimising their value such that the mathematical model fits the experimental data.

3.3. Experimental procedure

Four experiments of ABE upgrading are performed in a pressure vessel/Parr reactor (Parr Series 5500 HPCL Reactor and a 4848 Reactor Controller) with the volume of 300 mL, at autogenous 80 bar pressure and 300 °C (Ketabchi et al., 2020b).

Table 1. Conversion of reactants and yield of products having the highest concentration at different temperatures using Fe catalyst

Temperature (°C)		200	250	300
Conversion (%)	A	99.0	99.6	99.6
	B	-	-	95.9
	E	95.5	95.9	95.2
Yield (%)	D	6.45	10	6
	G	3.40	3.40	3.80
	L	-	0.60	0.64
	P	-	-	0.34
	J	-	-	0.40

The reactor is purged with N₂ to ensure an oxygen-free atmosphere, and filled with 0.5 g of reduced Fe catalysts and 83.5 mL of anhydrous ABE mixture (Sigma-Aldrich) with the mole ratio 3:6:1. After 18 hours, the catalysts are recovered through filtration and the liquid products analysed in a gas chromatograph through the SIM mode (Agilent HP6890 GC), using a DB-5 Capillary Analytical column and a flame ionisation detector (FID). Quantification of the reaction products is performed by establishing GC response

factors of the main products identified, by injection of samples with known concentrations prepared using standards. The results for the four experiments are shown in Tables 1 and 2.

Table 2. Experimental data used for obtaining kinetic data

<i>Experiment</i>	<i>Acetone (moles)</i>	<i>Butanol (moles)</i>	<i>Ethanol (moles)</i>
<i>K1</i>	0.306	0.601	0.2054
<i>K2</i>	0.153	0.601	0.1027
<i>K3</i>	0.306	0.300	0.1027
<i>Original</i>	0.306	0.601	0.1027

4. Results and discussion

Using the approach described in Section 3.2, the order and reaction rate constants summarised in Table 3 are calculated.

Table 3. Reaction rate parameters for the ABE upgrading network

<i>Reaction number</i>	<i>Total reaction order</i>	<i>Reaction rate value ($\times 10^{-6}$)</i>	<i>Unit</i>	<i>E_a (kJ/mole)</i>	<i>A</i>
<i>R1</i>	2	2200	<i>L/mole · s</i>	7.656.3626	17.16590×10^{-5}
<i>R2</i>	2	2500	<i>L/mole · s</i>	647.1950	66.21620×10^{-5}
<i>R3</i>	1.4	330	<i>L/mole · s</i>	284,122.6360	1.94199×10^{21}
<i>R4</i>	0	0.01300	<i>mole/L · s</i>	146,193.3760	16.09740×10^4
<i>R4x</i>	0	0.00310	<i>mole/L · s</i>	317,262.2400	2.03707×10^{-6}
<i>R-4x</i>	0	0.00135	<i>mole/L · s</i>		
<i>R5</i>	1	0.15300	<i>1/s</i>	89,358.8720	4.00504×10^{-30}
<i>R-5</i>	1.028	0.18600	<i>1/s</i>		
<i>R6</i>	1	0.15400	<i>1/s</i>	63,375.1278	9.46455×10^{-30}

It can be observed that at constant temperature the slowest reaction is R4x, as it has the lowest reaction rate and the highest activation energy, which is in good agreement with the observations by Narayanan et al. (1998) that Fe is not well suited for aldol condensation reactions involving hydrogen transfer. Interestingly, R2 has the highest reaction rate constant, showing that the Fe catalyst favours it. The model data is subsequently compared to the experiments, in concentration vs. time graphs (Figure 2), to assert the model accuracy. The expected trend of decreasing concentration with time can be seen with the reactants, while an increase is observed for the remaining components regarding the modelled data. However, one limitation of the model simplification should be discussed regarding components (such as Ethanol), which are both produced and consumed in the ABE upgrading. Production occurs also in

irreversible reactions, not considered in the final model. Therefore, for such cases, the model is only able to predict the overall trend.

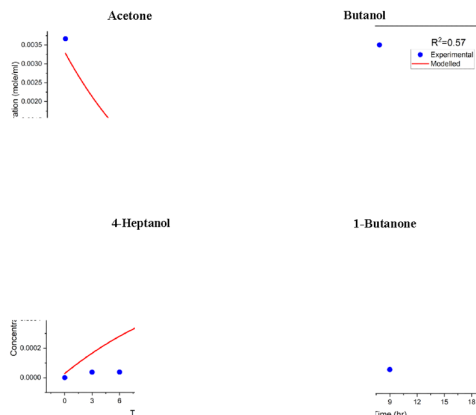


Figure 2. Comparison between experimental and modelled data

5. Conclusions

This contribution identifies a reaction network for the ABE upgrading using graph theory analysis, based on previous literature, as well as experimental data. Following this, the kinetic model parameters are determined and optimised for Arrhenius law-type of equations. The resulting model is compared with the experimental data through concentration-time graphs, demonstrating a decent degree of accuracy, despite a few components displaying abnormal behaviour. More experiments should be performed to obtain more data for these and for additional products to further improve the model and its accuracy.

References

- U. Gupta et al., 2016, Time scale decomposition in complex reaction systems: A graph theoretic analysis, *Comp Chem Eng* 95, pp. 170-181
- E. Ketabchi et al., 2020, Influence of reaction parameters on the catalytic upgrading of an Acetone, Butanol, and Ethanol (ABE) mixture: Exploring new routes for modern biorefineries, *Front Chem* 7 (1), pp. 1-14
- E. Ketabchi et al., 2020, Catalytic upgrading of Acetone, Butanol and Ethanol (ABE): A step ahead for the production of added value chemicals in biorefineries, *Renew Energy* 156 (1), pp. 1065-1075
- S. Narayanan et al., 1998, Acetone hydrogenation over co-precipitated Ni/Al₂O₃, Co/Al₂O₃ and Fe/Al₂O₃ catalysts, *J Chem Soc* 94, 8, pp. 1123-1128
- G. Onyestyák et al., 2015, Acetone alkylation with ethanol over multifunctional catalysis by a borrowing hydrogen strategy, *RSC Adv* 5, 120, pp. 99502-99509
- N. Qureshy et al., 2010, Production of butanol (a biofuel) from agricultural residues, *Biomass Bioenergy* 34, 4, pp. 566-571
- S. Sreekumar et al., 2015, Production of ABE mixture from *Clostridium acetobutylicum* and its conversion to high-value biofuels, *Nat Protoc* 10, 3, pp. 528-537

MOSKopt: A simulation-based data-driven digital twin optimizer with embedded uncertainty quantification

Resul Al^a, Gürkan Sin^{b,*}

^a*Novo Nordisk A/S, DK-2880 Bagsvaerd, Denmark*

^b*Process and Systems Engineering Center, Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*

gsi@kt.dtu.dk

Abstract

In this work, we present a new stochastic black-box optimizer—named MOSKopt—for optimization under uncertainty problems involving expensive-to-evaluate and arbitrarily complex simulation models subject to severe uncertainties and multiple design constraints. The optimizer relies on a novel surrogate-assisted, simulation-based, and data-driven optimization algorithm developed in our earlier work for multiple-constrained optimization of stochastic black-box systems using a stochastic Kriging (SK) type surrogate model, which has shown its promising potential in stochastic simulation optimization studies. MOSKopt replaces simulation replications with parallelized Monte Carlo simulations to characterize the resulting error distributions of the optimization objective and constraints in a more generalized way. In doing so, it broadly extends the applicability of SK-based frameworks to other optimization problems exposed to uncertain input data or model parameters—a commonly encountered paradigm in process simulations driving the digital twins. We benchmark the optimizer and its algorithms on three case studies with increasing complexity. The first two are benchmark test problems taken from simulation optimization literature, whereas the third optimizes the design and operations of a wastewater treatment plant (WWTP) subject to effluent quality constraints using a detailed plant-wide simulation model—the closest case to a digital twin—consisting of several high-fidelity bioprocess models (e.g., activated sludge and anaerobic digestion) calibrated for data obtained from full-scale wastewater treatment plants under uncertain influent scenarios. The results from case studies quantitatively demonstrate the superior efficacy of the MOSKopt algorithm over the literature methods not only in maintaining feasibility under multiple stochastic constraints but also in returning better objective function values. The results also show the potential of the MOSKopt’s entirely non-intrusive workflow, which makes it a suitable optimizer for applications involving digital twin projects with non-negligible uncertainties.

Keywords: simulation-based optimization, digital twins, optimization under uncertainty

1. Introduction

Industry is invested in building digital twins that are aimed at mirroring every facet of their physical counterparts (e.g., a pharmaceutical manufacturing process) as accurately as possible. At the core of this ambition lie high fidelity simulations driven by an integrated set of first principles/mechanistic, multiscale, and machine learning models that are continually kept up-to-date using data collected from sensors in real-time. There

is much to be done to realize the full potential of such ambitious projects. For instance, optimizing the design and operations of a physical twin requires a simulation-based and data-driven optimization engine that can interface with its virtual replica. To this end, simulation-based optimization algorithms hold an appealing promise due to their non-intrusive ability to seek a near-optimal set of decisions to arbitrarily complex process models along with the ability to incorporate stochastic objectives and constraints, thereby allowing for the quantification of uncertainties surrounding the decisions made from the use of the digital twin technology. In this work, we consider optimization under uncertainty problems whose objective and constraints are surrounded by uncertainties and can only be estimated by performing high fidelity simulations. For these problems, we comparatively evaluate sampling and simulation-based optimization methods.

2. Methods

2.1. Benchmark method: Exhaustive sampling

The exhaustive sampling method represents the standard Monte Carlo sampling-based approach to uncertain design space exploration problems. In this method, both the design space of the process parameters and the uncertainty space of uncertain input parameters are discretized using sampling [1]. At each design sample (x_i) the probability of satisfying active constraints under uncertainty is computed via Monte Carlo simulations performed within the uncertainty space. Table 1 shows several statistical metrics (e.g., mean, upper confidence interval of the mean, etc.) that can be used for hedging against the uncertainties. Despite being effective, this method often requires a large number of model evaluations, making it extremely intensive (computationally), if not prohibitive, especially for real practical applications, where the computational cost of running the simulation model can be very high. To reduce the required number of simulations and also to increase the efficacy of this method, simulation-based methods described next employ surrogate models that assist with performing a more informed exploration within the design space.

Table 1: Uncertainty hedging strategies used in this work for satisfying stochastic constraints.

Identifier	Explanation MCS: Monte Carlo simulations, cl : constraint limit
Mean	$\text{mean}(\text{MCS}) < cl$ for each constraint
UCI95	Upper confidence interval of $\text{MCS} < cl$ for each constraint
PF80	The probability of the feasibility of the $\text{MCS} > 80\%$ for each constraint
MeanPlusSigma	$\text{mean}(\text{MCS}) + \text{standard deviation}(\text{MCS}) < cl$ for each constraint

2.2. Simulation-based optimization

In our earlier work [2], we described the simulation-based optimization approach which employs stochastic Kriging [3] type surrogate models that are constructed from a dataset of initial observations of optimization objective and constraints. A Monte Carlo-based uncertainty analysis is also performed on each design sample in the initial dataset. The surrogate models are then used to find the next best candidate design points in an adaptive sampling stage, which performs an internal optimization using infill criteria. The designs proposed by this stage are validated by performing rigorous simulations, and the data obtained from these simulations are used to improve surrogate models of the objective, constraints, and their uncertainties. Lastly, after a certain computational budget is exhausted, the best feasible design among the simulated design points is returned as the near-optimal design. Since this design is obtained by taking into account the effects of uncertainties on the optimization objective and constraints, it is also the near-optimal

design under uncertainty. This workflow has been implemented in the solver MOSKopt with two infill criteria: feasibility enhanced expected improvement (FEI) [4] and multiple constrained FEI (mcFEI) [2]. Further details can be found in the original work [2].

3. Case studies

3.1. Case study 1: An illustrative test problem — Sasena

To demonstrate the benefits of the simulation-based approach, we first consider a widely-studied illustrative test problem named Sasena from literature [5]. With its two design variables and three active design constraints, Sasena readily allows for design space visualization and provides a case well representative of actual engineering design optimization problems subject to multiple stochastic constraints. To make this problem an optimization under uncertainty problem, we arbitrarily introduce four uncertain parameters entering the expressions of both the objective and the constraints as follows:

$$\begin{aligned}
 \min_x \quad & f(x) = -u_1(x_1 - 1)^2 - (x_2 - 0.5)^2 \\
 \text{s.t.} \quad & g(x) = [g_1(x), g_2(x), g_3(x)] \leq 0 \\
 & g_1(x) = (u_2(x_1 - 3)^2 + (x_2 + 2)^2)e^{-x_1^2} - 12 \\
 & g_2(x) = 10u_3x_1 + x_2 - 7 \\
 & g_3(x) = (x_1 - 0.5)^2 + u_4(x_2 - 0.5)^2 - 0.2
 \end{aligned} \tag{1}$$

$$0 \leq x_i \leq 1 \text{ for } i = 1, 2 \text{ and } u_j \approx \mathcal{N}(\mu, \sigma) \text{ for } j = 1, \dots, 4 \text{ and } \mu = 1, \sigma = 0.5$$

We solve this problem using the benchmark method. First, the two-dimensional design space of the problem is sampled using Latin hypercube sampling (LHS) with a varying number of samples—from a very coarse (10) to a very fine (10^5) design. Each sample in the sampling design is a candidate design for which an uncertainty analysis is performed using vectorized Monte Carlo simulations with 10^3 samples taken from the normally distributed input uncertainty space. The results are repeated 50 different times so as to alleviate the effects of the randomness in the initial sample locations. Table 2 shows the averaged (over 50 repetitions) results of the optimum found and its location. Expectedly, increasing the number of design samples results in improved objective values as the finer sampling cover the design space more rigorously. A design sample is declared feasible if the mean plus the standard deviations (obtained from Monte Carlo simulations) of each constraint satisfies the corresponding constraint limits. Similarly, the other hedging strategies, as was tabulated in Table 1, are also computed, and the obtained results are shown in Figure 1. Using only the mean for uncertainty hedging yields lower optimum values (-0.549 for 10^5 design samples), whereas more conservative approaches uplift the

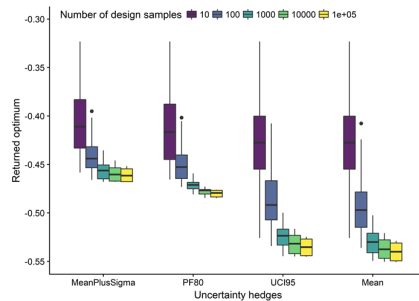


Figure 1: Uncertainty hedging strategies are compared along with the increasing number of design samples. The returned optimum shifts upwards for more conservative strategies.

optimum towards the positive axis. The same problem is also solved with the simulation-based optimizer (MOSKopt) using the two infill criteria (FEI and mcFEI) with a total of 100 samples in the design space (10 initial, 90 adaptive iteration samples). Table 2 shows the good agreement obtained among the methods.

Table 2: The average results of the optimum value of the objective and its location obtained using different methods. The uncertainty hedge is set to MeanPlusSigma.

Method	# of LHS design samples	Total # of calls to the model	Optimum value (f)	Location of the optimum (x)
Exhaustive	10^1	10^4	-0.395	[0.310, 0.662]
Exhaustive	10^2	10^5	-0.452	[0.204, 0.376]
Exhaustive	10^3	10^6	-0.464	[0.197, 0.465]
Exhaustive	10^4	10^7	-0.467	[0.202, 0.447]
Exhaustive	10^5	10^8	-0.468	[0.200, 0.453]
FEI	10^2	10^5	-0.451	[0.275, 0.932]
mcFEI	10^2	10^5	-0.458	[0.247, 0.445]

3.2. Case study 2: A high dimensional test problem — Rosen Suzuki

In order to both investigate the reliability and also compare the performance of both infill criteria in higher-dimensional problems, we solve the Rosen Suzuki test problem from literature [5]. Similar to the previous case, we introduce uncertain parameters in the expressions of both the objective and the constraints as follows:

$$\begin{aligned}
 \min_x \quad & f(x) = x_1^2 + x_2^2 + 2u_1x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \\
 \text{s.t.} \quad & g(x) = [g_1(x), 100g_2(x), 1000g_3(x)] \leq 0 \\
 & g_1(x) = x_1^2 + u_2x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 \\
 & g_2(x) = x_1^2 + 2x_2^2 + x_3^2 + 2u_3x_4^2 - x_1 - x_4 - 10 \\
 & g_3(x) = 2x_1^2 + u_4x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5
 \end{aligned} \tag{2}$$

$$-3 \leq x_i \leq 3 \text{ for } i = 1, \dots, 4 \text{ and } u_j \approx \mathcal{N}(\mu, \sigma) \text{ for } j = 1, \dots, 4 \text{ and } \mu = 1, \sigma = 0.25$$

Table 3 tabulates the results obtained from 50 different repetitions. The benchmark method used 10^5 design samples, whereas the MOSKopt solver used only 50 design samples (20 initial plus 30 infills) to locate the optimum, with the MeanPlusSigma

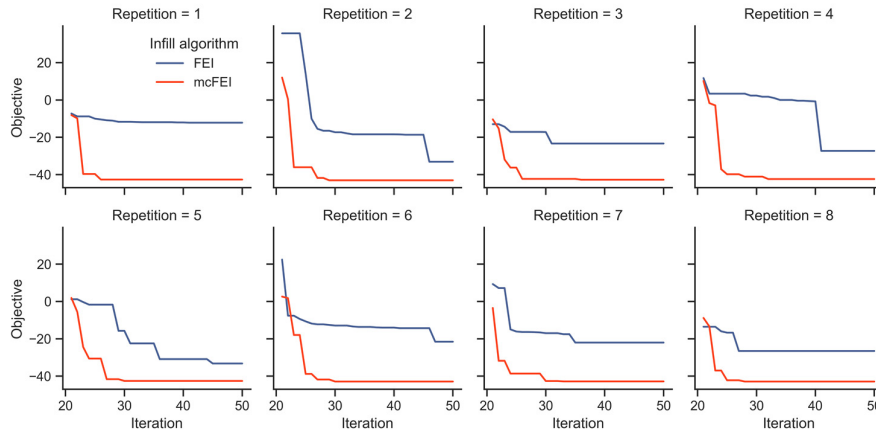


Figure 2: Comparison of FEI and mcFEI infill criteria for the Rosen Suzuki test problem. The multiple constraint modeling of the mcFEI criterion leads to better convergence.

hedging. Comparing the two infill criteria, Figure 2 shows convergences where the mcFEI criterion consistently outperforms the FEI criterion in returning better optimum solutions satisfying the stochastic constraints. This can be explained with the separate modeling of each constraint response and their uncertainty, as implemented in the mcFEI criterion, as opposed to the lumped approach of the FEI criterion. As each constraint has different scales, separate modeling helps to better identify the uncertain characteristics of the problem’s output space and hence leads to the better performance of the mcFEI criterion. The solver’s GitHub repository (<https://github.com/resulal/MOSKopt>) contains the working code examples of Sasena and Rosen Suzuki case studies.

Table 3: Results of simulation-based optimization for the uncertain Rosen Suzuki problem.

Method	# of LHS samples	Total # of calls to the model	Optimum value (f)	Location of the optimum (x)
Exhaustive	10^5	10^8	-42.152	[-0.026, 0.841, 1.951, -0.914]
FEI	50	50×10^3	-24.915	[-0.368, 0.607, 1.468, 0.128]
mcFEI	50	50×10^3	-42.701	[0.028, 0.608, 1.996, -0.996]

3.3. Case study 3: WWTP plant optimization under influent uncertainty

After demonstrating the motivations behind the simulation-based approach with the simple test problems, we now turn to the engineering design under uncertainty problems subject to multiple constraints. To this end, we build on the results of our earlier work [6], where we presented a Monte Carlo sampling-based design space exploration for two full-scale WWTPs: Avedøre (Denmark) and Valladolid (Spain). Both plants employ an activated sludge process-based wastewater treatment with anaerobic digestion of the sludge line. Four influent fractionations, namely soluble inert (*SI*), readily biodegradable substrate (*SS*), particulate inert organic matter (*XI*), and active heterotrophic biomass (*XBH*), are allowed to be uncertain around their nominal values with a coefficient of variation (σ/μ) of 10 %. The design objective is to minimize a plant-wide operational cost index (OCI) while the active design constraints are the regulated effluent quality metrics, e.g., chemical oxygen demand (COD), total nitrogen (TN), and ammonia (NH_4). A more detailed description of the mathematical models, plant-wide simulation strategy, and the calculation of plant performance indicators can be found in our earlier work [6]. The MOSKopt solver is called with the mcFEI criterion along with the MeanPlusSigma hedging and the plants are simulated to a steady-state. Figure 3 shows the progress of the objective value in both plants using 150 design samples (60 initial + 90 adaptive), whereas Table 4 shows the values of the decision variables obtained at the returned optimum. Compared to our earlier work [6], the results obtained here using the simulation-based solver show improvement in the returned best objective values, reducing the operational costs in both plants under uncertain influent scenarios.

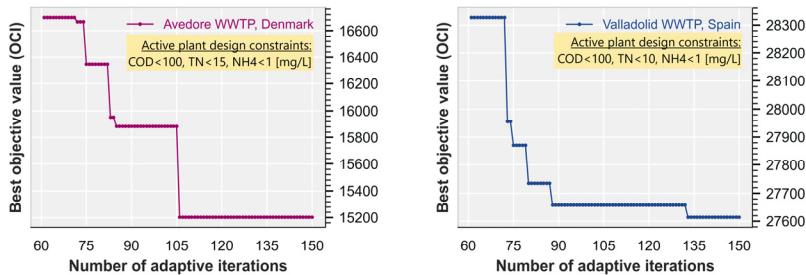


Figure 3: Results of simulation-based design space optimization in full-scale WWTPs.

Table 4: Optimization results for the WWTP design optimization case study.

Design and operational decision variables	Lower bound	Value at optimum Avedøre	Value at optimum Valladolid	Upper bound
Hydraulic residence time (HRT) [day]	0.125	0.125	0.125	0.5
Volume ratio of anoxic/aerobic tanks (VR)	0.2	0.49	0.52	0.8
Solids retention time (SRT) [day]	5	8.32	5.20	15
Dissolved oxygen setpoint (DO) [g/m ³]	0.7	0.70	0.89	1.5
Surface overflow rate (SOR) [m ³ m ⁻² day ⁻¹]	16	33	31.22	33
Internal recycle ratio (RRi)	1	2.75	2.78	3

4. Conclusions

This work has presented a novel generic stochastic black-box solver designed to meet a growing need from industry involved in optimization-based decision making coupled with the complex digital twin simulations. The objective was to showcase use cases for the solver by also comparing its performance to other literature methods. Results from three case studies highlight the importance of integrated uncertainty considerations in multiple design constraints in order to increase flexibility and resilience in the face of uncertainty. Using an embedded Monte Carlo simulations, the solver makes it possible to quantify uncertainties surrounding design constraints and thereby robustify design and operational decisions against the effects of possible variations in the influent wastewater compositions. The solver's entirely non-intrusive workflow makes it possible to plug in complex simulation models (e.g., via COM interfaces to CFD codes) without the further need for model modification and facilitates its wider adoption by non-specialist practitioners in other domains.

Acknowledgments

The authors acknowledge funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no.675251.

References

- [1] S. Kucherenko, D. Giamalakis, N. Shah, S. García-Muñoz, Computationally efficient identification of probabilistic design spaces through application of metamodeling and adaptive sampling, *Comput. Chem. Eng.* 132 (2019) 106608. <https://doi.org/10.1016/j.compchemeng.2019.106608>.
- [2] R. Al, C.R. Behera, K. V. Gernaey, G. Sin, Stochastic simulation-based superstructure optimization framework for process synthesis and design under uncertainty, *Comput. Chem. Eng.* 143 (2020) 107118. <https://doi.org/10.1016/j.compchemeng.2020.107118>.
- [3] B. Ankenman, B.L. Nelson, J. Staum, Stochastic Kriging for Simulation Metamodeling, *Oper. Res.* 58 (2010) 371–382. <https://doi.org/10.1287/opre.1090.0754>.
- [4] Z. Wang, M. Ierapetritou, Constrained optimization of black-box stochastic systems using a novel feasibility enhanced Kriging-based method, *Comput. Chem. Eng.* 118 (2018) 210–223. <https://doi.org/10.1016/j.compchemeng.2018.07.016>.
- [5] R.R. Carpio, R.C. Giordano, A.R. Secchi, Enhanced surrogate assisted framework for constrained global optimization of expensive black-box functions, *Comput. Chem. Eng.* 118 (2018) 91–102. <https://doi.org/10.1016/j.compchemeng.2018.06.027>.
- [6] C.R. Behera, R. Al, K. V. Gernaey, G. Sin, A process synthesis tool for WWTP – An application to design sustainable energy recovery facilities, *Chem. Eng. Res. Des.* 156 (2020) 353–370. <https://doi.org/10.1016/j.cherd.2020.02.014>.

Global Optimization of Mixed-Integer Polynomial Programs via Quadratic Reformulation

Tanuj Karia^a, Claire S. Adjiman^a and Benoît Chachuat^{a,*}

^a *Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom*

Abstract

Mixed-integer polynomial programs (MIPOPs) frequently arise in chemical engineering applications such as pooling, blending and operations planning. Many global optimization solvers rely on mixed-integer linear (MIP) relaxations of MIPOPs and solve them repeatedly as part of a branch-and-bound algorithm using commercial MIP solvers. GUROBI, one of the prominent MIP solvers, recently added the capability to solve mixed-integer quadratically-constrained quadratic programs (MIQCQPs). This paper investigates global optimization of MIPOPs via their reformulation as MIQCQPs followed by their solution to global optimality using GUROBI. The effectiveness of this approach is tested on 60 instances of MIPOPs selected from the library MINLPLib. The performance of the MIQCQP reformulation approach is compared to the state-of-the-art global solvers BARON, ANTIGONE and SCIP in GAMS. For the case of single threading, a reduction of 28% and 42% compared to SCIP and ANTIGONE respectively is observed. This approach, therefore, holds promise for integration into existing global solvers to handle MIPOPs.

Keywords: Deterministic global optimization, Polynomial programming, Mixed-integer nonlinear programming, MINLPLib

1. Introduction

Mixed-integer polynomial programs (MIPOPs) are a special case of mixed-integer nonlinear programs (MINLPs), whereby the objective and constraint functions are polynomial in the continuous and discrete variables. They are frequently encountered in chemical engineering applications, such as pooling & blending (Teles et al., 2013b), optimal scheduling of multiproduct plants (Castro and Novais, 2009), and short-term planning of integrated refinery-petrochemical complexes (Uribe-Rodriguez et al., 2020). Significant research effort has thus been devoted to solving MIPOPs to guaranteed global optimality. Generic global solvers based on spatial branch-and-bound search such as BARON (Kılınç and Sahinidis, 2018), ANTIGONE (Misener and Floudas, 2014) and SCIP (Gamrath et al., 2020), can all handle MIPOPs as special cases. Other algorithms include the reformulation-linearization technique (RLT) (Sherali and Tuncbilek, 1992), and the piecewise-linear relaxation approach (Teles et al., 2013a; Nagarajan et al., 2019). All of the aforementioned algorithms rely on linear (LP) or mixed-integer linear (MIP) relaxations of non-convex sub-expressions, which are solved using commercial MIP solvers. Recently, the capability to solve non-convex mixed-integer quadratically-constrained quadratic programs (MIQCQPs) was added to the GUROBI MIP solver (Gurobi Optimization LLC, 2020). In this paper, therefore, we investigate a novel solution approach, whereby MIPOPs are first reformulated as MIQCQPs and then

solved using GUROBI to global optimality. The performance of the reformulation approach is compared to the state-of-the-art global solvers BARON, ANTIGONE and SCIP. The rest of the paper is organized as follows: in Section 2 a brief description of the theory behind the reformulation, its implementation, and the test instances used is provided; the performance of our approach and comparisons with the aforementioned global solvers is analyzed in Section 3; finally, conclusions are drawn in Section 4.

2. Methodology

We consider MIPOPs in the form:

$$p_0(x, y) \tag{1}$$

$$s.t. \quad p_i(x, y) = 0, \quad i = 1, \dots, m_c \tag{2}$$

$$p_i(x, y) = 0, \quad i = m_c + 1, \dots, m \tag{3}$$

$$x \in [x^L, x^U] \subset R^{n_x}, \quad y \in [y^L, y^U] \subset Z^{n_y} \tag{4}$$

where x and y denote, respectively, the continuous and discrete variables; and the functions $p_i: R^{n_x} \times R^{n_y} \rightarrow R$, $i = 0, \dots, m$ are all polynomial in their arguments:

$$p_i(x, y) := \sum_{j=1}^{J_i} a_{i,j} q_{i,j}(x, y) \tag{5}$$

$$q_{i,j}(x, y) := x_1^{\alpha_{1,i,j}} \dots x_{n_x}^{\alpha_{n_x,i,j}} y_1^{\beta_{1,i,j}} \dots y_{n_y}^{\beta_{n_y,i,j}}, \quad j = 1, \dots, J_i \tag{6}$$

with real coefficients $a_{i,j} \in R$ and positive exponents $\alpha_{k,i,j}, \beta_{l,i,j} \in Z_0^+$, $k = 1, \dots, n_x$, $l = 1, \dots, n_y$.

2.1. Quadratic Reformulation

The main idea behind reformulating MIPOPs as MIQCQPs is that any polynomial function can be transformed into a quadratic function via the introduction of auxiliary variables. The reformulation proceeds by constructing a list B of monomials in the variables x and y , with the following properties (Rumschinski et al., 2010):

1. For every monomial q in Eqn. (6), there exists $b_1, b_2 \in B$ such that $q = b_1 b_2$
2. For every monomial $q \in B$ of degree higher than 1, there exists $b_1, b_2 \in B$ such that $q = b_1 b_2$
3. $1 \in B$

Notice that the list B is not unique in general.

From Property 1, the polynomial functions p_i in Eqn. (5) can be rewritten as quadratic forms:

$$p_i(x, y) = \xi^\top Q_i \xi, \quad i = 0, \dots, m \quad (7)$$

where the variable vectors ξ correspond to the elements of B ; and for suitable real symmetric matrices $Q_i \in S_+^{n_\xi}$ with $n_\xi = |B|$ obtained using Eqn. (7).

From Property 2, decomposing the elements in B with degree higher than 1 into products of lower-order monomials also in B gives rise to auxiliary quadratic constraints:

$$\xi^\top R_i \xi = 0, \quad i = 0, \dots, n_d \quad (8)$$

for suitable real symmetric matrices $R_i \in S_+^{n_\xi}$ and n_d is number of such decompositions.

Lastly, bounds on the variables ξ can easily be inferred, e.g., using interval calculations from the original variable bounds in Eqn. (4). These bounds can be represented by a set of linear constraints

$$A\xi \leq 0 \quad (9)$$

for a suitable real matrix $A_i \in R^{2n_\xi \times n_\xi}$.

Overall, an equivalent MIQCQP to the MIPOP in Eqns. (1) – (4) can be written in the form:

$$\min_{\xi \in R^{n_\xi}} \xi^\top Q_0 \xi \quad (10)$$

$$s. t. \quad \xi^\top Q_i \xi = 0, \quad i = 1, \dots, m_e \quad (11)$$

$$\xi^\top Q_i \xi \leq 0, \quad i = m_e + 1, \dots, m \quad (12)$$

$$\xi^\top R_i \xi = 0, \quad i = 0, \dots, n_d \quad (13)$$

$$SEQ \text{ Equation } \setminus * \text{ ARABIC } 13) \\ A\xi \leq 0, \quad \xi_1 = 1 \quad (14)$$

2.2. Implementation

A script was written in Python v3.7 that reads a GAMS scalar model and imports the list of all variables and variable bounds, the objective function, and the constraints. The quadratic reformulation described in Section 2.1 is automated in the library MC++ (Chachuat et al., 2015) by defining expression trees (DAG) from the problem information in Python. A dedicated class in MC++ returns the vector of monomials, sparse quadratic forms for all the polynomial objective and constraint functions (7), and sparse quadratic forms for the auxiliary constraints (8). Bounds on the auxiliary DAG variables are furthermore inferred using the verified interval library PROFIL (Knüppel, 1994). Ultimately, all of this information is collected by a C++ script that generates a lp file for the MIQCQP (10) – (14) and passes it to GUROBI (v9.0.3) for global optimization. This workflow is summarized in Fig. 1.

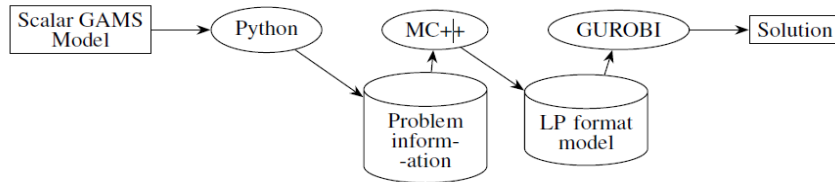


Figure 1: Workflow for reformulation of MIPOP as MIQCQP and solution using GUROBI

2.3. Testing

In order to test the effectiveness of this approach, 60 test instances of MIPOPs from MINLPLib (GAMS Development Corp., 2018) are considered, which correspond to a wide variety of applications. These instances comprise between 3--961 variables and 0--5329 constraints and have a degree ranging between 2--8. All of the computations are performed on a dual 12-core Intel® Xeon® CPU E5-2650 v4 @2.20GHz with 196GB of RAM and running Centos 7.8. The list of all instances used for testing and the corresponding '.lp' files can be accessed from https://github.com/tk3016/Test_instances_ESCAPE31.git.

In solving the reformulated MIQCQPs using GUROBI, only 1 thread is exploited, denoted as 'Gur-1'. This is to allow for a more transparent comparison with the state-of-the-art global solvers BARON (v20.4.14), ANTIGONE (v1.1) and SCIP (v7.0) in GAMS (v32.2) that only take advantage of multi-threading in solving the MIP relaxations. In this study, multi-threading is not exploited in state-of-the-art global solvers. For all the solvers considered, *Absolute termination tolerance* is set at 10^{-9} , *Relative termination tolerance* is set at 10^{-4} , *Feasibility tolerance* is set at 10^{-5} and a *Time limit* of 3600s or 1 hour is imposed.

Performance is assessed using performance profiles (Dolan and Moré, 2002). The shifted geometric mean (SGM) (Mittelmann, 2020) over all the test instances is also calculated for each solver based on the wall-time:

$$SGM = \exp\left(\frac{\sum_{i=1}^n \ln \ln((1, v_i + s_h))}{n}\right) - s_h, \quad (15)$$

where v_i is the wall-time taken to solve instance i , n is the total number of instances considered and s_h is a non-negative shift parameter, set to 5 in this paper.

3. Results and Discussion

As outlined in the previous section, first all the GAMS models are read using a Python script such that all relevant information about the problem can be collected and passed to MC++ for reformulation. It is noted that the time taken to process the reformulation is < 1 s for all instances except *transswitch0030r* (1.222 s), *transswitch0039r* (1.805 s) and *unitcommit2* (17.234 s). For *unitcommit2*, the time is unusually high because it

comprises 960 variables, 5329 linear constraints and a polynomial objective with 480 monomials of degree ≥ 2 and 720 monomials of degree 1. Since the time taken to process the re-formulation is negligible for majority of the instances, it is excluded from the discussion henceforth. Based on the performance profiles in Fig. 2, it is observed that the largest proportion of problems is solved within 1s in Gur-1, with the second largest proportion achieved with BARON. Within the maximum time limit of 1 hour, BARON solves 68% of the instances, with Gur-1 a close second at 65%. Similar behaviour is observed when considering the SGM in absolute and relative terms (Table 1), with the best performance achieved with BARON, followed by Gur-1, SCIP and ANTIGONE. Focusing on specific problem types, Gur-1 is found to perform worse than current state-of-the-art solvers for periodic scheduling of continuous multiproduct plants problems. Among solved instances of this type, BARON and SCIP can solve *multiplants_mtg1a* in approximately 1300s whereas Gur-1 and ANTIGONE are unable to solve the instance within 1 hour. BARON, ANTIGONE and SCIP can solve *multiplants_mtg2* in 1703s, 24s and 1375s respectively which is significantly faster than Gur-1 which solves it in 3573s. For unsolved instances of this type (*mtg1b*, *mtg1c*, *mtg5*, *mtg6*),

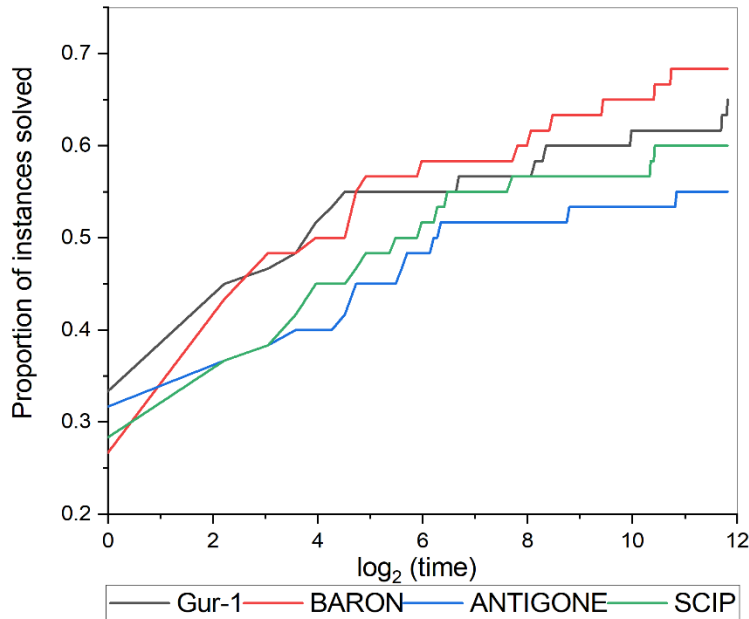


Figure 2: Performance profile for the test set showing proportion of instances solved as a function of wall-time on a log₂ scale

Solver	BARON	SCIP	ANTIGONE	Gur - 1
Shifted Geometric Mean	2.01	3.25	4.04	2.34
% reduction (Gur-1)	-16	28	42	-

Table 1: Shifted Geometric Mean of wall-time (s) over the test set for the five runs. % reduction (Gur-x) refers to the percentage change in SGM between Gur-x and the algorithm indicated in the relevant column, relative to that algorithm.

the relative gap ($|(Primal\ Bound - Dual\ Bound)/Primal\ Bound| \times 100$) is significantly larger for Gur-1 when compared to other global solvers. This can be attributed to weaker root node relaxations obtained by GUROBI than for other global solvers. The spatial branch and bound-based global solvers exploit several pre-processing techniques that reduce the bounds on variables significantly. Consequently, this leads to tighter relaxations and stronger cutting planes. Such pre-processing strategies are not employed with the reformulation approach, nor is an incumbent solution made available at the root node or at the start of the B&B tree. Despite this, the reformulation process is found to perform more effectively in most cases than direct solution of the MIPOP with state-of-the-art solvers. This demonstrates the merit of using this approach.

4. Conclusion

A reformulation of MIPOPs into MIQCQPs and their solution with a specialized algorithm has been proposed and tested for a varied set of problems. Comparison with direct solution of the MIPOPs indicates that the reformulation approach is a promising strategy. For the case of single-threading (Gur-1), the performance is better than ANTIGONE and SCIP but marginally worse than BARON. There is potential to further improve its effectiveness by incorporating pre-processing strategies. The approach can be further developed by automating the workflow and using pre-processing strategies will be investigated. Finally, this approach can potentially be integrated within global solvers used to solve MIPOPs since reformulation can generally be carried out quickly and efficiently.

Acknowledgements: TK is grateful for a PhD Scholarship from the Department of Chemical Engineering, Imperial College London. The authors would also like to thank Terrence Crombie for all his assistance.

References

- P. M. Castro, A. Q. Novais, 2009. Scheduling multistage batch plants with sequence-dependent changeovers. *AIChE Journal* 55 (8), 2122–2137.
- B. Chachuat, B. Houska, R. Paulen, N. Perić, J. Rajyaguru, M. E. Villanueva, 2015. Set-theoretic approaches in analysis, estimation and control of nonlinear systems. *IFAC-PapersOnLine* 48 (8), 981–995.
- E. D. Dolan, J. J. Moré, 1 2002. Benchmarking optimization software with performance profiles. *Mathematical Programming, Series B* 91 (2), 201–213.
- G. Gamrath, D. Anderson, et al, March 2020. The SCIP Optimization Suite 7.0. Technical report, OptimizationOnline. URL http://www.optimization-online.org/DB_HTML/2020/03/7705.html
- GAMS Development Corp., 2018. MINLPLib: A Library of Mixed-Integer and Continuous Nonlinear Programming Instances. URL <http://www.minlplib.org/http://www.minlplib.org>
- Gurobi Optimization LLC, 2020. Gurobi optimizer reference manual. URL <http://www.gurobi.com>
- O. Knüppel, 1994. PROFIL/BIAS-A fast interval library. *Computing* 53 (3-4), 277–287.
- M. R. Kılınc, N. V. Sahinidis, 5 2018. Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with BARON. *Optimization Methods and Software* 33 (3), 540–562.

- R. Misener, C. A. Floudas, 2014. ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations. *Journal of Global Optimization* 59 (2-3), 503–526.
- H. Mittelmann, 2020. Shifted Geometric Mean. URL <http://plato.asu.edu/ftp/shgeom.html>
- H. Nagarajan, M. Lu, S. Wang, R. Bent, K. Sundar, 2019. An adaptive, multivariate partitioning algorithm for global optimization of nonconvex programs. *Journal of Global Optimization* 74 (4), 639–675.
- P. Rumschinski, S. Borchers, S. Bosio, R. Weismantel, R. Findeisen, 2010. Set-based dynamical parameter estimation and model invalidation for biochemical reaction networks. *BMC Systems Biology* 4, 69.
- H. D. Sherali, C. H. Tuncbilek, 1992. A global optimization algorithm for polynomial programming problems using a reformulation-linearization technique. *Journal of Global Optimization* 2 (1), 101–112
- J. P. Teles, P. M. Castro, H. A. Matos, 2013a. Multi-parametric disaggregation technique for global optimization of polynomial programming problems. *Journal of Global Optimization* 55 (2), 227–251.
- J. P. Teles, P. M. Castro, H. A. Matos, 2013b. Univariate parameterization for global optimization of mixed-integer polynomial problems. *European Journal of Operational Research* 229 (3), 613–625.
- A. Uribe-Rodríguez, P. M. Castro, G. Guillén-Gosálbez, B. Chachuat, 2020. Global optimization of large-scale MIQCQPs via cluster decomposition: Application to short-term planning of an integrated refinery-petrochemical complex. *Computers & Chemical Engineering* 140, 106883.

A data-based approach for the development of asynchronous and distributed optimization algorithms in complex engineering applications

Ioannis Geroulis, Georgios Panayiotou, Antonis C. Kokossis,*

School of Chemical Engineering, National Technical University of Athens, Zografou, Athens 15772, Greece

Abstract

The paper outlines a generic optimization approach with extensive capabilities to distribute the computational work in the optimization, also to explore interim data generated by the optimization algorithms without penalizing search times and effort. The approach postulates a network of autonomous computing nodes and data pools (data repositories); nodes and pools respectively generate and host interim solutions that are redistributed in the network with a parallel use of data analytics that constitute asynchronous tasks. The approach is demonstrated in complex problems of (single and multiphase) reactor network synthesis. The algorithm is implemented using an infrastructure of multiple CPU cores. Results confirm significant reductions of computing times due to parallelization, also the benefit to explore cost-free analytics over data generated internally. Future work includes a more systematic and extensive use of machine learning, also the use of dynamic networks of nodes and pools.

Keywords: optimization, parallel and asynchronous search, data analysis, reactor networks.

1. Introduction and background

Optimization methods remain in the core of process systems engineering addressing a wide range of problems in process and product design (synthesis and process development, process integration and intensification) as well as in operations and process dynamics. Rather than relying on over-the-self algorithms, chemical engineering has pioneered novel optimization technology featuring advanced methods that reach out to the wider systems community. Floudas and Gounaris, 2009 offer an extensive list of methods presenting strong evidence that chemical engineering contributions have been major and lasting. Optimization algorithms include stochastic optimization algorithms invariably applied by means of Simulated Annealing, Tabu Search, Genetic algorithms and/or hybrid methods. The latter algorithms face the difficulty to choose parameter values so that convergence is guaranteed while deterministic algorithms face the challenge to transcend local optima and to ensure that the solution is globally optimum. Optimization problems in chemical reactor design rank among the hardest to tackle due to the complexity and nonlinearity of their state space as well as the numerous degrees of freedom to fathom (types of reactors, macro and micro-mixing, multiple processing paths and kinetics, trade-offs between mass transfer and chemical reactions etc.). Accordingly, reactor design sets up an excellent background to benchmark optimization methods especially as both deterministic (Achenie and Biegler, 1986) and stochastic methods (Ashley and Linke, 2004) are available to compare and test efficiency.

Most optimization algorithms, whether deterministic or stochastic, typically involve sequences of internal (local search) steps, namely tasks that repeat and iterate before convergence is pulled off once they meet appropriately selected criteria (e.g. KKT conditions and integrality gaps in deterministic methods; statistical targets in standard deviation for stochastic methods). Local search tasks generate significant volumes of interim data. Those are hardly explored by conventional algorithms notwithstanding an apparent potential for acceleration and efficiency: data can be explored to tune local search parameters (that are usually set by heuristics), share interpretation of the state space (that could be complex), and/or improve confidence in the solution quality. However, efficiency gains are debatable as any meaningful use of data analytics entails the introduction of expensive and interruptive steps, possibly more demanding than the standard local search tasks. Accordingly, reported efforts are so far restricted to offline analysis with a view to extract rules and knowledge from solution sets of stochastic optimization experiments (Ashley and Linke, 2004) or to merely reduce degrees of freedom in synthesis problems (Kokossis et al., 2016). Separate work has pointed out into better prospects to explore data structures in optimization (Cecelja et al, 2014) by completely reformulating the algorithm to the purpose. The paper responds to the challenge by presenting an algorithm based on a network of autonomous units that includes data pools and computing nodes. Numbers of pools and computing nodes are design parameters of the algorithm. Units generate, host, and redistribute interim solutions. Computing nodes are asynchronous and operate in parallel to generate interim solutions stored in data pools. Pools supply initial points for each local search while interim solutions may transfer across different pools based on their statistical properties and their solution quality. The algorithm is possible to demonstrate advantages provided there is access to a distributed network of CPUs, also on the assumption that the optimization problem to solve is reasonable complex or large. The following sections include the presentation of the methodological approach, the steps of its implementation and results in the application of the algorithm to optimize complex reactor networks and multiphase reaction systems.

2. Methodological Approach

The optimization problem is addressed in a general form:

$$\max f(x, y), \quad s. t. \{ \mathcal{H}(x, y) = 0, \mathcal{G}(x, y) \leq 0, x \in \mathbb{R}^n, y \in \{0,1\}^n \} \quad (1)$$

The set in Eq. (1) defines its feasible region

$$\mathcal{S} = \{(x, y) | \mathcal{H}(x, y) = 0, \mathcal{G}(x, y) \leq 0, x \in \mathbb{R}^n, y \in \{0,1\}^n\} \quad (2)$$

Let us further introduce

- a set $J = \{j\}$ of P_j computing nodes

$$\tilde{P} = \{P_j\}, \quad j = 1, 2, \dots, n_m \quad (3)$$

- a set $I = \{i\}$ of data pools \tilde{S}_i containing elements $s_{i,k}$ in \mathcal{S} so that

$$\tilde{S}_i = \{s_{i,k} \in \mathcal{S}, i \in I, k = 1, 2, \dots, n_L^p\} \quad (4)$$

where $n_L^p = \text{card}(\tilde{S}_i)$ is the maximum population of solutions in each pool and n_p and n_m are respectively the total number of pools and computing nodes. The nodes assigned to each pool \tilde{S}_i are denoted by $N_{i,l}$, $i = 1, \dots, n_p$ & $l = 1, \dots, n_c$, n_c is the total number of nodes corresponding to a single pool. Let $h_j: \mathcal{S} \rightarrow \mathcal{S}$ be a (local) search function assigned

at each computing node P_j . The application of h_j on a $s_{i,k} \in \tilde{S}_i$ yields a new element in \tilde{S}_i , namely

$$h_j: s_{i,k} \in \tilde{S}_i \xrightarrow{P_j} h_j(s_{i,k}) \in \tilde{S}_i \quad (5)$$

The successive application of h_j constitutes a Markov chain g^{L_j} with length L_j . Given the P_j , h_j and $s_{i,k} \in \tilde{S}_i$ the inclusion of $s'_{i,k}(\tilde{S}_i)$ will k -expand each pool \tilde{S}_i (Fig. 1) to

$$\tilde{S}'_i = \tilde{S}_i U\{h_j(s_{i,k})\} = \tilde{S}_i \quad (6)$$

Each element $s_{i,k,0} \in \tilde{S}_i$ is a candidate to process through available nodes P_j where successive applications of h_j yield new internal solutions along the lines:

$$g^{L_j}_{i,j,t}(s_{i,k,0}) = s_{i,k,0} \rightarrow s_{i,k,1} \rightarrow s_{i,k,2} \rightarrow \dots \rightarrow s_{i,k,L_j}, \quad t = 1, 2, \dots, L_j \quad (7)$$

The new interim solutions s_{i,k,L_j} are assessed by the Metropolis criteria (Metropolis et al, 1953).

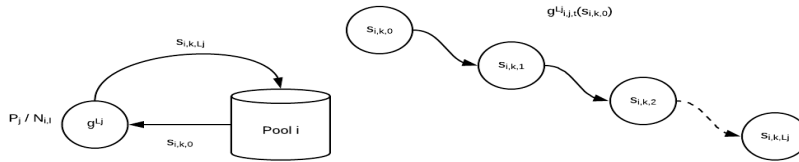


Figure 1: The k – expansion of pool \tilde{S}_i through a computing node and a Markov chain.

Each pool constitutes an autonomous system that can be connected to each other and it is also connected to a particular number of the available computing nodes. Local searches can be carried out in parallel and are asynchronous as there is no need to stall and wait for any other tasks to complete or start. Indeed, additional parallel tasks can be assigned to explore clustering, sorting, reduction, or any other type of data analytics. The analysis constitutes an asynchronous and independent step from the pool population expansion and the search for additional interim data. A clustering approach on \tilde{S}_i will partition the pool population into disjoint clusters

$$\tilde{S}_i = U_l \tilde{S}^P_{i,l} \quad (8)$$

$\tilde{S}^P_{i,l}$ denotes the disjoint partitions in \tilde{S}_i . For pools connected with each other partitions can be transferred across pools. Let two pools \tilde{S}_i , \tilde{S}_n with $i, n \in I$, which are connected through a path $(i, n) \in C_{i,n}$. Let the path $C_{i,n}$ indicates all permissible directions partitions can follow, essentially controlling data traffic from each pool \tilde{S}_i to \tilde{S}_n . Assuming a clustering approach is applied on \tilde{S}_i , a disjoint partition $\tilde{S}^P_{i,n}$ is termed as an *inflection* of \tilde{S}_i on \tilde{S}_n . A general formulation for the k -inflection is accordingly,

$$\tilde{S}'_i = U_l \tilde{S}^P_{i,l} - \tilde{S}^P_{i,n}, \quad \tilde{S}'_n = \tilde{S}_n U_l \tilde{S}^P_{i,n}, \quad l \neq n \quad (9)$$

Network connections are degree of freedom of the approach but is defaulted to serial connections in the illustrations that follow. A design parameter T_i is assigned to account for the level of acceptance allowed for new entries in each pool. The network structure of

Figure 2 is accordingly the default network tested in all illustrations and is defined by \mathbb{C} as

$$\mathbb{C} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_{n_p}\}, \quad T_i < T_l, \quad \forall i > l \quad (10)$$

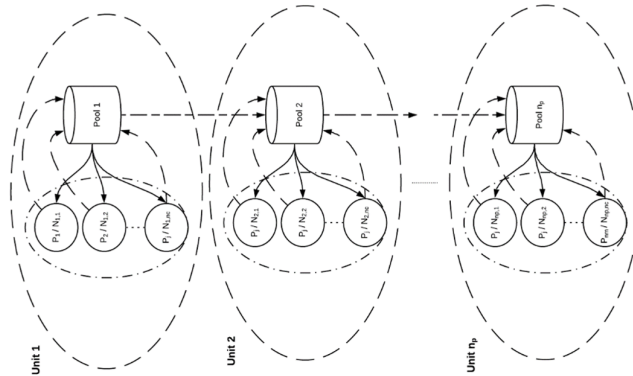


Figure 2: Mapping between autonomous units of the network \mathbb{C} .

3. Implementation, illustration and results

Design parameters for the optimization algorithm included

- The number of pools and nodes: the algorithm used 5-10 pools assigning 1-16 cores per pool. Markov chains ranged from 30-100 iterations.
- Pool parameters (mainly T_i): assignments ranged from 10^3 down to 10^{-4} (using logarithmic scales between successive assignments)
- Type and frequency of sorting and clustering tasks: solutions were sorted and partitioned discarding the lower 10% and transferring the top 20% over to pools with lower T_i
- Upper and lower bounds for the population in the pool: 15-100 solutions
- Parameters associated with convergence criteria

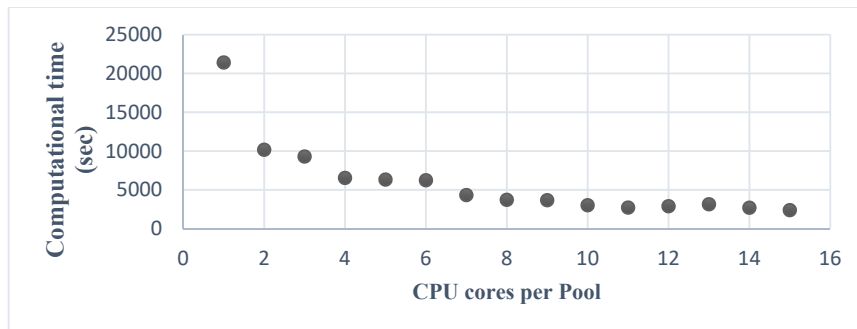


Figure 3: Computational time vs number of cores

The algorithm is implemented on python 2.7 using SQLite3 to create a local database (refraining the need for an internet connection) while computations relied on the National Infrastructures for Research and Technology (GRNET). GRNET operates ARIS (Advanced Research Information System), a high-performance system to support that

offers 80 cores with Intel Xeon E5-2680v2 10C 2.8GHz processors. Termination criteria included upper limits on the computational time, upper and lower limits on the number of tasks, progress in the objective function overall and individually at each pool (best stored and average). The method has been tested against complex problems in reactor network design (Mehta and Kokossis, 1997) using a superstructure approach. The engineering problem features multiphase chemical reactors, mass exchange units, mixers, and splitters. The set of equations consists of nonlinear equations with few discrete options. The domain addresses both single-phase and multi-phase applications and can be extended to reactive separations systems. The algorithm has been compared with standard annealing algorithm that is unable to parallelize but is empowered to allocate up to ~10,000 stages of calculation stages with fixed acceptance criteria (e.g. equivalent to the pools proposed in the algorithm).

Illustration 1. The purpose has been to test the efficiency of the algorithm to converge but also the parallelization that can be achieved. The reaction system included the following reactions and initial feed concentrations: $A \rightarrow B$ (1), $r_1 = k_1 C_A$, $k_1 = 0.05 \text{ s}^{-1}$, $A_{in,(l)} = 1 \frac{\text{kmol}}{\text{h}}$; $B \rightarrow C$ (2), $r_2 = k_2 C_B$, $k_2 = 0.03 \text{ s}^{-1}$, $B_{in,(l)} = 0 \frac{\text{kmol}}{\text{h}}$; $C \rightarrow A$ (3), $r_3 = k_3 C_C$, $k_3 = 2 \text{ s}^{-1}$, $C_{in,(l)} = 10 \frac{\text{kmol}}{\text{h}}$. The algorithm managed to converge with only 10 improvements per pool (while results of the parallelization are summarized in Fig. 3. The reason behind the faster performance has been the significant number of unsuccessful simulations for both algorithms. The computational time drops as more cores are added, sharper initially (as 2 and 3 cores are assigned per pool) and continuing to decrease, almost linearly to the point in which 10 cores are assigned per pool (e.g. at a level equal to the maximum number of iterations) and decreasing further even as the bound is surpassed. The distribution of internal solutions is presented in Figure 4. Markov chains ranged from 30-100: in shorter Markov chains (L=30) the new algorithm matched standard annealing but proved 5-6 times slower; it has been faster for higher chains (L=100) where it also converged to consistently better solutions.

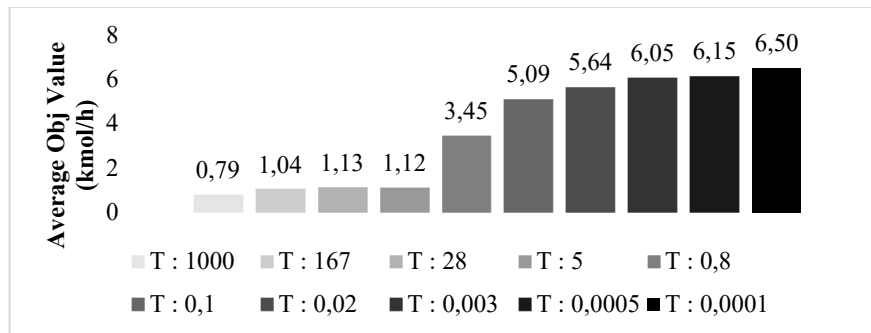


Figure 4: Population of internal solutions for Illustration 1

Illustration 2. The purpose of the illustration has been to test the algorithm in a much more demanding problem from the literature that involved two phases: (g)-gas and (l)-liquid. The reaction system included the following reactions and initial feed concentrations: $A + F \xrightarrow{k_1} B + H$ (1), $r_1 = k_1 C_A C_F$, $k_1 = 10 \frac{\text{m}^3}{\text{kmol h}}$; $A + 2F \xrightarrow{k_2} D + 2H$ (2), $r_2 = k_2 C_A C_F^2$, $k_2 = 5 \frac{\text{m}^3}{\text{kmol h}}$; $B + F \xrightarrow{k_3} C + H$ (3), $r_3 = k_3 C_B C_F$, $k_3 = 2 \frac{\text{m}^3}{\text{kmol h}}$; $B + 2F \xrightarrow{k_4} E + 2H$ (4), $r_4 = k_4 C_B C_F^2$, $k_4 = 1 \frac{\text{m}^3}{\text{kmol h}}$; $F_{in,(g)} = 200 \frac{\text{kmol}}{\text{h}}$, $A_{in,(l)} = 100 \frac{\text{kmol}}{\text{h}}$.

The distribution of internal solutions is presented in Figure 5. As in all other examples, a higher concentration of solutions is observed in the lower pools. Markov chains ranged from 40-100: in all Markov chains the new algorithm has been slower but consistently producing 5-20% better solutions than those reported in the literature.

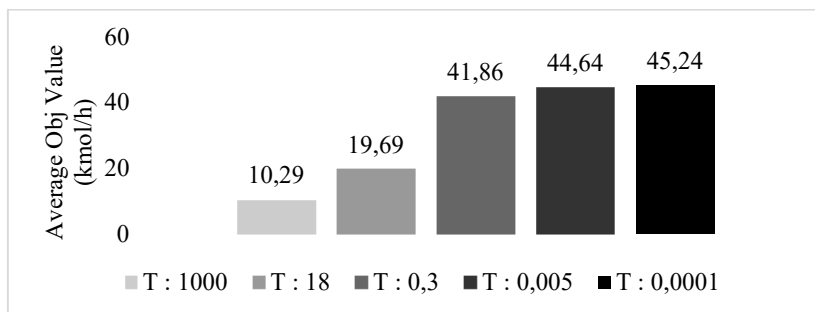


Figure 5: Distribution of internal solutions in Illustration 2

4. Conclusions

In all examples the analysis of the interim data produced in the pools remained largely unexplored. By a similar token, the options to connect the pools with computer nodes in all-inclusive networks has not been researched. The level of analytics employed has remained basic and could benefit from machine learning methods available to cluster solutions, extract problem features and/or adjust the algorithm in a dynamic fashion. Even at its basic form, the algorithm managed to outperform over-the-self developments and improve literature results. Its advantages include a significant degree of parallelization and its potential to invite data analytics without essential penalties on the search.

Acknowledgement

The authors gratefully acknowledge financial support from Marie Curie European Research Program, RENESING II(H2020-MSCA-RISE).

References

- Achenie, L. E., & Biegler, L. T. (1986). *Industrial & engineering chemistry fundamentals*, 25(4), 621-627.
- Ashley, V. M., & Linke, P. (2004). *Chemical Engineering Research and Design*, 82(8), 952-960.
- Cecelja, F., Kokossis, A., Du, D., & Yang, S. (2014). *Computers & chemical engineering*, 66, 276-289.
- Floudas, C. A., & Gounaris, C. E. (2009). *Journal of Global Optimization*, 45(1), 3.
- Kokossis, A., Labrador-Darder, C., & Cecelja, F. (2016). *Computers & Chemical Engineering*, 93, 64-86.
- Mehta, V. L., & Kokossis, A. (1997). *Computers & chemical engineering*, 21, S325-S330.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). *The journal of chemical physics*, 21(6), 1087-1092.

31ST EUROPEAN SYMPOSIUM ON
COMPUTER AIDED PROCESS
ENGINEERING

PART B

31ST EUROPEAN SYMPOSIUM ON
COMPUTER AIDED PROCESS
ENGINEERING

PART B

Edited by

Metin Türkay
Koç University

Rafiqul Gani
PSE for SPEED



ELSEVIER

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

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

Copyright © 2021 Elsevier B.V. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

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

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

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

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN (Part B): 978-0-323-98326-9

ISBN (Set) : 978-0-323-88506-5

ISSN: 1570-7946

For information on all Elsevier publications visit our website at <https://www.elsevier.com/>



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

Typeset by STRAIVE

Contents

106.	How tools and technologies help drive sustainability <i>Julien de Beer</i>	669
107.	A homotopy continuation-based branch and bound algorithm for process synthesis and design using rigorous unit operation models <i>Yingjie Ma, Nan Zhang, Jie Li</i>	677
108.	ROmodel: A Python Robust Optimization Modeling Toolbox <i>Johannes Wiebe, Ruth Misener</i>	683
109.	Modelling -- Mechanistic reductionism extended with holistic empirical components <i>Heinz A. Preisig</i>	689
110.	Thermodynamic Optimization of a Biomass-based Integrated Gasification Combined Cycle with Post Combustion Carbon Capture using Potassium Carbonate <i>Ikhlas Ghiat, Ahmed AlNouss, Tareq Al-Ansari</i>	695
111.	The state of play towards modeling of air quality in the city of Rabat <i>Soraya el Hamdouni, Essediya Cherkaoui, Abderrahmane Nounah and Mohamed Khamar</i>	703
112.	Machine learning applications and process intelligence for cement industries <i>Christos Chatzilenas, Thanasis Gentimis, Theodore Dalamagas, Antonis C. Kokkosis, Alexandros Katsiaboulas, Ioannis Marinos</i>	711
113.	Process Modelling and Simulation of Volatile Organic Compound (VOC) Recovery from Pharmaceutical Gas Emission Streams <i>Vasiliki E. Tzanakopoulou, Daniel Castro-Rodriguez, Dimitrios I. Gerogiorgis</i>	717
114.	Multi-zonal compartmentalization methodology for surrogate modelling in continuous pharmaceutical manufacturing <i>Pooja Bhalode, Marianthi Ierapetritou</i>	725
115.	ProMo variable/equation Ontology-based system modelling <i>Niloufar Abtahi, Heinz A. Preisig</i>	733
116.	Multi-scenario Design Optimization using ADMM of a Thermal Energy Storage System <i>Sandeep Prakash, Dinesh Krishnamoorthy, Johannes Jäschke</i>	739

117. Performance comparison of different optimization methods for the wastewater treatment plant model based on the Activated Sludge Model No. 3 and design of the plant control	
<i>Melinda Simon-Várhelyi, Vasile Mircea Cristea, Alexandra Veronica Luca, Marius Adrian Brehar</i>	747
118. Comparison of Group-Contribution and Machine Learning-based Property Prediction Models with Uncertainty Quantification	
<i>Adem R. N. Aouichaoui, Resul Al, Jens Abildskov, Gürkan Sin</i>	755
119. Intelligent Management Platform for Material Exchange Optimization and Industrial Symbiosis	
<i>Theodore Chatzidimitriou, Thanasis Gentimis, Christos Michalopoulos, Antonis C. Kokossis, Theodore Dalamagas</i>	761
120. Towards a Software Prototype for Synthesis of Operable Process Intensification Systems	
<i>Shivam Vedant, Mary Rivera Atencio, Yuhe Tian, Vaishnav Meduri, Efstratios N. Pistikopoulos</i>	767
121. Data-driven modelling of choke valve erosion using data simulated from a first principles model	
<i>Jan Henrik Jahren, Jose Matias, Johannes Jäschke</i>	773
122. A Metamodelling Approach for Air Pollution Source Estimation	
<i>Kuang Cheng, Yi Cao, Shuanghua Yang</i>	779
123. Uncertainty analysis in life-cycle assessment of early-stage processes and products: a case study in dialkyl-imidazolium ionic liquids	
<i>Husain Baaqel, Jason P. Hallett, Gonzalo Guillén-Gosálbez, Benoît Chachuat</i>	785
124. Interval-Valued Data Learning for Robustness of Energy Recovery Systems	
<i>Rachid Ouareta, Pascal Floqueta, Jean-Pierre Belauda, Stéphane Negny</i>	791
125. Rules for Predicting Benefits in Simultaneous versus Sequential Heat Integration	
<i>Hendrik Pötting, Avinash S. R. Subramanian, Truls Gundersen</i>	799
126. Processing systems design considering resilience	
<i>Ákos Orosz, Jean Pimentel, Ferenc Friedler</i>	807
127. Applying quality assurance concepts from software development to simulation model assessment in smart equipment	
<i>Jonathan Mädler, Isabell Viedt, Leon Urbas</i>	813
128. Studying the Synergy between Dimension Reduction and Clustering Methods to Facilitate Fault Classification	
<i>Melisa Mollaian, Gyula Dörgő, Ahmet Palazoglu</i>	819

- 129. Computational Fluid Dynamics Modelling of Phenol Oxidation in a Trickle-Bed Reactor using 3D Eulerian Model**
Tladi J. Makatsa, Siwela J. Baloyi, Cornelius M. Masuku 825
- 130. Enhanced deterministic approach for heat exchanger network synthesis**
Zekun Yang, Nan Zhang, Robin Smith 833
- 131. Modelling Biogas and H₂S in Anaerobic Digestion**
Bozzano Giulia Luisa, Moretta Federico, Rizzo Eleonora and Manenti Flavio 839
- 132. Implementation of first-principles surface interactions in a hybrid machine learning assisted modelling of flocculation**
Nima Nazemzadeh, Rasmus Fjordbak Nielsen, Krist V. Gernaey, Seyed Soheil Mansouri, Martin P. Andersson 845
- 133. Optimal physical property data for process simulations by optimal experimental design**
Lorenz Fleitmann, Jan Pyschik, Ludger Wolff, André Bardow 851
- 134. On the practical identifiability of kinetic models of hydroxymethylfurfural hydrogenation in batch reaction systems**
Philipp Deussen, Federico Galvanin 859
- 135. Continuous-Effort Approach to Model-Based Experimental Designs**
Kennedy Putra Kusumo, Kamal Kuriyan, Salvador García-Muñoz, Nilay Shah, Benoît Chachuat 867
- 136. Application of Machine Learning and Global Sensitivity Analysis for Identification and Visualization of Design Space**
Sergei Kucherenko, Oleksiy Klymenko, Nilay Shah 875
- 137. A Hybrid Model Coupling Advanced Oxidation Processes (AOP) and Conventional Bio-processes for the Removal of Recalcitrant Contaminants in Wastewaters**
Kourosh Nasr Esfahani, Montserrat Pérez-Moya, Moisès Graells 883
- 138. Analysis of the dissolution mechanism in the phosphoric acid manufacturing process: modelling and simulation**
Sanae Elmisaoui, Abderrazak M. Latifi, Lhachmi Khamar, Mohamed Salouhi 891
- 139. Focusing experiments in the early phase process design by process optimization and global sensitivity analysis**
Stefanie Kaiser, Tabea Menzel, Sebastian Engell 899
- 140. Design Space Approximation with Gaussian Processes**
Panagiotis Demis, Sergei Kucherenko, Oleksiy V. Klymenko 905

- 141. Employing Adaptive Just-In-Time-Learning in a Transfer Learning Frame for Soft-Sensor Design**
Burak Alakent 913
- 142. AC-Optimal Power Flow Solutions with Security Constraints from Deep Neural Network Models**
Zachary Kilwein, Fani Boukouvala, Carl Laird, Anya Castillo, Logan Blakely, Michael Eydenberg, Jordan Jalving, Lisa Batsch-Smith 919
- 143. The impact of CO₂ pricing in SC Resilience – An optimisation model**
João Pires Ribeiroa, Ana Barbosa-Póvoa 927
- 144. Machine learning-based approach to identify the optimal design and operation condition of organic solvent nanofiltration (OSN)**
Changsu Kim, Chanhee You, Do Thai Ngan, Minseong Park, Dongjun Jang, Sungju Lee, Jiyong Kim 933
- 145. Development of System Dynamic Model for Sustainability driven Technology Adoption in Indian Transport Sector**
Nandita Saraf, Yogendra Shastri 939
- 146. Multi-Period Design Optimization of Flexible Fixed-Bed Reactors by Stoichiometry-Based Model Reduction**
Ronny T. Zimmermann, Jens Bremer, Kai Sundmacher 947
- 147. Optimization based sampling for gray-box modeling using a modified upper confidence bound acquisition function**
Joschka Winz, Sebastian Engell 953
- 148. How to design cooperative sustainable industrial systems?**
Marianne Boix, Florent Mousqué, Ludovic Montastruc, Stéphane Négny 959
- 149. Population Balance Modelling of Pan Granulation Processes**
Eric Otto, Robert Dürr, Mateusz Przywara, Dorota Antos, Achim Kienle 965
- 150. Estimation of cutpoint temperature under uncertain feed composition and process conditions using artificial intelligence methods**
Junaid Shahzad, Iftikhar Admad 971
- 151. Application of Constrained EKF based State Estimation to a Coiled Flow Inverter Copolymerization Reactor**
Robin Semrau, Filippo Tamagnini, Alexandru Tatulea-Codrean, Sebastian Engell 977
- 152. Towards Automated Information Retrieval of Process Data and Knowledge from Academic Databases**
Fabian Lechtenberg, Javier Farreres, Ana Somoza-Tornos, Adrián PachecoLópez, Antonio Espuña, Moisès Graells 983

- 153. Prediction of Bioconcentration Factors (BCF) using Graph Neural Networks**
E. I. Sanchez Medina, S. Linke, K. Sundmacher 991
- 154. Data and environment based approach for process systems engineering**
Nancy Prioux, Jean-Pierre Belauda, Rachid Ouareta, Gilles Hétreux 999
- 155. A Data-based Predictive Model for Distillation Column of Bio-based 2, 3-Butanediol**
Yeongryeol Choi, Hyundo Park, Jiwon Roh, Jongkoo Lim, In-su Han, Duk-Ki Kim, Sangjun Jeon, Hee-Geun Nam, Il Moon, Hyungtae Cho, Junghwan Kim 1005
- 156. Data-driven Process Design Exemplified on the Steam Methane Reforming Process**
Laurens Lueg, Dominik Schack, Evrim Örs, Robin Schmidt, Patricia Bickert, Martin von Kurnatowski, Patrick Otto Ludl, Michael Bortz 1013
- 157. Input transformation for linearization, decoupling and disturbance rejection with application to steam networks**
Cristina Zotică, Sigurd Skogestad 1021
- 158. SEARCH: A Symptom-based Expert for Advanced Response to Chemical Hazards**
Sangwoo Yoo, Eunji Shin, Dongil Shin 1029
- 159. A Cloud Computing Application for the Supercritical Carbon Dioxide Extraction Using Coffee Grounds Silverskin**
Andrea Galeazzi, Rita Nasti, Giulia L. Bozzano, Luisella Verotta, Stefania Marzorati and Flavio Manenti 1035
- 160. Multi-objective optimization of the shape of a fixed-bed reactor**
Alexis Courtais, François Lesage, Yannick Privat, Cyril Pelaingre, Abderrazak M. Latifi 1041
- 161. Artificial Intelligence Based Prediction of Exergetic Efficiency of a Blast Furnace**
Muhammad Salman Arif, Iftikhar Ahmad 1047
- 162. AI System for Supporting Generation of Optimal Synthetic Pathways Based on Chemical Reaction Big Data**
Nagyeong Lee, Joonsoo Jeong, Dongil Shin 1053
- 163. Application of data science to study fluorine losses in the phosphate industry**
Houda Ariba, Paul Vanabelle, Salah Benaly, Thomas Henry, Cédric R. André, Grégoire Leonard 1059
- 164. Data-driven Branch-and-bound Algorithms for Constrained Simulation-based Optimization**
Jianyuan Zhai, Sachin Shirpurkar, Fani Boukouvala 1067

165. **Reactor scale modelling of gas hydrate-based CO₂ capture (HBCC) process**
Hossein Dashti, Gerald Fraser, Amirpiran Amiri 1073
166. **SMITH: differential homotopy and automatic differentiation for computing thermodynamic diagrams of complex mixtures**
Olivier Cots, Joseph Gergaud, Nataliya Shcherbakova 1081
167. **Dynamic modeling of a three-phase gas-solid-liquid fluidized bed absorber for CO₂ capture**
Ana-Maria Cormos, Flavia-Maria Ilea, Simion Dragan 1087
168. **Steady and Dynamic State Simulation-based Study of the Coupling of Electrolysis and Methanation Processes within Power to Methane (PtM) Context**
Maria Paula Novoa, Camilo Rengifo, Manuel Figueredo 1093
169. **Upgraded Meta-Learning based Surrogate Selection Paradigm (LEAPS2v2)**
Maaz Ahmad, Iftekhar A. Karimi 1099
170. **Flexibility analysis of chemical processes considering dependencies between uncertain parameters**
Christian Langner, Elin Svensson, Simon Harvey 1105
171. **Computational Characterization of a Pore Network Model by Using a Fast Nitrogen Porosimetry Simulation**
G. Ledezma, J.J. Verstraete, L. Sorbier, D. Leinekugel-Le Cocq, E. Jolimaitre, C Jallut 1111
172. **Acyclic modular flowsheet optimization using multiple trust regions and Gaussian process regression**
E.I. Sanchez Medina, D. Rodriguez Vallejo, B. Chachuat, K. Sundmacher, P. Petsagkourakis, E.A. del Rio-Chanona 1117
173. **Systematic generation and targeting of chemical recycling pathways: A mixed plastic waste upcycling case study**
Adrián Pacheco-López, Ana Somoza-Tornos, Antonio Espuña, Moisès Graells 1125
174. **Calibration of an EPANET model by using monitoring points in the bulk distribution of Sedibeng Water, for predicting disinfectant decay**
Charl van der Walt, Prof Evans Chirwa, Danie Traut 1131
175. **Compartmental Modelling of Shell Side Fouling in a Shell and Tube Heat Exchanger**
Renat Taurgalinov, Federico Lozano Santamaria, Sandro Macchietto 1137
176. **Optimization-based digital design of a commercial pharmaceutical crystallization process for size and shape control**
Ayşe Eren, Botond Szilagyí, Justin L. Quon, Charles D. Papageorgiou, Zoltan K. Nagy 1143

PROCESS OPERATIONS AND CONTROL

- 177. Condition Number As A Quantitative Measure Of Flexibility In A Process**
*César Ramírez-Márquez, Luz Gabriela Arreola-Nájera,
 Juan Gabriel Segovia-Hernández* 1149
- 178. Nonlinear Model Predictive Control of Food Sterilization Processes**
Chrysovalantou Ziogou, Margaritis Kostoglou, Michael C. Georgiadis 1155
- 179. Design and control of triple-column pressure swing distillation process for acetone-methanol-hexane separation**
Tan Akinciturk, Devrim B. Kaymak 1161
- 180. Design and control of a Biobutanol Purification Process through IBE Fermentation: Basic Design Configuration**
Ilayda N. Oksal, Devrim B. Kaymak 1167
- 181. Fault Diagnosis of Electrical Power Systems with Hybrid Quantum-Classical Deep Learning**
Akshay Ajagekar, Fengqi You 1173
- 182. Data-Driven Robust MPC for Controlled Environment Agriculture**
Wei-Han Chen, Fengqi You 1181
- 183. Initialization strategy for the dynamic optimization of a batch distillation column with a rate-based model**
*Mayra Margarita May-Vázquez, Fernando Israel Gómez-Castro,
 Edna Soraya Rawlings, Vicente Rico-Ramírez, Mario Alberto Rodríguez-Ángeles* 1189
- 184. Adversarial Autoencoder Based Nonlinear Process Monitoring**
Kyojin Jang, Minsu Kim, Hyungjoon Yoon, Jonggeol Na, Il Moon 1195
- 185. Machine learning for constrained self-optimizing control**
Hongxin Su, Chenchen Zhou, Yi Cao, Shuanghua Yang 1203
- 186. Driving Force Based Design and Control Performance Analysis of Reactive Distillation Columns**
*Ashfaq Iftakher, Seyed Soheil Mansouri, Ahaduzzaman Nahid, Anjan K. Tula,
 M. A. A. Shoukat Choudhury, Jay Hyung Lee, Rafiqul Gani* 1209
- 187. Generating an optimal control policy via apprenticeship and reinforcement learning**
Max Mowbray, Robin Smith, Antonio Del Río Chanona, Dongda Zhang 1215

188. Systematic model-based dynamic optimization of a combined cooling and antisolvent multistage continuous crystallization process <i>Jiaxu Liu, Brahim Benyahia</i>	1221
189. Predicting Microclimate of a Closed Greenhouse Using Support Vector Machine Regression <i>Farhat Mahmood, Rajesh Govindan, Tareq Al-Ansari</i>	1229
190. Adjustable Robust Optimal Control for Nonlinear Chemical Process under Uncertainty <i>Enzhi Liang, Zhihong Yuan</i>	1235
191. Adaptive Optimal Control of Baker's Yeast Fermentation Process with Extreme Learning Machine and Recursive Least Square Technique <i>Kazeem Alli, Jie Zhang</i>	1241
192. Data-based Industrial Soft-sensor Design via Optimal Subset Selection <i>Martin Mojto, Karol Lubušký, Miroslav Fikar, Radoslav Paulen</i>	1247
193. Economic MPC with Modifier Adaptation using Transient Measurements <i>Erika Oliveira-Silva, Cesar de Prada, Daniel Navia</i>	1253
194. Attack Detection Using Unsupervised Learning Algorithms in Cyber-Physical Systems <i>Divyang Deep Tiwari, Sourjya Naskar, Arasavelli Siva Sai, Venkata Reddy Palleti</i>	1259
195. Application of Self-Optimizing Control in a Real Heat Exchanger Network <i>Ahmet Coşgun, Oğuz Uzun, Murathan Bağdat,</i>	1265
196. Data-Driven Control Strategies for the Autonomous Operation of the Pharmaceutical Crystallization Process <i>Merve Öner, Gürkan Sin</i>	1271
197. Dynamic optimization of a stand-alone alkaline water electrolyzer for future operation using a predicted solar forecast <i>Haider Niaz, J. Jay Liu</i>	1277
198. A Mixed Integer Dynamic Optimization Approach for a Hybrid-Stand Alone Solar and Wind Powered Alkaline Water Electrolyser for Renewable Hydrogen <i>Haider Niaz, J. Jay Liu</i>	1285
199. A framework for adaptive online model-based redesign of experiments in dynamic systems <i>Arun Pankajakshan, Panagiotis Petsagkourakis, Federico Galvanin</i>	1293

- 200. Investigation of the use of transient process data for steady-state Real-Time Optimization in presence of complex dynamics**
Rodrigo Curvelo, Pedro de A. Delou, Maurício B. de Souza Jr., Argimiro R. Secchi 1299
- 201. Evaluating Control Room Operator Training Outcomes Through Eye Gaze Augmented Multi-Scale Data**
Mohammed Aatif Shahab, Babji Srinivasan, Rajagopalan Srinivasan 1307
- 202. Multi-layer Multivariate Statistical Analysis for Process Monitoring**
Simin Li, Shuanghua Yang, Yi Cao 1313
- 203. Digitalization, Control and Optimization for Cement Plants**
Zhanhao Zhang, Marcus K. Nielsen, Steen Hørsholt, Guruprasath Muralidharan, John B. Jørgensen 1319
- 204. Process Monitoring with Advanced Analytics for Improved Plant Efficiency in the Pharmaceutical Industry**
Deniz Koç, Daniel Castro-Rodriguez, Dimitrios I. Gerogiorgis 1325
- 205. A Study of Spectral Envelope Method for Multi-Cause Diagnosis using Industrial Data**
Gizem Kuşoğlu Kaya, Pelin Döloğlu, Çağla Odabaşı Özer, Ocan Şahin, Ahmet Palazoğlu, Murat Külahçı 1331
- 206. Dynamic Model Based Detection of Cyberattacks in Industrial Facilities**
Kathrin Reibelt, Hubert B. Keller, Veit Hagenmeyer, Jörg Matthes 1339
- 207. Evaluation of control strategies in CO₂ capture unit**
Pavan Kumar Veldandi, Abhisek Roy Chowdhary, Anuradha Durvasula, Rajkumar Krishnamoorthy 1345
- 208. Modelling and Temperature Uncertainty Analysis of a Fermentation Bioreactor at an Industrial Brewery**
Wilfredo Angulo, Dany De Cecchis, Santiago D. Salas, Yris Gozález, Carolina Massay 1351
- 209. Symbolic transfer entropy for root cause analysis of process disturbances**
Ching-Mei Wen, Yuan Yao 1357
- 210. Quality-by-Control of continuous drug substance isolation: study on a novel unit for integrated filtration-drying**
Francesco Destro, Vivian Wang, Mesfin Abdi, Xin Feng, Erin Wood, Simon Coleman, Paul Firth, Alastair Barton, Massimiliano Barolo, Zoltan K. Nagy 1363
- 211. Control of Batch and Continuous Crystallization Processes using Reinforcement Learning**
Brahim Benyahia, Paul Danny Anandan, Chris Rielly 1371

How tools and technologies help drive sustainability

Julien de Beer^a

^aAVEVA, High Cross, Madingley Rd, Cambridge CB3 0HB, United Kingdom

Julien.deBeer@aveva.com

Abstract

The chemicals industry is facing a highly dynamic environment. Demand is continually fluctuating, and the pressure for new sustainable processes and products is rising. New environmental regulations are always expanding and becoming less harmonized globally, and customers are demanding sustainable products that are friendly to the environment. To follow all the market changes and remain competitive, companies need to invest in capital projects to develop new processes and products. However, since it's a capital-intensive industry, new investments in capital projects need to be carefully managed.

Today, companies need to find ways to be more efficient in the execution of the engineering projects. One way is to compress engineering cycles and adapt processes and products to comply with sustainable KPIs and new demands. The Unified Engineering methodology is one option enabled by the latest technologies and tools available. For decades, engineering to design and build industrial plants has been developed through projects with a complex and highly iterative workflow, using siloed solutions. The entire process is very time consuming, and the final deliverable to the owner of the operation is a set of documents (P&IDs, datasheets, 3D model, etc.) with disperse data, with no guarantee of consistency.

In order to compress engineering cycles, a data-centric approach must be used. With this approach, documents and applications are always kept up to date with the latest validated data. The data-centric approach is the first step to apply the Unified Engineering methodology, which will evolve later to the plant Digital Twin. The Unified Engineering methodology uses a single source of information that is available for all the teams involved in the project. Drawings, steady-state and dynamic process simulations, line lists, datasheets, 3D models, and isometrics, among others, are regularly updated with the data available in the unified source of information. Engineers become more efficient and work with reliable information as documents and models are updated in a controlled way as soon as any change is made. For even greater sustainability and effectiveness, the best way is to use the Unified Engineering methodology in the cloud. The cloud enables higher degrees of flexibility and collaboration among people working from different locations, which means that always the most appropriate resources can be allocated for each part of the project.

Unified Engineering reduces capital project costs, risks, and delays enabling shorter engineering cycles required to deliver new sustainable projects. By minimizing engineering errors and accelerating project execution, companies can get 50% faster FEED stage, 30% increase in engineering efficiency, and 5% reduction in TIC (Total Installed Cost).

Keywords: sustainable products, compressed engineering cycles, unified engineering

1. Introduction

As an industrial software provider we support chemical companies achieve superior performance in their quest to make sustainable products, align with the circular economy and demonstrate product stewardship throughout the product life cycle. By digitally connecting assets, process, and people, our solutions empower companies to run safe and responsible operations, mitigating EHS risks, and moving toward more circular systems, while remaining profitable.

With more than 50 years of industrial software innovation, AVEVA enables 13 of top 15 chemical companies and most of the world's petrochemical crackers, with the most comprehensive portfolio that ties profitability to sustainability goals.

Research typically identifies the following Critical Sustainability Drivers:

- Demand higher transparency on a company's environmental, social & governance (ESG) performance by stakeholders
- Manage a complex environment and provide safety to employees, processes, products and local communities.
- Minimize energy and utilities consumption, so as emissions.
- Manage liquid effluents, waste and suppliers, seeking the circular economy.
- Commitments to global and local regulations
- Portfolio management towards innovation and sustainable products

Today's technology allows suppliers to develop better solutions to those markets that are in constant change. The Digital Twin technologies, initially adopted mainly by the automotive and aerospace industries, are now promoting big changes in how chemical plants are operated and managed. This type of technology can change the decision-making process since more reliable information is available in real time.

Significant step was taken recently in terms of process simulation driven by two major sustainability industry trends. For the circular economy industry trend, chemical companies are developing new chemical processes that yield materials that can be recycled rather than used one time. For the hydrogen economy industry trend, companies will substitute hydrogen fuels to reduce CO₂ emission into the atmosphere. Both the chemical and hydrogen economy trends lead to corresponding advancement in process simulation including modeling and thermodynamics.

Now it is possible for the engineering and operating companies to build the Digital Twin of the process plant. The Digital Twin is built on a simulation platform that will support the entire plant lifecycle, from design to operation, while also addressing new hydrogen and circular economy requirements.

2. The Digital Twin

The new generation of process simulation uses a platform approach that evolves the simulation model from the conceptual engineering to the operation optimization. It allows a new approach to be implemented so companies can transition from the conventional scenario to the use of the process simulation Digital Twin, expanding benefits to the entire plant lifecycle. The same platform is used for process simulation and process utilities (cooling water, flare, steam and others), allowing engineers to further evaluate how each system impacts the other. Heat and material balances can be re-evaluated after equipment and pipeline sizing, since that information is in the simulation from

the beginning, as a result, little or no extra engineering effort is required. Once sizing is validated, the simulation is switched to dynamic mode, in which control loops are included to the simulation model to validate the process control strategy. As it is easier to shift the simulation to dynamic mode, rather than build a completely new model using the conventional approach or converting a model that cannot be taken back to the steady state mode, dynamic studies are performed earlier in the project lifecycle. This promotes savings in equipment acquisition and in operating costs, since control logic responses are evaluated in earlier stages. Plus, when something doesn't respond as expected, simulation is taken back to steady state mode, for re-evaluation of heat and material balance and re-sizing. The ability to go back and forth between steady state and dynamic modes is critical to increase efficiency in the project lifecycle, leading to huge savings in engineering effort. Table 1 summarizes how the new platform approach affects the plant lifecycle and its results.

Lifecycle Phase	Model	Benefits
Conceptual Engineering	The Digital Twin for the process plant is "first born"	<ul style="list-style-type: none"> • Fast evaluation of design alternatives due to continuously solved and flexible specifications • A native cloud application that protects IP to reduce IT costs • Open modeling for first-of-a-kind processes and equipment
Front End Engineering and Design (FEED)	The Digital Twin further develops to represent all process equipment	<ul style="list-style-type: none"> • One product with one learning curve for multiple applications, such as process, process utilities, and relief and flare • Integrated Asset Modeling of interacting but separate systems, such as, an oil field gathering and topsides processing, or the process and its flare system • Automated population of an engineering database • Automated creation of FEED engineering deliverables • Multi-user collaboration of a single simulation
Detailed Engineering	Simulation-Driven Engineering: The Digital Twin grows to also represent the mechanical design and the control strategy	<ul style="list-style-type: none"> • Other disciplines, such as controls, mechanical, piping, all contribute to the engineering database • Simulation takes information from the engineering database to test the Digital Twin continuously as it is designed • Process engineering trends towards new agile software engineering practices with a test-driven development now made possible because of the existence of the Digital Twin

Lifecycle Phase	Model	Benefits
Startup and Commissioning	The Digital Twin is used for Operator Training and Controls Checkout	<ul style="list-style-type: none"> • The actual DCS logic can be integrated to the Digital Twin • Operators are trained without a separate operator training simulator investment
Operations	The Digital Twin is a master simulation model for process improvement, equipment monitoring, optimization, and more	<ul style="list-style-type: none"> • One master Digital Twin model can be spawned to many applications, such as, training, equipment monitoring, and real time optimization to reduce the sustainment costs associated with separate point solutions • No longer need to maintain several process simulation models for a plant – design model, operator training simulator (OTS) model, unit performance monitoring and real-time optimization (RTO)
Revamp	The Digital Twin is used for new evaluations since it is always up to date.	<ul style="list-style-type: none"> • When a modernization or performance improvement is required the Digital Twin has current information and allows calculations and analysis without the need of re-engineering for an up to date simulation.

Table 1 – How a maturing Digital Twin develops and provides benefits for each phase

3. Unified Engineering

Companies are beginning to make progress on their digitalization journey, finding the right applications for digital transformation and seeing increasingly better returns on their investment. While the age-old market environment challenges (such as supply and demand, cost and price) haven't gone away, competitive pressures are making the digital transformation opportunity more pressing than ever. Many have already started to leverage the latest data-centric technology and work processes for their workforce to collaborate and take control of their data, reducing the risk for errors, delays and increased project cost throughout the asset lifecycle. By doing so they are in a stronger position to become more competitive, increase their margins and win new business.

Unified Engineering is a new proposition to break down the silos between FEED and Detailed Design to minimize risk and maximize return on Capital Investment.

Unified Engineering enables global multi-discipline teams to work concurrently in a common data-centric environment, controlling and managing change across the entire project. This breaks down the silos between FEED and detailed design. The simulation data created in FEED is readily available for use in detailed design and is checked and validated in real-time, increasing efficiency, minimizing risk, and maximizing return on investment on your Capital Projects.

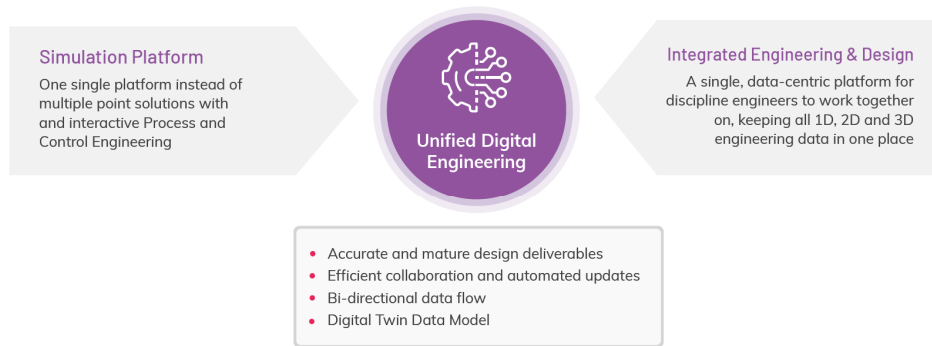
Never have the stakes been higher for companies when it comes to making improvements to their engineering work processes to maximize ROI on Capital Projects. Productivity has not developed in decades – the average Capital Project schedule lags by 20 months and goes over budget by 80%¹.

In many of today's Capital Projects, there is a disconnect between FEED and Detailed Design. In response, AVEVA are the first industrial software provider to pioneer a new solution to break down the silos between these engineering disciplines.

Unified Engineering consists of two main components, the Unified Lifecycle Simulation Platform (one model), and Integrated Engineering and Design (one database). The two are combined to form a robust process model and an engineering database that is able to synchronize through bi-directional flow of all 1D, 2D and 3D data on one platform. The bi-directional integration of a steady state and dynamic process model with an engineering database makes the process seamless and eliminates the need for MS Excel or other intermediate steps to transfer information between tools.

With the Unified Engineering model you can have:

- Verification that the plant will operate as expected, and that controls are properly configured
- Verification that equipment and piping are properly sized
- One single version of the truth that remains up-to date



The process model and the engineering database synchronizes through bi-directional flow from a single platform

Figure 1 - Unified Engineering principles

4. Process Simulation lifecycle

Process simulators are irreplaceable tools for every process engineer. Since the nineteen seventies, process simulators have found widespread adoption within operating companies in oil & gas, refining and chemical industries, as well as the engineering companies and equipment manufacturers that service these industries. The tools available in the market today have incrementally improved over the years to provide more features and functionality. However, they

trace their origins to legacy architectures, operating systems and aftermarket user interfaces, which create inherent limitations:

- They cannot support the full plant lifecycle as they are limited by their single-purpose architecture such as steady state process simulation, dynamic simulation, optimization, or flow network analysis for which they were originally designed
- Extending their functionality can be performed by a very small number of software developers with chemical engineering knowledge, software programming skills, and/or knowledge of that particular specialized program
- They are often based on decades old programming code that cannot leverage the more recent technological developments within the software industry

Today's simulators typically only support a single phase of the lifecycle and are often based on thermodynamics of different simulation vendors and different calculation methods. This not only leads to lack of trust in the results, but causes substantial rework by having to build a new simulation model in each new tool. And the results are hard to compare.

The technological limitations of incumbent process simulation tools forestall improvements in engineering workflows. Collaboration only occurs outside the simulation, because the software cannot accommodate it. Engineering departments send analysis and optimization questions to outside specialists because they cannot easily perform advanced simulations with their in-house tools and software expertise. These complications trap engineering workflows in a waterfall project management paradigm where development is forced into a linear process. Iteration is to be avoided because legacy software makes it cumbersome, error-prone, and tedious.

Global competition, pricing pressure and energy alternatives are now driving the need for a new approach. The oil & gas industry has seen high volatility and the lower price level of today is seen as the "new normal". The chemicals industry has a continuous need to innovate for greater agility and lower costs.

The next generation of workers also expects a modern, scalable and easy to use solution with technology they now take for granted – high speed internet access, mobile devices, touch screens and virtual reality. New concepts like the Industrial Internet of Things (IIoT), Industry 4.0, and Artificial Intelligence have created greater opportunities with a new next generation platform that provides a "Digital Twin" of the plant through the process lifecycle that cannot be provided with today's tools.

A next generation process simulation platform means that one process model is extended throughout the entire lifecycle of the plant, from concept through to operations. This requires a process design mode, a fluid flow/rating mode and a dynamic mode, in combination with the ability to toggle back and forth between modes. Optimization may be provided to any mode. Table 1 describes each phase of the project lifecycle, and how a maturing Digital Twin develops and provides benefits for each phase.

A single, easy-to-use simulation platform will allow engineers to move seamlessly between questions of design, analysis, and optimization. Engineers will be able to assess the impact of design and specification changes quickly and with a holistic view of multiple disciplines. Intensive collaboration becomes commonplace. In this environment, organizations will be able

to adopt agile engineering workflows based on smaller pieces of work with continuous integrated testing to reduce development cost while eliminating surprises at the end of the project.

5. Process Simulations for Sustainable Technology

Physics-based, first-principles models are a critical component of the asset digital twin. As industry develop new process technologies to serve sustainable industry trends such as the circular economy and the hydrogen economy, process simulation programs must adapt to become useful tools for process engineers to design sustainable processes. While academia research how to simulate model reactions or chemical thermodynamics, this modeling and thermodynamics must be integrated within industrial simulation products so that the unified engineering workflow may be applied to develop the digital twin of plants as they are engineered.

There are many Hydrogen-based projects under consideration around the world. Blue, Green, and Gray Hydrogen projects need common process simulation advancement that need to be modeled in processes that support the hydrogen economy.

Green hydrogen is produced by the electrolysis of water. Process simulators will include new electrolyzer equipment models. The Digital Twin for the operation of these plants may use an integrated model to optimize the use of wind and solar power generation for electrolysis.

Blue and gray hydrogen is produced by splitting the methane in natural gas into Hydrogen and CO₂ by steam methane reforming or auto thermal reforming. Development of integrated process plant simulation is required for removal of the CO₂ for capture and storage to produce blue hydrogen. New membrane adsorption models will separate hydrogen for greater purity for blending with natural gas. The carbon dioxide created must be removed using new generation of amines thermodynamics for CO₂ capture by companies that manufacture the amines and companies that use the amines for the separation of CO₂ for carbon capture.

Due to the low volumetric energy density of gaseous hydrogen when used as a fuel, plants must liquify the hydrogen to use it as a transportation fuel. Process simulators will include thermodynamic models appropriate for cryogenic hydrogen. In addition, ammonia is used as a transition fuel with models and thermodynamics required for ammonia pipelines and hydrogen conversion.

Circular economy trends are driving chemical companies to produce materials that may be recycled more readily. This drives the development of new thermodynamic methods and component data to study new process simulation. Chemical and energy companies are replacing fossil based raw materials with renewable feedstocks driving research and development of new reaction technology.

Finally, to make all processes more sustainability by reducing their energy consumption footprint, new simulation products will include the ability to calculate the cost of utilities such as steam, cooling water, and electricity, to minimize operating costs.

6. Conclusion

Lifecycle process simulation has been a vision for process simulation providers and their customers for a long time. Today's simulators cannot leverage the rapid developments occurring in the software industry due to legacy architecture.

Looking at the industry's increasing demand for higher transparency, this can only be achieved using a data-centric Digital Twin approach. This data-centricity enables an ideal platform for new product and process development to create new models and include the management of complex environments bringing together both steady-state and dynamic simulation with constant iteration and constant solving capabilities.

The integration of the process analysis and simulation with other disciplines also allows to breakdown the silos that were typically existing previously. Connecting and remotely controlling previously unconnected processes will increase sustainable operations and improve business efficiency in a sustainable environment. Over the long term the impact will drive resilience and sustainable performance through technologies.

References

¹ [Reinventing Construction - a route to higher productivity](#)

McKinsey, Rice Global E&C Forum Roundtable, September 8th 2017

T.G. Jayanth, Expert – Capital Projects & Infrastructure, McKinsey & Co

[Building a process Digital Twin for your plant with Unified Lifecycle Simulation - Whitepaper](#)

AVEVA, July 2020,

Ian Willetts, Vice President Process and Simulation, AVEVA

Cal Depew, Product Owner, AVEVA

[Unified Engineering: A new proposition to break down the silos between FEED and Detailed Design to minimize risk and maximize return on Capital Investment - Whitepaper](#)

AVEVA, July 2019,

Portfolio Marketing group, AVEVA

A homotopy continuation-based branch and bound algorithm for process synthesis and design using rigorous unit operation models

Yingjie Ma, Nan Zhang, Jie Li*

Centre for Process Integration, Department of Chemical Engineering and Analytical Science, The University of Manchester, Manchester M13 9PL, UK

jie.li-2@manchester.ac.uk

Abstract

Synthesizing chemical processes using rigorous unit operation models often leads to a large-scale strongly nonconvex mixed-integer nonlinear programming problem, which is difficult to solve. In this work, we propose a homotopy continuation-based branch and bound (HCBB) algorithm in which an adaptive variable-step length HC method is used to navigate the NLP subproblem to gradually approach a feasible solution. The computational results indicate that the proposed HCBB algorithm provides a robust convergence to a higher-quality locally optimal solution from different initial points for an example of process synthesis compared to existing general MINLP solvers.

Keywords: Process synthesis, homotopy continuation, branch and bound algorithm, rigorous models

1. Introduction

Optimisation-based chemical process synthesis and design problem using rigorous models usually leads to a large-scale strongly nonconvex mixed integer nonlinear programming (MINLP) problem, which is rather difficult to solve (Biegler et al., 1997). It is known that the global solvers such as BARON (Tawarmalani & Sahinidis, 2005) and ANTIGONE (Misener & Floudas, 2014) often cannot converge or even fail to provide a feasible solution for such MINLP problems within acceptable computational effort (Gopinath et al., 2016). The branch and bound (B&B) algorithm (Land & Doig, 1960) and the outer approximation (OA) algorithm (Duran & Grossmann, 1986) are often used to find a locally optimal solution within acceptable computational time (Viswanathan & Grossmann, 1993; Zhang et al., 2018). As both OA and B&B algorithms require to solve a series of strongly non-convex NLP subproblems, which are vulnerable to divergence. Therefore, they also often fail or converge to a locally optimal solution with low quality.

In this work, we propose a homotopy continuation-based B&B (HCBB) algorithm to solve the strongly nonconvex MINLP problem resulting from a superstructure-based synthesis of reaction-separation-recycle processes using rigorous models (Ma et al., 2019). The homotopy continuation (HC) method (Ficken, 1951) is used to improve the performance of the B&B algorithm. At each node, the generalised reduced gradient (GRG) algorithm (Drud, 1985) is first employed to find a feasible solution for the NLP subproblem using the solution from the previous node as the initial point. Once it fails, an adaptive variable-step length HC method is used to gradually approach a feasible

solution, which is then used to initialise the GRG algorithm to generate a locally optimum. The synthesis and design of a hydrodealkylation (HDA) process (Douglas, 1985) is solved to illustrate the capability of the proposed HCBB algorithm. It is shown that the HCBB algorithm is able to always converge to the same local optimum of 4.959 M\$ y⁻¹ in 5 minutes even from different initial points, whilst BARON, ANTIGONE and DICOPT solvers in GAMS (GAMS Development Corporation, 2014) fail in 1 h. The SBB/GAMS solver and the successive relaxed MINLP (SRMINLP) algorithm (Ma et al., 2019) converged to a worse solution of 4.713 M\$ y⁻¹ from some initial points.

2. Problem statement

The problem is stated as follows: for given raw material, product and production requirements (e.g., production rates and product purities), we want to get the optimal flowsheet structure and reactor types, optimal operating conditions, optimal sizes of reactors, distillation columns and heat exchangers. The objective is to minimize total annualised cost (TAC) or maximize profit. This is realised through a superstructure-based method. The superstructure is composed of a reactor network and a separation network for ternary separation. The superstructure for reactor network is from (Lakshmanan & Biegler, 1996). There are several stages in the reactor network. At each stage, different types of reactors can be used. At most one reactor can be selected at each stage. Distillation is assumed to be the only separation technology considered in the separation network since it is widely used for separation. The separation network is represented using the state equipment network (SEN) representation (Yeomans & Grossmann, 1999). The distillation columns are modelled by rigorous equilibrium stage model. The detailed model of the whole process synthesis problem can be found in (Ma et al., 2019).

3. Homotopy continuation-based branch and bound algorithm

3.1. Overall framework of the proposed HCBB algorithm

The proposed HCBB algorithm is shown in Fig. 1. It is based on the classical B&B algorithm. At each node, an NLP subproblem requires to be solved using the GRG algorithm due to its good convergence. However, the GRG algorithm requires a feasible solution before entering optimisation phase, which is difficult to provide (Drud, 1985). In the proposed algorithm, the homotopy-continuation method is used to get a feasible solution to initialise the GRG algorithm, which is shown in the following section.

3.2. Homotopy continuation method for the derivation of a feasible solution

In the B&B algorithm, \mathbf{x} denotes continuous variables and \mathbf{y} denotes binary variables. At a node i , there are \mathbf{S} ($s = 0, 1, 2, \dots, S$) binary variables in total including \mathbf{S}_F^i fixed binary variables and \mathbf{S}_R^i relaxed binary variables. That is $\mathbf{S}_F^i \cup \mathbf{S}_R^i = \mathbf{S}$. Then all binary variables at this node i are denoted as $\mathbf{y} = \{y_s | s \in \mathbf{S}\}$ with $\mathbf{y}_F^i = \{y_s | s \in \mathbf{S}_F^i\}$ to denote the fixed binary variables and $\mathbf{y}_R^i = \{y_s | s \in \mathbf{S}_R^i\}$ to denote the relaxed binary variables. Based on these notations, an NLP subproblem at a node i can be denoted as $P^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$. All the NLP subproblems $P^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$ have the same constraints and objective function but differ in \mathbf{y}_R^i and \mathbf{y}_F^i .

To get a feasible solution of $P^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$, we need to solve a corresponding feasibility

problem, which is denoted as $FP^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$. Note that $FP^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$ has the same constraints as $P^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$, but its objective function is a constant (e.g., 0). We take the optimal solution or a feasible solution from the previous node (e.g., the parent node of the current node) as an initial point in which we use \mathbf{x}^0 and \mathbf{y}^0 to represent the values of continuous variables and binary variables, respectively. In other words, $\mathbf{y}_R^{i,0} = \{y_s^0 | s \in \mathcal{S}_R^i\}$, and $\mathbf{y}_F^{i,0} = \{y_s^0 | s \in \mathcal{S}_F^i\}$. Since the initial point is an optimal solution from a previous node, $(\mathbf{x}^0, \mathbf{y}_R^{i,0})$ must be a solution of $FP^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$ where \mathbf{y}_F^i is fixed to $\mathbf{y}_F^{i,0}$. To gradually reach to the solution of $FP^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$ from the solution of $FP^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^{i,0})$, i.e. the initial point, we can sequentially solve a series of feasibility problems, $FP^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^{i,m})$ ($m = 0, 1, 2, \dots$) with $\mathbf{y}_F^{i,m}$ gradually approaching \mathbf{y}_F^i from $\mathbf{y}_F^{i,0}$, which is the HC process. For \mathbf{y}_F^i and $\mathbf{y}_F^{i,0}$, some of their components may be equal, while the other components are different. Thus, we only need to change the components that are different, which we call homotopy variables, and are notated as \mathbf{y}_{FH}^i and $\mathbf{y}_{FH}^{i,0}$, respectively. The common components are notated as \mathbf{y}_{FC}^i . Following this, the feasibility problems during homotopy can be represented by $FP^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_{FH}^{i,m}, \mathbf{y}_{FC}^i)$ ($m = 0, 1, 2, \dots$) with $\mathbf{y}_{FH}^{i,m}$ changing from $\mathbf{y}_{FH}^{i,0}$ to \mathbf{y}_{FH}^i . For convenience, we introduce a homotopy parameter t ($0 \leq t \leq 1$), and set $\mathbf{y}_{FH}^{i,m} = t \cdot \mathbf{y}_{FH}^i + (1-t) \cdot \mathbf{y}_{FH}^{i,0}$. Thus, the feasibility problems during homotopy can be represented as $FP^i(\mathbf{x}, \mathbf{y}_R^i; t^m, \mathbf{y}_{FH}^{i,0}, \mathbf{y}_{FH}^i, \mathbf{y}_{FC}^i)$ with t^m changing from 0 to 1. During HC, t^m is updated by the step length Δt^m though applying $t^{m-1} + \Delta t^m$ or $t^{m-2} + \Delta t^m$. To improve both convergence and efficiency, Δt^m is allowed to vary adaptively in the following rules:

Algorithm 1. Update Δt^m and t^{m+1} according to solution history during homotopy.

- 1 **Input:** convergence status of a solution \mathbf{Z}^f , Δt^{m-1} , Δt^{m-2} , t^{m-1} and t^m .
- 2 **If** \mathbf{Z}^f is infeasible then
- 3 $\Delta t^m = 0.5 \cdot \Delta t^{m-1}$
- 4 $t^{m+1} = t^{m-1} + \Delta t^m$
- 5 **Else if** $m = 1$ or $\Delta t^{m-2} \neq \Delta t^{m-1}$ then
- 6 $\Delta t^m = \Delta t^{m-1}$
- 7 $t^{m+1} = t^m + \Delta t^m$
- 8 **Else**
- 9 $\Delta t^m = 2 \cdot \Delta t^{m-1}$
- 10 $t^{m+1} = t^m + \Delta t^m$
- 11 **End if**

With above discussion, the HC method for solving NLP subproblems is as followings:

Algorithm 2. Solve an NLP subproblem using the homotopy continuation method.

- 1 **Input:** Select the solution of a previous NLP subproblem as an initial point \mathbf{Z}^0 , set $\mathbf{y}_{FH}^{i,0}$ and \mathbf{y}_{FH}^i , set $0 < \Delta t^0 < 1$, $0 < t^1 < 1$, $0 < \Delta t^{min} < 1$, and $M > 0$.
- 2 **For** $m = 1, 2, 3, \dots$
- 3 Solve $FP^i(\mathbf{x}, \mathbf{y}_R^i; t^m, \mathbf{y}_{FH}^{i,0}, \mathbf{y}_{FH}^i, \mathbf{y}_{FC}^i)$ from \mathbf{Z}^0 and get the solution \mathbf{Z}^f .
- 4 **If** \mathbf{Z}^f is optimal and $t^m = 1$ then
- 5 Solve $P^i(\mathbf{x}, \mathbf{y}_R^i; \mathbf{y}_F^i)$ using \mathbf{Z}^f as an initial point and get the solution \mathbf{Z}^* .
- 6 Break #Get the optimal solution of the NLP subproblem
- 7 **Else if** $m \leq M$ then

```

8       Update  $\Delta t^m$  and  $t^{m+1}$  according to Algorithm 1.
9       If  $\Delta t^m < \Delta t^{min}$  then
10          Break #Fail to find a solution
11       End if
12     Else
13       Break #Fail to find a solution
14     End if
15 End for

```

In this work, $\Delta t^0 = 0.5$, $t^1 = 0.5$, $\Delta t^{min} = 0.01$ and $M = 20$. The HCBB algorithm is implemented in Python (Python Software Foundation, 2016) through interfacing CONOPT solver (Drud, 1985) in GAMS for NLP solution.

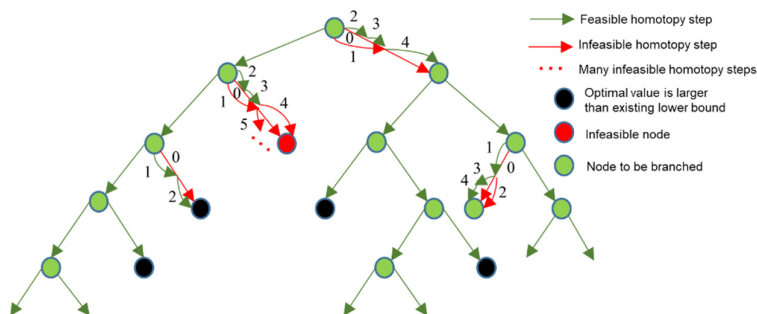


Figure 1. Homotopy-continuation-based branch and bound algorithm.

4. Case study

The HDA process using toluene and hydrogen to produce benzene and by-product diphenyl from (Douglas, 1985) is used to illustrate the capability of the proposed HCBB algorithm. The superstructure is given in Fig. 2, where the reactor can be either adiabatic or isothermal and there are three distillation columns. The required benzene molar purity is 99.97% with a production rate of 124.8 kmol h⁻¹. The objective is to maximize economic profit which is calculated by the revenue of benzene and diphenyl minus annualised capital cost and operating cost. All parameters for chemical reaction, thermodynamic calculation and economic evaluation are from (Kocis & Grossmann, 1989). While (Kocis & Grossmann, 1989) used short-cut models (i.e., Fenske-Underwood-Gilliland method for distillation columns), we use rigorous equilibrium stage model for distillation columns. There are 8142 constraints, 8643 continuous variables, and 13 binary variables.

Besides the HCBB algorithm, five other algorithms are also used, including DICOPT implementing OA algorithm (Duran & Grossmann, 1986), SBB implementing conventional B&B algorithm (GAMS Development Corporation, 2014), SRMINLP algorithm from (Ma et al., 2019), BARON (Tawarmalani & Sahinidis, 2005) and ANTIGONE (Misener & Floudas, 2014). The maximum computational time is set as 1 hour. We generate five different initial points, with which two local optima are found at the root using CONOPT. These two optima are used to initialize the six algorithms. While SBB, SRMINLP and HCBB can solve the problem from both initial points with the results

provided in Table 1, BARON, ANTIGONE and DICOPT fail in 1 h. From Table 1, SBB find a worse local optimum of 4.713 M\$ y⁻¹ from the root node 2 than HCBB. This is because around half of the NLP subproblems are infeasible as shown in Table 1, which misses the better optimal solution. Similarly, SRMINLP also finds a worse local optimum with TAC of 4.711 M\$ y⁻¹ from the root node 2 because it is easy to be trapped in the local optimum. This is because the initial point at root node 2 is close to the worse local optimum and the subproblem becomes more and more ill-conditioned and isolated with the tightness of the relaxed parameters. HCBB finds nearly the same optimal solution from both initial points because it significantly improves the success rate of solving NLP subproblems. The HC calculation has been invoked for 3 and 5 NLP subproblems respectively due to normal GRG algorithm fails when starting from root nodes 1 and 2. The optimal design of HDA process from the HCBB is presented in Fig. 3.

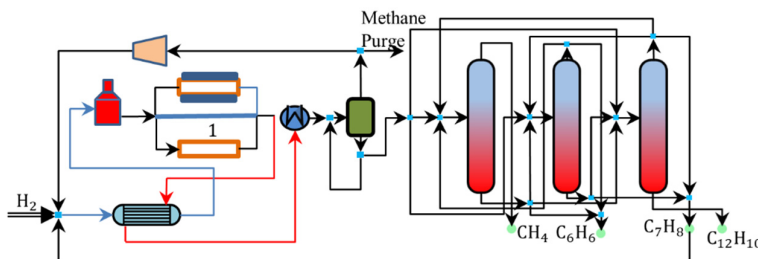


Figure 2. The superstructure for the HDA process.

Table 1. Comparative design results and computational times of different algorithms

	Profit (M\$ y ⁻¹)		Computational Time (s)		No. infeasible NLP/total No. of NLP	
	Init 1	Init 2	Init 1	Init 2	Init 1	Init 2
SBB	4.959	4.713	63	51	3/10	3/7
SRMINLP	4.956	4.711	120	63	0/2	0/2
HCBB	4.958	4.957	101	126	0/8	1/9

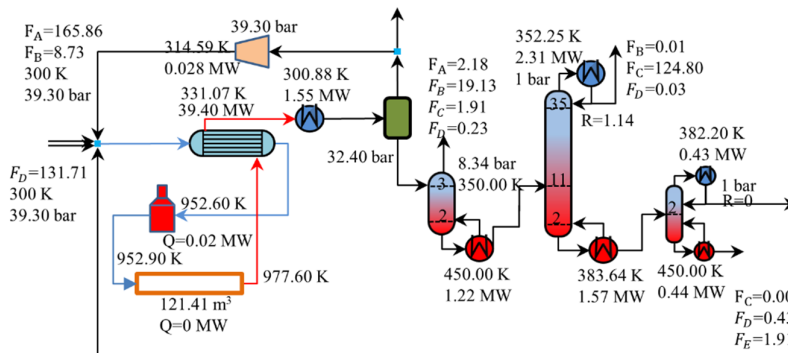


Figure 3. Optimal design of HDA process where the unit of flow rate is kmol h⁻¹ and A, B, C, D and E represent H₂, CH₄, C₆H₆, C₇H₈ and C₁₂H₁₀ respectively.

5. Conclusions

In this work, we proposed a HCBB algorithm to solve strongly nonconvex MINLP problems resulting from process synthesis and design problems using rigorous models. In this HCBB algorithm, an adaptive variable-step length HC method was used to navigate the NLP subproblem to approach a feasible solution. The computational results demonstrate that the proposed HCBB algorithm is able to converge to the same local optimum with a profit of 4.959 M\$ y⁻¹ in 5 minutes from different initial points, whilst BARON, ANTIGONE and DICOPT fails to find a feasible solution in 1 h. SBB and SRMINLP converged to a worse solution with 5% less profit.

References

- Biegler, L. T., Grossmann, I. E., & Westerberg, A. W. (1997). *Systematic Methods Of Chemical Process Design*. Upper Saddle River, NJ, USA: Prentice Hall PTR.
- Douglas, J. M. (1985). A hierarchical decision procedure for process synthesis. *AIChE J.*, 31, 353-362.
- Drud, A. (1985). CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems. *Math. Program.*, 31, 153-191.
- Duran, M. A., & Grossmann, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math. Program.*, 36, 307-339.
- Ficken, F. A. (1951). The continuation method for functional equations. *Commun. Pure Appl. Math.*, 4, 435-456.
- GAMS Development Corporation. (2014). *GAMS User's Guide Release 24.3*.
- Gopinath, S., Jackson, G., Galindo, A., & Adjiman, C. S. (2016). Outer approximation algorithm with physical domain reduction for computer-aided molecular and separation process design. *AIChE J.*, 62, 3484-3504.
- Kocis, G. R., & Grossmann, I. E. (1989). Computational experience with dicopt solving MINLP problems in process systems engineering. *Comput. Chem. Eng.*, 13, 307-315.
- Lakshmanan, A., & Biegler, L. T. (1996). Synthesis of Optimal Chemical Reactor Networks. *Ind. Eng. Chem. Res.*, 35, 1344-1353.
- Land, A. H., & Doig, A. G. (1960). An Automatic Method of Solving Discrete Programming Problems. *Econometrica*, 28, 497-520.
- Ma, Y., El-Khoruy, A., Yang, Z., Sun, L., Zhang, N., Li, J., & Xiao, X. (2019). Simultaneous Synthesis and Design of Integrated Reaction-Separation Systems Using Rigorous Models. In S. G. Muñoz, C. D. Laird & M. J. Realff (Eds.), *Comput. Aided Chem. Eng.* (Vol. 47, pp. 371-376): Elsevier.
- Misener, R., & Floudas, C. A. (2014). ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations. *J. Glob. Optim.*, 59, 503-526.
- Python Software Foundation. (2016). *Python Language Reference Version 3.6*. In.
- Tawarmalani, M., & Sahinidis, N. V. (2005). A polyhedral branch-and-cut approach to global optimization. *Math. Program.*, 103, 225-249.
- Viswanathan, J., & Grossmann, I. E. (1993). Optimal feed locations and number of trays for distillation columns with multiple feeds. *Ind. Eng. Chem. Res.*, 32, 2942-2949.
- Yeomans, H., & Grossmann, I. E. (1999). A systematic modeling framework of superstructure optimization in process synthesis. *Comput. Chem. Eng.*, 23, 709-731.
- Zhang, X., Song, Z., & Zhou, T. (2018). Rigorous design of reaction-separation processes using disjunctive programming models. *Comput. Chem. Eng.*, 111, 16-26.

ROmodel: A Python Robust Optimization Modeling Toolbox

Johannes Wiebe,^a Ruth Misener,^a

a Imperial College London, London, UK

j.wiebe@imperial.ac.uk

Abstract

We introduce ROmodel, a Python package that extends the modeling capabilities of the popular modeling language Pyomo to robust optimization problems. ROmodel contains a library of commonly used uncertainty sets which can be generated using their matrix representations, but it also allows the definition of custom uncertainty sets using Pyomo constraints. The resulting models can be solved using ROmodels solvers which implement both the robust reformulation and cutting plane approach. We apply the problem to a number of instances of three case studies and show some results.

Keywords: robust optimization, optimization under uncertainty, software tools

1. Introduction

Robust optimization has become a method of choice for optimization under uncertainty in process systems engineering with applications ranging from production scheduling to flexible chemical process design (Janak and Floudas, 2005; Li and Ierapetritou, 2008; Zhang et al., 2015; Ning and You, 2017; Shang and You, 2018; Grossmann et al., 2016). This has been accompanied by considerable development of new techniques: distributionally and adjustable robust optimization can reduce solution conservatism (Grossmann et al., 2016), data-driven robust optimization designs application specific uncertainty sets based on available data (Bertsimas et al., 2018), and approximate robust optimization bridges the gap between classic robust optimization and semi-infinite programming, making non-linear problems more tractable (Houska and Diehl, 2013). While this is undoubtedly a positive development, the large number of approaches and the required domain knowledge can discourage practitioners from making the transition from deterministic optimization to optimization under uncertainty. Furthermore, the lack of a platform that allows the easy implementation and application of new algorithms means that it is difficult to compare different approaches (Marc and Schöbel, 2016)

In this paper, we introduce ROmodel, a python package that extends the popular, Pythonbased modeling language Pyomo (Hart et al., 2017) to facilitate modeling of robust optimization problems and implementation of robust optimization algorithms. ROmodel combines intuitive modeling of robust optimization problems with the richness of Pyomo's solver interfaces and methods for model transformations and Python's data processing capabilities. It supports both automatic reformulation and cutting plane algorithms and can be extended to incorporate other approaches. Uncertainty sets can be chosen from a library of common geometries, or custom defined using Pyomo constraints. A number of similar software packages for modeling robust optimization problems have been introduced in the past: JumPeR extends Julia's modeling language Jump to robust optimization problems (Dunning, 2016), ROME provides a matlab interface (Goh and Sim, 2011), and Vayanos et al. (2020) recently proposed the C++ robust optimization

solver ROC++. AIMMS also has some capabilities for modeling robust optimization problems. One advantage of ROmodel is that it is based on Python and Pyomo. Python is becoming increasingly popular in many fields and particularly in data analytics and machine learning. ROmodel allows these techniques to be integrated seamlessly with robust optimization methods. Pyomo is a very powerful modeling environment and is also open source. Because ROmodel is entirely based on Pyomo, it can take advantage of many of its solver interfaces, model transformations, and more. ROmodel is open source and available on Github (Wiebe and Misener, 2020).

The rest of this paper is structured as follows. In Section 2 we introduce the new modeling objects which ROmodel provides and show how they can be used to easily model robust optimization problems. Section 3 introduces the three solvers which ROmodel contains: a reformulation based solver, a cutting plane solver, and a nominal solver for obtaining nominal solutions of robust problems. Section 4 introduces the three case studies we use to evaluate ROmodel and presents a number of results.

2. Modeling

In order to model robust optimization problems within Pyomo, we introduce two new modeling objects:

1. `UncSet`: A class based on Pyomo's `Block` class used to model uncertainty sets.
2. `UncParam`: A class similar to Pyomo's `Param` and `Var` class used to model uncertain parameters.

The two new modeling objects are sufficient for modeling quite generic robust optimization problems. Uncertainty sets can be defined in two ways: using Pyomo constraints or, for common types of geometries, using their matrix representation.

2.1. Generic uncertainty sets

The `UncSet` class inherits from Pyomo's `Block` class and can be used largely in the same way. One can use this property to construct generic uncertainty sets by adding Pyomo constraints to the `UncSet` object. The following example shows how a polyhedral set can be modeled using this approach:

```
# Define uncertainty set
m.U = UncSet()
m.w = UncParam(range(2), uncset=m.U, nominal=[0, 0])
# Add constraints
m.cons1 = Constraint(expr=m.w[0] + m.w[1] <= 1)
m.cons2 = Constraint(expr=m.w[0] - m.w[1] <= 1)
...
```

This way of modeling uncertainty sets is very flexible, however not every set that can be modeled using this approach can necessarily also be solved. We discuss which types of uncertainty sets can be solved in Section 3.

2.2. Library uncertainty sets

For standard uncertainty sets which are commonly used, the above approach is unnecessarily complicated. Instead we have implemented custom classes which can be used to define them using their matrix representation. Polyhedral sets of the form $U =$

$\{w: Pw \leq b\}$ can be constructed by passing the matrix P and the right hand side b to the `PolyhedralSet` class, e.g.:

```
# Polyhedral set
m.U = PolyhedralSet(mat=[[ 1,  1],
                        [ 1, -1],
                        [-1,  1],
                        [-1, -1]],
                    rhs=[1, 1, 1, 1])
```

Ellipsoidal sets of the form $(w - \mu)^T \Sigma^{-1} (w - \mu) \leq 1$ can be constructed using the `EllipsoidalSet` class, the covariance matrix Σ and the mean vector μ , e.g.:

```
# Ellipsoidal set
m.U = EllipsoidalSet(cov=[[1, 0, 0],
                        [0, 1, 0],
                        [0, 0, 1]],
                    mean=[0, 0, 0])
```

Implementing additional sets simply requires a new Python class collecting the necessary data and a function which, given a Pyomo constraint and an uncertainty set returns its robust counterpart as a Pyomo block.

2.3. Uncertain parameters

Once an uncertainty set is defined, we can utilize the `UncParam` class to easily construct uncertain constraints. Consider the following deterministic Pyomo constraint:

```
# deterministic
m.x = Var(range(3))
c = [0.1, 0.2, 0.3]
m.cons = Constraint(expr=sum(c[i]*m.x[i] for i in m.x) <= 0)
```

If the coefficients c are uncertain, we can model the robust constraint $c^T x \leq 0 \forall c \in U$ as:

```
# robust
m.x = Var(range(3))
m.c = UncParam(range(3), nominal=[0.1, 0.2, 0.3], uncset=m.U)
m.cons = Constraint(expr=sum(m.c[i]*m.x[i] for i in m.x) <= 0)
```

2.4 Swapping uncertainty sets

One of the advantages of ROmodel is, that it makes trying different uncertainty sets very easy. A robust model can simply be resolved with a different uncertainty set:

```
# Uncertain parameter
m.c = UncParam(range(3), nominal = [...], uncset = m.U1)
# Solving the model with m.U1
solver = SolverFactory('romodel.cuts')
solver.solve(m)
```

```
# Swapping uncertainty sets and resolving with m.U2
m.coef.uncset = m.U2
solver.solve(m)
```

3. Solvers

We implement three solvers within RModel: A robust reformulation based solver, a cutting plane based solver, and a nominal solver.

3.1 Reformulation

The reformulation-based solver implements standard duality based techniques for reformulating robust optimization problems into deterministic counterparts (Bertsimas and Sim, 2004). The solver is illustrated in Fig. 1. It essentially consists of a pre-processing:

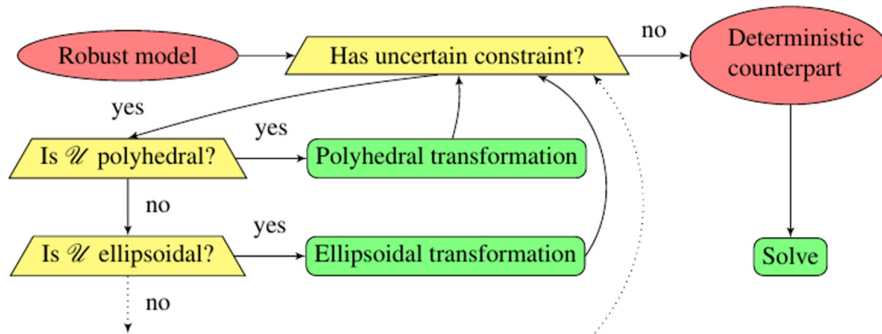


Figure 1: Schematic of reformulation solver

First, it detects every constraint containing uncertain parameters. Second, it checks the structure of each uncertain constraint and the corresponding uncertainty set to determine if a known reformulation is applicable. Finally, it applies a model transformation and solves the resulting problem using one of the available solvers in Pyomo.

RModel currently implements standard reformulations for ellipsoidal and polyhedral uncertainty sets and linear uncertain constraints, but it can easily be extended to include other reformulations. Adding new reformulations requires two Python functions: a function which detects whether a constraint and uncertainty set have the required structure and a function which generates the robust counterpart.

3.2 Cutting planes

The cutting plane solver, outlined in Fig. 2, implements an iterative strategy for solving robust optimization problems (Mutapcic and Boyd, 2009). It replaces each uncertain constraint by a `CutGenerator` object which initially just contains the nominal constraint. The solver then iteratively solves the master problem and generates cuts to cut off solutions which are not robustly feasible. RModel's cutting plane solver can generally be applied to any convex uncertainty set.

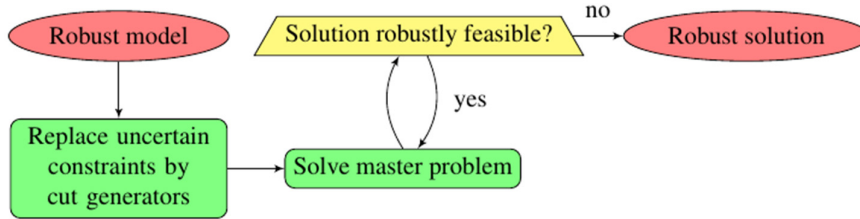


Figure 2: Schematic of cutting plane solver

3.3 Nominal

ROmodel also includes a nominal solver. This solver replaces all occurrences of the uncertain parameters by their nominal values and solves the resulting deterministic problem. The advantage of this is that it is not necessary to have separate implementations of the nominal and robust problem. An implementation of the robust model can easily be used to obtain the solution of the nominal problem:

4. Results

We use ROmodel to model and solve robust versions of three literature problems: a portfolio problem (Bertsimas and Sim, 2004), a knapsack problem, and a pooling problem instance (Adhya et al., 1999). We solve each problem with both the reformulation and cutting plane solver for ellipsoidal and polyhedral uncertainty sets and using both the library approach to generating uncertainty sets as well as the generic, Pyomo constraint-based approach. We solve each instance for 30 different uncertainty set sizes.

		Reformulation	Cuts	Overall
Knapsack	Polyhedral	71	365	84
	Ellipsoidal	65	248	83
Pooling	Polyhedral	99	469	119
	Ellipsoidal	60171	607	712
Portfolio	Polyhedral	69	402	92
	Ellipsoidal	65	40927	81
Overall		71	406	244

Table 1: Median time in milliseconds taken to solve the three example problems with different uncertainty set geometries using the reformulation and cutting plane solvers

Table 1 shows the median time in milliseconds taken to solve each problem for a given uncertainty set geometry and solver. The reformulation solver generally outperforms the iterative cutting solver. An exception is the the non-linear, non-convex pooling problem with an ellipsoidal set. For this instance, the cutting plane solver achieves significantly better results, which is in line with previous research (Wiebe et al., 2019).

5. Conclusion

ROmodel can be used to easily formulate robust versions of common optimization problems. The modeling environment it provides makes robust optimization methods more readily available to practitioners and makes trying different solution approaches and uncertainty sets very easy. ROmodel is open source and available free of charge and could play a vital role as a platform for prototyping novel robust optimization algorithms and comparing them to existing approaches.

6. Acknowledgements

This work was funded by the Engineering & Physical Sciences Research Council (EPSRC) Center for Doctoral Training in High Performance Embedded and Distributed Systems (EP/L016796/1), an EPSRC/Schlumberger CASE studentship to J.W. (EP/R511961/1, voucher 17000145), and an EPSRC Research Fellowship to R.M. (EP/P016871/1).

References

- N. Adhya, M. Tawarmalani, N. V. Sahinidis, 1999. A Lagrangian Approach to the Pooling Problem. *Ind. Eng. Chem. Res.* 38 (5).
- D. Bertsimas, V. Gupta, N. Kallus, 2018. Data-driven robust optimization. *Math. Program.* 167.
- D. Bertsimas, M. Sim, 2004. The price of robustness. *Oper. Res.* 52.
- I. R. Dunning, 2016. Advances in robust and adaptive optimization: Algorithms, software, and insights. Ph.D. thesis, Sloan School of Management, MIT.
- J. Goh, M. Sim, 2011. Robust optimization made easy with rome. *Oper. Res.* 59.
- I. E. Grossmann, R. M. Apap, B. A. Calfa, P. García-Herreros, Q. Zhang, 2016. Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty. *Comput. Chem. Eng.* 91.
- W. E. Hart, C. D. Laird, J.-P. Watson, D. L. Woodruff, G. A. Hackebeil, B. L. Nicholson, J. D. Siirola, 2017. *Pyomo — Optimization Modeling in Python*. Vol. 67. Springer International Publishing.
- B. Houska, M. Diehl, 2013. Nonlinear robust optimization via sequential convex bilevel programming. *Math. Program.* 142.
- S. L. Janak, C. A. Floudas, 2005. Advances in robust optimization approaches for scheduling under uncertainty. *Comput. Chem. Eng.* 20 (C).
- Z. Li, M. G. Ierapetritou, 2008. Robust Optimization for Process Scheduling Under Uncertainty. *Ind. Eng. Chem. Res.* 47 (12).
- A. G. Marc, Schöbel, 2016. Algorithm engineering in robust optimization. *Algorithm Engineering: Selected Results and Surveys*.
- A. Mutapcic, S. Boyd, 2009. Cutting-set methods for robust convex optimization with pessimizing oracles. *Optim. Method. Softw.* 24.
- C. Ning, F. You, 2017. A data-driven multistage adaptive robust optimization framework for planning and scheduling under uncertainty. *AIChE Journal* 63 (10).
- C. Shang, F. You, 2018. Distributionally robust optimization for planning and scheduling under uncertainty. *Comput. Chem. Eng.* 110.
- P. Vayanos, Q. Jin, G. Elissaios, 2020. Roc++: Robust optimization in c++. arxiv.org/abs/2006.08741.
- J. Wiebe, I. Cecilio, R. Misener, 2019. Robust optimization for the pooling problem. *Ind. Eng. Chem. Res.*
- J. Wiebe, R. Misener, 2020. Romodel 0.1. github.com/johwiebe/romodel.
- Q. Zhang, I. E. Grossmann, C. F. Heubeger, A. Sundaramoorthy, J. M. Pinto, 2015. Air separation with cryogenic energy storage: Optimal scheduling considering electric energy and reserve markets. *AIChE Journal* 61 (5).

Modelling – Mechanistic reductionism extended with holistic empirical components

Heinz A Preisig

*Norwegian University of Science and Technology
Department of Chemical Engineering
NO-7491 Trondheim, Norway
heinz.preisig@chemeng.ntnu.no*

Abstract

We discuss the underlying aspects of our computational engineering suite, implementing application-specific reductionism. Our Process Modelling suite (ProMo) supports complex multi-discipline and multi-scale models using discipline-specific basic building blocks. An application-domain-specific ontology captures the involved disciplines and their respective basic entities being the smallest granulates in each field. Each basic entity is primarily equipped with a mechanistic mathematical input/output representation. We augment this mechanistic approach with a holistic component by allowing multiple definitions of these behaviour descriptions, thereby enabling empirical models. ProMo's ontology reasoner handles the resulting multi-directed variable/equation graphs representing the entity models. The discipline-related model components are linked together over two mechanisms depending on the nature of their interaction. The concept of "tokens" provides the necessary abstraction, allowing for a network of networks description of multi-disciplinary models.

Keywords: Ontology, multidisciplinary, multi-scale, simulations, digital twin.

1. Mathematical models – modelling

If one takes modelling at its face value, it is omnipresent, and people use the term *model* in many different contexts. The term *model* is "loaded" with many different meanings and interpretations. In experimental work, one often plays with geometrical scales: one works on a small model process instead of the "real" process. In process systems engineering, we mainly focus on mimicking the process' input/output behaviour through mathematical modelling and simulation. Computational engineering is the primary activity of a process engineer, and while one also "plays" with *scales*, the term has different interpretations.

For process engineers, *models* are mathematical representations of systems' input/output behaviours. One applies reductionism, and the general idea is to describe the process as a conglomerate of interacting parts. In physics-related science subjects, one typically proclaims the applicability of the reductionism paradigm. In contrast, one also accepts that one does not always have enough knowledge or insight that enables one to provide all fundamental parts or the ability to describe all interactions required to capture the observed behaviour and one has to resort to a holistic approach.

Here we explore the possibility of combining the two approaches.

2. The Science Viewpoint

We are careful in defining "science" as the fundamental approach characterised by reductionism's ultimate application. Careful, because this view constitutes the belief that nature builds on fundamental entities. Physics, as theoretical and fundamental science, does have the objective to find those fundamental entities. The approach is to break the object of interest recursively down into smaller and smaller bits until no smaller "bits" can be identified or detected. Once in a dead-end, one declares the finest-grained entities as the current fundamental entities, and one continues the search on the next lower scale.

Models must only be good enough for the application and no better – a saying which well-known and lesser-known people have repeated many times. People construct models for a *purpose* (Apostel (1960)). The quality of the model one defines indirectly, namely in the context of the *application*. The process of generating a model is therefore recursive and ranges over design, realisation and application. Based on this consideration, one defines the termination points of the granularisation on the *appropriate* level in the break-down process, which, in a first instance, is a matter of judgment, making modelling to be seen as an art.

Once one has identified the smallest required entities, one *declares* them as the fundamental entities, and builds the models using them as building blocks.

3. Different complexities

Following those lines of thought, one distinguishes different kinds of complexities:

Many-complexity: The first is on a given level of granularity; the process behaviour consists of many items. Examples are easy to find. One requires many molecules to represent a material's behaviour appropriately. Sand piles consist of many sand particles, and catalytic beds have many active sites, many surfaces, etc. In these cases, one looks into thousands to millions of separate objects.

Structur-complexity: In contrast, the second type of complexity is about several scales. Here just a hand full is considered "many".

Discipline-complexity: Having elements from different disciplines is yet another type of complexity and again "many" is in the order of a hand full.

4. Scales

The "frame" in which a physical process lives, is defined by the four-dimensional space-time and we distinguish between two types of scales, namely time scales and length scales. Material science is the home of length scales: macro, meso, micro and nano are used in this context. Not without discussion though, as particularly meso is only defined as the scales between micro and macro or nano and macro.

Time scales are not often put into the limelight of modelling discussions. The reason may be that they are linked to the length scales. Obviously, a smaller part of a whole reacts quicker to a change on its boundary than the larger of the same kind. Though if one views the modelling process again from the point of not getting too detailed, it is useful to consider the time-scales.

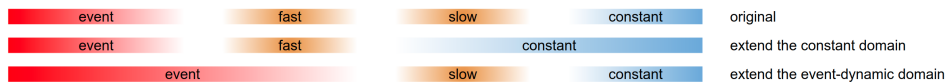


Figure 1: Time-scale motivated model reduction: Shrinking and expanding time scales

We claim that each physical process model assumes splitting the time scale into three sections: constant, dynamic and event-dynamic. The constant part divides into a set of reservoirs providing the driving forces for the embedded dynamic process, which is a Carnot-type of view for a process' operations. The event part is considered to "just happen" in the time-scale of the dynamic process. Typical chemical engineering example is the pressure wave propagation in a plant. Some model reduction methods operate on time scales. Keywords are singular perturbation with perturbation parameters being assumed small, which results in merging fast parts of the dynamic system into the event-dynamic domain. At the same time, on the other end of the dynamic spectrum, large capacities may be merged into the constant environment, see Figure 1.

5. Disciplines & Interactions

Let us first define *discipline* in the context of modelling. Here we use the term for a *domain* in which the same "things" are being handled, exchanged, processed ... The term is thus circularly defined with interactions. The declared basic entities process the *tokens* are exchanged between the entities. It is these *tokens* that indirectly represent the domain of the discipline. To give an example: declaring mass, energy and momentum as the *tokens* define a class of physical systems. One may classify further into particulate and continuous systems. Thus *tokens* are per se not the only criteria for the declaration of a discipline, but we may use it to declare domains of entities that interact with a class of *tokens*.

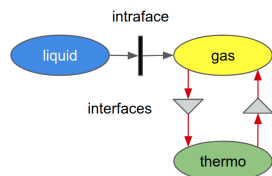


Figure 2: The liquid and the gas exchange tokens, while the gas and the thermo domain exchange information.

Domains characterised by different classes of *tokens* may interact only by exchanging information. Examples are the interaction between the physical domain and the domain capturing control. Others may be associated with analysing the designed process such as life-cycle or techno-economical evaluations, optimisation or the like.

Two domains of different length scales do not interact by exchanging *tokens* but information. As an example: a macroscopic system may require information about the material properties. Using molecular simulation technology is one possible method to compute the necessary properties. The input provided by the macroscopic model is the vector of canonical variables, such as temperature, pressure and molar composition. The molecular code returns

the requested property. The macroscopic level provides state information, mostly in the form of intensive properties, and in return, it obtains the requested physical quantity information. Exchanging the molecular code by any empirical model or a database does not change the nature of information exchange.

6. On Boxes

The use of *boxes* is a typical thinking pattern when describing the behaviour of a system. Generically, a *box* represents the input/output behaviour of a modelled system. In the context of reductionism, this implies that a network of the smallest identified entities, the basic *boxes*, the basic building blocks, describes the modelled plant's behaviour. These networks may become very large, defining the *many-complexity*, which is approached using recursion and hierarchical representation of the model. Thus the picture is interacting *boxes* in *boxes* or networks in networks.

Boxes may be classified. The terminologies of using *white box* for mechanistic models and *black box* for empirical models are widely used in engineering and science¹. *Glass-box* is intuitively the better term for *white-box* models, as it expresses the fact that one can see what is inside thereby providing information about the internal state. While mechanistic models are classically based on the reductionism view, black-box models are not built on insight but represent only the input/output behaviour without considering the internal state. They are based on a holistic view.

Many will view the use of the terms *white box* and black box as traditional. Though things tend to change with time and the meaning of terms undergoes adaptation to new or non-traditional fields, like this is the case with the renewed emerging field of artificial intelligence. For example, in AI literature, one finds explanations like: *White-box models are the type of models which one can clearly explain how they behave, how they produce predictions and what the influencing variables are.* (see Tannam (2020), or SciForce (2020)).

7. ProMo's network of network approach

NTNU's Process Systems Engineering group's ProMo project is based on reductionisms. It constructs multi-disciplinary models from the different discipline's base entities, where base entities are defined in the context of the model and its granularity.

The European Commission financed the development of the European Materials & Modelling Ontology (EMMO) (EMMC (2020)) and processes are part of the EMMO. The extension of the EMMO with the respective entity behaviours is NTNU's PSE groups' contribution.

Our Process Modelling suite *ProMo* provides the ontology and simulation construction tools. It is designed to handle large models that may stretch over several scales and employ different disciplines. Thus *ProMo* handles all three types of complexities: *many-complexity*, *structure-complexity* and *discipline-complexity*. As a basis, we use a network of networks approach and recursion within the networks. The "networks" represent the domains with its interacting *boxes* with the *boxes* being the networks' nodes. The *boxes* may again contain networks thereby allowing for a hierarchical representation of domains with a large many-complexity on the corresponding granularity level Preisig (2010). The top network links different discipline-specific networks. One has typically macroscopic continuous physical systems, microscopic particulate systems, control systems, material models, and databases for chemical engineering processes. The *ProMo* ontology provides

¹A well-formulated definition can be found in Wiener (1961).

the information of what disciplines are involved by defining a discipline tree.

ProMo's ontology provides a set of fundamental *boxes*, a kind of Lego-set from which the user constructs more complex *boxes* and finally complete process models. It is *ProMo*'s expert section that provides the necessary tools to define the model building instruction and the basic boxes (Preisig (2020)). For physical systems, the fundamental behaviours are based on the conservation and the balancing principles. For mechanical systems, including the molecular level, the Hamiltonian framework serves as a basis. In contrast, it is the contact geometry that defines the configuration space for thermodynamic systems.

8. Need for Empirical Models

So let us take a step backwards: The underlying philosophical concept is reductionism. Its idea is to break the object recursively into smaller and smaller parts until a level is reached, which can be considered the smallest essential component. The relative "size" of the smallest elements defines the model's granularity, with the required details depending on the application. Latter implies that the nature of the entities is defined in the context of the application. However, this approach is often not feasible, simply because one has too little knowledge about what the lower-level entities are and how one can describe their functionality. Another reason is that one would require a myriad of elements to capture the system's behaviour appropriately. The way out is to use another approach, namely local *holistic* approach. The holistic approach replaces a (highly) detailed or ill-defined entity with a surrogate. Surrogates go under various names, but the term "black-box" does capture it quite well: one replaces a mechanistic description, in this terminology called a white-box or glass-box, with a set of interacting functions. This network of functions is then fitted to input/output data yielding a modelled physical unit's simulated behaviour. This idea is by no means new. These models were termed black-box or metamodels, neural nets, and symbolic regression models (Cozad and Sahinidis (2018)), to mention a few. They are all functional networks, and so are the mechanistic descriptions. The result is that one defines multiple representations of the same input/output behaviour.

9. *ProMo*'s Handling of Multiplicity of Representations

ProMo's expert section has the tools for defining the equations representing the entity input/output behaviour, allowing for a multiplicity of definitions. An application-tailored reasoner² to handle this task. The modeller, called the *translator*, does though want to have a unique behaviour description. *ProMo*'s reasoner allows selecting an alternative in each case. The MoDeNa platform (Karolius et al. (2016, 2017)), uses a dynamic allocation of alternative models. An approach that is new in the domain of what today is termed "Open System Platforms" adds a great degree of flexibility in computing multi-scale processes.

10. Conclusions

ProMo enables the construction of complex models using a few discipline-specific basic building blocks. By providing a basic set of building blocks, the art of modelling is

²Term reasoner is used in the computer-science's ontology domain - usually implementing a first-order logic analysis.

moved from the writing of the behaviour equations to defining the model's structure. So the user's main task is selecting the structure of the model, selecting the right level of granularity, and considering the time scale in which the model must represent the modelled object. Model structures are information-exchanging, discipline-specific networks, each characterised by transferring and processing *tokens*.

The base entities' definition is done by experts in the respective field, thus removing the need for the user to acquire the necessary knowledge. A mechanistic, deductive, description may not always appropriately capture all behaviours, but one may have to succumb to an empirical representation rich enough to capture the observed behaviour. Those models are holistic and mostly empirical or semi-empirical. The multiplicity of representing individual entities is not a problem to handle. *ProMo*'s reasoner is designed to process the resulting multi-bipartite variable/equation graphs.

The approach taken in the design of *ProMo* splits the modelling activity into sections, where experts are providing the behaviour description of the basic building blocks input/output behaviours. A problem translator constructs the complex models using these building blocks. All input/output relations are automatically compiled and processed internally, without the user's need for interaction. The approach eases the construction of complex models. The problem translator tasks focus on the model's structure and not the representation of the basic blocks. The overall result is a reduced time from problem formulation to simulation, thus increasing the ability to explore alternatives.

Acknowledgments:

ProMo research was funded in parts by: (i) Bio4Fuels Research Council of Norway (RCN) project 257622 (ii) MARKETPLACE H2020-NMBP-25-2017 project 760173. (iii) VIPCOAT H2020-NMBP-TO-IND-2020 project 952903T (iv) MODENA FP7-NMP- Specific Programme "cooperation": Nanosciences, Nanotechnologies, materials and new Product Technologies Grant agreement ID: 604271

References

- L. Apostel, 1960. Towards the formal study of models in the non-formal sciences from the concept and the role of the model in mathematics and natural and social sciences. In: H. Freudenthal (Ed.), *The concept and the role of the model in mathematics and natural and social sciences*. D.Reidel Publishing Company, Dordrecht, The Netherlands.
- A. Cozad, N. V. Sahinidis, 2018. A global minlp approach to symbolic regression. *Mathematical Programming* 170 (1), 97–119.
- EMMC, 2020. Emmo: European materials & modelling ontology. <https://github.com/emmo-repo/EMMO>.
- S. Karoliuss, H. A. Preisig, H. Rusche, 2016. Multi-scale modelling software framework facilitating simulation of interconnected scales using surrogate models. *Computer Aided Chemical Engineering* 38, 463–468.
- S. Karoliuss, H. A. Preisig, H. Rusche, 2017. Sequential multi-scale modelling concepts applied to the polyurethane foaming process. *Computer Aided Chemical Engineering* 27, 487–492.
- H. A. Preisig, 2010. Constructing and maintaining proper process models. *Comp & Chem Eng* 34(9), 1543–1555.
- H. A. Preisig, 2020. Promo – a multi-disciplinary process modelling suite. *Computer-Aided Chemical Engineering*, 571–576.
- SciForce, 2020. Introduction to the white-box ai: the concept of interpretability. <https://medium.com/sciforce/-introduction-to-the-white-box-ai-the-concept-of-interpretability-5a31e1058611>.
- E. Tannam, 2020. What are the benefits of white-box models in machine learning? <https://www.siliconrepublic.com-/enterprise/white-box-machine-learning>.
- N. Wiener, 1961. *Cybernetics: or the Control and Communication in the Animal and the Machine*. MIT Press.

Thermodynamic Optimization of a Biomass-based Integrated Gasification Combined Cycle with Post Combustion Carbon Capture using Potassium Carbonate

Ikhlas Ghiat,^a Ahmed AlNouss,^a Tareq Al-Ansari^{a,b*}

^a*Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.*

^b*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar.*

*talansari@hbku.edu.qa

Abstract

Biomass proved promising outcomes as a sustainable alternative to fossil fuels for energy generation. With the mounting concerns on the environment and the increasing pressures on natural resources, there is an impetus to investigate new biomass feedstocks and further develop efficient biomass conversion processes and technologies. Gasification, amongst other thermochemical conversion processes, is considered as one of the most efficient and clean processes, generating syngas which can be fed to gas turbines for electricity production. To enhance the process' efficiency, the integrated gasification combined cycle is adopted which utilises the exhaust heat from the gas turbine to operate a joined steam turbine. In this study, a biomass based integrated gasification combined cycle (BIGCC) with CO₂ recovery is modelled, simulated and optimised. The carbon recovery segment consists of a post-combustion chemical absorption carbon capture from the exhaust flue of the BIGCC using potassium carbonate as solvent. The presented integrated system is able to produce electricity from date pits at a negative carbon credit. Thus, this system aims at recycling waste for the generation of valuable commodities, reducing GHG emissions and diversifying the energy portfolio, all feeding into the concept of sustainability. The proposed system is simulated as a thermodynamic equilibrium model for the BIGCC and a rate-based model for the carbon capture segment in Aspen Plus software. The energy and exergy efficiencies for the overall system are calculated and optimised. Several parametric studies were conducted to select impactful operating parameters and their ranges for the optimisation model. Outcomes of this study indicate that the proposed system generates 420 kW of electricity and captures 80% of the CO₂ emitted from the BIGCC. The integrated system achieves maximum energy and exergy efficiencies of 54% and 61% respectively.

Keywords: Biomass, BIGCC, Carbon capture, Potassium carbonate, Optimisation.

1. Introduction

The rapid population growth has led to increasing demand on energy, water and food. As a consequence, the use of fossil fuel resources for energy generation has grown, and is now posing significant concerns on the environment because of the high and cumulative greenhouse gas emissions (GHG) associated with burning these fuels (AlNouss et al., 2019). A major contributor to greenhouse gas emissions is CO₂, which has reached an amount of 33.1 Gt in 2018 amounting from energy generation activities (IEA, 2019). Mitigation efforts, and target plans have been set to keep the temperature rise below 1.5°C to 2°C, agreed upon during the Paris agreement and put towards actions plans in the national development goals. Pathways reflecting these efforts include as one of the priority scenarios the implementation of bioenergy with carbon capture and storage (BECCS) as a solution (IPCC, 2018). The integration of greenhouse gas control technologies has become a necessity to mitigate the ever increasing global GHG emissions while sustaining the population needs in terms of energy, water and food. A study integrated greenhouse gas control technologies with an energy, water and food nexus system with the aim of evaluating the corresponding environmental impact of different scenarios. The achievement of negative GHG emissions was proven in this study by using bioenergy with carbon capture and storage technology (Al-Ansari, 2017). The optimisation of BECCS technology can lead to less GHG emissions, and can make the technology less energy intensive and ready for commercialisation. The optimisation can be done at the different process levels of the integrated system. For example, a study conducted an optimisation of the syngas produced through determining the optimal biomass blend used as well as the optimal operating configurations for the gasifier (AlNouss, 2020). Others studied the efficient integration of biomass integrated gasification combined cycle (BIGCC) with the carbon capture segment, extracting heat from one system and feeding it to the other. For example, a study investigated the use of heat from the low pressure steam turbine to regenerate the solvent used for carbon capture Monoethanolamine, reaching an energy efficiency of 47.03% (Dinca, 2018). Others focused on studying the carbon capture process separately like Oyenekan and Rochelle (2009), where they investigated the use of piperazine promoted potassium carbonate for CO₂ removal using a rate-based model. In this study, an optimisation model is proposed for maximizing the energy and exergy efficiencies of a BIGCC with post combustion carbon capture using potassium carbonate (K₂CO₃) as a solvent. To run the optimisation, a prior rigorous sensitivity analysis is conducted to select the impactful operating parameters on the efficiencies and determine their ranges. In the next section, the proposed integrated system is described and baseline operating parameters are presented. Then, in the third section the analysis and optimisation model for the system are described.

2. Model development

In this study, a model is developed of a biomass based integrated gasification combined cycle with a post combustion carbon capture using potassium carbonate as a solvent. The system is able to generate power, and capture CO₂ while reutilising low value heat streams from the multiple subsystems for a better efficiency. Data points are chosen as a biomass and is fed to the gasification unit at a baseline rate of 100 kg/h. The ultimate and proximate analyses are presented in Table 1 (AlNouss, 2018).

Table 1: Ultimate and proximate analyses of date pit biomass

Proximate Analysis	Volatile matter	Fixed carbon	Ash	Moisture	Ultimate Analysis	C	H	O	N	S
	77.7	16.31	0.98	5.01		50.84	6.83	37.88	4.45	0

The gasifier operates at a temperature of 800°C and a pressure of 1 bar. The produced syngas is then sent to a heat recovery unit to cool the syngas before compression and combustion. The combusted gas is then expanded in a gas turbine for power generation, and is discharged at a pressure of 32 bar. The heat value of the syngas is utilised by the dual-pressure steam turbine along with the exhaust heat from the gas turbine for more power production. Superheated steam is fed to the high pressure turbine at 538°C and 124 bar, and to the low pressure turbine at 500°C and 32 bar. After utilising the heat from the gas turbine effluent, the stream is transferred to a carbon capture unit along with potassium carbonate. The gas from the BIGCC called EX-GAS is fed at the bottom stage of the absorber, and the solvent called LEANIN is fed to upper stage producing a countercurrent flow. The aqueous potassium carbonate solution (40 wt%) then absorbs CO₂ and exists the absorber as RICHOUT1 and goes through an expansion to decrease its pressure before entering the stripper at stage two with a pressure of 150 mbar. The stripper is set to a column pressure drop of 10 mbar. The CO₂ is removed from the rich solution and exits the stripper as ACIDGAS, while the remaining solvent exits from the bottom of the stripper as LEANOUT1 and is pumped for regeneration as shown in Figure 1.

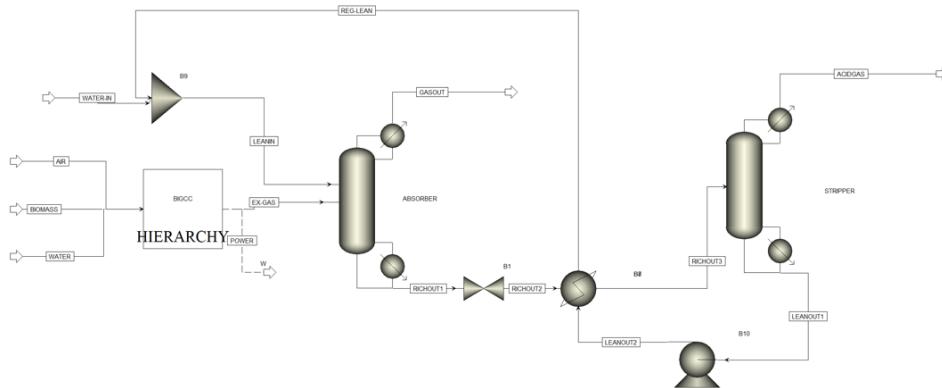
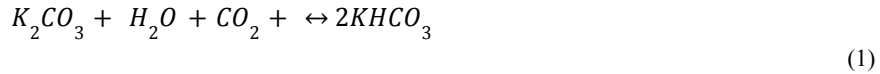


Figure 1: System's flowsheet in Aspen Plus

3. Analysis and optimisation

Aspen Plus software is used in simulating the above presented model for the BIGCC and carbon capture integrated processes. The BIGCC segment is simulated following a thermodynamic equilibrium model, while the carbon capture unit is simulated using a based model that takes into consideration the reaction kinetics for more accurate results. Assumptions taken in this simulation include steady state and steady flow conditions,

isothermal conditions for the decomposition and separation stages with atmospheric conditions taken as 25 °C and 1 bar, adiabatic conditions for the combustion, tar formation is neglected, and char is considered as 100% carbon. The Peng-Robinson equation of state with Boston-Mathias modifications is used for the BIGCC unit and the Electrolyte Non-Random Two Liquid (NRTL) property method is used for the carbon capture unit. The overall reaction of carbon capture by potassium carbonate is described in equation (1).



The kinetics of the reaction mechanisms are calculated using the Power Law as shown in equation (2).

$$Kinetic\ factor = kT^n \cdot e^{-E/RT} \quad (2)$$

The model is simulated initially for a baseline scenario following the operating parameters described in the model development section. The overall energy and exergy efficiencies are then calculated following equations (3) and (4) respectively. A sensitivity analysis is then run to see the effect of varying operating configurations on key parameters of the efficiencies such as the effect of gasification temperature on power and CO₂ mass flowrate, the effect of biomass flowrate on power production and CO₂ mass flowrate, the effect of solvent temperature and pressure and stripper pressure drop on reboiler heat duty and CO₂ captured. This sensitivity analysis works as a basis for defining a range to be studied for the operating parameters. The efficiencies are then optimised by changing the studied ranges of the operating parameters and determining simultaneous optimal values for these parameters that will maximise the energy and exergy efficiencies.

Variations in each operating parameter are studied through the simulation in Aspen plus. Results of the sensitivity analyses are transferred to Matlab to build regression models describing the energy and exergy efficiencies corresponding to changes in each operating parameter. Energy and exergy efficiency functions for each operating parameter are developed. The weighted average of these functions is computed to represent the objective functions of the proposed optimisation model. A pareto front is developed using Genetic Algorithm (GA) in Matlab to generate optimal solutions for the simultaneous computing objectives of energy and exergy efficiencies.

Variables:

\dot{W}_{net} : net power (W)

\dot{m}_{CO_2} : CO₂ flowrate in acid gas (kg/s)

h_{CO_2} : CO₂ enthalpy in acid gas (J/kg)

ex_{CO_2} : CO₂ total exergy in acid gas (J/kg)

$ex_{biomass}$: total exergy of biomass (J/kg)

$\dot{Q}_{heaters}$: heat duty of heaters in BIGCC (W)

$\dot{Q}_{reboiler}$: reboiler heat duty in stripper (W)

Decision variables:

$T_{gasification}$: gasification temperature (°C)

$\dot{m}_{biomass}$: biomass flowrate (kg/h)

$T_{solvent}$: lean solvent temperature (°C)

$\dot{m}_{solvent}$: lean solvent flowrate (m^3/h)

$\Delta P_{stripper}$: stripper pressure drop (mbar)

Objective function:

$$\text{Maximise: } \eta_{en} = \frac{\dot{W}_{net} + \dot{m}_{CO_2} h_{CO_2}}{\dot{m}_{biomass} LHV + \dot{Q}_{heaters} + \dot{Q}_{reboiler}} \quad (3)$$

$$\text{Maximise: } \eta_{ex} = \frac{\dot{W}_{net} + \dot{m}_{CO_2} ex_{CO_2}}{\dot{m}_{biomass} ex_{biomass} + \dot{Q}_{heaters} (1 - \frac{T_0}{T_s}) + \dot{Q}_{reboiler} (1 - \frac{T_0}{T_s})} \quad (4)$$

Constraints:

$$700^\circ\text{C} \leq T_{gasification} \leq 900^\circ\text{C}$$

$$100\text{kg/h} \leq \dot{m}_{biomass} \leq 1000\text{kg/h}$$

$$40^\circ\text{C} \leq T_{solvent} \leq 80^\circ\text{C}$$

$$5\text{ m}^3/\text{h} \leq \dot{m}_{solvent} \leq 30\text{ m}^3/\text{h}$$

$$3\text{ mbar} \leq \Delta P_{stripper} \leq 10\text{ mbar}$$

4. Results and discussion

The results of the sensitivity analyses are presented in Figures 2-6. The effect of gasification temperature on power and CO₂ mass flowrate presented in Figure 2 demonstrates an increasing trend for both energy and exergy efficiencies with the increase in gasification temperature. This is due to the increase in power output and the decrease in CO₂ generation and reboiler duty. Whereas, the increase in biomass flowrate presented in Figure 3 indicates an increase in exergy efficiency and a decrease in energy efficiency. This is linked to the important increase in power generation and that of the energetic heat of the reboiler, but the small increase in the exergetic reboiler heat.

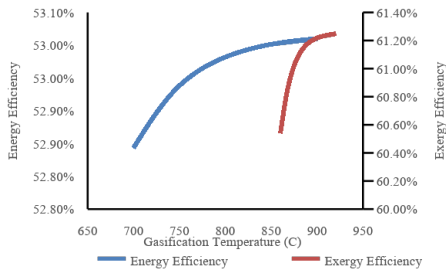


Figure 2: Effect of gasification temperature

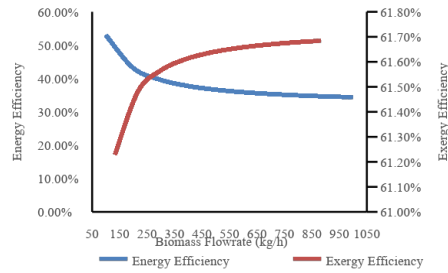


Figure 3: Effect of biomass flowrate

The effect of solvent flowrate presented in Figure 4 indicates a decreasing trend for both energy and exergy efficiencies with the increase in solvent flowrate until a certain point approximately at 14 m³/h where beyond it the trend starts to increase. This can be linked to the effective increase in CO₂ captured starting a solvent flowrate of 14 m³/h. Figure 5 illustrates the effect of solvent temperature where a decreasing trend in both energy and exergy efficiencies is demonstrated with the increase in solvent temperature due to the decrease in CO₂ capture.

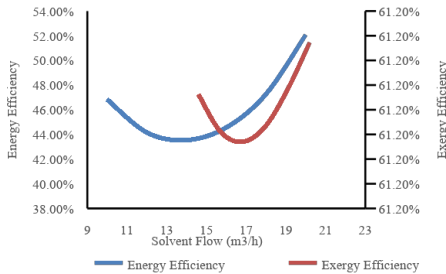


Figure 4: Effect of solvent flowrate

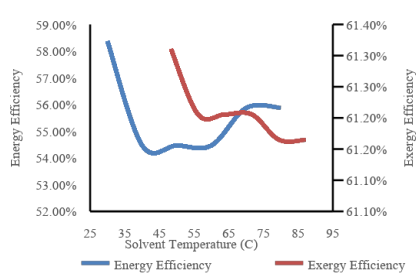


Figure 5: effect of solvent temperature

The effect of stripper pressure drop presented in Figure 6 indicate a decreasing trend in both energy and exergy efficiencies with the increase in stripper pressure drop. This is due to the fact that the effectiveness of CO₂ capture process decreases with the increase in column pressure drop and the increase in reboiler duty requirement.

The results of the optimisation problem considering these five sensitive decision variables are illustrated in Figure 6. The optimisation suggests gasification temperature approximately at 900°C, lean temperature at 30°C, lean flowrate of around 20 m³/h and stripper pressure drop of 3 mbar for both maximisation cases of energy and exergy efficiencies. Whereas, the optimisation recommends two different biomass flowrates with 315 kg/h for exergy efficiency maximisation case and 100 kg/h for energy efficiency maximisation case. This is due to the distinct behaviour both showed with the increase in biomass flowrate.

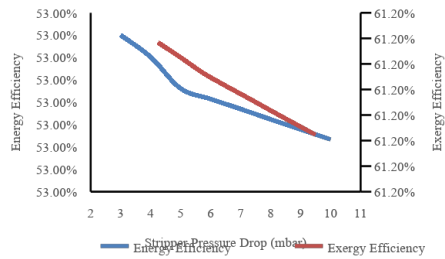


Figure 6: Effect of stripper pressure drop

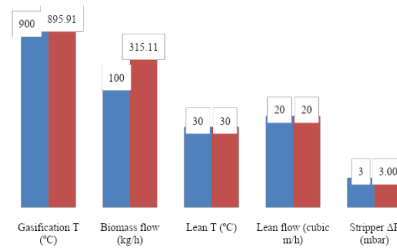


Figure 7: Optimisation results

5. Conclusions

BECCS technology is considered as a viable strategy and solution to reduce CO₂ concentration levels from the atmosphere. In efforts to make this technology applicable to large scale processes and economically profitable, it is necessary to improve its efficiency and comprehensively study the integration of all the subsystems. Hence, in this study a proposed model of a biomass integrated gasification combined cycle (BIGCC) with post combustion carbon capture using potassium carbonate is analysed and optimised thermodynamically in terms of energy and exergy efficiencies. The first step of this study consisted of modeling and simulating a baseline scenario of the integrated system. Then, a sensitivity analysis was conducted to see the effect of varying operating parameters on the efficiencies. The most impactful operating parameters were then selected and inputted as decision variables to the optimisation model. The objective of the optimisation was to maximise the energy and exergy efficiencies subject to some constraints defined by the allowable ranges of the operating parameters set in the sensitivity analysis. Results show that the maximum energy and exergy efficiencies achieved by this optimisation are 54% and 61% respectively.

References

- A. AlNouss, S. Namany, G. McKay, and T. Al-Ansari, 2019, Applying a Sustainability Metric in Energy, Water and Food Nexus Applications; A Biomass Utilization Case Study to Improve Investment Decisions, *Computer Aided Chemical Engineering*, 46, 205-10.
- IEA, 2018, *Global Energy and CO₂ Status Report: The latest trends in energy and emissions in 2019*.
- IPCC, 2018, *Summary for Policymakers*.
- T. Al-Ansari, A. Korre, Z. Nie, & N. Shah, 2017, Integration of greenhouse gas control technologies within the energy, water and food nexus to enhance the environmental performance of food production systems, *Journal of Cleaner Production*, 162, 1592-1606.
- A. AlNouss, G. McKay, T. Al-Ansari, 2020, Production of syngas via gasification using optimum blends of biomass, *Journal of Cleaner Production*, 242, 118499.
- C. Dinca, N. Slavu, C. Cormoș, A. Badea, 2018, CO₂ Capture from Syngas Generated by a Biomass Gasification Power Plant with Chemical Absorption Process, *Energy*, 149, 925-936.
- B. Oyekan, G. Rochelle, 2009, Rate modeling of CO₂ stripping from potassium carbonate promoted by piperazine, *International Journal of Greenhouse Gas Control*, 3, 121-132.
- A. AlNouss, G. McKay, T. Al-Ansari, 2018, Optimum Utilization of Biomass for the Production of Power and Fuels using Gasification, 28th European Symposium on Computer Aided Process Engineering, 43, 1481-1486.

The state of play towards modeling of air quality in the city of Rabat

Soraya EL HAMDOUNI,^a Essediya CHERKAOUI,^b Abderrahmane NOUNAH^c and Mohamed KHAMAR

Civil Engineering and Environment Laboratory (LGCE); Materials, Water and Environment Team; Higher School of Technology of Salé; Mohammed V University in Rabat, Morocco

Abstract

Rabat, the capital of Morocco, a green and sustainable city, a radiant city on a regional and national scale, faces a serious problem of air pollution caused by the daily flow of vehicles. Faced with the inaccessibility of the data on this issue, the study focuses on the quantification of the CO₂ emission rate by vehicles and their contributions to air pollution and the choice of station for monitoring CO₂ by a sensor. Once the database is completed, it will be shared through an application/website accessible to the population in order to make them aware of this problem.

Keywords: Pollution, treatment, modeling

Located in the extreme northwest of the African continent, Morocco is a country bordered by the Mediterranean Sea to the north and the Atlantic Ocean to the west. Inland, the Atlas mountain range runs through the center of the country, between which lie vast plateaus of grassland. The natural border in the south is the Sahara Desert. It covers an area of 446,550 km² with a population of 34 million inhabitants, this population is spread over 12 regions according to the new administrative division of 2015.

The prefecture of Rabat, the object of our study is the administrative capital of the Kingdom, it belongs to the region of Rabat Salé Kénitra and is the capital of its region. It covers an area of 118 km². The prefecture is bounded to the north by the prefecture of Salé, to the south by the prefecture of Skhirat-Témara, to the east by the two prefectures Salé and Skhirat-Témara, to the west by the Atlantic Ocean. Rabat is made up of two municipalities, the urban municipality of Rabat which includes five districts (Agdal-Riad, El Youssoufia, Hassan, Souissi, Yacoub El Mansour), and the urban municipality of Touarga where the royal palace is located. It benefits from a crossroads location (accessible to all the provinces in the region), as it represents an attractive hub on a regional and national scale given its administrative weight.

We produced a situation map of the prefecture of Rabat at the national, regional, and provincial levels using ArcGIS software on the basis of the new administrative division of 2015 (Figure 1).

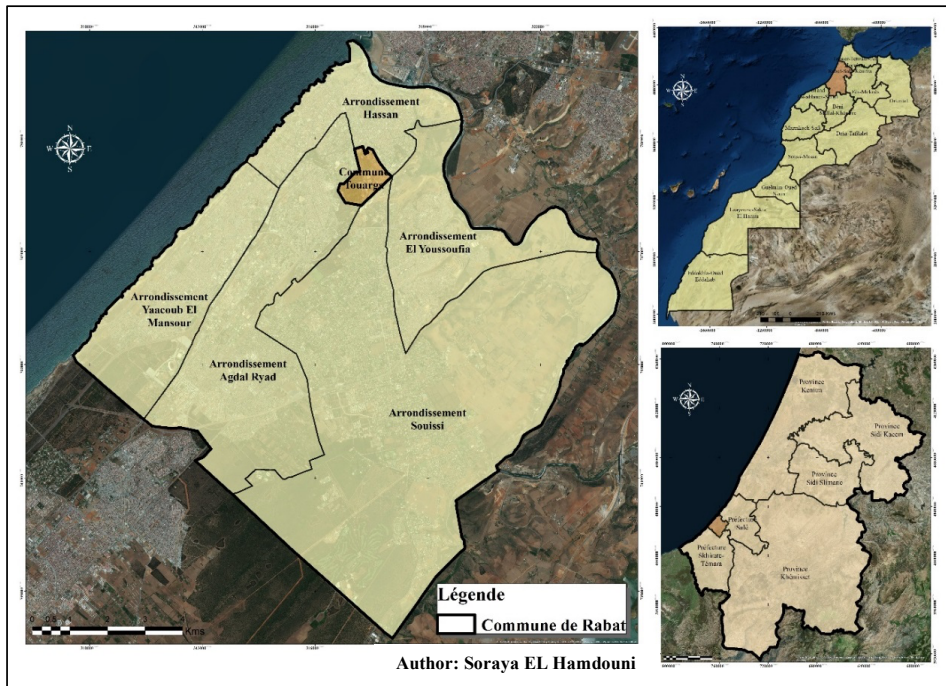


Figure 1: Situation map of the prefecture of Rabat

Demographically, Rabat is home to a population of 572,717 inhabitants, or 12.6% of its region and 1.7% of the whole country.

since 2004 the city of Rabat has been losing its population. This regression is due to several factors, including (Figure 2)

- The logic of social filtration following slum relocation operations (departure of a large demographic mass);
- Geographic constraints (the prefecture is bounded by the Bouregreg wadi, the Atlantic Ocean, and the green belt) which slow down urban sprawl;
- Land constraints that do not allow planned urbanization;
- A political will to keep the city of Rabat a lower town, less dense and elitist;
- A mismatch between supply and housing needs in terms of stock, price, and quantity offered;
- Lower and lower birth rate;
- High standard of living compared to neighboring municipalities;

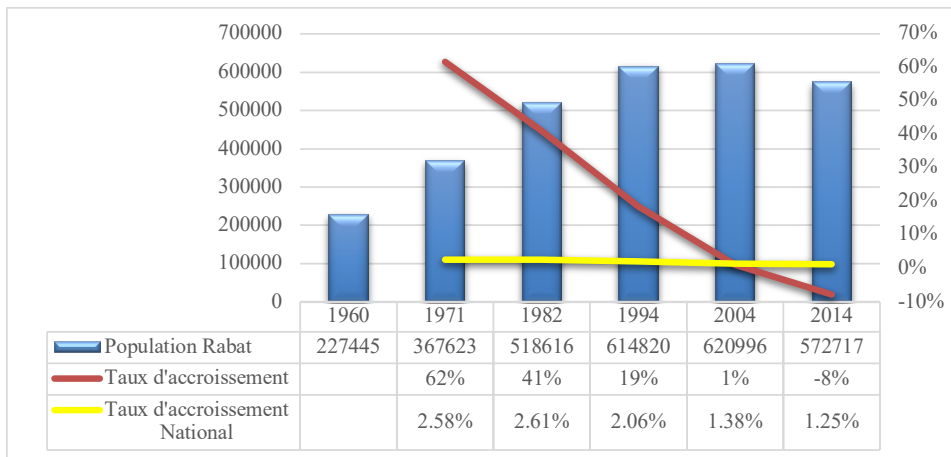


Figure 2: Evolution of the population of the prefecture of Rabat

Regarding the economic aspect, precisely the branches of activity, Rabat knows a dominance of the tertiary sector with a percentage of 38%, this dominance is justified by the concentration of all the administrations and public and private services, this concentration is not other than the result of the centralization adopted previously in Morocco, followed by services (19%), trade occupies third place (14%) and lastly agriculture (1%) and the water branch, electricity, and energy (1%). (figure 3)

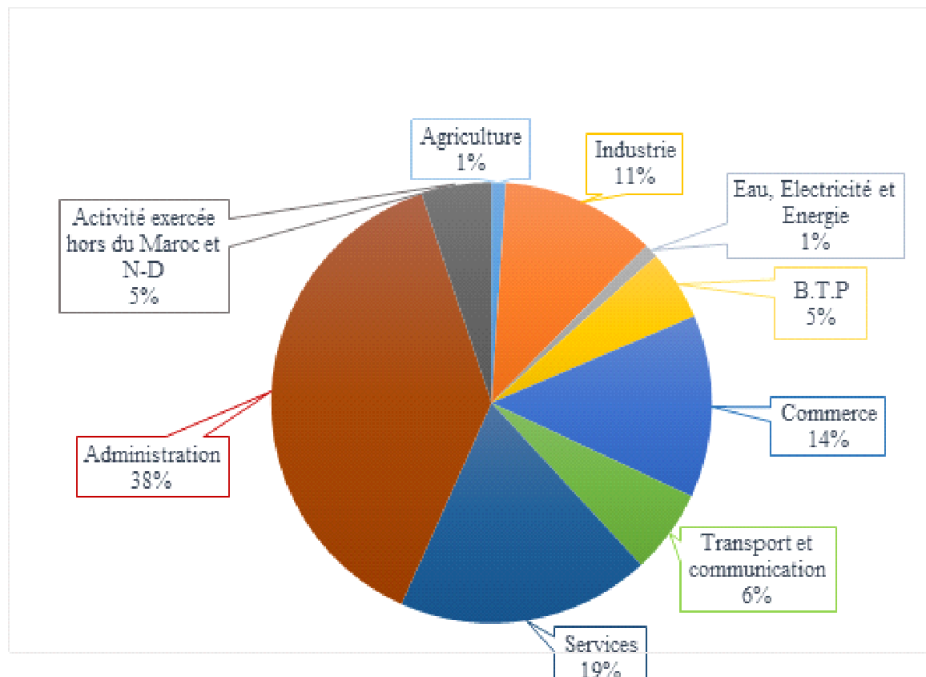


Figure 3: The branches of activity at the prefecture of Rabat

As for the industrial component which represents 11%, according to the Chamber of Commerce, Industry, and Service, Rabat has 3 industrial zones,

- The Takkadoum industrial zone with 37 companies in the field of Textile, IMM, Automotive, and Construction.
- The Al Fadila industrial zone with 24 companies in the field of Agri-food, Textile, IMME, Pharmaceuticals, and Construction.
- The Vita industrial zone with 4 companies in the chemicals sector, IMM and a Showroom (76% of units), and car dealerships.

Although Rabat is labeled a green city, thanks to the doubling of the global average of green spaces for each individual,) its industrial zones which do not really impact the quality of the air, its delegated management of solid and efficient waste with a collection is around 99%, a liquid waste management system by a WWTP which has been in operation since 2011.

The major problem that the capital faces on a daily basis and which has a direct impact on air quality is transportation. Being an elitist city, attractive and radiant at the national level, Rabat really suffers from this phenomenon in environmental and urban terms, this problem is strongly linked to the phenomenon of commuting between Rabat and the neighboring municipalities.

To respond to this problem, Rabat has opted for various solutions, such as the tramway, which is still not sufficient to speak of Rabat as a city free from air pollution.

In order to better understand the consequences of commuting, fieldwork was carried out (interrupted several times by COVID 19) at the level of the entrances and exits of the city of Rabat aiming at the quantification by category of vehicles in rush hour in order to show the contribution rate of vehicles to daily CO₂ emissions.

To do this, we opted for the quantification of vehicles on all the entrances and exits of Rabat on the side of the Salé prefecture to the north-east and on the side of the Skhirate Témara prefecture to the south-east which are the two prefectures that delimit Rabat.

the fieldwork was spread over 12 peak hour quantification stations from 7:30 am to 9:30 am and 3:30 pm to 5:30 pm. After the database was developed, it was processed using ArcGIS software in order to find the station with the most traffic (Figure 4).

We made the map at ArcGIS, to locate the 12 stations with the number of vehicles at peak hours.

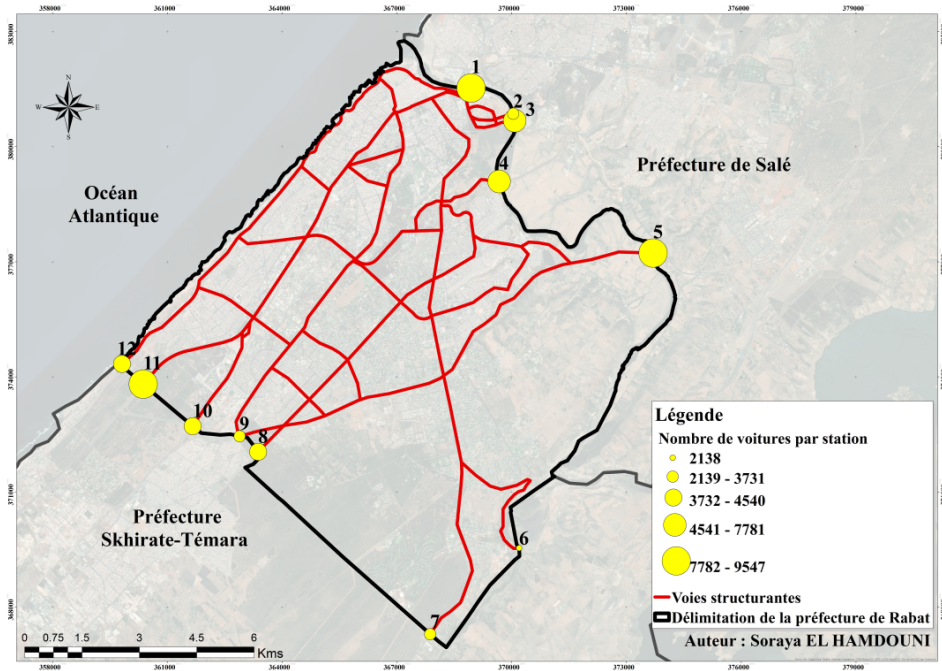


Figure 4: Location map of vehicle census stations

Based on the number of cars passing through each station and the quantity of CO₂ in g / km released by types of the car according to their technical datasheet, it was found that the maximum quantity is observed at station 1 and the minimum is at station level 6 (Table 1)

Station	1	2	3	4	5	6	7	8	9	10	11	12
CO ₂ g/km	1307	560	1167	1069	1279.5	321	548	634.5	960	681	1432	609

Station 1 (Hassan II Bridge) representing the north side of the prefecture and station number 11 on the south side representing Avenue Hassan II are the stations most marked on the one hand by the greatest flow of cars and other the CO₂ emission quite high.

This is why we chose the two stations which will be monitored by gas sensors of the quantities of CO₂ released by the means of transport into the air. In our case study, it is station 1 (Pont Hassan II) on the Salé side and station 11 (Avenue Hassan II) on the Témara side with the largest number of vehicles. (Figure 5)

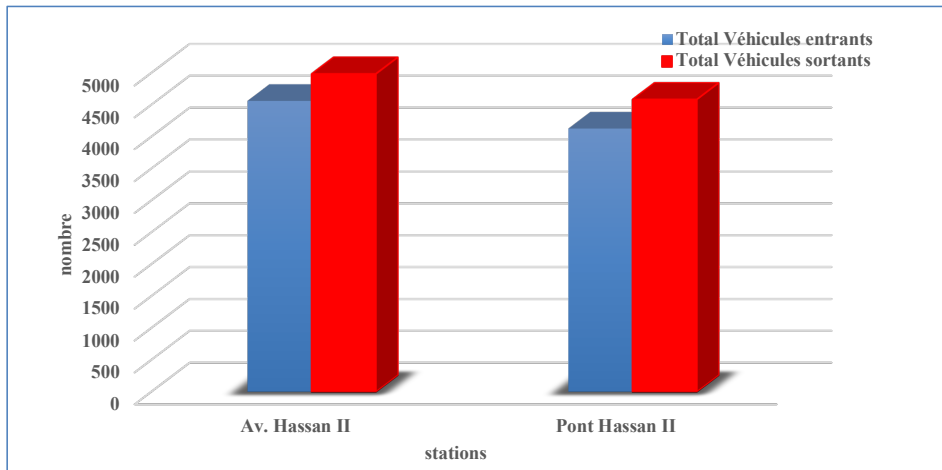


Figure 5: Incoming and outgoing vehicles at the two stations under study

The quantification of vehicles at the entrances and exits of Rabat was accompanied in parallel with a study on the CO₂ emission rate of each brand of vehicle, to highlight the contribution rate of vehicles to pollution of air daily.

The graph below groups together all the vehicle brands with increasing order of CO₂ emissions. According to the figure, the least polluting brand is Citroën and the most polluting brand is Toyota. As for Morocco, the best-selling brand of vehicles is the Dacia, which is ranked 5th in terms of CO₂ emissions in the air. (Figure 6)

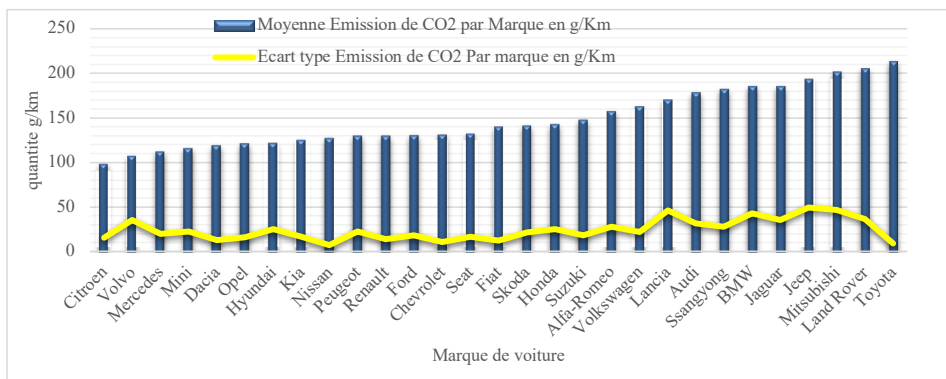


Figure 6: CO₂ emissions by vehicle brand

The data collected from the two sensors will be the subject of typical modeling and pilot for future sensors. As they will serve as a roadmap for the realization of an effective, real and achievable project.

It is in this logic that the present proposal relating to the creation of an application / Website accessible to the population and to the local authority of the city of Rabat is important and useful in the fight against pollution of the air caused by vehicles in a participatory approach.

References

- High Commission for Planning, General Population and Housing Census, 2014
- Ministry of the Interior, Monograph of the Rabat Salé Kénitra Region, 2015
- High Commission for Planning, General Census of Population and Housing, 1960, 1971, 1982, 1994, 2004, 2014.
- High Commission for Planning, General Population and Housing Census, 2004
- BaseMap, Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES / Airbus D S, U S D A, U S G S, AeroGRID, IGN, and the GIS User Community Arcgis 10.2.2
- Car brands website, brochures, 2020

Machine learning applications and process intelligence for cement industries

Christos Chatzilenas^a, Thanasis Gentimis^a, Theodore Dalamagas^a, Antonis C. Kokossis^c, Alexandros Katsiaboulas^b and Ioannis Marinos^b

^a*SYMBIOLABS Circular Intelligence LC, Artemidos 6 & Epidavrou str., Marousi 15125, Greece*

^b*TITAN SA, 22A Halkidos Str., Athens 11143, Greece*

^c*School of Chemical Engineering, National Technical University of Athens, 9 Iroon Poly-techniou str, Zografou 15780, Greece*

Abstract

Estimating energy consumption in cement mills is critical for the cement industry. Following data science practices and adopting machine learning (ML) technologies, we developed energy consumption prediction models for a cement mill of TITAN SA plant in Kamari Viotia. The models exploit historical sensor measurements and operational data and give predictions for energy consumption with accuracy better than almost one order of magnitude compared to existing baseline methods.

Keywords: machine learning, energy consumption, prediction, cement industry

1. Introduction

Cement grinding is a key process in the cement industry. During this phase, gypsum and mineral or artificial raw materials (e.g., pozzolan, ash, limestone) are added to the clinker (basic raw material for cement production) and, then grinded in special mills until a very small grain size is achieved. During the grinding process, corrective actions are constantly taken, in both the quota of the ingredients of the recipe and the operating parameters of the mill in order to achieve the appropriate fineness of the mixture. Large amount of electrical energy is consumed for mill operation. Therefore, it is essential for a cement industry to estimate its energy consumption in its cement mills. Estimation of the energy consumption can be used in a number of processes, such as assisting in the process of compliance with the requirements of ISO 50001 for energy efficiency, detecting changes in the operating parameters of the mill to save energy and executing scenarios for the energy consumption footprint of new mixtures.

To this end, we propose a model that estimates the energy consumption of a cement mill. We followed a data-driven approach, utilizing historical data recorded during the grinding process, and we designed and implemented machine learning models for energy prediction. Our models have been applied in TITAN SA plant in Kamari Viotia. Results show that we achieved an improvement of accuracy for energy consumption prediction by one order of magnitude, based on baseline method.

2. Background

TITAN uses a proprietary methodology for creating a model for estimating energy consumption in the company's cement mills. This methodology takes into consideration a number of variables, such as the cement mix composition and the product's blaine. At the core of this modelling approach is a linear function with weighted fractions of clinker, gypsum, slag, fly ash, pozzolan, limestone and other components in the product. The weights assigned to each component of the mix are estimated for each of TITAN's cement mills using linear regression on past data. The methodology also adds to the model corrections for over-grinding of the softer components, and also takes into consideration the operational parameters for each type of cement mill (e.g., mill type, separator type etc). The model created by this methodology, calculates the estimation of power consumption for a standard blaine value, and then, using an adjustment formula, corrects it to the final estimation based on the product's blaine.

In our approach we were granted full access to the proprietary methodology adopted by TITAN (not to be disclosed in this paper), in order to use it as a baseline method and compare the results of our methodology.

3. Dataset

3.1. Description

We collected hourly operations data from TITAN Kamari cement plant, near Athens, from January 2018 to August 2019. The processing of the data resulted in a dataset with 15,592 records and 43 features. Data is comprised of features for cement mix composition (e.g., the percentage of, clinker, gypsum, fly ash, etc), and operational parameters (e.g, like separator speed, fan speed, etc). Furthermore, measurements from sensors (temperature in various stages of the mill process, etc), quantities of additives and qualitative characteristics of the mix (blaine, etc) are some extra sets of features that were included in the data.

3.2. Data Exploration and Feature Correlation

Using statistical methods and data visualization, we got a better understanding of the data. Specifically, the correlation table indicated that attributes did not have high correlation, so there was no need of dimensionality reduction. Regarding data preprocessing, we deleted records with zero running time of the mill, as well as zero production and real blaine. Records that were associated with incorrect measurements or outliers were also deleted. Further preprocessing involved One-Hot Encoding for categorical attribute cement type. The final dataset consisted of 5,010 records and 39 features.

4. Methodology

Our methodology consisted of the following three steps.

- We designated which prediction methods we will apply in our experiments and we came up with 7 models.
- We examined which methods will boost our predictions and we concluded that cross-validation and feature selection was those that we are going to study.
- We considered various metrics, which are used in the existing bibliography, in order to compute the performance of our regression models.

4.1. Prediction Methods

We implemented a group of models using 7 prediction methods, grouped in 3 major categories, (a) Multiple Linear Regression, (b) Ensemble Methods, and (c) Hyperplanes.

Multiple Linear Regression includes the classic form of linear regression that calculates the coefficients of a linear function by minimizing the mean square error between the predicted and the actual values. Furthermore, Ridge and Lasso regressions both are a convention of the linear regression which minimize the mean square error plus one regularization term to keep the coefficients at a low range. However, these regularization terms are different as the Lasso regularizer leads to more aggressive reduction of the range of coefficients. This often generates zero coefficients corresponding to characteristics that have a small effect on the target value. Practically, it performs a feature selection.

In the Ensemble Methods, AdaBoost Regression (Freund and Schapire, 1995) (Drucker, 1997) and XGBoost Regression (T. Hastie and Friedman, 2009) the estimation of the target is done by combining estimates of many individual prediction models based on decision trees. On the contrary, in the Random Forest Regression (Breiman, 2001) the estimation of the target value is done by combining average estimation values of several individual prediction models based on classifying decision trees for a number of subsets of the data set. The difference between AdaBoost Regression and XGBoost Regression is how the combination of the estimations is done. In AdaBoost Regression the combination of estimations is done serially, so each new model corrects the previous one and is trained giving more weight to the training data for which the previous model showed underfitting behaviors. On the other hand, in XGBoost Regression every new model is trained to estimate the prediction errors of the previous model, so that the prediction of the previous model is corrected based on the new one.

Finally, Hyperplanes included Support Vector Regression-SVR which is based on the support vector machine classification model, which has been modified to predict continuous values. The main difference between SVR and linear regression is that the latter aims to minimize the mean square error between the predicted values and the actual values of the target value, while the SVR aims to limit the error to a range of values.

4.2. Feature Selection and Cross-Validation

Feature selection serves three basic purposes. To start with, it makes model training faster, since the size of training set is reduced. Also, since it performs dimensionality reduction, the risk of overfitting is being overcome. Lastly, it increases the accuracy of the prediction since features that add noise are not taken into account. We followed the feature selection method implemented in Guyon et al. (2002). This approach ends up with a feature ranking through recursive feature elimination and cross-validated selection of the best number of features. More particularly, starting with a list of all features, the method recursively deletes one feature in every repetition. Finally, we came up with a list of features that returned the best results during the evaluation of the model based on the validation set. Table 1 shows the number of features recommended in every prediction model.

Prediction methods	Number of features selected
Linear Regression	30
Ridge Regression	28
Lasso Regression	25
AdaBoost Regression	22
XGBoost Regression	16
Random Forest Regressor	23
Support Vector Regression	22

Table 1. Number of features selected in each prediction method

Dataset records	Training set	Validation set	Folds	Testing set
5,010	4,008	1,002	4	1,002

Table 2. Dataset splits

A training set is a set of data with records that give us the target value along with the values of other attributes that affect it. Usually, such a set results from data files in which each record relates to the same reference period. Conventionally, for the implementation of machine learning projects the 70-80% of the whole dataset is used as the training set and the rest as testing set which is used for the evaluation of the model. This method often leads to overfitting, with the model performing extremely well in predicting values when the input belongs to the testing set, but it fails if the input is foreign to the testing set.

In our case, model implementation was performed with the method of cross-validation (T. Hastie and Friedman, 2009). Specifically, we applied K-Fold cross-validation. This method uses different parts of data to fit and test the model. In particular, the method consists of the following steps:

1. Data is splitted into K equal-sized folds.
2. For the k-th fold, the model is fitted to the other K-1 folds.
3. Prediction error of the fitted model is calculated when predicting the k-th fold.
4. The second and third step are repeated for all folds and the K estimates of prediction error are combined.

For the k-fold cross-validation, we have split the training set into 4 folds with suffling. Table 2 shows how we split the whole dataset in training, validation and testing sets.

4.3. Evaluation Metrics

For every dataset that we used in our methodology (training, validation and testing), we calculate evaluation metrics to understand the model's behavior and its' predictive ability.

Prediction methods	Datasets	R-squared	MAE	MSE	RMSE
Baseline	Training set	-0.1497	4.0933	43.3750	6.5859
	Validation set	-	-	-	-
	Testing set set	-0.1497	4.2088	48.6345	6.9738
AdaBoost Regression	Training set	0.6847	1.7114	11.9253	3.4533
	Validation set	0.6645	1.7218	12.997	3.6052
	Testing set	0.5775	1.7777	17.963	4.2383
Random Forest Regression	Training set	0.9429	0.4854	2.1615	1.4702
	Validation set	0.7251	0.8029	10.5898	3.2541
	Testing set	0.6282	0.8349	15.8105	3.9762
XGBoost Regression	Training set	0.9999	0.005	0.0001	0.0128
	Validation set	0.7370	0.671	10.1639	3.188
	Testing set	0.6564	0.6613	14.608	3.822

Table 3. Results in training, validation and testing sets for baseline and 3 best predictions methods

A simple metric is Mean Absolute Error (MAE) which is the sum of the average of the absolute difference between the predicted and actual values. Another metric used is the Mean Squared Error (MSE), which is the same as the MAE, but the difference is that it

squares the difference between actual and predicted values before summing them all. Root Mean Squared Error (RMSE) is taking the square root of MSE. Also, we used R2 which quantifies the quality of fit of a set of predicted output values to the actual output values.

5. Results

We performed our experiment on a Dell PowerEdge R430 Intel(R) Xeon(R) CPU E5-2609 v4 @ 1.70GHz with 14 cores and 16G RAM running Ubuntu 18.04 LTS, kernel 4.15.0-99. For model training, we used two open-source Python libraries, scikit-learn (Buitinck et al., 2013) which provides tools for model fitting, data preprocessing, model selection and evaluation, and XGBoost (Chen and Guestrin, 2016) which is an optimized distributed implementation of Gradient Boosting.

As mentioned above, in our experiment the K-Fold cross-validation we have split the training set into 4 folds with shuffling. Initially, we examined the evaluation metrics in a validation set. For Multiple Linear Regression methods, the R-squared and MSE in every fold of the validation set had high standard deviation, indicating poor performance of these methods. Hence, it was not worthwhile to check their evaluation metrics in the testing set. The same behavior was observed for Support Vector Regression. In contrast, Ensemble methods performed very well, so we continued by exploring the evaluation metrics, for those methods, in the testing set. Table 3 shows the results in training, validation and testing sets for Ensemble methods.

The baseline method had a poor performance, indicated by the negative R-squared, and high values of MAE, MSE and RMSE. Among Ensemble methods, AdaBoost Regression had the worst performance with an R-squared in the testing set not exceeding 0.6, and MAE almost 1.8. Concerning Random Forest Regression, even though the results of the method in the training set seemed promising, its performance in the validation and the testing sets wasn't as good as expected. However, we detected a significant improvement in the MAE which was reduced in half. Finally, XGBoost Regression delivered the best results among all methods, regarding all evaluation metrics, since the R-squared in the testing set fluctuated near 0.65, the MAE was reduced even more and the MSE was the lowest observed among all three methods. Comparing the results from the XG-Boost Regression to those from the baseline method, we observe a decrease both in MAE and MSE. In general, Ensemble methods delivered improved results according to the R-squared. To this end, as results demonstrate, the machine learning methods used in this study, outperformed the baseline method.

6. Conclusions and Future Work

Using data science practices and adopting machine learning (ML) technologies, we have developed energy consumption prediction models for a cement mill of TITAN SA plant featuring superior performance over conventional models used in the past. The models exploit historical sensor measurements and operational data and give predictions for energy consumption achieving an improvement of accuracy for energy consumption prediction by one order of magnitude, based on baseline method.

Specifically, the top three models reached MAE less than 1.77 (the existing baseline method has a MAE >4.00) and R2 more than 0.57 (the existing baseline method has a R2 <0.2). The models produced have so far exploited data streams, not first principles. Future work will exploit such principles to improve the performance towards the development of hybrid process engineering models, to generalize the approach, and to

enable the analysis with a potential to extract knowledge from the data and support decisions.

References

- L. Breiman, Oct 2001. Random forests. *Machine Learning* 45 (1), 5–32.
- L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, G. Varoquaux, 2013. API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*. pp. 108–122.
- T. Chen, C. Guestrin, 2016. Xgboost: A scalable tree boosting system. *CoRR* abs/1603.02754.
- H. Drucker, 08 1997. Improving regressors using boosting techniques. *Proceedings of the 14th International Conference on Machine Learning*.
- Y. Freund, R. E. Schapire, 1995. A decision-theoretic generalization of on-line learning and an application to boosting. In: P. Vitányi (Ed.), *Computational Learning Theory*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 23–37.
- I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Jan 2002. Gene selection for cancer classification using support vector machines. *Machine Learning* 46 (1), 389–422.
- R. T. T. Hastie, J. Friedman, 2009. *Elements of Statistical Learning* Ed. 2. Springer.

Process Modelling and Simulation of Volatile Organic Compound (VOC) Recovery from Pharmaceutical Gas Emission Streams

Vasiliki E. Tzanakopoulou,^a Daniel Castro-Rodriguez,^b Dimitrios I. Gerogiorgis^{a,*}

^a*School of Engineering (IMP), University of Edinburgh, Edinburgh, EH9 3FB, UK*

^b*GlaxoSmithKline (GSK), Montrose, Angus, DD10 8EA, UK*

**D.Gerogiorgis@ed.ac.uk*

Abstract

Volatile Organic Compounds (VOCs) are a class of solvents extensively used for the production of Active Pharmaceutical Ingredients (APIs) in the pharmaceutical industry. To ensure adherence to environmental regulations, adsorption is widely implemented as an effective VOC emission abatement method. The present work relies on an adsorption model which accounts for mass transfer by advection-dispersion in the gas phase and pore diffusion in the solid phase. Kinetics are described via the Linear Driving Force (LDF) model with a lumped mass transfer coefficient; the adsorption equilibrium concentration is expressed via the Langmuir isotherm (single-component adsorption) and the Extended Langmuir Isotherm (binary mixture adsorption). Dynamic simulation results highlight the breakthrough behaviour of key VOCs and mixtures towards improved process efficiency.

Keywords: Process simulation, Volatile Organic Compound/VOC, emissions, adsorption

1. Introduction

Primary (upstream) pharmaceutical manufacturing is characterized by the consumption of large solvent volumes, key in reactions and separations: Volatile Organic Compounds (VOCs) are prominent but also problematic, as their high volatility at standard conditions has catastrophic effects on both human health and environment (Perez-Vega et al., 2013). Adsorption is an effective abatement method, exploiting the potential of porous materials for selective VOC capture (Yang et al., 2019); others (membranes, cryo-condensation, thermal/catalytic oxidation) can be costlier (Belaissaoui et al., 2016; Chung et al., 2019). Widely employed for high volumetric flows at low VOC concentrations (Das et al., 2004), adsorption offers high efficiency, low energy demand and easy installation/maintenance. However, adsorbent purchase and VOC-saturated bed regeneration costs pose challenges.

The present paper describes the use of a published fixed bed adsorption model to analyse the operation of an activated carbon bed for adsorption of VOCs from pharma emissions. The first-principles dynamic model is used to simulate adsorption of solvents and binary mixtures (methanol, ethanol, acetone, dichloromethane, chloroform, benzene, toluene). Specifically, the adsorbate mass balance describes the VOC concentration distribution between gas and solid phase in the fixed bed, the Linear Driving Force (LDF) kinetic model approximates VOC adsorption rate onto particles,

and the Langmuir isotherm is employed to quantify the adsorbate equilibrium distribution onto the solid carbon phase. Breakthrough curves for single-component and binary VOC system adsorption on various adsorbents can critically assist towards informed operational decisions in regard to raw feed handling (e.g. mixing, sequencing, conditions) to achieve OpEx cost minimization.

2. Dynamic Process Modelling: Assumptions and Structure

The mathematical model used in the present study relies on the following assumptions:

1. Isothermal conditions are ensured throughout the bed, with negligible pressure drop and constant fluid velocity due to low VOC concentrations considered (Das et al., 2004).
2. Radial concentration gradients and carrier gas adsorption are considered negligible.
3. The gas behaviour can be considered to obey the ideal gas law (Tefera et al., 2013).
4. The VOC mass transfer in the activated carbon bed considered for pharma emission treatment occurs by advection-dispersion in gas phase and by diffusion in solid phase.
5. The mass transfer rate in the solid phase is approximated by the LDF model which is characterized as “simple, analytical and physically consistent” (Sircar and Hufton, 2000).
6. Equilibrium obeys the Langmuir model for pure substances (Tefera et al., 2013) and the Extended Langmuir model (Jain & Snoeyink) for binary mixtures (Tefera et al., 2014).

$$q_{e,1} = \frac{(q_{m,1} - q_{m,2})b_1C_1}{1 + b_1C_1} + \frac{q_{m,2}b_1C_1}{1 + b_1C_1 + b_2C_2} \quad (13)$$

$$D_{ax,i} = \left(\alpha_0 + \frac{Sc_i Re_p}{2} \right) \frac{D_{AB,i}}{\varepsilon_b} \quad (14)$$

$$\varepsilon_b = 0.379 + \frac{0.078}{\left(\frac{D_b}{d_p} - 1.8 \right)} \quad (15)$$

$$D_{AB,i} = 10^{-3} T^{1.75} \frac{\sqrt{\frac{M_A + M_B}{M_A M_B}}}{P \left(\left(\sum \nu \right)_A^{0.33} + \left(\sum \nu \right)_B^{0.33} \right)^2} \quad (16)$$

$$-\Delta H_{ad,i} = 103.2 + 1.16a_i + 0.76\Delta H_{vap,i} - 3.87(IP_i) - 0.7\gamma_i - 26.19w_{mic} \quad (16)$$

$$\frac{\partial q_i}{\partial t} = k_{LDF,i} (\rho_p q_{e,i} - q_i) \quad (17)$$

$$C_i(z=0, t) = C_{0,i} \quad (17)$$

$$k_{LDF} = \frac{60\varepsilon_p C_{o,i} D_{eff,i}}{q_{o,i} \tau_p d_p^2} \quad (18)$$

$$C_i(z, t=0) = 0 \quad (18)$$

$$\varepsilon_p = V_{pore} \rho_p \quad (19)$$

$$q_i(z, t=0) = 0 \quad (19)$$

$$\tau_p = \frac{1}{\varepsilon_p^2} \quad (8) \quad \frac{\partial c_i}{\partial z}(z = L, t) = 0 \quad (20)$$

$$\rho_p = \frac{\rho_b}{1 - \varepsilon_b} \quad (9) \quad q_{o,i} = \rho_p q_{e,i} \quad (21)$$

$$\frac{1}{D_{eff,i}} = \frac{1}{D_{AB,i}} + \frac{1}{D_{k,i}} \quad (10) \quad Re_p = \frac{d_p V_s \rho_m}{\mu_m} \quad (22)$$

$$D_{k,i} = 97 r_p \sqrt{\frac{T}{M_A}} \quad (11) \quad Sc_i = \frac{\mu_m}{\rho_m D_{AB,i}} \quad (23)$$

$$q_e = \frac{q_m b C}{1 + b C} \quad (12) \quad V_s = \frac{4Q}{\pi D_b^2} \quad (24)$$

A short description follows: on Eq. (1) C , $D_{ax,i}$, u , ε_b , q are the gas phase VOC concentration (kg m^{-3}), axial dispersion coefficient ($\text{m}^2 \text{s}^{-1}$), interstitial velocity (m s^{-1}), bulk bed porosity and adsorbed phase VOC concentration (kg m^{-3}), respectively. Eq. (2) of Dantas et al. (2011) and Eq. (3) further introduce Sc_i , Re_p , $D_{AB,i}$, α_0 , D_b and d_p as the Schmidt number of component i , the Reynolds number (adsorbent particle), the molecular diffusivity ($\text{m}^2 \text{s}^{-1}$), the empirical mass diffusion correction factor (20), the internal bed diameter (m) and the (considered) unidisperse particle diameter (0.00075 m) respectively.

In Eq. (4), Σv is the atomic diffusion volume (A: VOC, B: air), T is temperature (K), P is pressure (atm) and M is molecular weight (g mol^{-1}). Eq. (5) features $k_{LDF,i}$ as LDF mass transfer coefficient (s^{-1}) and $q_{e,i}$ is the inlet PT adsorbent equilibrium capacity (kg kg^{-1}). In Eq. (6), ε_p , $C_{o,i}$, $D_{eff,i}$, $q_{o,i}$ and τ_p are the particle porosity, inlet concentration (kg m^{-3}) and effective diffusion coefficient of component i ($\text{m}^2 \text{s}^{-1}$), the concentration at the solid phase which is in equilibrium with C_o (kg m^{-3}), and the particle tortuosity, respectively. Eqs. (7)-(11) present ρ_p , ρ_b , V_{pore} , $D_{k,i}$ and r_p as particle/bed density (kg m^{-3}), pore volume ($\text{m}^3 \text{kg}^{-1}$), Knudsen diffusivity coefficient ($\text{m}^2 \text{s}^{-1}$) and particle radius (m) respectively.

Beyond the Eqs. addressing macroscopic phenomena, we need microscopic descriptions. Thus, Eqs. (12)-(15) introduce $q_{e,i}$, $q_{m,i}$, b_i , $b_{o,i}$, and $\Delta H_{ad,i}$ as the equilibrium and maximum adsorption capacity of component i (kg kg^{-1}), the Langmuir affinity coefficient ($\text{m}^3 \text{kg}^{-1}$), the pre-exponential constant ($\text{m}^3 \text{kg}^{-1}$) and the heat of adsorption (kJ mol^{-1}), respectively. Eq. (16) presents α_i , $\Delta H_{vap,i}$, IP_i , γ_i and w_{mic} as polarizability (10^{-24} cm^3), heat of vaporization (kJ mol^{-1}), ionization potential (eV), VOC surface tension (mN m^{-1}) and average micropore width (nm) respectively. Finally, Eqs. (22)-(24) provide dimensionless number definitions with V_s , ρ_m , μ_m and Q denoting the superficial velocity (m s^{-1}), system density (kg m^{-3}), system viscosity (Pa s) and volumetric flowrate ($\text{m}^3 \text{s}^{-1}$), respectively.

3. Dynamic Model Parameters for Adsorption Systems

The dynamic model is used to simulate VOC adsorption on activated carbon, as organic emissions from distillation and pressure filtration units occur during API manufacturing. The original set of PDEs is reduced to a set of ODEs using the Method of Lines for space (central finite difference) discretisation and subsequent solution in MATLAB v.

2020a. The adsorption of acetone, benzene and ethanol with air, as well as toluene with either air or N₂ as the carrier gas (Tefera et al., 2013; Delage et al., 2000) has been examined; we also explore adsorption of two binary mixtures (acetone-ethanol, toluene-ethanol) at the same (P = 1 atm, T = 300 K) conditions, on another (BAC) adsorbent (Tefera et al., 2014). Benzene and air are used for model validation purposes, not due to industrial relevance.

The system viscosities are determined through Wilke's equation (Dantas et al., 2011) for binary mixtures (air: T = 293 K, 300 K, VOCs: T = 293 K, 298 K, 300 K); moreover, system densities are computed from pure component data via mixing rules (NIST, 2020).

Air as a carrier is a binary mixture (N₂:O₂=79:21% v/v); x_{VOC} is the VOC molar fraction. Tables 1 and 2 summarise all essential parameter values for the determination of the axial dispersion coefficient ($D_{\text{ax},i}$) for all single components and binary mixtures, respectively.

Table 1: Parameters for axial dispersion coefficient calculation (single-component adsorption).

System	ρ_m (kg m ⁻³)	μ_m (Pa s)	Re	Sc	$D_{\text{AB},i}$ (m ² s ⁻¹)	x_{VOC}
Acetone–Air	1.95	$2.15 \cdot 10^{-5}$	62.17	1	$1.10 \cdot 10^{-5}$	0.001
Benzene–Air	2.04	$2.25 \cdot 10^{-5}$	62.15	1.19	$9.28 \cdot 10^{-6}$	0.001
Ethanol–Air	1.99	$7.62 \cdot 10^{-5}$	2.72	3.11	$1.23 \cdot 10^{-5}$	0.021
Toluene–Air	2.03	$2.25 \cdot 10^{-5}$	61.93	1.33	$8.35 \cdot 10^{-6}$	0.001
Toluene–N ₂	2.00	$1.98 \cdot 10^{-5}$	69.14	0.27	$3.73 \cdot 10^{-5}$	0.001

Table 2: Parameter values for axial dispersion coefficient calculation (binary mixture adsorption).

System	ρ (kg m ⁻³)	μ (Pa s)	Re	Sc	D_{AB} (m ² s ⁻¹)	x_{VOC}
Acetone–Ethanol	1.94	$2.24 \cdot 10^{-5}$	59.45	1.05	$1.10 \cdot 10^{-5}$	0.0005
			59.45	0.90	$1.28 \cdot 10^{-5}$	0.0005
Toluene–Ethanol	1.98	$2.33 \cdot 10^{-5}$	58.33	1.41	$8.35 \cdot 10^{-6}$	0.0005
			58.33	0.92	$1.28 \cdot 10^{-5}$	0.0005

Table 3 summarizes the model parameter values of each system. The adsorbents modelled are beaded activated carbon (Tefera et al., 2013) and Picactiv NC 60 granular carbon ($\rho_b = 430$ kg m⁻³) (Delage et al., 2000). For ethanol, we have computed $\Delta H_{\text{ad}} = -28700$ J/mol.

Table 3: Parameter values for single-component and binary mixtures used in VOC adsorption.

System	$D_{\text{ax},i}$ (m ² s ⁻¹)	L (m)	u (m s ⁻¹)	q_m (kg kg ⁻¹)	b (m ³ kg ⁻¹)	ε	C_0 (kg m ⁻³)	k_{LDF} (s ⁻¹)
Acetone–Air	0.0015	0.12	2.405	0.41	267.02	0.38	0.00236	$6.7 \cdot 10^{-5}$
Benzene–Air	0.0014	0.12	2.405	0.42	814.64	0.38	0.00317	$4.0 \cdot 10^{-5}$
Ethanol–Air	0.0007	0.20	0.348	0.39	170.07	0.40	0.04700	$3.2 \cdot 10^{-4}$
Toluene–Air	0.0013	0.12	2.405	0.42	3412.18	0.38	0.00374	$3.4 \cdot 10^{-5}$
Toluene–N ₂	0.0009	0.12	2.405	0.42	3412.18	0.38	0.00374	$3.5 \cdot 10^{-5}$
Acetone–Ethanol	0.0015	1	2.405	0.41	267.02	0.38	0.00118	$5.87 \cdot 10^{-5}$
	0.0016	1	2.405	0.39	129.19	0.38	0.00094	$1.43 \cdot 10^{-4}$
Toluene–Ethanol	0.0013	1	2.405	0.42	3412.18	0.38	0.00187	$1.87 \cdot 10^{-5}$
	0.0016	1	2.405	0.39	129.19	0.38	0.00094	$7.49 \cdot 10^{-4}$

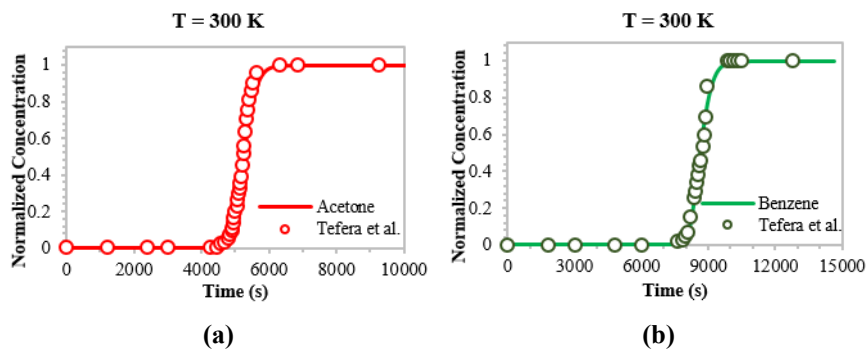
4. Results and Discussion

Breakthrough curves of VOC components in all examined systems have been normalised by adsorbate inlet concentration and illustrated in Fig. 1, for acetone, benzene, ethanol, toluene (two carriers) and two binary mixtures (acetone-ethanol, toluene-ethanol). Dynamic model results have been validated vs. published experimental single-component breakthrough curves which dictated our inputs (Tefera et al., 2013; Delage et al., 2000). For all pure component adsorbates (air or N₂ carrier) and operating conditions considered, our computed breakthrough are found to be in great agreement with experimental data.

Benzene and toluene demonstrate similar adsorption (breakthrough onset and duration) behaviour, despite the differing LDF mass transfer coefficient values shown on Table 3; these observations are in excellent agreement with previous findings (Tefera et al. 2013).

Acetone, on the other hand, shows markedly earlier (and somewhat faster) adsorption. Ethanol also demonstrates fast adsorption (the fastest of all substances studied here), due to its high LDF mass transfer coefficient. The lower operating temperature (T = 293 K) used for comparison vs. experimental data (Delage et al., 2000) also affects breakthrough curve features, as adsorption is exothermic. Further investigating the magnitude of radial dispersion contributions but also the effect of non-isothermal operation for either pure or mixture feeds will provide significant benefits towards ensuring high VOC model fidelity. The carrier gas effect is examined by simulating toluene adsorption with N₂ vs. air: the axial dispersion coefficient is computed as 31% lower for the pure N₂ carrier gas case ($D_{ax}/N_2 = 0.0009$ vs. $D_{ax}/air = 0.0013$ m² s⁻¹), resulting in a steeper breakthrough curve.

Binary mixtures adsorbing on BAC appear very similar: ethanol breakthrough starts much faster than the other (acetone, toluene) component, as per its pure-component behaviour and due to favourable adsorption kinetics, and a likely weaker affinity for the adsorbent.



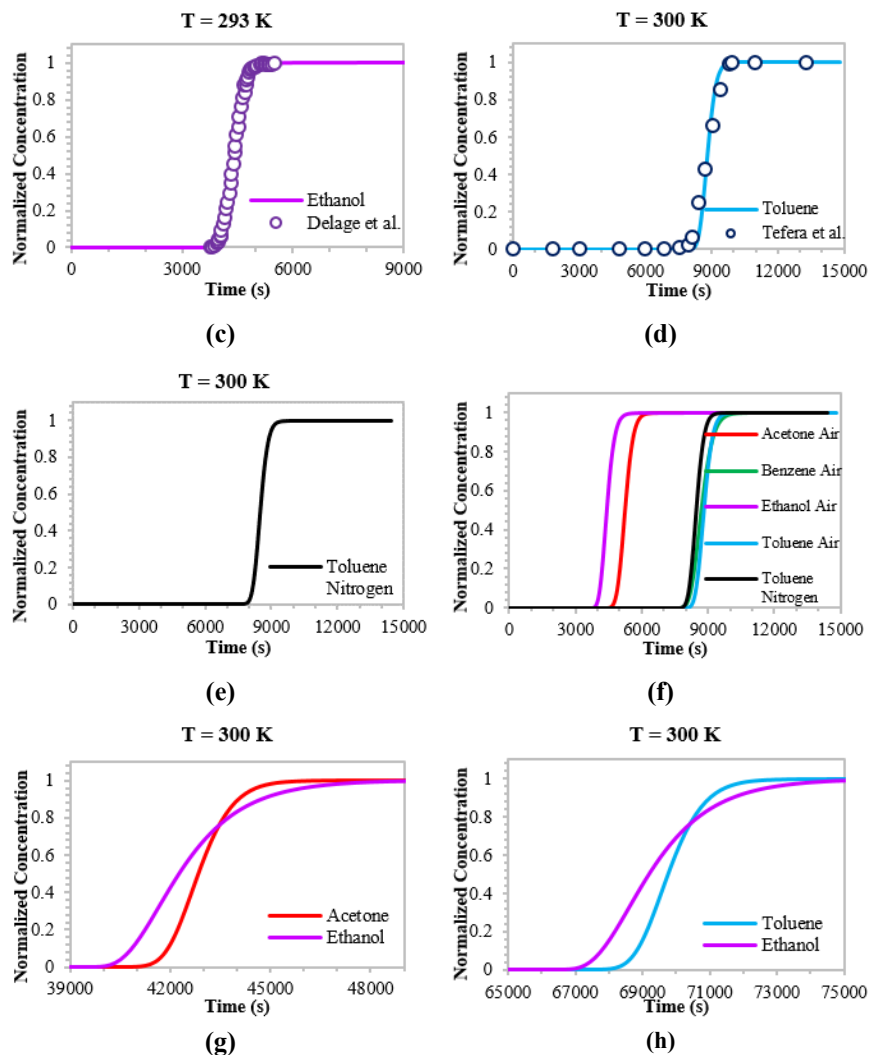


Figure 1. Breakthrough curves vs. published experiments (Tefera et al., 2013; Delage et al., 2000).

5. Conclusions

Upstream pharmaceutical manufacturing inevitably relies on significant solvent volumes for many (especially reaction/separation) uses, with great environmental repercussions. Efficiently capturing Volatile Organic Compounds (VOC) from pharma emission streams is a strategic priority for advanced manufacturing striving to embrace greener practices, hence inducing the research challenge of ensuring plant longevity and cost-effectiveness.

Adsorption of VOC pollutants on activated carbon beds is an industrially implemented solution for lowering emissions, so overcoming its efficiency limitations is a necessity. Modelling pharma solvent adsorption systems is tackled by many literature studies, yet without a concerted effort for a unified platform encompassing both technical (feed

variability, adsorption kinetics) and economic (maintenance, scheduling) problem vistas. This paper relies on an adsorption model for pharma VOC recovery, which considers axial dispersion and Linear Driving Force mass transport under isothermal conditions. Dynamic simulations of the adsorption behaviour of several VOC-air systems show clear differences between polar (acetone, ethanol) and non-polar (benzene, toluene) solvents. The binary systems examined demonstrated no competitive adsorption, which however may occur if multi-component raw organic streams are fed to the same adsorption bed. The vastly different breakthrough timescales (in the order of hours) vary by mixture composition but also by adsorbent (e.g. activated carbon vs. Picaktif vs. BAC) indicate that high-fidelity PDE models are critical for economically relevant process optimization.

Acknowledgements

The authors gratefully acknowledge the SRPe-NMIS IDP PhD Scholarship awarded to V.T., as well as a Royal Society Short Industrial Fellowship (2020-21) awarded to D.I.G.

References

- B. Belaissaoui et al., 2016. Energy efficiency of a hybrid membrane/condensation process for VOC recovery from air: A generic approach, *Energy*, **95**(2), 291–302.
- W.-C. Chung et al., 2019. Removal of VOCs from gas streams via plasma and catalysis, *Catalysis Reviews Science and Engineering*, **61**(2), 270-331.
- T.L.P. Dantas et al., 2011. Modeling of the fixed-bed adsorption of CO₂ and CO₂-N₂ mixture on zeolite 13X, *Brazilian Journal of Chemical Engineering*, **28**, 3, 533-544.
- D. Das et al., 2004. Removal of volatile organic compound by activated carbon fiber, *Carbon*, **42**, 14, 2949-2962.
- F. Delage et al., 2000. Mass transfer and warming during adsorption of high concentrations of VOCs on an activated carbon bed, *Environmental Science & Technology*, **34**, 22, 4816-4821.
- S. Giraudet et al., 2006. Estimation of adsorption energies using the physical characteristics of activated carbons and the molecular properties of VOCs, *Carbon*, **44**, 12, 2413-2421.
- National Institute of Standards & Technology, *NIST Chemistry WebBook*, NIST Std. Ref. Database Number 69, Gaithersburg, MD, USA - <https://doi.org/10.18434/T4D303>, (Nov. 8, 2020).
- S. Perez-Vega et al., 2013. A system view of solvent selection in the pharmaceutical industry: Towards a sustainable choice, *Environment, Development and Sustainability*, **15**, 1, 1-21.
- S. Sircar and J.R. Hufton, 2000. Why does the linear driving force model for adsorption kinetics work? *Adsorption*, **6**, 2, 137-147.
- D.T. Tefera et al. 2013. Two-dimensional modeling of volatile organic compounds adsorption onto beaded activated carbon, *Environmental Science & Technology*, **47**, 20, 11700-11710.
- D.T. Tefera et al. 2014. Modeling competitive adsorption of mixtures of volatile organic compounds in a fixed-bed of beaded activated carbon, *Environmental Science & Technology*, **48**, 9, 5108-5117.
- C. Yang et al., 2019, Abatement of various types of VOCs by adsorption/catalytic oxidation: A review, *Chemical Engineering Journal*, **370**, 1128–1153.

Multi-zonal compartmentalization methodology for surrogate modelling in continuous pharmaceutical manufacturing

Pooja Bhalode ^a, Marianthi Ierapetritou ^{b,*}

^a*Rutgers University, 98 Brett Road, Piscataway, NJ, USA 08901*

^b*University of Delaware, 150 Academy Street, Newark, DE, USA 19716*

mgi@udel.edu

Abstract

Detailed process understanding of powder systems is important for ensuring regulation compliant manufacturing of pharmaceutical solid-based drug products. However, this can get challenging due to the current limitations in experimental and computational approaches, constraining the available knowledge of critical quality attributes. To address this drawback, we propose the use of computationally efficient and mechanistically informed surrogate models developed using multi-zonal compartmentalization methodology. In this work, the proposed modelling strategy is demonstrated for continuous powder blenders, wherein, first principle-based discrete element simulation is used to obtain an inter-connected network of zones for detailed evaluation of mixing within the system. The obtained surrogate models are further validated using the original simulation, illustrating the accuracy of the proposed methodology.

Keywords: multi-zonal compartmentalization, continuous pharmaceutical manufacturing, surrogate modelling, continuous powder blender, radial mixing.

1. Introduction

Given the significance of ensuring product quality of pharmaceutical solid-based drug products, the regulatory agencies like U.S. Food and Drug Administration have proposed several initiatives to guide and regulate drug product manufacturing. Quality-by-Design (QbD) (Yu et al. 2014) is one such initiative focusing on designing the process such that end product quality is maintained. However, to successfully implement this initiative, it is important to obtain detailed mechanistic information of the overall process. Several research efforts (Ierapetritou et al. 2016) over the past years have contributed towards this goal of improving process understanding of pharmaceutical unit operations through use of experimental and simulation techniques. However, for some unit operations like continuous powder blender, the experimental techniques like particle tracking (Portillo et al. 2010) can get very expensive to capture powder mixing within the blender. Furthermore, researchers (Muzzio et al. 1997) have observed difficulties like disruption of powder bed and identification of adequate number of samples while using experimental technique for measurement of powder

mixing through use of sampling probes. These challenges hinder the accurate assessment of radial mixing and blend uniformity which is a critical quality attribute. On the other hand, first principle-based simulation techniques like discrete element modelling (Cundall and Strack 1980) can get computationally intensive as it involves evaluation of particle contacts and contact forces for all particles at each time step. To address the above challenges for acquiring assessment of powder mixing within these systems in reasonable computation time, we propose the use of surrogate model using multi-zonal compartmentalization methodology for modelling pharmaceutical unit operations. In this work, the proposed methodology is demonstrated for continuous powder blender. The following sections highlight the details of the different techniques utilized in this work, followed by results of the compartmentalization methodology and model validation. Lastly, section 6 focuses on the conclusions of the proposed work.

2. Discrete element modelling

Discrete element modelling (DEM) is a first principle-based technique to simulate particle dynamics in powder systems using Newton's laws of motion and contact laws based on discrete particle physics. Details of the equations involved in DEM are described in literature (Dubey et al. 2011; Bhalode and Ierapetritou 2020b) and has been widely applied for powder systems (Ketterhagen et al. 2009). Following the construction of the desired geometry and addition of particles within DEM, this method involves detection of particle contacts, evaluation of contact forces and corresponding particle trajectories for all particles at each time step, thus making it computationally intensive. In this work, we implement DEM for simulation of continuous powder blender using Hertz-Mindlin contact model, wherein, we simulate a periodic section (Gao, Yijie et al. 2012; Bhalode and Ierapetritou 2020a) of the blender. Here, a central section of the blender is used with periodic boundaries to replicate the powder behaviour as observed in the blender. In this work, two types of spherical, cohesionless particles are added to the periodic section at the start of the simulation, with a total of 3000 particles for each particle type, and the periodic section is simulated for 20 seconds with a blade speed of 100 RPM.

3. Multi-zonal compartmentalization methodology for surrogate modelling

Multi-zonal compartmentalization methodology is a hybrid technique to model complex systems. This method captures multi-scale information using computationally efficient and mechanistically informed surrogate models, while combining systemic or global and local process information (Jourdan et al. 2019). It is widely used in literature for different applications like bioreactors, fluidized columns and crystallizers. Multi-zonal compartmentalization transforms a high-fidelity simulation like DEM model of a process unit into an inter-connected network of compartments, where individual DEM grids are classified into compartments, such that the process variables of interest of each compartment are maintained within user-defined tolerance limits. This inter-connected network acts as a computationally efficient surrogate, while replicating the observed process behaviour. In this work, we implement this methodology for the continuous powder blender to develop surrogate representation of the DEM model, which can later be incorporated within process flowsheets. The proposed methodology is illustrated in Figure 1, where we simulate a periodic section of the blender in DEM. This section is

divided into 10 slices along the axial direction, and each slice is further divided into 20x20 grids along Y and Z directions. The averaged particle velocities are extracted for all grids in these slices and compartmentalized into multiple zones. These zones include the empty zone where no particles are present, and filled zone where particles are present. The filled zone is further divided into compartments corresponding to particle velocities being greater than 0, equal to 0, less than 0 and the dead zone. This zoning strategy when applied to Y and Z directions is termed as radial compartmentalization and when applied to the X direction, is termed as axial compartmentalization.

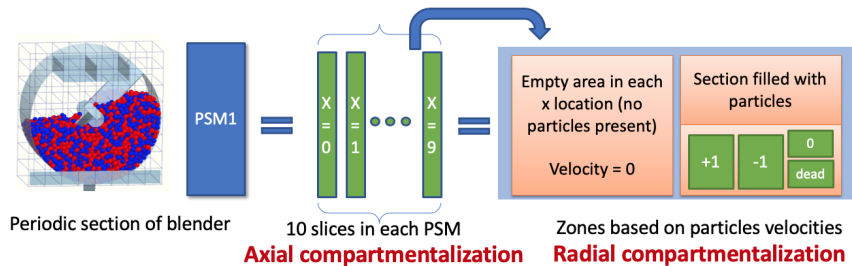


Figure 1: Modelling strategy for multi-zonal compartment modeling of periodic section of the continuous powder blender, illustrating axial and radial compartmentalization

4. Results

4.1. Radial compartmentalization

Based on the averaged velocities of individual grids extracted from DEM, the radial compartmentalization is developed for each slice along Y and Z directions. The individual grids are categorized into '+1', '0' and '-1' zones based on their averaged velocities and radial compartments are obtained as shown in Figure 2A.

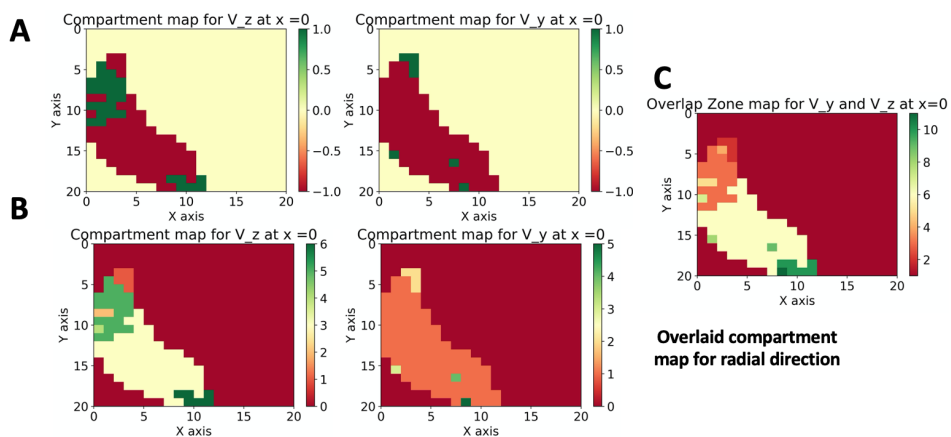


Figure 2: Radial compartment maps for slice $x = 0$. (A) the radial compartment maps for Y and Z directions, (B) corresponding distinct compartments for Y and Z directions, (C) the overlaid compartment map for radial direction.

Each compartment in the compartment maps is labelled separately to create distinct compartments as shown in Figure 2B. This is important for development of flowrates

(Delafosse et al. 2014) between the proposed inter-connected network of zones. The distinct compartments along Y and Z directions are combined to develop an overlaid compartment map (Tajsoleiman et al. 2019) for radial direction as shown in Figure 2, for slice corresponding to $x = 0$. Following the evaluation of overlaid compartment map, the inter-connection flowrates between compartments (F_{ij}) are evaluated using the interfacial area between compartments (A_{ij}) and averaged velocities of grids (v_i, v_j) along specific direction (k), as shown in Eq. (1) (Delafosse et al. 2014) where (i) and (j) correspond to different compartments. The flowrates are evaluated separately for Y and Z directions to convert the compartment map into an inter-connected network of zones.

$$F_{ij}(k) = A_{ij} * (v_i(k) - v_j(k)) \quad (1)$$

4.2. Axial compartmentalization

A similar strategy is implemented for axial compartmentalization of all slices along X direction in the periodic section. Here, the compartments are developed for each slice based on averaged velocities of grids along the X axis and categorized into compartments marked by '+1', '0', and '-1', as defined in section 3. The empty grids with no particles are classified into a separate section marked by '-2'. The axial compartment maps thus developed, are shown in Figure 3 for slices corresponding to $x = 0$ and $x = 1$. Following the axial compartmentalization, the inter-connection flowrates between these compartments are evaluated using Eq. (1), with interfacial area equal to the total area occupied by each compartment. The flowrates thus evaluated, provide forward and backward flows for each slice, resulting in a net forward flow of particles for the periodic section.

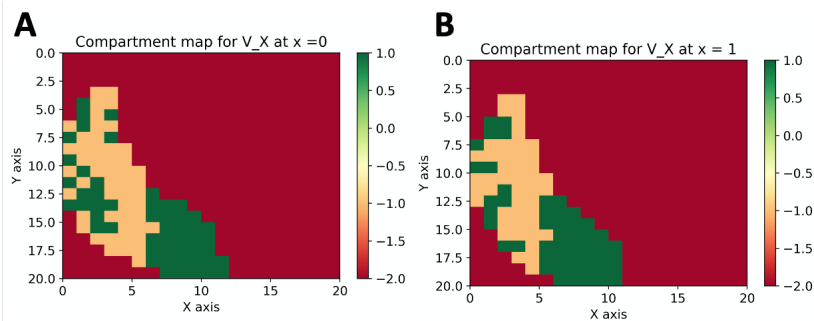


Figure 3: Axial compartments developed for slices $x = 0$ and $x = 1$ in the periodic section

Thus, combining the axial and radial compartmentalization strategies, we can obtain a surrogate model based on developing inter-connected network of zones, to simulate the powder flow as observed in original DEM model.

5. Validation of multi-zonal compartment model

5.1. Validation of radial compartmentalization

The proposed radial compartmentalization is validated based on the evaluation of powder mixing within the periodic section. In this study, a variance-based index,

relative standard deviation (RSD) (Gao, Yijie et al. 2012), is used to compare the surrogate model based on compartments with the original DEM model and evaluated using Eq. (2). For compartment model, C_i corresponds to the concentration of individual compartments, with C_{av} being the averaged concentration, and N equal to the total number of compartments. This equation is also applied to evaluate the RSD for DEM model where, C_i is the concentration of individual grids, with N being the total number of grids. Following the evaluation of RSD, T_{95} mixing time (Bhalode and Ierapetritou 2020a) is evaluated to obtain a quantitative comparison of the two models. This corresponds to the time it takes for the system to reach 95% of total mixing in the system, shown in Eq. (2).

$$RSD = \left(\frac{1}{C_{av}} * \sqrt{\frac{\sum_{i=1}^N (C_i - C_{av})^2}{N-1}} \right), T_{95} = \left(-\frac{1}{\gamma} \right) * \ln \ln \left(\frac{1-0.95}{A} \right) \quad (2)$$

Here, the mixing index profile is normalized, and the parameter estimation is performed to obtain A and γ . Following this, the T_{95} mixing time is calculated, as shown in Table 1, evaluated for different blade speeds. The low error margins validate the proposed strategy for radial compartmentalization.

Table 1: T_{95} mixing times for DEM and CM model

Blade speed	T 95 from DEM	T 95 from CM	Absolute relative error
50 RPM	14.812	14.5485	0.0178
75 RPM	12.366	12.129	0.0191
100 RPM	8.911	9.2684	0.0401

5.2. Validation of axial compartmentalization

The axial compartmentalization is validated based on residence time distribution (RTD) profiles (Gao, Yijie et al. 2012). The residence time profiles obtained from DEM correspond to the time it takes for all particles to travel from one end of the periodic section to another, calculated separately for both particle types (P1_DEM, P2_DEM), whereas the residence time for the surrogate model (RT_CM) corresponds to the time it takes for the particles to travel the length of the section based on averaged velocities extracted from the developed compartments. The obtained profiles shown in Figure 4 illustrate that the residence time of the surrogate model is fairly close to that obtained from DEM models, thus, validating the proposed axial compartmentalization strategy for surrogate models.

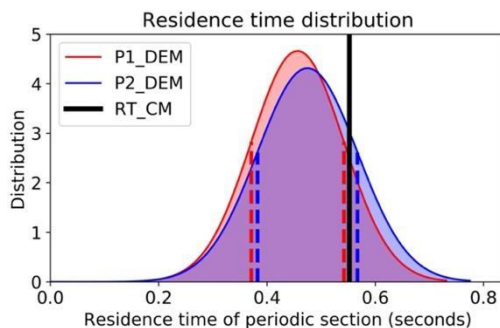


Figure 4: Residence time distribution profiles for validation of axial compartmentalization.

6. Conclusions

We propose the use of multi-zonal compartmentalization methodology for surrogate modelling of pharmaceutical continuous powder blender using periodic sections. A computationally efficient surrogate model consisting of inter-connected network of zones is developed for the DEM simulation, to quantify powder mixing within the system and validated with the original DEM using radial mixing time and axial residence time profiles. Thus, the proposed methodology is shown to capture axial and radial mixing of powder blender in a computationally efficient manner. For future work, we aim to integrate the proposed surrogate model within process flowsheets for integration of overall manufacturing lines.

7. Acknowledgements

The authors would like to acknowledge funding from U.S. Food and Drug Administration (FDA) through grant 1U01FD006487-01.

References

- Bhalode P, Ierapetritou M. A review of existing mixing indices in solid-based continuous blending operations. *Powder Technology*. Elsevier; 2020a Aug 1;373:195–209.
- Bhalode P, Ierapetritou M. Discrete element modeling for continuous powder feeding operation: Calibration and system analysis. *International Journal of Pharmaceutics*. Elsevier; 2020b May 28;585:119427.
- Cundall PA, Strack ODL. A discrete numerical model for granular assemblies. *Géotechnique*. 1980 Sep;30(3):331–6.
- Danckwerts PV. Continuous flow systems: Distribution of residence times. *Chemical Engineering Science*. Pergamon; 1953 Feb 1;2(1):1–13.
- Delafosse A, Collignon M-L, Calvo S, Delvigne F, Crine M, Thonart P, et al. CFD-based compartment model for description of mixing in bioreactors. *Chemical Engineering Science*. Elsevier; 2014;106:76–85.
- Dubey A, Sarkar A, Ierapetritou M, Wassgren CR, Muzzio FJ. Computational Approaches for Studying the Granular Dynamics of Continuous Blending Processes, 1 – DEM Based Methods. *Macromolecular Materials and Engineering*. Wiley-Blackwell; 2011 Mar 28;296(3-4):290–307.
- Gao, Yijie, Ierapetritou M, Muzzio FJ. Periodic section modeling of convective continuous powder mixing processes. *AIChE Journal*. John Wiley & Sons, Ltd; 2012 Jan 1;58(1):69–78.

- Ierapetritou M, Muzzio FJ, Reklaitis G. Perspectives on the continuous manufacturing of powder-based pharmaceutical processes. *AIChE Journal*. Wiley-Blackwell; 2016 Jun 1;62(6):1846–62.
- Jourdan N, Neveux T, Potier O, Kanniche M, Wicks J, Nopens I, et al. Compartmental Modelling in chemical engineering: A critical review. *Chemical Engineering Science*. 2019;210.
- Ketterhagen WR, Ende MTA, Hancock BC. Process modeling in the pharmaceutical industry using the discrete element method. *Journal of Pharmaceutical Sciences*. Wiley-Blackwell; 2009 Feb 1;98(2):442–70.
- Muzzio FJ, Robinson P, Wightman C, Dean Brone. Sampling practices in powder blending. *International Journal of Pharmaceutics*. 1997 Sep;155(2):153–78.
- Portillo PM, Vanarase A, Ingram A, seville J, Ierapetritou M, Muzzio FJ. Investigation of the effect of impeller rotation rate, powder flow rate, and cohesion on powder flow behavior in a continuous blender using PEPT. *Chemical Engineering Science*. Pergamon; 2010 Nov 1;65(21):5658–68.
- Tajsoleiman T, Spann R, Bach C, Gernaey KV, Huusom JK, ruhne U. A CFD based automatic method for compartment model development. *European Journal of Pharmaceutics and Biopharmaceutics*. 2019;123:236–45.
- Yu LX, Amidon G, Khan MA, Hoag SW, Polli J, Raju GK, et al. Understanding Pharmaceutical Quality by Design. *AAPS J*. Springer US; 2014 Jul 1;16(4):771–83.

ProMo variable/equation Ontology-based system modelling

Niloufar Abtahi^a and Heinz A. Preisig^b

^{a, b,*} *Norwegian University of Science and Technology, Department of Chemical Engineering, NO-7491 Trondheim, Norway*
niloufarabtahi@gmail.com

Abstract

The description of materials in the context of thermodynamics is an essential part of the chemical engineer's models. The representation of thermodynamic relations is the objective of this study. By making use of ProMo variable/equation builder, we construct all the thermodynamics quantities based on the configuration space of contact geometry. The configuration space builds on the quantities internal energy (U), entropy (S), Volume (V) and mass species (n) adding the equations of state being the derivatives yielding the temperature, the negative pressure and the chemical potential.

ProMo expert domain provides the facilities to construct variable/equation sets from the conserved basic physical quantities, namely the ones defined above. It follows the strict rule of defining new variables as a function of existing variables. This ProMo building regulation is an excellent match to the contact geometry's fundamental approach to define thermodynamics. It generates a consistent set of variables and equations and provides a formal description that builds on pure mathematical terms. This mathematical representation also facilitates and simplifies automatic multiscale simulation modelling.

The use of well-defined variables and equations for the base ontologies system domain, facilitate an integrating factor for the application domain both for chemical engineers as well as the material multidisciplinary modelling community.

Keywords: Multidisciplinary and multi-scale simulations, Ontology-based framework, ProMovariable/equation builder

1. Introduction

Nowadays, having a generic stand-alone software for modelling and simulation (digital twin), data-driven reasoning, and knowledge-guided decision making are considered a significant success which will bring a digital revolution into industrial science and technology. Such platform optimises the dependence on trial and error based experimentation cycles for implementation in products and for manufacturing chemical processes development. Besides, it will accelerate and facilitate the integration of process and material design which is leading to plant and product design. However, the automation and industrialisation of this approach have many challenges to master.

Models are usually used to mimic and map the behaviours of a system into objects. It also gives freedom to the mind, allows tampering and testing, playing with what could become

real before it has real-world consequences beyond the use of modelling and simulation time. Models are thus central to any exploratory work that has the objective to exploit the object's behaviour.

However, the modelling term is highly knowledge-intensive activities (nearly all engineering and science activities) which covers a wide range of contexts from chemical engineering and more specifically, process systems engineering to manufacturing materials engineering. Models are multi-scale, and multi-dimensional capturing a variety of physical phenomena from classical mechanics and thermodynamics to quantum mechanics. Therefore for the integration of process and material design, selecting the right physical domain, scale, dimension, and parameter, besides, to guarantee the interoperability proper linking across the multi-scale models is not an easy task. Latter raises also the question of interoperability when realising the computational code.

There also exists a gap in modelling and simulation, which demands the representation and coupling of knowledge. Recently digitalisation of the industries increases the need for multidisciplinary simulation, thereby generating a demand for an extension towards the integration of different disciplines. Ontologies provide a perfect vehicle for the representation and coupling of knowledge. So the first important step is to introduce ontologies on all levels of the broader domain of simulation software systems, and to construct a generic simulation ontology-based framework. Because the used models and relations are mathematical descriptions of reality, the theoretical concepts of such frameworks are captured in the discipline-specific ontologies. In this respect, the use of well-defined variables and equations for the base ontologies system domain, facilitate an integrating factor for the application domain both for chemical engineers as well as the material multidisciplinary modelling community [6, 3].

ProMo, which is overarching the above concepts, is the focus of this paper. It is a software suite that realises an ontology-based computational engineering environment. It has a capability to provides the modelling and simulation of large and complex, multi-scale, multidisciplinary, and dynamic processes.

In this letter, the representation of thermodynamic relations by making use of **ProMo** is studied in detail. We also will see the correlation between the **ProMo** expert domain and the **ProMo** building domain.

- Section. 2 introduces a general review on **ProMo** and its objective. Moreover, an overview of the different parts of **ProMo** is provided.
- Section. 3 provides a brief explanation of the **ProMo** base ontology editor.
- Section. 4 illustrates how we construct all the thermodynamics quantities based on the configuration space of contact geometry by making use of the **ProMo** variable/equation builder.
- Section. 5 wraps up this manuscript with some concluding remarks.

2. The ProMo process modelling suite – an overview

ProMo is constructed on a series of modules, each of which considers as separate tasks (executable program, App, or application). Its design enforces a context-free approach,

which means, there is no loop in the definition space, but knowledge is built in a strict sequence: "Things" are put strictly on top of each other requiring a tiny set of fundamental "things". Besides, **ProMo** consists of three sets of tools, **expert**, **translator**, and **user**, which stay in contact with each other due to the transformation of information via directed graph structure.

ProMo consists of three sections, namely a section for generating the ProMo ontology, a section dedicated to generating process models, and a section that aims at the simulation and its utilisation [7, 8, 9]. **ProMo** not only create simulation code, but it also provides the base models as an extension to the EMMO (European Materials Modelling Ontology) in Ontology Web Language OWL form. It, in turn, can be fed into the Semantic Network providing the information to construct the software components automatically realising inter-task interoperability.

The **expert**'s section first constructs the ProMo base ontology. Then, in the second stage, adds the variable/equation system to the ontology-defined domain spaces to capture the mathematical behaviour of the disciplines' base entities. The third module links the now-defined entity models to graphical elements, which the **translator** uses with the graph-based visual editor of ProMo's second section.

The second section aims at the **translator** assists with the assembly of a process/material model given the set of base building blocks. Any composed model can be made part of the model library thereby extending the library of basic entity behaviours with composite models, and the task-building process generating the application simulating the modelled process. The instantiation tool adds the numerical information specifying a case, thereby defining a mathematical/numerical problem.

The third is for the **user** who will finally map the model into an executable code by **ProMo**'s task factory producing the digital twin. The user can then change the simulation conditions, but not the process structure.

3. ProMo's ontology builder

It is essential to structure the ontology hierarchically so that the knowledge is inherited down into the branches. In this way, domain ontology act as a vehicle to capture the knowledge of the interested domain. This is only possible when the ontology structure has the right form in a specific domain. The right design here means domain tree structure, as discussed above. Figure. 1 shows a hierarchical domain structure generated by the ProMo ontology editor. With this design, we make sure that integration and exchanging of the information is controlled [9].

It is the domain **experts producing** who produce the entity blocks for the different disciplines. It starts with the definition of the taxonomy and the structure captured in the **ProMo** base-ontology (see fig. 1) thus defines the syntax for the **ProMo** suite.

ProMo two global types of variable classes: (i) one of the type port variables, (ii) one that derives from port variables. The set of port variables is the base set of the fundamental variables, including constants, in the specific area of physics, which allows the expert to build all of the physics equations as a function of these variables. We build our entity-behaviour ontology for physical process based on classical mechanics, thermodynamics,

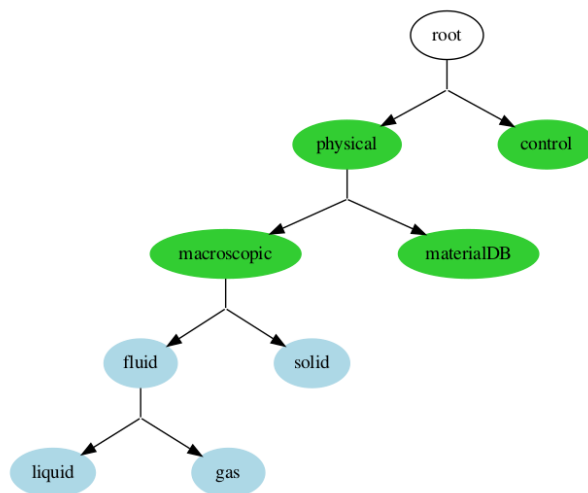


Figure 1: A graphical representation of the domain tree

particle physics, and quantum mechanics.

ProMo's ontology editor also generates all the containers for the variable/equation multi-bipartite graph and the indices. Indices play a main role in the **ProMo**. We define indices for nodes and arcs as well as for the tokens, which in the case of physics, are the conserved quantities. Further, indices capture the existing conversions and species. Product sets are used where variable-length index sets appear, like for species in nodes or arcs, etc [5].

From this perspective, by making use of ProMo variable/equation builder, we can now take a look at the construction of all the thermodynamics quantities based on the configuration space of contact geometry.

4. ProMo's equation editor

ProMo's **expert** section provides the facilities to construct variable/equation sets from the conserved basic physical quantities. The equation editor implements the strict rule of defining new variables as a function of existing variables based on a (minimal) set of defined port variables. The result is a consistent set of variables and equations that provide a formal description that builds on pure mathematical terms. The ProMo building regulations are an excellent match to the contact geometry's fundamental approach to define thermodynamics.

To construct the variables and equations, one first has to choose the domain in the ProMo domain tree. The ProMo ontology implements an inheriting policy: each variable defined on the selected level is inherited into the branch below. The process starts with the definition of the port variables. For each of them, the physical units and the dimensionality inform of index sets are defined. The expert defines a new variable as a function of existing variables. The editor disables the use of not-existing variables. Each operator is designed to compute the units and the dimensionalities following a rule defined for each

operator. This construction rules impose a lower-triagonal variable/equation structure with consistent variable dimensions and units. Some mathematical entity models require the solution of implicit equations. ProMo sets the rule that the expert has to define first an explicit expression for the new variable. Only in a second stage, a second equation can be added for this variable, which is a solution to a root computation. This approach resolves the problem of keeping consistency for both the units and the dimensionality of the variable.

4.1. Thermodynamics configuration space

There exist an analogy between classical mechanics and thermodynamics. Such that the equations of state are first-order partial differential equations, which are defining the thermodynamics mathematical models of the system and they are pretty similar to Hamilton's equations. Therefore classical mechanics and thermodynamics have both symplectic structures. With this resemblance in mind, we can construct the thermodynamics configuration space on the quantities internal energy (U), entropy (S), Volume (V) and mass species (n) adding the equations of state being the derivatives yielding the temperature, the negative pressure and the chemical potential [4, 2, 1]. By making use of the contact geometry concept, one can write the configuration space as:

$$U = (S, V, n, T, -p, \mu) \quad (1)$$

where $T = \frac{\partial U}{\partial S}$, $p = -\frac{\partial U}{\partial V}$, and $\mu = \frac{\partial U}{\partial n}$ indicate temperature, pressure, and chemical potential respectively. With this in mind, we can define the thermodynamics potentials as follows.

Enthalpy is obtained as:

$$H = U + pV \quad (2)$$

Helmholtz free energy is written as:

$$F = U - TS \quad (3)$$

Gibbs free energy is defined as:

$$G = H - TS \quad (4)$$

Grand potential is obtained as:

$$\phi = F - \mu n \quad (5)$$

Maxwell relations are given by:

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V, \quad \left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p \quad (6)$$

$$\left(\frac{\partial p}{\partial T}\right)_V = \left(\frac{\partial S}{\partial V}\right)_T, \quad \left(\frac{\partial V}{\partial T}\right)_p = -\left(\frac{\partial S}{\partial p}\right)_T \quad (7)$$

In this way, we can describe the thermodynamics behaviour of materials and processes in details by generating the variable/relation (equation) system based on the port. In this

letter, the thermodynamics port variables are (U, S, V, n) , which we declare as of class "state variables". Next we generate the equations for $(T, -p, \mu)$ as a function of the just-defined state variables and port variables and declare them belonging to the class effort variables¹. Once we defined state variables and their units and indices, we cannot change them as the whole equation system will build on them. Once we have defined the right-hand-side expression defining the declared variable, it will be compiled. The compiler computes the units and indices. If compilation fails, the expression is rejected and the user is informed about the detected problem. With this sophisticated method, we can build all thermodynamics equations based on the minimum set of port variables.

ProMo compiler can compile all variables and equations into Matlab, python, C++, as well as generating OWL file for variables and relations. Besides, it compiles a Latex file to show the documented results.

5. Concluding remarks

In this letter, we have studied how to build and modify the thermodynamics variable/equation system based on configuration space methodology. In the physics domain, the port variables are the base variables for defining the configuration spaces of the mechanical and thermodynamic systems. The former is governed by the Hamiltonians and the latter by Contact Geometry. The construct will provide a solid fundamental framework of variables and equations. The approach guarantees proper propagation of physical units and dimensionality of the mathematical object. It also defines for each variable an input/output representation, which we allow to be substituted by any alternative model, thereby enabling surrogate models.

References

- [1] A. Bravetti, 2017. Contact hamiltonian dynamics: the concept and its use. *Entropy* 19 (10), 535.
- [2] G. Chen, N.-s. E. Transport, 2005. Conversion: A parallel treatment of electrons, molecules, phonons, and photons.
- [3] P. Mason, C. R. Fisher, R. Glamm, M. V. Manuel, G. J. Schmitz, A. K. Singh, A. Strachan, 2017. Proceedings of the 4th World Congress on Integrated Computational Materials Engineering (ICME 2017). Springer.
- [4] M. A. Peterson, 1979. Analogy between thermodynamics and mechanics. *American Journal of Physics* 47 (6), 488–490.
- [5] H. A. Preisig, 2010. Constructing and maintaining proper process models. *Computers & chemical engineering* 34 (9), 1543–1555.
- [6] H. A. Preisig, 2019. The abc of modelling. In: *The ABC of Modelling*. NTNU.
- [7] H. A. Preisig, 2019. Base-ontology structure for mathematical representation of multi-disciplinary simulations. NTNU - MarketPlace.
- [8] H. A. Preisig, 2019. Ontology-adding the equations. NTNU - MarketPlace.
- [9] H. A. Preisig, A. T. Elve, S. Karolius, 2019. More power to the individual modeller using ontologies. In: *Computer Aided Chemical Engineering*. Vol. 46. Elsevier, pp. 1099–1104.

¹The term "effort variable" was coined in the context of the Bond-graph theory. Physics often uses generalized forces in a thermodynamic they are the conjugates to the potentials.

Multi-scenario Design Optimization using ADMM of a Thermal Energy Storage System

Sandeep Prakash,^a Dinesh Krishnamoorthy,^a Johannes Jäschke^{a*}

^a*Norwegian University of Science and Technology, Trondheim 7050, Norway*

johannes.jaschke@ntnu.no

Abstract

Using nonlinear models to represent multi-scenario design optimization problems can lead to very large NLPs that can become intractable to be solved centrally due to the available memory in the computing device used. In this paper, we consider a simple but general approach for partitioning the large problem into smaller NLPs by adding consensus constraints. A distributed algorithm is then developed by applying the Alternating Direction Method of Multipliers (ADMM) to solve the partition problems separately and overcome this memory limitations. The approach is demonstrated using a simple case study and compared against the solution obtained by solving the problem centrally.

Keywords: ADMM, thermal energy storage, optimal design.

1. Introduction

Thermal energy storage (TES) systems help manage the asynchronous behaviour between supply and demand of thermal energy that occur in many industrial processes. They store excess energy during off-peak periods and discharge it during peak demand, thus reducing reliance on external utilities and decrease operating costs. During the design stage, the profiles for the supply and demand for energy under which the system is expected to operate in are uncertain and needs to be accounted for. This uncertainty can

be represented by a set of discrete scenarios $S := \{1, \dots, S\}$ with cost weights ω_s to represent the likelihood of scenario s being realized. Stochastic programming approaches can be used to cast the optimal design problem as a two-stage optimization problem where

the design variables w^{des} are first-stage (here and now) decisions while the operations variables w_s^{oper} are second stage (wait and see) decisions for each scenario (Thombre et al., 2020).

The dynamics of the system in each scenario can be discretized into N equally spaced sampling intervals represented by the set $K := \{0, \dots, N-1\}$ and the optimal design problem cast as an NLP in the centralized form as ()

$$\min_{w^{des}, w^{oper}} \phi^{capex}(w^{des}) + \sum_{s \in S} \omega_s \left[\sum_{k \in K} l(x_{k,s}, u_{k,s}) \right] \quad (a)$$

$$s.t \quad g(x_{k,s}, u_{k,s}, \theta_{k,s}, w^{des}) \leq 0 \quad \forall k \in K, \forall s \in S \quad (b)$$

$$x_{k+1,s} = h(x_{k,s}, u_{k,s}, \theta_{k,s}, w^{des}) \quad \forall k \in K, \forall s \in S \quad (c)$$

$$x_{0,s} = x_0 \quad \forall s \in S \quad (d)$$

The vectors x , u and θ represent the differential states, control inputs applied to the plant during operations, and time-varying parameters respectively. The subscript $(\cdot)_{k,s}$ is used to represent the s^{th} scenario at timestep k while the initial condition for all scenarios is x_0 . We define the operations variable for scenario s built by stacking the vectors in each timestep k as, $(w_s^{oper})^T := [(x_{0,s})^T, \dots, (x_{N,s})^T, (u_{0,s})^T, \dots, (u_{N-1,s})^T]$. Functions $\phi^{capex}(\cdot)$ and $l(\cdot)$ represent the capital costs and operating costs. The function $g(\cdot)$ represents inequality constraints and function $h(\cdot)$ is used to represent the dynamics of the system. When considering many scenarios and longer time horizons, solving the optimal design problem in the central form as in (b) can become computationally intractable due to the limited memory available in the computing device used. In this paper, we explore the option of dividing this problem into smaller partition problems which can be solved separately by multiple smaller machines to overcome the memory limitations. The partition problems are then solved iteratively with a coordination step in between to reach a solution to the central problem.

2. Methodology

2.1. Reformulating the design problem as a general form consensus optimization problem

We can divide the two-stage dynamic optimization problem into P partitions denoted by the set $P := \{1, \dots, P\}$ in a very flexible manner. For example, besides considering each scenario as a partition, we could also form partitions within the prediction horizon of a particular scenario, or bundle together similar parts from multiple scenarios into a partition. Each partition p then is a separate optimization problem with its own local variables for design (w_p^{des}) and operations (w_p^{oper}) and can be solved separately. An illustration of a problem with 2 scenarios divided into 6 partitions is shown in Figure 1.

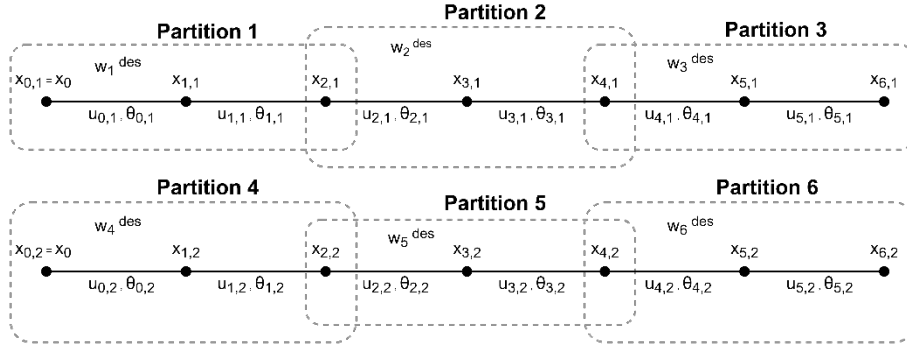


Figure 1: Illustration of dividing the optimal design problem with 2 scenarios into 6 partitions.

The solutions from the individual partitions put together is a solution to the centralized problem () when,

- All the partitions achieve consensus between their local design variables
- Adjacent partitions achieve consensus of the differential state variables shared between them (at the edge of the partitions)

We introduce a global copy of all the variables that must reach consensus into a vector ν and the consensus requirements can then be imposed as a constraint in each partition p . We can thus write the optimal design problem as a sum of smaller partition problems that are linked with consensus constraints as,

$$\min_{w_p, \nu} \sum_{p \in P} \Phi_p^{capex}(w_p^{des}) + \Phi_p^{oper}(w_p^{oper}) \quad (a)$$

$$s.t \quad g_p(w_p^{oper}, \theta_p, w_p^{des}) \leq 0 \quad \forall p \in P \quad (b)$$

$$h_p(w_p^{oper}, \theta_p, w_p^{des}) = 0 \quad \forall p \in P \quad (c)$$

$$w_p^{des} - \nu^{des} = 0 \quad \forall p \in P \quad (d)$$

$$A_p w_p^{oper} - B_p \nu^{oper} = 0 \quad \forall p \in P \quad (e)$$

$$w_p^T = \left[(w_p^{des})^T, (w_p^{oper})^T \right] \quad \forall p \in P \quad (f)$$

A_p and B_p are selection matrices used to link a subset of the local variables of partition p to the corresponding sections in the global copy ν . The objective function terms in the partition problems are chosen appropriately to add up to the original objective in ().

2.2. Applying ADMM to get a distributed algorithm

The individual partition problems in () are not trivially separable due to the presence of constraints that enforce the consensus condition. We can solve this problem in a distributed approach using ADMM as described below. The primal residual of the constraints (d) and (e) can be denoted using the vector r_p as,

$$r_p = \begin{bmatrix} w_p^{des} - v^{des} \\ A_p w_p^{oper} - B_p v^{oper} \end{bmatrix} \quad \forall p \in \mathbf{P}$$

The partial augmented Lagrangian (AL) function of () can be formed by relaxing these constraints and adding it to the objective as,

$$L_R(w_p, v, \lambda_p) = \sum_{p \in \mathbf{P}} \Phi^{capex}(w_p^{des}) + \Phi_p^{oper} + (\lambda_p)^T r_p + \frac{1}{2} (r_p)^T R (r_p) \quad ()$$

The vector $(\lambda_p)^T := \left[(\lambda_p^{des})^T, (\lambda_p^{oper})^T \right]$ is built by stacking lagrange multipliers associated with the consensus constraints and R is a symmetric positive definite matrix. The AL function is additively separable except for the quadratic penalty terms. The

ADMM algorithm involves solving the partition problems while keeping v and λ fixed and then updating them by keeping the local variables w_p fixed in an alternating fashion until convergence. The i^{th} iteration of the ADMM algorithm thus takes the form

$$(w_p)^{i+1} = \arg \min_{w_p \in W_p} L_R(w_p, v^i, \lambda_p^i) \quad \forall p \in \mathbf{P} \quad (a)$$

$$(v)^{i+1} = \arg \min_v L_R(w_p^{i+1}, v, \lambda_p^i) \quad (b)$$

$$(\lambda_p)^{i+1} = (\lambda_p)^i + R (r_p)^i \quad \forall p \in \mathbf{P} \quad (c)$$

Step (a) involves solving an optimization problem for each partition while the global variable is kept constant. This is the computationally expensive step in the approach but can be solved using separate machines and in parallel to speed up convergence. Step (b) is the minimization of the AL function while the local variables in partitions are kept constant. This step for consensus problems reduces to finding the minimum of a quadratic function and can be shown to be the averaging operator (Rodriguez et al., 2018) as,

$$v^{j+1}(j) = \frac{1}{|\mathbf{P}_j|} \sum_{p \in \mathbf{P}} w_p^{j+1}(j)$$

where $\mathbf{P}_j \subseteq \mathbf{P}$ denotes the set of partitions connected to the j^{th} element of the global variable. Step (c) is the update to the lagrange multipliers of consensus constraints and can be carried out in each partition separately. The termination criteria for ADMM

iterations are that primal residual (r^{i+1}) and the dual residuals $(s_p^{i+1} = (\lambda_p)^{i+1} - (\lambda_p)^i)$ be reasonably close to zero as explained in Boyd et al. (2010).

3. The TES Design problem

We use flowsheet in Figure 2 to represent the heating section of a district heating network, where water is used in a closed-loop to satisfy the heating requirement in an area.

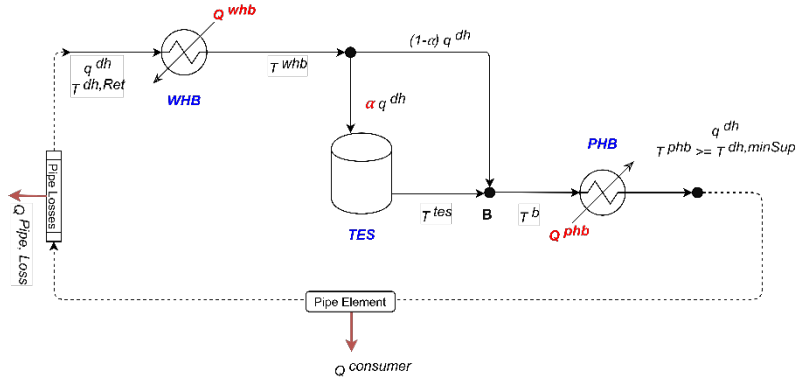


Figure 2: Simplified flowsheet of a district heating system.

The volumetric flow rate of water in the loop is q^{dh} which is determined by the number of consumers and their heating demands at any time. The temperature of water returned is $T^{dh,Ret}$ and determined by the heat losses in the system and is assumed to be correlated to ambient weather conditions. The operational objective of the heating system is to heat this water to a temperature T^{phb} that must be above a contractually specified temperature $T^{dh,minSup}$. A cheap source of heat is available from a process stream in an industrial process that needs to be cooled. It transfers heat to the district heating system using the heat exchanger WHB. Any additional heating required in the district heating side is met by using the peak heat boiler PHB. There is a temporal mismatch in the supply of heat from the process stream and the demand for heating from the consumers. To better manage this mismatch and decrease the reliance on external utilities, a thermal energy storage system in the form of a simple buffer tank is being considered. This simple TES system can charge/ discharge by raising/ lowering the temperature of the tank by manipulating the flow split α .

Mass and energy balance equations can be written out to model the dynamics of the system, details of which and the parameters used can be found in Prakash (2020). The differential states are $x^T = [T^{whb}, T^{tes}, T^{phb}]$ and the control inputs $u^T = [\alpha, Q^{phb}, Q^{dc}]$ and the uncertain parameters at each time step k are $\theta_k^T = [q_k^{dh}, T_k^{dh,Ret}, Q^{whb}]$. The design

problem is then to find the optimal volume of the TES tank (V^{tes}) that must be installed, given the uncertain profiles of future supply and demand of thermal energy.

4. Results and Discussions

We present a simulation study to demonstrate the distributed approach and compare it against solving the design problem as a single NLP centrally. Two equally likely scenarios are used to represent future operations (with $N = 60$). The profiles for Q^{whb} has a step change at timesteps $k = 20$ and $k = 40$, while all other parameters are held constant.

Quadratic functions are used to represent capital cost $\phi^{capex} = 0.001(V^{tes})^2$ and the operating cost $l = (Q^{phb})^2$. The design variables and operations variables were all scaled to be within zero and one while formulating the optimization problem. Scaling facilitated selecting the AL penalty parameters in R to tune the convergence in our distributed approach more easily. The penalty parameter influences the convergence speed of ADMM when applied to convex problems. In the case of nonconvex problems, poorly chosen values of this penalty parameter can make convergence harder or can even prevent convergence altogether. We use a diagonal matrix for R (with values 0.001 and 0.1 corresponding to the design and operation variables respectively) to roughly balance the magnitudes of the original objective and the added penalization term at the initial guess.

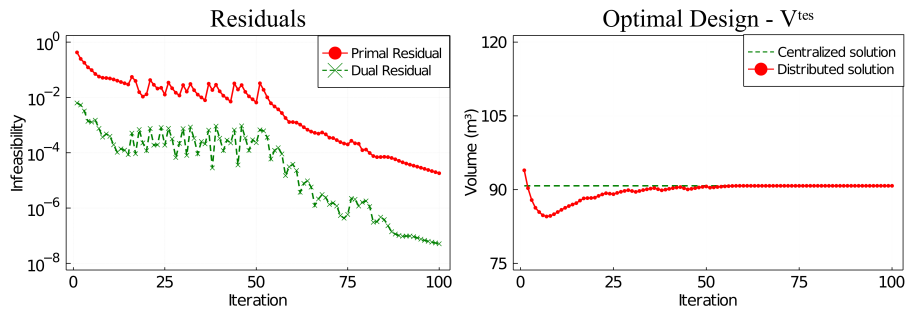


Figure 3: Results from solving the design problem centrally vs distributed approach.

In Figure 3, we can see that the primal and dual residuals become sufficiently small in the ADMM algorithm, indicating the convergence in our distributed approach. The optimal design variables can be seen to converge to the same solution as found by solving the problem centrally. An important point to note is that this behavior is not guaranteed by ADMM in the case of nonconvex problems. When applied to nonconvex problems, ADMM need not converge and even when it does converge, it need not converge to an optimal point and must be hence considered just as another local optimization method (Boyd et al., 2010). Although the convergence guarantees for ADMM in the case of complex nonconvex NLPs are poorly understood, it has been shown to perform satisfactorily in practice (Rodriguez et al., 2018). Three snapshots of the optimal differential state trajectory T^{tes} in the partitions of scenario 1 are shown in Figure 4.

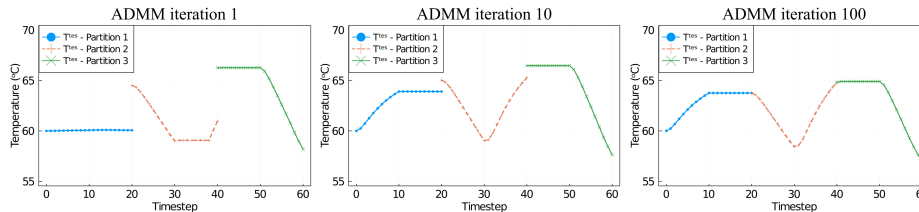


Figure 4: Snapshots at ADMM iterations ($i = 1, 10, 100$) in the partition problems.

An interesting observation in our approach is that all partitions apart from the leftmost partition has the initial condition of the dynamic optimization problem as a variable. Thus in Figure 4, we can see that during the initial iterations (left subplot), the optimal solution is to initialize the TES tank at a high temperature. The penalization terms added are then

updated to close the gap between the partitions and achieve consensus (middle and right subplots). In this aspect our approach shares similarities to the multiple shooting (MS) approach in dynamic optimization. The key difference is that MS is solved centrally and the state continuity (consensus constraint) is enforced explicitly as an equality constraint by the solver. We solve it in a distributed way where the consensus constraint is relaxed by forming the AL. The ADMM iterations are then able to enforce this constraint implicitly by minimizing the AL.

5. Conclusion

In this paper, we presented the optimal design of a simple TES system under uncertainty as a two-stage nonlinear dynamic optimization problem. Due to limitations in memory of solving the problem centrally in a single machine, an approach for forming smaller partition problems in a general fashion was shown. The ADMM algorithm was applied to coordinate between the subproblems which could be solved separately and in parallel. A simple simulation exercise was used to demonstrate the approach.

References

- S. Boyd, N. Parikh, E. Chu, B. Paeto, J. Eckstein, 2010, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends in Machine Learning*, 3 (1), 1-122.
- S. Prakash, 2020, Distributed optimization using ADMM for Optimal Design of Thermal Energy Storage systems, Master's Thesis, Norwegian University of Science and Technology, Norway.
- J. S. Rodriguez, B. Nicholson, C. Laird, V. M. Zavala, 2018, Benchmarking ADMM in nonconvex NLPs, *Computers and Chemical Engineering*, 119, 315-325.
- M. Thombre, S. Prakash, B. R. Knudsen, J. Jäschke, 2020, Optimizing the Capacity of Thermal Energy Storage in Industrial Clusters, 30th European Symposium on Computer Aided Process Engineering. *Computer Aided Chemical Engineering*, 48, 1459–1464.

Performance comparison of different optimization methods for the wastewater treatment plant model based on the Activated Sludge Model No. 3 and design of the plant control

Melinda Simon-Várhelyi,^a Vasile Mircea Cristea,^{a,*} Alexandra Veronica Luca,^a Marius Adrian Brehar^a

^a*Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany János Street, Cluj-Napoca 400028, Romania*

mircea.cristea@ubbcluj.ro

Abstract

Calibration of the wastewater treatment plant (WWTP) models is a challenging task, because reliable collection of measured data and good process knowledge are required in each particular case. This research presents the case study of a Romanian municipal WWTP with the Anaerobic-Anoxic-Oxic (A²O) configuration. Activated Sludge Model No. 3 (ASM3) is the core of the developed WWTP model. The proposed WWTP model calibration methodology assumes the selection and computation of a set of influent variables, associated to a group of process and settler parameters, such as they fit the model predictions to the plant measurements. The objective function used for calibration consists in the absolute difference between the measured and model predicted effluent data and its constrained minimization finds the decision variables as calibration targets. Three optimization approaches were selected and investigated: genetic, genetic-hybrid and multiobjective Pareto algorithms. The calibration performance of the different optimization methods was evaluated and compared. The genetic-hybrid optimization method showed the best performance as the model predicted data approximates well the effluent measured data both in steady and dynamic state. The properly calibrated WWTP model was complemented with nitrification and denitrification control loops, aimed to improve the municipal WWTP operation by enhancing the effluent quality and reducing the aeration and pumping energy.

Keywords: wastewater treatment, ASM3 modelling, calibration, optimization algorithms, control.

1. Introduction

Activated sludge technology is one of the most reliable and efficient purification methods used at the municipal wastewater treatment plants. This technology requires the presence of heterotrophic and autotrophic microorganisms for achieving the removal of carbon, nitrogen and phosphorus pollutants emerged from the household and industrial activities. The demands of the continuously stricter effluent regulation limits

and the need of reducing the operation costs ask for the development of efficient operating solutions. They are effectively investigated by complex modelling tools and new solutions are implemented by efficient control systems (Hauduc et al., 2013). Mathematical models support the understanding of the complex processes occurring at the wastewater treatment plants, the operators training and the elaboration of the control strategies (Ostace et al., 2011). Activated Sludge Models (ASMs) describe the bio-chemical process, while the Benchmark Simulation Models (BSMs) complement the physical separation processes taking place at the municipal WWTPs (Rieger et al., 2013). Activated Sludge Model No. 3 (ASM3) describes the organic carbon removal, the nitrification and denitrification processes, rectifying some weak points of the Activated Sludge Model No. 1 (ASM1). According to the ASM3, the growth of heterotrophic biomass is achieved in two steps: in the first step biodegradable substrate is stored as internal storage products and in the second step these internal cellular storage products are consumed for the growth of heterotrophic microorganisms (Gujer et al. 1999). The WWTP models need to be calibrated, because the influent wastewater concentration and flow rate profiles, associated to the geometrical characteristics of the purification units, vary from one WWTP to another (Mannina et al., 2011). Various nitrification and denitrification control configurations are investigated and applied at the municipal WWTPs (Várhelyi et al., 2019). Two main control loops are considered to control the dissolved oxygen (DO) concentration in the biodegradation tanks, by manipulating the air flow rate, and to control the nitrates-nitrites concentration in the anoxic basin, by acting on the flow rate of nitrate recirculation from the aeration tanks (Gernaey et al., 2014). This paper presents the calibration of the Anaerobic-Anoxic-Oxic (A²O) ASM3-based WWTP model by comparing the performance of three optimization methods: genetic, genetic-hybrid and Pareto multiobjective algorithms. They are followed by the implementation of the nitrification and denitrification the control system.

2. Municipal WWTP model

The developed municipal WWTP model comprises a nonreactive primary settler relying on Otterpohl primary clarifier model equations, five bioreactors (anaerobic, anoxic and three aerobic units) incorporating the biological equations of ASM3 and a nonreactive secondary settler described by the double-exponential velocity function settling model of Takács. The A²O layout of the municipal WWTP, with its anaerobic and anoxic reactors, is considered with the particular position of nitrate recirculation flow rate for transporting nitrates and nitrites from the aerobic zone into the anoxic bioreactor. The WWTP model was built in Matlab software and the Simulink graphical extension. For reducing the simulation time, the process equations were written in C programming language and were incorporated as compiled C files in Simulink S-function blocks.

3. Methods

3.1. Proposed calibration approach

The proposed calibration approach consists in five main steps: i) defining the calibration goals, targeted model variables and parameters, ii) collecting and analysing the construction and process measured data, iii) developing and software implementing the ASM3-based WWTP model, iv) calibrating the model in steady state, using different optimization methods, v) evaluating the calibration performance in steady and dynamic

state. The main calibration goal was to fit the model predicted effluent data to the effluent data measured at the municipal WWTP. The main pollutant components concentration in the influent and effluent streams, associated to the influent stream, recirculation and air flow rates, were measured with online sensors and were reconciled with laboratory measurements. The measured data was collected during May 2016 and the data set of the first 22 days of the month was considered for the WWTP model calibration (Simon-Várhelyi et al., 2020). In the calibration step the optimization problem was formulated and 10 decision variables containing four influent variables concentration (denoted as x_1 to x_4), three process parameters (x_5 to x_7) and three settler parameters (x_8 to x_{10}) were chosen in order to be calibrated. They were: readily biodegradable organic substrates (S_S , x_1), inert particulate organic material (X_I , x_2), heterotrophic biomass (X_H , x_3), nitrifying biomass (X_A , x_4), storage rate constant (k_{STO} , x_5), dissolved oxygen saturation constant (K_{O_2} , x_6), autotrophic maximum growth of the nitrifying organisms (μ_A , x_7), hindered zone settling parameter (r_h , x_8), flocculent zone settling parameter (r_p , x_9) and non-settleable fraction (f_{ns} , x_{10}). As objective function ($objfunc_{total}$) it was considered the weighted sum of the absolute differences between the effluent measured data and effluent model predicted data for: soluble part of the chemical oxygen demand (COD_{sol}), nitrates and nitrites (S_{NOX}), free and saline ammonia (S_{NH_4}) and total suspended solids (X_{SS}) concentrations. The lower (LB) and upper bounds (UB) of the decision variables were defined based on literature-scrutinized values. Equations (1-8) describe the proposed optimization problem, as they were formulated for the WWTP model calibration.

$$\min_X [objfunc_{total}(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})] \quad (1)$$

$$X = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}] \quad (2)$$

$$objfunc_{total} = objfunc_{COD_{sol}} + objfunc_{NOX} + 10 \cdot objfunc_{NH_4} + objfunc_{SS} \quad (3)$$

$$objfunc_{COD_{sol}} = abs(COD_{sol,eff,measured} - COD_{sol,eff,simulated}) \quad (4)$$

$$objfunc_{NOX} = abs(NOX_{eff,measured} - NOX_{eff,simulated}) \quad (5)$$

$$objfunc_{NH_4} = abs(NH_4_{eff,measured} - NH_4_{eff,simulated}) \quad (6)$$

$$objfunc_{SS} = abs(SS_{eff,measured} - SS_{eff,simulated}) \quad (7)$$

$$LB \leq X \leq UB \quad (8)$$

The calibration was performed using three different optimization algorithms. The genetic algorithm optimization approach was the first tested method. It was based on the *ga* Matlab function. The second considered method was the genetic-hybrid algorithm, as it combines two optimization methods, i.e. the genetic algorithm and the classical algorithm. The *hybrid* Matlab function was used for its implementation. The third investigated optimization method was the multiobjective optimization, which finds the Pareto solutions using the genetic algorithm. The *gamultiobj* function of Matlab was implemented. For each of the three optimization methods, the calibration was performed in steady state using the same objective function and constraints. The best calibration results were selected, implemented in the model and further tested by performing dynamic simulations. The WWTP dynamic model behaviour was compared with process measured data, based on a 10 minutes sampling time evaluation approach.

3.2. Proposed control approach

The calibrated ASM3-based WWTP model was complemented with two feedback control loops having Proportional-Integral controllers, as presented in Figure 1. The first one is intended for the nitrification (aeration) control, while the second one is implemented for the denitrification control. This control design also achieves the removal of carbon and phosphorous pollutants. Control of the aeration is based on the DO concentration control in the last aerated bioreactor. The manipulated variable is the air flow rate introduced in the aerobic bioreactors. The second control loop aims the control of nitrates and nitrites concentration (S_{NOx}) in the anoxic bioreactor by manipulating the nitrate recirculation flow rate.

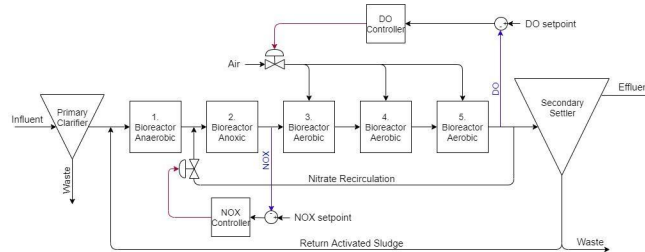


Figure 1. The proposed and investigated control loops.

Different setpoint values for the DO concentration in the last aerobic bioreactor and for the nitrates and nitrites concentration in the anoxic bioreactor were proposed and tested. The aeration energy (AE), pumping energy (PE) and effluent quality (EQ) were calculated based on the Eqs. (9-11) and compared to the actual performance of the municipal WWTP. The investigated control cases are presented in Table 1.

$$AE = C_{AE} \cdot \frac{SO_{sat}}{T \cdot 1.8 \cdot 1000} \int_0^T \sum_{\text{aerated bioreactor}} [V^{\text{bioreactor}} \cdot K_L a_i(t)] dt \quad (9)$$

$$PE = C_{PE} \cdot \frac{1}{T} \int_0^T [0.004 \cdot Q_{NR}(t) + 0.008 \cdot Q_{RAS}(t) + 0.05 \cdot Q_{waste}(t)] dt \quad (10)$$

$$EQ = \frac{1}{T \cdot 1000} \int_0^T [PU_{TSS}(t) + PU_{COD}(t) + PU_{BOD}(t) + PU_{TKN}(t) + PU_{No}(t)] Q_{effluent}(t) dt \quad (11)$$

Table 1. Investigated cases of the control loops with different setpoint values.

Case	Description
Case 0	WWTP actual operation (predominant manual control)
Case 1	DO setpoint value: 3 g O ₂ /m ³ , S _{NOx} setpoint value: 0.1 g N/m ³
Case 2	DO setpoint value: 3 g O ₂ /m ³ , S _{NOx} setpoint value: 0.05 g N/m ³
Case 3	DO setpoint value: 3 g O ₂ /m ³ , S _{NOx} setpoint value: 0.01 g N/m ³
Case 4	DO setpoint value: 2 g O ₂ /m ³ , S _{NOx} setpoint value: 0.1 g N/m ³
Case 5	DO setpoint value: 2 g O ₂ /m ³ , S _{NOx} setpoint value: 0.05 g N/m ³
Case 6	DO setpoint value: 2 g O ₂ /m ³ , S _{NOx} setpoint value: 0.01 g N/m ³

4. Results and discussion

4.1. Steady state calibration results

The steady state calibration results are presented in Tables 2-4, for all three considered optimization algorithms. Table 2 presents the calibrated decision variables, Table 3 comparatively shows the measured effluent data associated to the calibrated model predictions and Table 4 reveals the values of the optimization index. The calibrated decision variables show a good agreement with the values reported in the literature. The effluent simulated data approximates well the effluent measured data in steady state. The best results were achieved by the hybrid optimization method, as the overall objective function demonstrated the minimum value of 0.025. The calibrated decision variables obtained using hybrid optimization were used in the subsequent dynamic simulations.

Table 2. Calibrated values of the influent variables, process and settler parameters obtained with the different optimization methods.

Decision variable	Notation	Genetic algorithm	Hybrid method	Multiobjective method	Unit
Readily biodeg. organic substrates	S_s	49.02	15.37	74.2	g COD/m ³
Inert particulate organic material	X_i	36.21	62.86	32.81	g COD/m ³
Heterotrophic organisms	X_H	3.61	0.463	2.54	g COD/m ³
Nitrifying organisms	X_A	3.42	2.84	8.97	g COD/m ³
Storage rate constant	k_{STO}	2.24	3.35	0.91	g COD _{XS} / (g COD _{XH} day)
Saturation constant for DO	K_{O_2}	0.46	0.492	1.09	g O ₂ /m ³
Autotrophic maximum growth of the nitrifying organisms	μ_A	8.93	4.74	6.69	1/day
Hindered zone settling parameter	r_h	0.000799	0.00113	0.000687	m ³ /g SS
Flocculant zone settling parameter	r_p	0.00794	0.0125	0.0114	m ³ /g SS
Non-settleable fraction	f_{ns}	0.00449	0.006	0.00329	-

Table 3. Comparison between the average measured effluent data and the average simulated effluent data obtained by the calibration with different optimization methods.

Effluent variable	Notation	Measured data	Genetic algorithm	Hybrid method	Multiobjective method	Unit
Soluble chem. oxig. demand	COD_{sol}	4.84	4.84	4.84	4.84	g COD/m ³
Nitrates and nitrites	S_{NOX}	3.76	3.554	3.782	3.686	g N/m ³
Free and saline ammonia	S_{NH4}	0.17	0.1699	0.1698	0.0151	g N/m ³
Suspended solids	X_{SS}	12.00	12.00	12.00	11.83	g SS/m ³

Table 4. Comparison between the overall objective function values of the calibration results obtained by the different optimization methods.

	Genetic algorithm	Hybrid method	Multiobjective method
Overall objective function	0.211	0.025	1.797

The calibration based on the genetic algorithm optimization method also shows very good calibration performance, as only the reduced difference of 5 % for the effluent nitrates and nitrites concentration was observed. Results of the multiobjective optimization method show dissimilarities at the free and saline ammonia effluent concentration, but the simulated effluent soluble COD, nitrates and nitrites, suspended solids concentrations also fit to the measured effluent data.

4.2. Calibrated model results for the dynamic state

Dynamic simulations were performed using the fitted values obtained by the hybrid optimization method. Figures 2 and 3 show the comparison between the measured

effluent WWTP data and model predicted effluent data, for the concentration of the soluble COD and nitrates and nitrites (S_{NOX}) main variables.

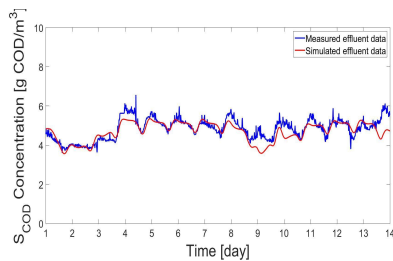


Figure 2. Measured vs model predicted effluent data for soluble COD.

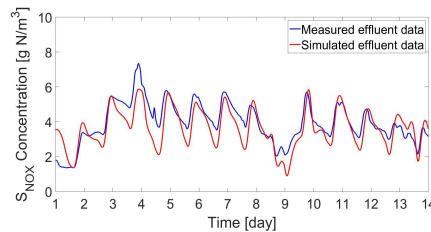


Figure 3. Measured vs model predicted effluent data for nitrates and nitrites concentrations.

Predictions of the calibrated model for the WWTP effluent approximates well the measured effluent data, both in steady and dynamic state.

4.3. Results of the WWTP control

Table 5 shows the values of the performance indices obtained by dynamic simulations using the calibrated model for the proposed control cases.

Table 5. Performance indices of the calibrated WWTP model with the control strategies.

Case	Aeration energy [kWh/day]	Pumping energy [kWh/day]	Effluent Quality [kg. P.U./day]
Case 0	13,720.03	8,793.29	20,440.06
Case 1	11,904.58	10,248.65	18,474.86
Case 2	11,680.42	7,942.05	18,768.56
Case 3	11,441.55	5,799.24	19,068.96
Case 4	10,357.93	14,138.25	16,694.57
Case 5	10,165.38	9,984.21	16,900.63
Case 6	9,940.05	6,260.02	17,138.97

As revealed by Table 5, the implementation of the proposed control improves the actual municipal WWTP operation when aeration energy and effluent quality are considered. However, depending on the case, pumping energy may be affected. The overall best performance among the investigated control cases was obtained for Case 6, i.e. by setting the DO setpoint value to 2 g O₂/m³ and the S_{NOX} setpoint value to 0.01 g N/m³. For this favourable control case, the aeration energy was reduced by 27.5 %, the pumping energy was diminished by 28.8 % and the effluent quality was improved by 16.1 %, when comparison was made to the actual WWTP operation.

5. Conclusions

The calibration of the municipal WWTP model based on the Activated Sludge Model No. 3 was accomplished using different optimization algorithms and their performance was comparatively presented. The hybrid optimization method demonstrated the best calibration performance. The calibrated model was used to implement control loops with different setpoint values, aimed to WWTP performance improvement. The most favourable case was presented and proposed for WWTP implementation as it showed energy reduction of about 28 % and effluent quality improvement of about 16 %.

References

K.V. Germaey, J.B. Copp, U. Jeppsson, I. Nopens, M.-N. Pons, P. A Vanrolleghem, 2014, Benchmarking of control strategies for wastewater treatment plants, IWA Publishing, London.

- W. Gujer, M. Henze, T. Mino, M. van Loosdrecht, 1999, Activated sludge model no. 3, *Water Science & Technology*, 39, 183-193.
- H. Hauduc, L. Rieger, A. Oehmen, M. C. M. van Loosdrecht, Y. Comeau, A. Héduit, P. A. Vanrolleghem, S. Gillot, 2013, Critical review of activated sludge modeling: state of process knowledge modeling concepts and limitations, *Biotechnology and Bioengineering*, 110, 24-46.
- G. Mannina, A. Cosenza, P. A. Vanrolleghem, G. Viviani, 2011, A practical protocol for calibration of nutrient removal wastewater treatment models, *Journal of Hydroinformatics*, 13 (4), 575-595.
- G. S. Ostace, V. M. Cristea, P. S. Agachi, 2011, Extension of Activated Sludge Model No. 1 with Two-step Nitrification and Denitrification Processes for Operation Improvement, *Environmental Engineering and Management Journal*, 10, 1529-1544.
- L. Rieger, S. Gillot, G. Langergraber, T. Ohtsuki, A. Shaw, I. Takács, S. Winkler, 2013, *Guidelines for Using Activated Sludge Models*, IWA Publishing, London.
- M. Simon-Várhelyi, V. M. Cristea, M. A. Brehar, 2020, Efficient calibration methodology of the wastewater treatment plant model based on ASM3 and application to municipal wastewater, *Desalination and Water Treatment*, 189, 108-118.
- M. Várhelyi, V. M. Cristea, M. Brehar, 2019, Improving Waste Water Treatment Plant Operation by Ammonia Based Aeration and Return Activated Sludge Control, 29th European Symposium on Computer Aided Process Engineering – ESCAPE29, 1165-1170.

Comparison of Group-Contribution and Machine Learning-based Property Prediction Models with Uncertainty Quantification

Adem R. N. Aouichaoui^a, Resul Al^b, Jens Abildskov^a, Gürkan Sin^{a,*}

^a*Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Kgs. Lyngby, DK-2800, Denmark*

^b*Novo Nordisk A/S, Bagsvaerd, DK-2880, Denmark*

gsi@kt.dtu.dk

Abstract

This study demonstrates the development of three modeling approaches for predicting thermophysical property with the ability to quantify the uncertainty in the prediction. The modeling approaches consist of a classical non-linear group-contribution (GC) model (GCM), Gaussian-Process regression (GPR), and a deep neural network (DNN) all applied to the first-order groups defined by Marrero and Gani as the molecular descriptor. The uncertainty was quantified using different methods: linear error propagation using the parameter covariance matrix for the GCM, the inherent uncertainty quantification of GPR models, and using a probabilistic layer able to learn the distribution of model outputs in DNN. The models have been applied to the lower flammability limit (LFL) at 298K. The model performance was evaluated using 5 folds cross-validation to ensure the models were exposed to all data and to detect potential overfitting,—a procedure frequently used within machine learning. The models obtained produce a good fit to the experimental data when applied to all available data with a coefficient of determination (R²) above 0.9 for all models, a maximum mean absolute error of 0.39 [%-vol], and a maximum mean squared error of 0.51.

Keywords: QSPR, Deep-Learning, Property Prediction, Uncertainty Analysis

1. Introduction

Thermophysical properties such as critical constants, enthalpic, and flammability properties play a vital role in many chemical engineering applications, e.g., phase equilibria calculations, process simulation, and risk assessment (Frutiger et al., 2016a). Predicting these properties in a fast yet computationally efficient way is therefore of paramount importance for many applications, especially in the absence or scarcity of experimental data and since it is impractical to perform measurements whenever the need arises. Traditionally, quantitative structure-property relationships (QSPRs) type models are employed to predict such properties from structural information of the chemical compounds (Austin et al., 2016). Recently, efforts have also been invested in applying emerging machine learning techniques as the new predictive tool in QSPRs (Zhang et al., 2018). Among them, deep neural networks (DNN) have drawn the utmost attention due to their ability to extract hidden features in large datasets and approximate the behavior of any function. This allows such models to compare, if not outperform, traditional QSPR

models. Some of the recent applications are the prediction of the perceived odor of molecules (Zhang et al., 2018) and critical properties of halogenated olefins (Mondejar et al., 2017). Few have however addressed the uncertainty of the predictions made by the models. This could potentially hinder the broader use of these models, especially in applications involving engineering decision-making. In this work, we develop and compare property prediction models with uncertainty quantification based on three modeling approaches: a GC-based non-linear regression (GCM), a machine learning-based Gaussian process regression (GPR), and a deep learning-based neural network (DNN) applied to the lower flammability limit property (LFL) of organic compounds.

2. Methodology

In developing QSPR models, two key modeling decisions must be addressed: how to describe the structural information of the chemical and how to relate the structure to the property of interest. The selection of these is discussed in the following.

2.1. Molecular Descriptor

The structural information of the molecule is transformed into a machine-readable molecular descriptor to be used as input to a mathematical model. One example of such expert-crafted descriptors is the segmentation in GC, where the molecule is described through the occurrence of specific predetermined atomic arrangements. The groups are defined to reflect a series of feature subsets of the molecule that are frequently occurring in the chemical design space (Marrero and Gani, 2001). Hukkerikar et al. (2012) developed a widely used GCM, where the molecule is described through three levels of increasing complexity to capture larger atomic arrangements. In this study, only the first-order groups defined by Marrero and Gani (2001) were used as input to all models. The exclusion of higher-order groups is to reduce the model complexity and because they do not always result in a reduction in the prediction error (Frutiger et al., 2016b).

2.2. Property data

The Lower Flammability Limit (LFL) was chosen to benchmark the models. The LFL is used in quantitative risk assessment and is defined as the lowest concentration of a chemical for which a flammable mixture with air is formed. The LFL is measured at 298K and 1 atm and is expressed in [%-vol]. The LFL was obtained from the AIChE DIPPR 801® database considering only experimental and accepted data (Rowley et al., 2019). A

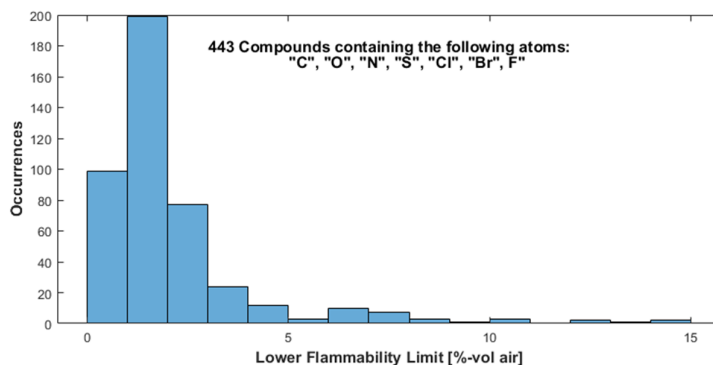


Figure 1: Histogram of the LFL

2.3. Non-linear regression

A GCM for predicting the LFL of organic molecules has previously been developed in (Frutiger et al., 2016a) and can be seen in Eq. (1).

$$f(LFL) = \log \left(\frac{LFL}{LFL_{const}} \right) = T\theta \quad (1)$$

Where T is the group occurrence matrix and θ are the group contributions. The contributions and the model constant are determined by minimizing the sum of squared error (SSE) between the prediction and the experimental value. The model uncertainty is determined by linear error propagation using the parameter covariance matrix as described in (Frutiger et al., 2016b).

2.4. Machine-learning model: Gaussian Process Regression (GPR)

Gaussian process regression (GPR) is a probabilistic machine learning (ML) algorithm that, unlike many other ML models, allows prediction of the underlying uncertainties in its predictions. It does so by using kernels to explain a given model response as a realization of a random function of the following shape:

$$y^{GPR} = \beta f(x) + \sigma^2 z(x, w) \quad (2)$$

where the first term models the mean, σ^2 the variance and $Z(x)$ is a zero-mean, unit-variance stochastic Gaussian process. β denotes regressed coefficients and $f(x)$ a set of basis functions. The stochastic part $Z(w, x)$ introduces latent variables w to correlate observations x and new points x' in the input space using kernel functions, for which a variety of alternatives exist in both literature and available software packages. The most commonly used kernel types include squared exponential kernel, exponential kernel, “matern 3/2”, “matern 5/2” and rational quadratic kernel. In this work, we use MATLAB’s GPR fitter (`fitrgp`), which additionally provides model prediction error uncertainty and confidence intervals.

2.5. Deep-learning model

Deep neural networks (DNN) is a class of machine learning algorithms similar to the artificial neural network and aims to mimic the information processing of the brain. DNNs have more than one hidden layer (l) situated between the input and output layers (Goodfellow et al., 2016). Each layer contains a given number of units (neurons) that apply a certain functional transformation to the input. These types of models can approximate the behavior of any function (universal approximation theorem). The output (y) of a unit (i) in layer (l) is related to the output (x) of the earlier layer (k) with J outputs through a set of weights ($w_{i,k}$), a bias (b) and a non-linear activation function f.

$$y_i^l = f \left(\sum_{k=1}^J (w_{i,k} x_k + b_i) \right) \quad (3)$$

To fairly compare the various models in this work, only feed-forward layers were used. Despite the considerable interest, DNNs have gained in regression applications, few have dealt with the uncertainty in the prediction. This could be due to the complexity required to perform such an analysis. In this work, we use a probabilistic machine learning technique by introducing a probabilistic layer after the dense layers that can learn the distribution over the weights in the network. This layer learns the probability distribution of the outputs (y) related to the inputs (x) through weights (w). This makes it possible to model the loss function as the negative log-likelihood. The DNN was implemented using

the TensorFlow framework along with the TensorFlow distributions to model the probabilistic layer (Dillon et al., 2017). Grid-search was used to optimize the DNN hyperparameter systematically by varying the number of neurons, and dropout rate between [8,16,32,64], and [0.2-0.6] respectively. This resulted in a DNN comprised of: three dense layers (64) separated by three dropout layers (0.5) with a learning rate of $1e^{-4}$ and using the “Adam” optimizer and the “ReLU” activation function. The architecture was determined using an 80:10:10 training, validation, and testing split.

2.6. Training, validation, and model uncertainty

To validate the model performance and its ability to generalize to unseen data, 5-fold cross-validation was performed. The mean-squared-error (MSE), mean absolute error (MAE), and coefficient of determination (R^2) are reported as an average with a standard deviation across all folds. The uncertainty quantification of GCM and GPR is done using all data for the fitting following the approach in (Frutiger et al., 2016b; Hukkerikar et al., 2012), while for the DNN the average across folds was used.

3. Results

The model predictions and the 95% confidence intervals (CI) have been determined for the best performing model for each approach and are visualized in Figure 1. The average performance metrics across all folds and using all data can be seen in Table 1.

Table 1: Average statistical performance across all folds using all data

Model	MSE			MAE			R^2		
	train	test	total	train	test	total	train	test	total
GCM	0.33	1.60	0.51	0.33	0.70	0.39	0.95	0.74	0.90
GPR	0.00	1.23	0.25	0.01	0.57	0.12	0.99	0.73	0.94
DNN	0.10	1.10	0.26	0.12	0.59	0.20	0.98	0.78	0.94

All obtained models conform well to the experimental data when considering all available data for the evaluation (test and training data). Depending on the metric chosen, the models show different ability to generalize to unseen data. The DNN shows the highest R^2 and the highest MSE for the test data, however, it shows a higher MAE compared to GPR. And while the GCM shows a better R^2 score for the test data than GPR, it produces the highest MAE and MSE among all models. To produce a common ground for comparison and to evaluate the models' ability to characterize the chemical design space, the metrics have also been evaluated using all available data. Based on this, the GPR had the highest R^2 , alongside the DNN and the lowest MAE, and MSE. Although none of the models presented match the performance of the model produced in (Frutiger et al., 2016a) ($R^2=0.99$), it is important to note that the previous work used all Marrero Gani group orders and performed outlier treatment. Besides, all data were used for the training of the model resulting in the high R^2 value (Frutiger et al., 2016a). In this study we are applying an established cross-validation methodology for testing models and therefore assess the ability of the models to *predict* the property of interest, this is quite different from earlier approaches where all the data is used for GCM building. Using the k-fold cross validation also ensures that the model has been exposed to all available groups at least once. This is an effort to overcome the fact that not all groups appear with the same frequency. The alternative would be to select a training set that has all available groups at least once. However, this approach might produce a model that is biased towards one class of

compounds. In general, it seems the machine-learning and deep-learning approaches generalize better to unseen data.

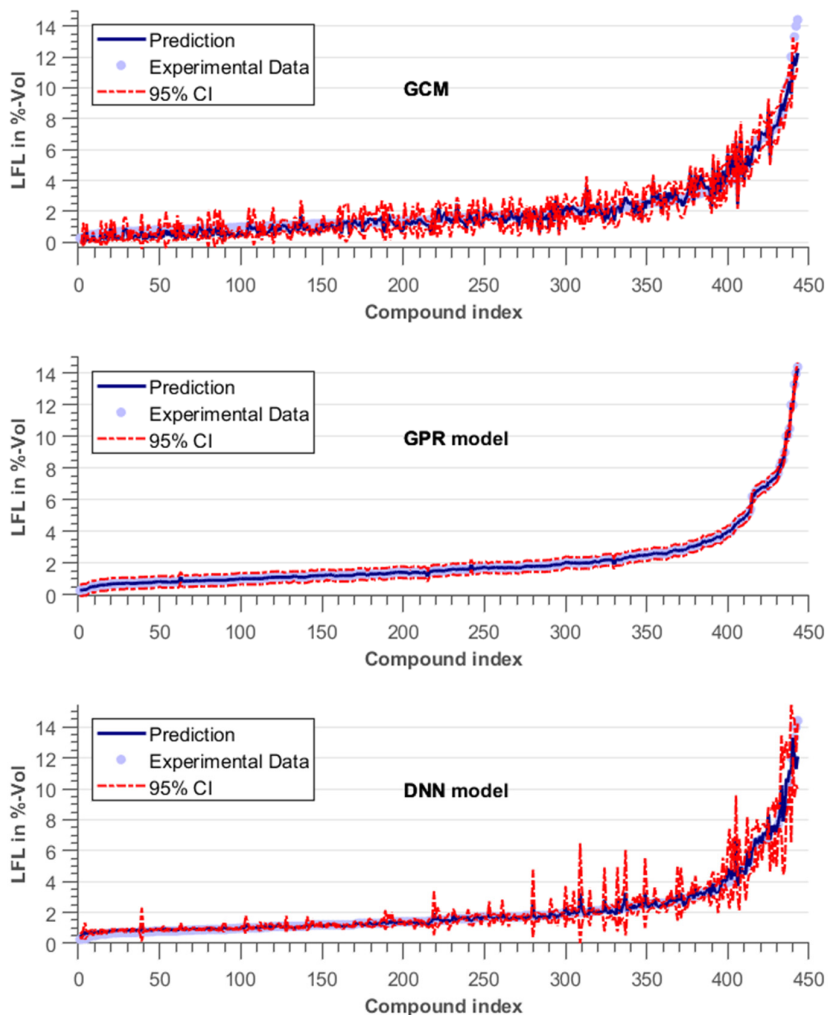


Figure 2: LFL model predictions and the 95% lower and upper CI interval

While the GCM underpredicts the LFL values in the lower end (lower than 0.5%), the DNN model is overpredicting these values. The GPR model does not show any of these tendencies. The uncertainty bounds resulting from the GPR model are very smooth in comparison to the ones produced by the GCM and the DNN model. The uncertainty bounds produced by the covariance matrix for the GCM provides more noisy confidence bounds compared to the two other models. The DNN model seems to struggle for predicting the uncertainty bounds for the data points at the higher end of the LFL spectrum (above 4%) and the confidence bounds are larger for these values and noisier. This could be due to the amount of data in these regions being much lower than for LFL values below 4%. The skewness of the data at the larger values of LFL could be the reason some of

the models struggle in predicting values in this region. From practical applications, the GCM approach with an asymptotic approximation of prediction uncertainty as well as the GPR approach seems in agreement with each other and could be used in process safety analysis, while further refinement of the DNN approach is needed. Moreover, assessing further the quality of the uncertainty predictions from different approaches would require the availability of a comprehensive measurement of the uncertainty of experimental data points. While qualitative information is provided in the DIPPR database, a quantitative analysis of the measured data uncertainty is particularly needed.

4. Conclusions

The uncertainty quantification has been illustrated for a thermo-physical property prediction model to predict the LFL through three different modeling approaches. All approaches are not only capable of providing the point prediction of the property but can also provide meaningful information about the models' confidence in its prediction. The uncertainty quantification using the covariance matrix for the GCM provides more noisy confidence bounds compared to the smooth prediction for the GPR model. The confidence of the DNN in the prediction suffers greatly for higher LFL values.

References

- Austin, N.D., Sahinidis, N. V., Trahan, D.W., 2016. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. *Chem. Eng. Res. Des.* 116, 2–26. <https://doi.org/10.1016/j.cherd.2016.10.014>
- Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., Saurous, R.A., 2017. TensorFlow Distributions.
- Frutiger, J., Marcarie, C., Abildskov, J., Sin, G., 2016a. Group-contribution based property estimation and uncertainty analysis for flammability-related properties. *J. Hazard. Mater.* 318, 783–793. <https://doi.org/10.1016/j.jhazmat.2016.06.018>
- Frutiger, J., Marcarie, C., Abildskov, J., Sin, G., 2016b. A Comprehensive Methodology for Development, Parameter Estimation, and Uncertainty Analysis of Group Contribution Based Property Models-An Application to the Heat of Combustion. *J. Chem. Eng. Data* 61, 602–613. <https://doi.org/10.1021/acs.jced.5b00750>
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep Learning*. MIT Press.
- Hukkerikar, A.S., Sarup, B., Ten Kate, A., Abildskov, J., Sin, G., Gani, R., 2012. Group-contribution + (GC +) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilib.* 321, 25–43. <https://doi.org/10.1016/j.fluid.2012.02.010>
- Marrero, J., Gani, R., 2001. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* 183–184, 183–208. [https://doi.org/10.1016/S0378-3812\(01\)00431-9](https://doi.org/10.1016/S0378-3812(01)00431-9)
- Mondejar, M.E., Cignitti, S., Abildskov, J., Woodley, J.M., Haglind, F., 2017. Prediction of properties of new halogenated olefins using two group contribution approaches. *Fluid Phase Equilib.* 433, 79–96. <https://doi.org/10.1016/j.fluid.2016.10.020>
- Rowley, R.I., Wilding, W.V., Oscarson, J.L., Giles, N.F., 2019. DIPPR Data Compilation of Pure Chemical Properties.
- Zhang, L., Mao, H., Liu, L., Du, J., Gani, R., 2018. A machine learning-based computer-aided molecular design/screening methodology for fragrance molecules. *Comput. Chem. Eng.* 115, 295–308. <https://doi.org/10.1016/j.compchemeng.2018.04.018>

Intelligent Management Platform for Material Exchange Optimization and Industrial Symbiosis

Theodore Chatzidimitriou,^a Thanasis Gentimis,^{a,*} Christos Michalopoulos,^b
Antonios C. Kokossis,^{a,b} Theodore Dalamagas,^a

^a*SYMBIOLABS Circular Intelligence LC, Artemidos 6 & Epidavrou str., Marousi 15125, Greece*

^b*School of Chemical Engineering, National Technical University of Athens, Athens, Greece*

Abstract

Developing Industrial Symbiosis (IS) networks is a knowledge-intensive practice, where information is of crucial importance in order to discover the potential of IS connections between facilities engaging in exchange of materials and/or energy. In this work, we provide methods and tools that assist modelling and systemic use of existing IS knowledge to analyse potential IS networks and synergies.

Keywords: industrial symbiosis, knowledge graphs, graph databases.

1. Introduction

Industrial Symbiosis (IS) is an innovative approach to create industrial networks (Isenmann and Chernykh, 2009) for economic, environmental and social benefits. IS brings together companies from all business sectors through material trading and sharing assets to add value, reduce costs and benefit the environment (Lehtoranta et al., 2011). Developing IS networks is a knowledge-intensive practice, where information is necessary to discover potential IS connections between flows of materials, waste and/or other resources related to industrial facilities. Therefore, the efficiency of ICT solutions for facilitating IS depends heavily on collecting, analysing and integrating highly-diverse quantitative and qualitative data (Grant et al., 2010). The key objective of our work in the SYMBIOICT project (co-financed by the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH-CREATE-INNOVATE) is to overcome two major barriers for detecting large-scale IS networks such as: (a) focus on “1-1” connections, (b) lack of modelling and systemic use of existing knowledge. By applying data analytics technologies we produced a knowledge graph of existing and potential IS connections, matching input/output flows of industrial facilities. Each connection can be weighted using parameters such as the distance of the facilities (transport cost), the facility size and the cost savings from the proposed waste prevention. We demonstrate how knowledge graphs can be effectively used for matching (a) material to process, (b) material to company, (c) company to company. Using the knowledge graph we accomplished an effective analysis of potential IS networks and synergies that would otherwise be extremely difficult to identify. These results were also visualized on a map, displaying “hot spots” of existing IS synergies as well as cases with high potential of forming new IS connections.

2. Background

2.1. Knowledge Graphs

A knowledge graph captures the semantics of a domain using a set of definitions of concepts, their properties, and the relations between them. Knowledge graph has become a term that is recently ubiquitously used yet does not have a well-established definition. Typically, a knowledge graph: (i) mainly describes real world entities and their interrelations, organized in a graph, (ii) defines possible classes and relations of entities in a schema, (iii) allows for potentially interrelating arbitrary entities with each other, and (iv) covers various topical domains (Paulheim, 2017).

Since its inception by Google in 2012 (powered in part by the Freebase knowledge base) and until today knowledge graphs have been an intrinsic part of organizing knowledge in a distributed way for web applications, especially after the “second wave” of the Semantic Web with the introduction of Linked Data and the adoption of Resource Description Framework for information modelling on the web (Ji et al., 2020).

2.2. Industrial Symbiosis facilitation

In EU countries, the promotion of IS is usually supported by teams of experts or practitioners that engage in networking sessions with industries and other stakeholders for the purpose of the development of IS projects (Artola et al., 2018). In these sessions, brokers usually apply straightforward techniques (without the use of ICT tools) to detect the possibility of resource (raw materials, waste, energy) exchange among the participating stakeholders. This approach can be restrictive, leading to serendipitous discovery of cooperation opportunities and IS networks of limited scale and complexity. The success of such efforts relies heavily on the skills and experience of each mediator and not on a systemic use of information and knowledge. Other efforts, such as the SMILE Resource Exchange platform launched in 2010 in Ireland as a waste trading system to promote IS, have later transitioned towards a more hands-on facilitated approach focusing on “1 on 1” solutions before utterly coming to a close in 2018.

In recent years, a few digital services have been developed focusing on supporting decision making for waste valorisation and reduction. For example, International Synergies (www.international-synergies.com) has developed SYNERGie, a proprietary ICT resource management database and platform, enabling organizations to identify resource reuse opportunities. Even though these digital solutions incorporate knowledge of best practices for waste valorisation, they often follow the same approach of searching for “1 on 1” connections. In this way they fail to take into consideration data on raw materials, waste and industrial processes, the analysis of which can highlight new possibilities for establishing more complex and less obvious IS networks.

3. Dataset and Model

3.1. Description

We adopt a data-driven approach for the facilitation of IS, proposing knowledge graphs as the tool for incorporating IS related data, such as industrial processes combined with flows of materials, waste and other resources related to industrial facilities. With this representation, the user can model large and complex IS networks involving heterogeneous industries and a wide range of by-products exchanged. Logic built into knowledge graphs allows us to reason about the information contained within, and to make implicit information in the graph explicitly accessible.

The challenge of creating a knowledge graph for waste valorization solutions lies in the fact that the necessary information is usually fragmented and located in many

heterogeneous sources. Therefore our first step was the extraction of the data related to IS from various sources, and then we moved on to the transformation and interconnection of them. In this way we managed to construct a knowledge graph and perform further analysis on the data.

In particular, the knowledge graph can be used for matching waste to process, waste to company, company to company or querying using only desired properties. We can demonstrate examples in agricultural activities, rural and municipal communities and ports, with reliable results. The graph can be enriched as more technologies are developed. Connections can be either implemented or potential, depending on the development stage. Furthermore quantitative data can be used for evaluating possible synergies with mathematical optimization techniques.

3.2. Data Collection

Data used for the development of the graph database were mainly mined from official repositories of IS best practices as well as research papers and news items. In particular, our main sources were best practices implemented in Sweden: (industrialsymbiosis.se), Finland: (www.industrialsymbiosis.fi/home-en-gb/) and Denmark (Kalundborg Eco-industrial Park). The Nordic countries have a long tradition in sustainable development and thus they offer a breadth of available data and best practises, which could be replicated in other less circular countries. Other novel examples we took into consideration include Guitang group in China as well as IS networks established in Greece (symbiosisproject.eu), which proves that even in countries where circular economy initiatives are still at an early stage, IS networks can be developed, given that the necessary knowledge is available.

4. Methodology

4.1. Knowledge Graph Model

Nodes and edges (or relationships) were used in order to represent entities involved and their in between connections. In the proposed model nodes can have one of three different labels: Company, Material or Process which connect to each other with relationships (edges) having one of the following three labels: Is_Input, Has_Output or Has_Process. The nodes, the edges and the attributes are presented below in more detail. Regarding nodes, three different labels were used.

- Set of Companies C: {c|c company} Nodes with label: “Company” represent the geographic location where the various processes take place.
- Set of Materials M : {m|m material} Nodes with label: “Material” represent materials in any state (solid, gaseous, liquid) and can be waste materials, raw materials, byproducts etc.
- Set of Processes P : {p|p process} Nodes with label: “Process” describes actual industrial processes that transform raw materials into useful products, but also activities such as farming etc.

Regarding edges, three different labels were used.

- Set of Is_Input : {i|i Is_Input} Relationships with type: “Is_Input” relate Material nodes with Process nodes.
- Set of Has_Output: O : {o|o Has_Output} Relationships with type: “Has_Output” relate Process nodes with Material nodes.
- Set of Has_Process H: {h|h Has_Process} Relationships with type: “Has_Process” relate Company nodes with Process nodes.

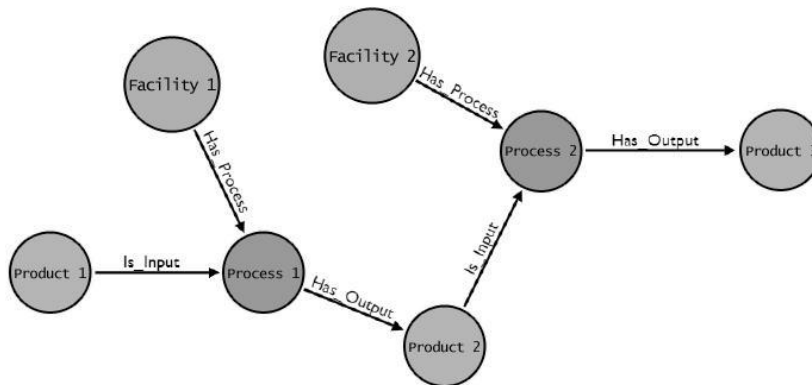


Figure 1. Sample graph with the three different types of nodes and three different types of edges

The node labeled Company is connected with the node labeled Process through the edge labeled Has_Process, whereas the node Process and Material can connect with edges labeled Is_Input or Has_Output depending on the direction. Both edges and nodes have attributes which vary depending on the label. Properties regarding materials can be physical, chemical, biological etc. whereas properties regarding processes can be CAPEX, OPEX etc. In the next section we demonstrate how the knowledge graph is used to suggest connections between companies, waste and processing technologies.

5. Results

5.1. Implementation

In our approach we utilize the neo4j platform (neo4j.com), a native graph database built from the ground up to leverage data and also data relationships. Using neo4j we created a graph database for storage, automation and visualization of IS networks. Using the data collected we created a Knowledge Graph containing: 49 industrial facilities, 99 industrial processes, and 154 different materials. The resulting Knowledge Graph can suggest connections between companies, waste and processing technologies.

5.2. Query Example

Using our graph we can perform a variety of queries for facilitating IS planning and networking, offering matches between (a) materials to process, (b) materials to company, (c) company to company. E.g., a question such as: "My facility produces Ethanol by Fermentation of Molasses. Which other materials can be used instead of Molasses in my Fermentation process to produce Ethanol?" is expressed as follows in neo4j:

```

match p=(n:Material)-[*]->(a:Material)
where any(node_on_p in NODES(p)
  where node_on_p.ProcessName='Fermentation' ) and n.MaterialName <>
'Molasses' and a.MaterialName='Ethanol' and n.MaterialRole
in['final','exchange']
return p
  
```

By executing the above query, neo4j presents a result set with materials that can lead to the production of Ethanol by Fermentation, based on the data of the Knowledge Graph. We also get the corresponding paths for the chain of processes in order to achieve it. In

this example we get: Bread Residues (through Hydrolysis), Sugar Beet (directly), Wheat Waste (directly), Starch (directly), and Municipal Solid Waste (through Gasification).

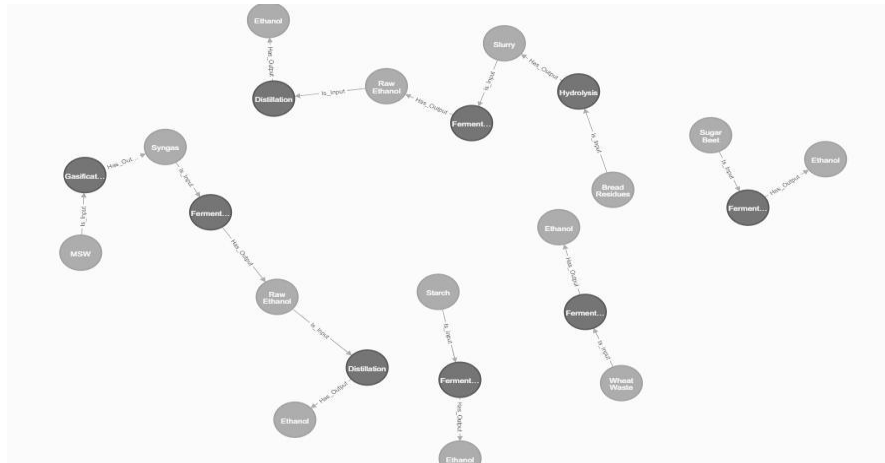


Figure 2. Result set of query on Knowledge Graph.

5.3. Visualization of IS potential

Combining information extracted from the knowledge graph with data regarding location and activity of industrial facilities in Greece (e.g., location, magnitude, industrial NACE codes), we were able to detect cases with high potential of forming successful new IS connections. We visualized these proposed IS connections on a digital platform (snf-252687.vm.okeanos.grnet.gr), where industrial facilities are connected based on the potential exchange of materials and/or other resources. Each IS connection is weighted based on the distance between the industrial facilities involved combined with the facility magnitude. Results are visualised on an interactive map, indicating proposed IS connections as well as “hot spots” with high potential of IS regional synergies.

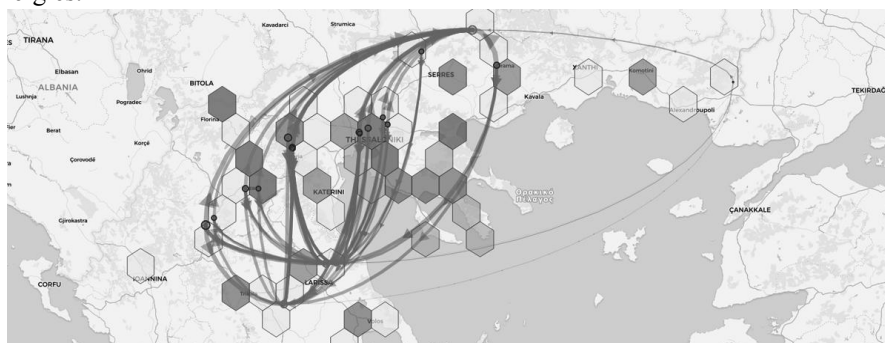


Figure 3. Map of proposed IS connections and “hot-spots” using the Knowledge Graph.

6. Conclusions

In conclusion, the developed model allows real time analysis about potential IS networks which include several different partners and synergies that would otherwise be

extremely difficult to identify. In this way, a user of the knowledge graph can use even limited information as parameters in order to obtain suggestions on relative applied industrial symbiosis practices, proposed technologies that can process their waste, and generally identify possible IS opportunities. Finally, we demonstrate a digital platform that exploits IS knowledge graphs present IS hotspots and potential new connections.

Limitations of this methodology, which actually will be taken into consideration in our future work plans to improve our methods, include the data collection process, which is not automated, as well as the lack of data regarding the economic value of the exchanges. Also, the graph cannot prioritize potential matches or compare them with regard to their economic or environmental performance. Finally, at this stage, we do not provide narrow searches involving constraints in equipment and processes, or even involving company location information, but such type of data has already been collected (see also Figure 3) and will be soon exploited to provide such type of searches.

References

- I. Artola, A. Doranova, T. Domenech, L. Roman, M. Smith, 2018. Cooperation fostering industrial symbiosis market potential, good practice and policy actions : final report.
- G. B. Grant, T. P. Seager, G. Massard, L. Nies, 2010. Information and communication technology for industrialsymbiosis. *Journal of Industrial Ecology* 14 (5), 740–753.
- R. Isenmann, K. Chernykh, 2009. The role of ict in industrial symbiosis projects – environmental ICT applications for eco-industrial development. In: V. Wohlgemuth, B. Page, K. Voigt (Eds.), *Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools*. Shaker Verlag, Aachen
- S. Ji, S. Pan, E. Cambria, P. Martinen, P. S. Yu, 2020. A survey on knowledge graphs: Representation, acquisition and applications.
- S. Lehtoranta, A. Nissinen, T. Mattila, M. Melanen, 2011. Industrial symbiosis and the policy instruments of sustainable consumption and production. *Journal of Cleaner Production* 19 (16), 1865 – 1875, pro-moting Transformation towards Sustainable Consumption and Production in a Resource and Energy Intensive Economy - the Case of Finland.
- H. Paulheim, 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8, 489–508, 3

Towards a Software Prototype for Synthesis of Operable Process Intensification Systems

Shivam Vedant^{a,b}, Mary Rivera Atencio^{a,b}, Yuhe Tian^{a,b}, Vaishnav Meduri^{a,b}, and Efstratios N. Pistikopoulos^{a,b}

^a*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, United States*

^b*Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843, United States*
stratos@tamu.edu

Abstract

In this work, we present a software prototype for the generation of Process Intensification (PI) systems integrated with safety, flexibility, and control analysis. The prototype comprises three major suites: (i) Synthesis Suite - which systematically generates promising PI configurations based on a phenomenological process synthesis approach (i.e. Generalized Modular Representative Framework), (ii) Simulation Suite - which is integrated with a PI model library to translate the PI synthesis results into corresponding equipment-based process alternatives, and (iii) Operability Suite - which performs model-based safety, flexibility, and controllability analysis to ensure the actual operational performance of the resulting PI systems under varying conditions. A user interface has been built to coordinate the functionalities of the suites, allowing them to work in tandem to create an environment that generates safe and operable PI flowsheets without pre-postulation of potential process schemes. A case study on the metathesis reaction of pentene is presented to demonstrate the potential of the prototype in deriving safe and operable PI systems.

Keywords: Process Intensification, Process Synthesis & Optimization, Process Control, Software Prototype.

1. Introduction

Process Intensification (PI) has attracted an emerging interest in the chemical engineering research community and the chemical process industry owing to its potential ability to drastically increase process profitability and efficiency (Moulijn et al., 2008; Tian et al., 2018). The past few decades have witnessed significant advances in the field of PI (Keil, 2018; Segovia-Hernández and Bonilla-Petriciolet, 2016). However, a widely used process intensification commercial software is still lacking (Tula et al., 2019; Skiborowski, 2018). Driven by several recent national initiatives such as the RAPID Institute (Bielenberg and Palou-Rivera, 2019), the academic community has initiated attempts to develop software prototypes for computer-aided process design, synthesis, and intensification leveraging state-of-the-art process systems engineering approaches (Tula et al., 2017; Demirel et al., 2017; Chen et al., 2018; Miller et al., 2018).

In this paper, we present the development towards a software prototype based on our recently proposed framework for the synthesis of operable process intensification systems (Tian et al., 2020). The prototype systematically generates intensified process systems by utilizing a novel phenomenological synthesis approach with embedded process operability, safety, and explicit model predictive control analysis criterion. Section 2 of the paper elucidates the prototype architecture and the function of embedded suites. In section 3, a case study of a reactive separation system is presented to demonstrate the working of the prototype. The remaining section covers conclusions and future directions.

2. Software Prototype

The prototype platform consists of three suites, namely: (i) Synthesis Suite, (ii) Simulation Suite, and (iii) Operability & Control Suite. As illustrated in Figure 1, the user interface (UI) built on Python brings these suites together in a seamless manner to provide a consolidated environment while allowing user input and interaction. The different functional suites take advantage of various commercially available software packages, such as GAMS[®] for process synthesis and optimization, ASPEN PLUS[®] for steady-state process simulation, and gPROMS[®] for dynamic modeling and control analysis. These suites work in tandem to provide key insights to the user and are coordinated through the UI as detailed below.

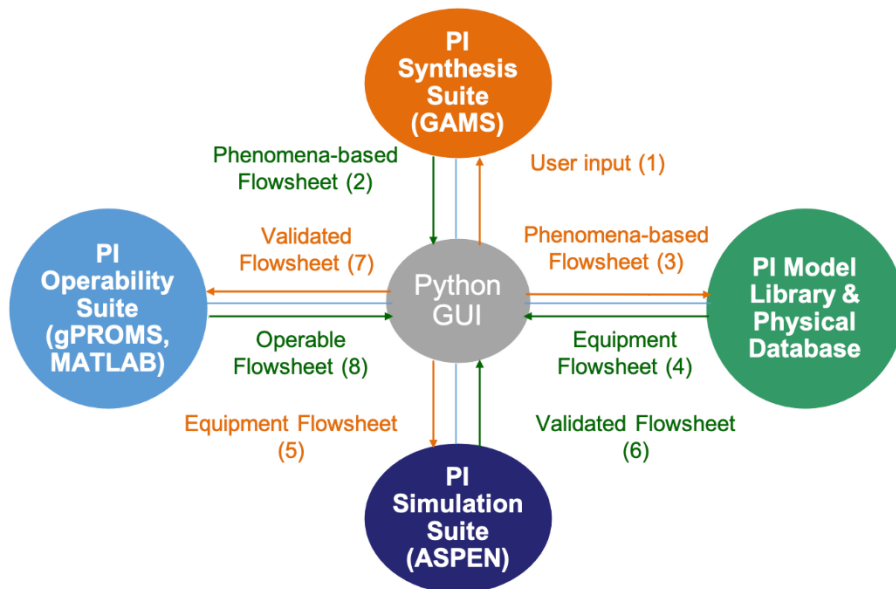


Figure 1: Information flow chart for the software prototype platform

2.1. User Interface

The interface is programmed to facilitate communication and the transfer of data between the suites of the software prototype. Through the UI, users can select individual suites for targeted equipment/flowsheet intensification, design, or analysis. However, the UI can also guide users to navigate between different suites in a step-by-step manner

for integrated design with operability, safety, and/or control. To this end, the UI acts as a central node for information flow as shown in Figure 1 and provides key input and output information to the user at every stage.

2.2. *Synthesis Suite*

The synthesis suite provides the tools required for the systematic generation of optimal and intensified process solutions. The technical foundation of the suite is laid by the Generalized Modular Representative Framework (GMF) (Papalexandri and Pistikopoulos, 1996). To characterize various chemical processes, GMF employs two phenomenological modules, namely, the pure heat exchange module and the multi-functional mass/heat exchange module. Interlinked configurations of these modules, generated as a result of solving an optimization problem, can form the basis for the creation of conventional and/or novel unit operations. The optimization problem is formulated as a mixed-integer nonlinear programming problem (MINLP) and solved using the Generalized Benders Decomposition algorithm in GAMS[®]. The parameters of the model are assumed to be deterministic and provided by the user through the UI. To exchange information between the UI and the synthesis suite, Python Application Programming Interface (API) is used. More detail on GMF modular representation and model formulation can be found in Tian and Pistikopoulos (2020).

2.3. *Simulation Suite*

To validate the resulting configuration, the GMF module based flowsheet is converted to the corresponding equipment based flowsheet. This translation is achieved through a model library which comprises of information pertaining to various process equipment, and a set of rules governing the assignment of equipment to a module or a group of modules. The user can avail from a library of models and suggestions provided by the database to choose the equipment. This allows the prototype to account for equipment constraints and lends the flexibility to generate alternative flowsheets. It is imperative to note that the accuracy of translation will depend on the extensiveness of the library database. Novel equipment will have to be appended to the library in prior to achieve the desired translation. Subsequently, the equipment based flowsheet is simulated using ASPEN PLUS[®] to perform steady-state validation. Furthermore, high fidelity models are developed to fully capture and analyze the process dynamics.

2.4. *Operability Suite*

Model-based analyses are currently enabled for the following PI operational considerations: (i) flexibility analysis to ensure feasible operation under process uncertainty, (ii) risk analysis to evaluate the inherent safety performance of the resulting process configuration at the conceptual design stage, and (iii) explicit/multi-parametric model predictive control to deliver optimal dynamic operation strategies under disturbances following the PAROC (PARAMetric Optimisation and Control) framework (Pistikopoulos et al., 2015). These operability analysis approaches can also be integrated with the above Synthesis Suite and Simulation Suite to simultaneously generate optimal and intensified process designs with guaranteed operability, safety, and control performance (Tian et al., 2020).

3. Case Study: Pentene Metathesis Reaction

3.1. Problem Statement

In this section, we revisit the problem of 2-Pentene (C_5H_{10}) metathesis to form 2-Butene (C_4H_8) and 2-Hexene (C_6H_{12}) which is an equilibrium limited reaction adapted from Tian and Pistikopoulos (2020). At atmospheric pressure, the reaction takes place in the liquid phase and can be described by ideal vapor-liquid equilibrium (VLE). The production target is to obtain 50 kmol/h of 98% butene and 50 kmol/h of 98% hexene from a saturated liquid feed stream of 100 kmol/h pure pentene at atmospheric pressure. The objective is to design a process with the minimum utility cost.

3.2. User Input

The UI window is built on Python with the help of the *Tkinter* package. It allows the creation and management of window attributes like panels, buttons, and entry fields. The UI window comprises of dedicated panels for showing input/output data, run time data, superstructure layout, and comparison between alternatives of the generated flowsheet. To allow for the dynamic updates of UI panels, they are created as objects of python defined classes. Users can access all functionalities of the prototype with the help of the button attribute. For the case study, the required input information includes physical properties like molecular weight and molar mass, thermodynamic information such as Antoine and heat capacity coefficients, reaction information such as stoichiometry and standard Gibbs energy of formation, utility costs, and input feed conditions.

3.3. Process Synthesis

To generate the module based flowsheet, the maximum number of modules was set to 5. This number can be increased to acquire more information about the identified unit operations or decreased to reduce the computational time. The optimal MINLP solution includes values for the number of pure heat and mass/heat exchanger modules (integer variables), the presence of connections between stream and modules (binary variables), and the operating conditions (continuous variables). The results are exported to a database file (.db) to create modular structures using Python scripts. To visualize the solution, conditional statements are used to read the binary values and create their corresponding module or stream. For example, a module exists if the associated binary variable is assigned the value of 1, or 0 otherwise.

The resulting graphical module based flowsheet with the information on the constituent phenomena is shown in Figure 2. The modules and text in the flowsheet are sized according to the number of modules, and the size of the allocated window. Additionally, the python scripts account for the optimal layout to display process streams without over-lapping with the modules. The red and blue lines show hot and cold stream conditions respectively. The pure heat exchange module represented by white blocks, at the top and bottom can be translated into condenser and reboiler respectively. The mass/heat exchange modules can be translated into trays of a reactive distillation column.

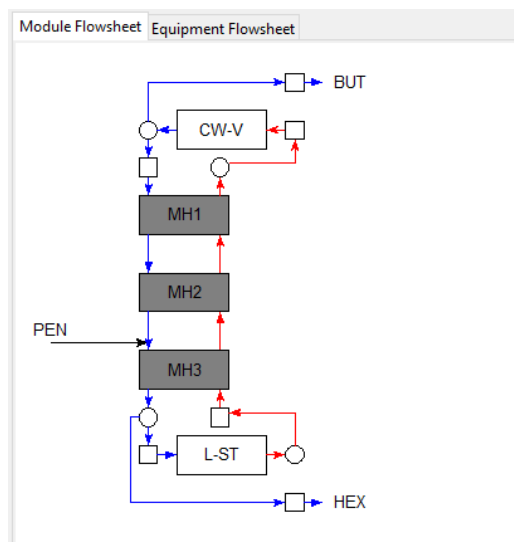


Figure 2: Module flowsheet for the pentene metathesis reaction

3.4. Remarks

The resulting GMF modular solutions are then identified and translated to equipment-based process alternatives with the help of a specialized PI model library developed as part of the SYNOPSIS Project (Pistikopoulos et al., 2020). The model library consists of validated rigorous and short-cut models for various intensified reaction and/or separation systems, including but not limited to, advanced distillation columns and reactive distillation columns. Moreover, the model library integrates different commercial software platforms (e.g., Aspen, gPROMS, Python) to leverage the existing unit operation models as well as to enable the flexible use of models for different computational purposes (e.g., simulation, optimization, control) in a platform-independent manner.

The safety, operability, and control performance of the PI systems can be further analyzed using the model-based metrics introduced in Section 2.4. This can be achieved either via posterior operational analysis for a given intensified process design, or via integrated process design optimization with operability, safety, and control considerations to systematically generate optimal process structures with desired operational performances.

4. Conclusion and Future Work

In this paper, we have presented our work towards the development of a software prototype to integrate process synthesis and process simulation with process operability, safety, and control analysis of the PI design to deliver validated operable PI system flowsheets. The connection of different parts of the prototype is explained and demonstrated using a pentene metathesis reaction example. Ongoing work addresses

the creation of the model library and establishing a connection of the UI with the simulation and operability suites.

5. Acknowledgment

We acknowledge the financial support from the Texas A&M Energy Institute, Shell, and RAPID SYNOPSIS Project (DE-EE0007888-09-03, Partner Organizations: Texas A&M University, Georgia Institute of Technology, Auburn University, Shell, Dow Chemical Company, Process Systems Enterprise Limited).

References

- J. Bielenberg, I. Palou-Rivera, 2019. The RAPID Manufacturing Institute – Reenergizing US efforts in process intensification and modular chemical processing. *Chemical Engineering and Processing - ProcessIntensification* 138, 49 – 54.
- Q. Chen, E. S. Johnson, J. D. Siirola, I. E. Grossmann, 2018. Pyomo.GDP: Disjunctive Models in Python. *Computer Aided Chemical Engineering* 44, 889–894.
- S. E. Demirel, J. Li, M. F. Hasan, 2017. Systematic process intensification using building blocks. *Computers & Chemical Engineering* 105, 2–38.
- F. J. Keil, 2018. Process intensification. *Reviews in Chemical Engineering* 34 (2), 135–200.
- D. C. Miller, J. D. Siirola, D. Agarwal, A. P. Burgard, A. Lee, J. C. Eslick, B. Nicholson, C. Laird, L. T. Biegler, D. Bhattacharyya, N. V. Sahinidis, I. E. Grossmann, C. E. Gounaris, D. Gunter, 2018. Next generation multi-scale process systems engineering framework. *Computer Aided Chemical Engineering* 44, 2209– 2214.
- J. Moulijn, A. Stankiewicz, J. Grievink, A. Górak, 2008. Process intensification and process systems engineering: A friendly symbiosis. *Computers & Chemical Engineering* 32, 3–11.
- K. P. Papalexandri, E. N. Pistikopoulos, 1996. Generalized modular representation framework for process synthesis. *AIChE Journal* 42 (4), 1010–1032.
- E. N. Pistikopoulos, N. A. Diangelakis, R. Oberdieck, M. M. Papathanasiou, I. Nascu, M. Sun, 2015. PAROC –An integrated framework and software platform for the optimization and advanced model-based control of process systems. *Chemical Engineering Science* 136, 115 – 138.
- E. N. Pistikopoulos, M. M. F. Hasan, J. S. Kwon, M. J. Realff, F. Boukouvala, M. R. Eden, S. Cremaschi, B. J. Tatarchuk, J. B. Powell, L. Spanu, R. Bindlish, S. Leyland, 2020. SYNOPSIS – Synthesis of Operable Process Intensification Systems. RAPID Institute Project 9.3, DE-EE0007888-09-03.
- J. G. Segovia-Hernández, A. Bonilla-Petriciolet, 2016. *Process Intensification in Chemical Engineering*. Springer.
- M. Skiborowski, 2018. Process synthesis and design methods for process intensification. *Current opinion in chemical engineering* 22, 216–225.
- Y. Tian, S. E. Demirel, M. F. Hasan, E. N. Pistikopoulos, 2018. An overview of process systems engineering approaches for process intensification: State of the art. *Chemical Engineering and Processing - ProcessIntensification* 133, 160–210.
- Y. Tian, I. Pappas, B. Burnak, J. Katz, E. N. Pistikopoulos, 2020. A systematic framework for the synthesis of operable process intensification systems – Reactive separation systems. *Computers & Chemical Engineering* 134, 106675.
- Y. Tian, E. N. Pistikopoulos, 2020. Towards an envelope of design solutions for combined/intensified reaction/separation systems. *Industrial & Engineering Chemistry Research* 59 (24), 11350–11354.
- A. K. Tula, D. K. Babi, J. Bottlaender, M. R. Eden, R. Gani, 2017. A computer-aided software-tool for sustainable process synthesis-intensification. *Computers & Chemical Engineering* 105, 74–95.
- A. K. Tula, M. R. Eden, R. Gani, 2019. Computer-aided process intensification: Challenges, trends and opportunities. *AIChE Journal* 66 (1).

Data-driven modelling of choke valve erosion using data simulated from a first principles model

Jan Henrik Jahren^a, Jose Matias^a and Johannes Jäschke^{a*}

^aDepartment of Chemical Engineering, Norwegian University of Science and Technology, Sem Sælands vei 4, 7034 Trondheim Norway

Abstract

Maintenance of subsea operating equipment is a significant part of the operational costs of running an oil production system. For instance, on the Norwegian continental shelf, it amounts to 60 billion NOK in operational expenses. One of the principal mechanisms of degradation in subsea process equipment is erosion by sand, which is a very complex process and, thus, difficult to model using physical domain knowledge. Because of this difficulty, we propose in this paper the use of data-driven approaches for modelling erosion in critical equipment of a subsea oil production rig. In such systems, a multitude of available process measurements such as flowrates, gas lift injection rates, pressures etc. can be combined to a soft sensor for component degradation. This approach could save significant amounts of resources by allowing fewer cost intensive inspections and monitoring schemes. A soft-sensor approach was tested with simulated data of the sand degradation of a choke valve in a gas lifted oil production system with three wells. In this paper, we present results from soft-sensor methods like multiple linear regression, regression trees, ensembles methods and kernel methods. The approaches were tested and compared in two case studies, the first with constant sand outflow from the reservoir for initial exploration of the data driven approaches, and the second with a more realistic profile in which the sand rate is increasing. In both cases, we compare the soft sensor modelling techniques in terms of their basic requirements of accuracy as well as transparency and interpretability.

Keywords: equipment degradation, data-driven modelling, machine learning

1. Main Text

Failure to detect faults in large scale, expensive or critical equipment can have immense consequences. Both financial and in the most extreme cases, loss of life. One of the main mechanisms in equipment degradation in subsea oil extraction, is sand erosion. Accurately modelling this process is vital for monitoring of equipment health (Si et al. (2012)). Erosion by sand is a very complex process and, thus, difficult to model using physical domain knowledge. Because of this difficulty, we propose in this paper the use of data driven approaches for modelling erosion in critical equipment of a subsea oil production rig. In such systems, a multitude of available process measurements, such as flowrates and pressures, can be combined to a soft sensor for component degradation. This approach could save significant amounts of resources by allowing fewer cost intensive inspections and monitoring schemes as well as improving safety. To investigate the usability of such a data-driven modelling approach models will be tested on simulated data using the subsea gas lifted oil well network model proposed by Krishnamoorthy et al. (2016) and adapted by Verheyleweghen and Jäschke (2018).

2. Gas lifted oil well network

In some cases, oil wells do not have sufficient reservoir pressure to lift fluids to the topside facility. Gas lift, the injection of compressed gas through the annulus, can be used to overcome this. The annulus is the void between the well piping and the casing. This leads to a reduction in the fluid mixture density. That, in turn reduces the hydrostatic pressure loss in the well. Consequently, the pressure at the well bottom decreases, compensating for the low reservoir pressure.

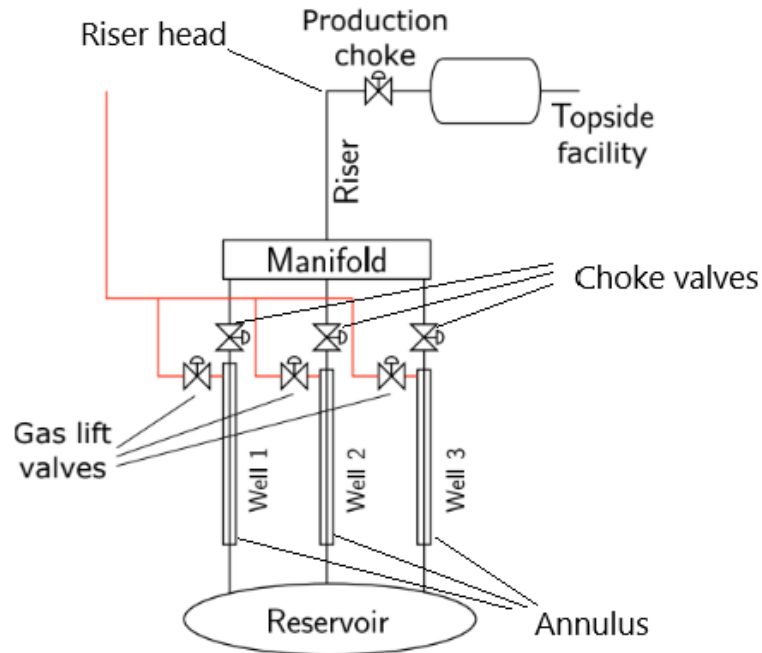


Figure 1: Illustration of a gas lifted oil production system with three wells showing how production goes from reservoir to topside facility (Verheyleweghen and Jäschke (2018)).

In Figure 1 the different parts of a gas lifted oil well network is shown. The oil flows, from a single reservoir into three wells, then, into a manifold before the riser takes the fluid mixture to the surface. Maturing fields experience a significant increase over time in sand production, which can be approximated by an exponential function (Hettema et al. (2006)).

3. Method

3.1. Simulations

A small additional adaption to the model was made to incorporate a varying sand production rate (SPR). For each case study datasets of 200 time series of 500 days were simulated containing the total erosion length of the wells choke valves. The gas lift rate was varied randomly every 50 days within a given range:

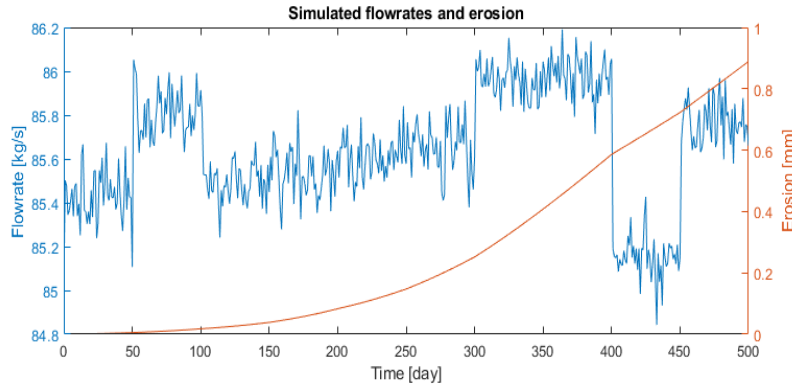


Figure 2: Illustrative plot of cumulative erosion [mm] (right axis) and flowrate [kg/s] (leftaxis) from one 500 day time series.

Constant sand production rate yielding a greatly simplified model, but a good starting point for initial exploration of the data-driven methods. This is equivalent to: $m_{\text{sand}}(t) = m_{\text{sand}}(0)$.

Exponential sand production rate yielding a more realistic model of a field increasing sand production as it matures. An exponential function, $m_{\text{sand}}(t) = m_{\text{sand}}(0)e^{0.05t}$, was chosen to emulate a qualitative description of the sand production rate in a field over time (Hetteema et al. (2006)). Figure 2 shows the cumulative erosion of the choke valve for one simulated time series in addition to the total well production flowrate. Note that the choke valve erosion is a function of the well production flowrate, as expected, but the dominating force is the sand production flowrate. The more sand produced by the well, the faster the choke valves erode.

Process variables were measured every time step (day), with added noise to process measurements. Erosion, gas lift injection rate and sand production rate were recorded directly without noise. The pressure is measured in the annulus, well head (top of each well), riser head (top of the riser) and in the manifold connecting the wells. The production rates of gas and oil in the top of the riser as well as the top of each individual well are also measured. In addition to the process measurements, the sand production rate of the reservoir and the gas lift injection rate (control input of the optimisation) are used to simulate the erosion in the choke valves atop each well. The simulated data is then split in two parts of equal size, a training set and an independent test set. Due to the large availability of data for validation, since simulated data is used, we choose to use holdout validation, holding out 30% of the training set for model selection and validation. The final data sets contained 100 time series for training and validation and 100 time series for an independent test set.

3.2. Pre-processing

The gradient of the simulated cumulative erosion is used as response variable (i.e., the erosion rate), this is chosen as the response variable as the cumulative erosion has significant autocorrelation and is a time series, while the erosion rate at any given point only depends on the process variables at that point in time. Sand production rate is usually.

not a constantly monitored variable, as such it was supplied to the models with a realistic sampling interval of 50 days, but the impact of lowering this to 30 days is also studied. Models were tested on normalized data with unit variance and mean zero. Tests were also done using principal component (PC) scores as predictors, in this case PC's explaining > 95% variance.

3.3. Data-driven models

For obtaining the soft sensor for erosion degradation, multiple traditional statistical learning methods are implemented and compared for the use cases, a brief introduction is given below but the interested reader is referred to Friedman et al. (2017). All models were implemented in MATLAB using the Statistics and Machine Learning toolbox (The MathWorks (2020)).

Multiple linear regressions (MLR) are the simplest of all statistical models, where we assume a linear relationship between predictors (or combinations of predictors when including interactions). When such a relationship exists, these models yield good interpretability and room for extrapolation. For the linear regression models, least squares loss is used, and a stepwise approach (Stepwise MLR) is applied for model selection, with iterative addition of predictors (linear terms and interaction terms) to a null model until additional predictors no longer lead to significant reduction in loss.

Ensemble methods essentially aggregate several simple models, in this case regression trees to yield a better prediction than individual models can. In this work bootstrap aggregation (Bagged trees) and gradient boosting (Boosted trees) were used. Ensemble methods can provide very good models of linear and non-linear phenomena, but this comes at a significant cost in transparency as interpreting relationships from these methods is very difficult.

Support vector regression (SVR), which is a kernel method adapted from the famous classification algorithm support vector machines, was used. This method uses the support vector machine (SVM) algorithm to create a separating hyperplane in the data space which will allow prediction on new data samples. SVM's are particularly suited for high dimensional data and are robust against outliers (Drucker et al. (2003)). The reported results are from Cubic kernel SVR, which had the strongest overall performance.

In this work the loss function for parameter optimisation, model selection and model validation is always the mean square error of prediction (MSE). Hyperparameters of the ensemble and kernel methods were tuned using Bayesian optimization, using the expected improvement as acquisition function.

4. Results and discussion

Learning methods were tested on both the data simulated from constant sand production rate and exponential sand production rate. The performance of the methods is measured by independent test set mean squared error of prediction (MSE). The performance results are given in Table 1, where the first two columns give results for the first case study and the last for the second case study. There are some considerations outside of model performance to be taken into account when selecting a model for actual applications. There are significant differences in model training time, with the outlier being cubic kernel SVR, requiring significantly more training than bagged trees, boosted trees and

stepwise MLR. When dealing with safety critical equipment and infrastructure we are not only interested in a high accuracy, but also a high level of trust in the model predictions which in turn means that transparent and interpretable models are preferred. None of the models considered here are completely black box, in general models with transparent coefficients showing how a prediction is made are easier to interpret. Linear regressions are a class of such methods, allowing easy analysis of coefficients and, consequently its predictions. As such, when performance is close the preferred methods will be MLR models due to fast training times and superior interpretability.

Method	Const. SPR	Const. w/ PCA	Exp. SPR	Exp. w/ PCA
Stepwise MLR	0.0096	0.0166	0.0182	0.0191
Bagged trees	0.0091	0.0162	0.0184	0.0192
Boosted trees	0.0125	0.0320	0.0184	0.0239
Cubic kernel SVR	0.0120	0.0129	0.0164	0.0169
30 Day Sampling Rate				
Stepwise MLR	<i>N/A</i>	<i>N/A</i>	0.00740	0.00946
Bagged trees	<i>N/A</i>	<i>N/A</i>	0.00537	0.0134
SVR	<i>N/A</i>	<i>N/A</i>	0.00687	0.00879

4.1. Constant sand production rate

All the methods show a small MSE value when tested on unseen data. Since the underlying phenomenon of erosion behaves linearly when sand production rate is held constant as in this test case. Very good fit is, thus, expected. The usefulness of this test case was primarily for initial exploration on a very simplified system. Additionally, we note that the performance of all models is degraded when principal components analysis (PCA) pre-processing is applied. This could be because the overall degrees of freedom afforded to the model is lowered. If multiple variables are mainly represented in one principal component, the interactions between them cannot be properly modelled by the interaction terms in a linear regression model, similar arguments hold for the other methods.

4.2. Exponential sand production rate

For the data simulated with exponential sand production rate, a significant decrease in the performance was observed for all methods, as expected with a more complex phenomenon being emulated. Similarly, to the initial test case, there is a drop in performance when PCA is applied, but this effect is relatively weaker for the exponential data. In this case cubic kernel SVR proved to have the strongest performance, but in general bagged trees, boosted trees and Stepwise MLR all provided accurate predictions of the erosion rate. The sampling rate of the sand production rate measurement as expected has a very significant impact on the model accuracy with exponential distribution, increasing sampling to once every 30 days instead of 50 for example reduces the MSE of an optimised bagged ensemble from 0.0066 to 0.0184. Similar trends are seen in the other methods as well, with model performance improving significantly when

sampling rate is increased. Such an effect is expected, as the models are working with more accurate data.

5. Conclusion

It is observed that for constant sand production rate, very accurate predictions of the erosion rates are made. Additionally, a significant degradation of accuracy is seen for all the constant sand production rate models except for the kernel methods when PCA is used. This could be due to a lower reliance on variable interactions which to some extent is hidden when PCA is applied. This provided a useful initial exploration and foundation for the second case study. On the data simulated from an exponential sand production rate the performance is overall worse, which is expected since the phenomenon that the models are attempting to reproduce is more complex. The methods are still relatively accurate with under 0.02 MSE on normalised unseen test data. However, when the slope of the sand production rate profile gets very steep (i.e time increases) the models suffer quite significantly from the sampling rate. With current model performance there is a strong case to be made for selecting linear regression-based methods as they provide superior model transparency and interpretability. Having observed that simulated data can be predicted well using statistical models, further investigation on real world data is merited to ascertain applicability to real industrial facilities.

6. Acknowledgements

This work was carried out as a part of SUBPRO, a Research-based Innovation Centre within Subsea Production and Processing. The authors gratefully acknowledge the financial support from SUBPRO, which is financed by the Research Council of Norway, major industry partners, and NTNU.

References

- Drucker, H., C, C., Kaufman, L., Smola, A. and Vapnik, V. (2003), ‘Support vector regression machines’, *Advances in Neural Information Processing Systems* 9.
- Friedman, J., Hastie, T. and Tibshirani, R. (2017), *The Elements of statistical learning, Data Mining, Inference, and Prediction*, Springer.
- Hettema, M. H., Andrews, J. S., Papamichos, E. and Blaasmo, M. (2006), *The relative importance of drawdown and depletion in sanding wells: Predictive models compared with data from the staffjord field*, in ‘Proceedings of the SPE International Symposium and Exhibition on Formation Damage Control, Lafayette, 15 – 17 February 2006’, Society of Petroleum Engineers.
- Krishnamoorthy, D., Foss, B. and Skogestad, S. (2016), ‘Real-time optimization under uncertainty applied to a gas lifted well network’, *Processes* 4, 52.
- Si, X., Wang, W., Hu, C., Zhou, D. and Pecht, M. G. (2012), ‘Remaining useful life estimation based on a nonlinear diffusion degradation process’, *IEEE Transactions on Reliability* 61(1), 50–67.
- The MathWorks, I. (2020), *Statistics and Machine Learning Toolbox*, Natick, Massachusetts, United State. URL: <https://www.mathworks.com/help/stats/>
- Verheyleweghen, A. and Jäschke, J. (2018), ‘Oil production optimization of several wells subject to choke degradation’, *IFAC-PapersOnLine* 51(8), 1–6.

A Metamodelling Approach for Air Pollution Source Estimation

Kuang Cheng^a, Yi Cao^{a,b}, Shuanghua Yang^{a,b}

^a*Zhejiang University, Hangzhou, Zhejiang Province, China 310027*

^b*Institute of Zhejiang University-Quzhou, Quzhou, Zhejiang Province, China 324000*

Abstract

The fast and accurate identification of air pollutant release effectively improved the emergency management for fine chemical industrial parks. The pollutant emission parameters could be estimated with the observed pollutant and meteorology data using an air dispersion model. Most of the fast response source term estimation methods are based on the Gaussian plume model for the low computation complexity. The concentration profile estimated by the Gaussian plume model can hardly represent the observation result affected by the complicated terrain with buildings in dense industrial area. Besides, the dispersion models with well characterized building profile like computation fluid dynamic (CFD) are time consuming.

In this paper, a metamodelling method with downwash effect of buildings in the domain to estimate the air pollutant emission is proposed. The dispersion model (Aermod) with combined plume rise and building downwash (PRIME) model preferred by U.S. Environmental Protection Agency is used to establish an emission inventory which covers most possible release scenarios in a dedicated area such as a chemical industrial park. The initial emission inventory is then fed to a neural network. The emission inventory will be updated as well as the neural network with an increasing number of scenarios until the trained neural network meets the pre-set criteria. The well-trained neural network is then used to predict the source emissions with observed data collected in a fine chemical industrial park.

A case study is performed with a localized Aermod model with geometry and surface character data collected from an industrial park. The results show that the building down wash effect in the industrial parks is responsible for some source estimating error produced by the inverse Gaussian dispersion model. It shows the advance of time consumption and competitive prediction accuracy compared to typical least square inversion methods.

Keywords: Source Term Estimation, Metamodelling, Aermod, Air Pollution, Emergency Management

1. Introduction

Chemical industrial parks gather a large number of industrial enterprises in limited space, challenging daily and emergency management for air pollution control. The development of emission determining has attracted great attention recently.

There are two major types of source tracing methods for chemical industrial parks: (1) the receptor-based model source tracing such as chemical mass balance (CBM) and molecular markers (MM) and (2) the source-based model source tracing such as least square (LS) technique and Bayesian inference-based methods e.g. Markov chain Monte Carlo (MCMC). CMB and MM have been proved with a better performance for source

tracing via studies implemented in a number of locations in USA (Schauer and Cass, 2000), while the prior knowledge of the source profile required by CBM and MM is hardly to be satisfied due to the complexity of pollutant composition and confidentiality (Jaeckels et al., 2007). The source-based model source tracing method such as LS has an important role in various of scenarios. Concentration measurements from sensors can be used for source terms estimation (STE). With the application of sensors measuring the concentration of relative pollutant the source term in the dispersion model can be estimated with the measured data from sensors with regard to source term estimation (STE) (Xu and Xu, 2017). Issartel (2003) used an inverse dispersion model to address the reconstruction of space-time geometry without statistical frame. An experiment data with released tracing gas measured with 14 stations in a characteristic spacing of 500 km was used for evaluation purposes. Lushi and Stockie (2010) developed a STE method based on the inverse Gaussian plume dispersion model. With the meteorology and deposition data from a lead-zinc smelting operation they estimated the total emission and produced a reasonable result. Other than optimization based methods, the Bayesian inference based methods of STE provide the probabilistic considerations for uncertainty analysis with not only one solution but a probability density function (Xu and Xu, 2017).

Apart from the algorithms used for STE problems, the performance of dispersion model is critical for source tracing. Yee et al. (2014) addressed the STE problems with Bayesian inference framework using different dispersion models indicating the significance of model errors caused by the difference between the dispersion model and the fact in real situation (Yee et al., 2014). A limitation of the Gaussian dispersion model (Wawrzynczak et al., 2014) is that Gaussian dispersion model can hardly estimate the source strength simulated by a Gaussian puff model (SCIPUGG).

American Meteorological Society (AMS) and U.S. Environmental Protection Agency (EPA) Regulatory Model (Aermod) enhanced the performance of Gaussian model not only by characterizing the vertical structure of wind profile but also describing the effect of building downwash with incorporated plume rise enhancement (PRIME) algorithms (Perry et al., 2005; Schulman et al., 2000). Although the model performance has been evaluated by many researchers, there are rare applications to STE problems due to the relative high computation load compared with Gaussian dispersion model.

Artificial neural network (ANN) based metamodel has been widely used for nonlinear mapping for the ability to quickly learn from data while limited applications in air pollution source tracing are proposed. In this paper, to overcome the problem of Aermod to be used for STE, an ANN based metamodel is adopted to allow the STE problems to be solved with less computation time but maintaining a reasonable degree of accuracy. The rest of the paper is organized as follows, Section 2 describes the problems and methods used in this paper in detail, Section 3 introduces the numerical experiment setup followed by Section 4 and 5 with result discussions and conclusions respectively.

2. Methodology

2.1. Problem description

Based on the advection-diffusion equation, a point source in location $\mathbf{x}(x, y)$ with releasing rate q produces concentrations measurement c at receptors with coefficient

$a(\mathbf{x})$. Considering n sources with unknown emission rates $\mathbf{q} \in \sim^n$, the concentration measurements at m receptors $\mathbf{c} \in \sim^m$ are dependent on emission rate linearly as follows:

$$\mathbf{c} = \mathbf{A}\mathbf{q} + \boldsymbol{\varepsilon} \tag{1}$$

$$\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

Where matrix $\mathbf{A} \in \sim^{m \times n}$ is a sensitive matrix and its value is related to \mathbf{x} . The vector $\boldsymbol{\varepsilon} \in \sim^m$ represents noise and model errors. The problem is now to estimate \mathbf{q} and \mathbf{x} to minimize Euclidean norm:

$$J = \frac{1}{2} \sum_{i=1}^m \varepsilon_i^2 = \frac{1}{2} \sum_{i=1}^m (c_i - \sum_{j=1}^n a_{ij} q_j)^2 = \frac{1}{2} (\mathbf{c} - \mathbf{A}\mathbf{q})^T (\mathbf{c} - \mathbf{A}\mathbf{q}) \tag{2}$$

For source locations \mathbf{x}_k , with fixed unit emission rate \mathbf{q}_k , coefficient matrix \mathbf{A}_k can be calculated with an atmospheric dispersion model. Source strength \mathbf{q}_k and $\boldsymbol{\varepsilon}_k$ can be estimated with calculated \mathbf{A}_k and concentration measurement \mathbf{c} .

$$\mathbf{q}_k = \mathbf{A}_k^+ \mathbf{c}, \quad \boldsymbol{\varepsilon}_k = (\mathbf{I} - \mathbf{A}_k \mathbf{A}_k^+) \mathbf{c} \tag{3}$$

To minimize $J_k = \frac{1}{2} \boldsymbol{\varepsilon}_k^T \boldsymbol{\varepsilon}_k$, a necessary condition is satisfied for a stationary point:

$$\frac{\partial J_k}{\partial \Delta \mathbf{x}_k^T} = \left(\frac{\partial \boldsymbol{\varepsilon}_k}{\partial \Delta \mathbf{x}_k} \right)^T \boldsymbol{\varepsilon}_k = 0 \tag{4}$$

The location \mathbf{x}_k at stationary point indicates the locations of the unknown sources.

After the estimation of source location, the sensitive matrix \mathbf{A} is available and the source strength \mathbf{q} for unknown source can be estimated according to Eq. (3).

There are a number of optimization algorithms in the literature to estimate the variable \mathbf{q} and \mathbf{x} such as Levenberg-Marquardt (Marquardt, 1963) and Newton-Raphson (Loke and Barker, 1996). These methods are diverse with iterative techniques and initial values of target parameters.

The iterative techniques may be time consuming. One of reasons that the Gaussian model is widely used for STE problems is the balanced performance and computation cost. Different from Gaussian model based least square method, the iteration of Aermოდ model is extraordinarily time consuming that hardly to be used for STE problems, particularly for emergency emission source tracing. To overcome this deficit, an Aermოდ model-based ANN metamodelling (ANN-Aermოდ) approach is proposed.

2.2. Source terms estimation using ANN

This metamodelling approach is not simply to replace Aermოდ to make the computation fast, but actually to solve it as an inverse model, hence to replace the LS optimization as well.

The methodology is essentially realized in three steps. In the first step, a dedicated industrial area, the Aermოდ model localized with parameters like albedo, Bowen ratio, surface roughness and terrain elevation, is used to generate emission scenario inventory. The inventory contains full cross design samples to cover the entire domain dimension by dimension.

In the second step, the samples in the inventory are passed to a typical neural network. In terms of the input and output of the ANN, the input includes:

(1) a set of concentration \mathbf{c} simulated to represent the values recorded by the measurement system which dimension is depends on the number of sensors in the industrial park (2) wind speed v and wind direction d (3) sources emission rate q (4) ambient temperature T . Other than these variables, the Aermოდ model requires more to process the data, namely, location of measurement system, albedo, Bowen ratio and surface roughness. These variables are typically unchanged, therefore, we do not need to consider these parameters in the ANN model. The output of the ANN is three-dimensional set $\{x, y, q_n\}$ corresponding to the source locations in x and y -axis coordinate and source emission rate q_n . The input dataset IN and output dataset OU are described as:

$$IN = \{c_m, v_m, d_m, T_m\}_{m=1}^{M_i}, \quad OU = \{x_n, y_n, q_n\}_{n=1}^{N_i} \quad (5)$$

Where M_i is number of concentration measurements and N_i is number of source locations.

The third step is to estimate the source locations and emission rate with ANN. The sampled measurements dataset $D_i = \{c_i, v_i, d_i, T_i\}_{i=1}^{M_i}$ is the input of the ANN and the output is source locations and emission rate: $D_o = \{x_\tau, y_\tau, q_\tau\}_{\tau=1}^{N_i}$.

Table 1 Range of variation of variables considered in source emission inventory

x_n, y_n	q_m	u_m	d_m	T_m
2 sources	min: 0[g/s]	min: 1[m/s]	min: 10°	min: 0°C
	max: 1[g/s]	max: 3[m/s]	max: 360°	max: 30°C

3. Experimental study

A numerical experiment is undertaken to evaluate the performance of features of the proposed method. The Aermოდ model is localized with building geometry, measurement locations, terrain elevation and surface roughness data collected from a chemical industrial park in Shangyu, China whose layout is demonstrated in Figure 1.

There are 11 receptors around the industrial park denoted by triangles in Figure 1. The circles are potential two source locations corresponding to the highly risk area. Other variables used to generate the inventory and their value are shown in .

Table 1. In this work, a scenario inventory with the total 1.3×10^5 records is processed for training. A back propagation neural network with three hidden layers is used to correlate the input and output. For performance evaluation purposes, a numerical

simulation is undertaken using the least square method with either Gaussian model (LS-Gaussian) or Aermod model (LS-Aermod). Addition to the test data for three methods, initial guess is required for LS-Gaussian and LS-Aermod method.

4. Results and discussion

The numerical simulation is implemented using Aermod (version 19191) running on a computer with Intel Core i5 (3.00GHz), 24.0GB RAM. The emission scenario data collection used for performance evaluation is not included in the training data. The prediction results are shown in Table 2. The test data simulated an emission in location (-

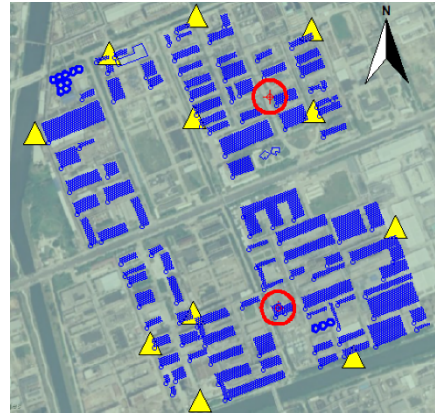


Figure 1 The geometry of an industrial park in Shangyu

175.00, 640.00), denoted by the circle in the south of the industrial park in Figure 1, is implemented. In Table 2, the computation time 0.402s indicating the advance of metamodelling method compared to 29.157s (LS-Gaussian) and 4123.617s (LS-Aermod). The location error of LS-Gaussian is 801.22 m, 3.50 m for LS-Aermod and 9.48 m for the ANN-Aermod method. The details of true and estimated source locations are shown in fourth and fifth column in Table 2. The results of LS-Gaussian and LS-Aermod indicated the Aermod model provided a different concentration profile in building complexity area compared to the Gaussian model. The building downwash effect plays an important role in chemical industrial parks.

Although the LS-Aermod method demonstrated less retrieval error, the relatively high time consumption is hardly to be accepted for emergency emission estimation. The LS-Gaussian method costs a ratio of 0.7% computation time of LS-Aermod, while the high estimation error of location is away of the region of high-risk area shown as circles in Figure 1. The ANN-Aermod model comparing with others methods, cost a ratio of 0.001% computation time of LS-Aermod method, with an estimation error ratio of 1.18% of LS-Gaussian method, could process a valuable result with the least time consumption.

Table 2 Numerical experiment using test data simulated with Aermod model. The first column shows the source tracing method. The second column is total computation time for all trials. The third column is Euclidian distance E_L corresponding to location estimation error. The fourth and fifth column show the true release location and retrieved location respectively.

Method	Computation Time(s)	E_L (m)	Release Location	Retrieval Location
LS-Gaussian	29.157	801.22	(-175.00, 640.00)	(-157.12, -161.02)
LS-Aermod	4123.617	3.50	(-175.00, 640.00)	(-171.50, 640.00)
ANN-Aermod	0.402	9.48	(-175.00, 640.00)	(-175.00, 647.32)

5. Conclusion

This paper proposed an efficient and accurate source term estimation technique for chemical industrial parks. The prediction error ratio is 1.18% to Gaussian model based least square method and the computation time ratio is 0.7% of Aermod model-based method. The accurate prediction of source location shows the competitive performance of source tracing in industrial parks with complex constructions. Thus, accurate predictions can be made in a limited time, providing plenty of valuable time for emergency responses.

This source tracing method runs free of initial guess of source locations and therefore provides a more practical solution for source tracing in emergencies.

Acknowledgements

This work was supported by National Key Research and Development Project of China (No.2018YFC0214102)

References

- J. J. Schauer and G. R. Cass, "Source apportionment of wintertime gas-phase and particle-phase air pollutants using organic compounds as tracers," *Environ. Sci. Technol.*, 2000, doi: 10.1021/es981312t.
- J. M. Jaekels, M. S. Bae, and J. J. Schauer, "Positive matrix factorization (PMF) analysis of molecular marker measurements to quantify the sources of organic aerosols," *Environ. Sci. Technol.*, 2007, doi: 10.1021/es062536b.
- J. Xu and L. Xu, "A review of source term estimation methods for atmospheric dispersion events using static or mobile sensors," *Integr. Syst. Heal. Manag.*, vol. 36, pp. 101–156, 2017, doi: 10.1016/b978-0-12-812207-5.00003-1.
- J. P. Issartel, "Rebuilding sources of linear tracers after atmospheric concentration measurements," *Atmos. Chem. Phys.*, 2003, doi: 10.5194/acp-3-2111-2003.
- E. Lushi and J. M. Stockie, "An inverse Gaussian plume approach for estimating atmospheric pollutant emissions from multiple point sources," *Atmos. Environ.*, vol. 44, no. 8, pp. 1097–1107, 2010, doi: 10.1016/j.atmosenv.2009.11.039.
- E. Yee, I. Hoffman, and K. Ungar, "Bayesian Inference for Source Reconstruction: A Real-World Application," *Int. Sch. Res. Not.*, vol. 2014, no. 1, pp. 1–12, 2014, doi: 10.1155/2014/507634.
- A. Wawrzynczak, P. Kopka, and M. Borysiewicz, "Sequential Monte Carlo in Bayesian assessment of contaminant source localization based on the sensors concentration measurements," 2014, doi: 10.1007/978-3-642-55195-6_38.
- S. G. Perry et al., "AERMOD: A Dispersion model for industrial source applications. Part II: Model performance against 17 field study databases," *J. Appl. Meteorol.*, vol. 44, no. 5, pp. 694–708, 2005, doi: 10.1175/JAM2228.1.
- L. L. Schulman, D. G. Strimaitis, and J. S. Seire, "Development and evaluation of the prime plume rise and building downwash model," *J. Air Waste Manag. Assoc.*, vol. 50, no. 3, pp. 378–390, 2000, doi: 10.1080/10473289.2000.10464017.
- D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *J. Soc. Ind. Appl. Math.*, 1963, doi: 10.1137/0111030.
- M. H. Loke and R. D. Barker, "Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method," *Geophys. Prospect.*, 1996, doi: 10.1111/j.1365-2478.1996.tb00142.x.

Uncertainty analysis in life-cycle assessment of early-stage processes and products: a case study in dialkyl-imidazolium ionic liquids

Husain Baaqel,^a Jason P. Hallett,^a Gonzalo Guillén-Gosálbez^b and Benoît Chachuat^{a,*}

^a*Department of Chemical Engineering, Sargent Centre for Process Systems Engineering, Imperial College London, London, United Kingdom*

^b*Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 1, Zürich, Switzerland*

Abstract

This paper presents a methodology for combining foreground and background uncertainty in the life-cycle assessment (LCA) of processes and products at a low technology-readiness level. We compare the LCA of two ionic liquids, 1-butyl-3-methyl-imidazolium tetrafluoroborate [bmim][BF₄] and 1-butyl-3-methyl-imidazolium hexafluorophosphate [bmim][PF₆]. The nominal scenario predicts that [bmim][BF₄] generates lower end-point environmental impacts than [bmim][PF₆]. However, the uncertainty ranges around these nominal predictions overlap significantly, with [bmim][BF₄] causing higher impacts than those of [bmim][PF₆] in up to 30% of the uncertainty scenarios. On top of this, accounting for uncertainty in the foreground data more than doubles the estimated impact ranges in several damage categories. This case study, therefore, demonstrates the need for combining foreground and background data uncertainty for more reliable life-cycle assessments.

Keywords: uncertainty analysis, life cycle assessment, ionic liquids, process simulation

1. Introduction

Life-cycle assessment (LCA) is a popular methodology to quantify the environmental impacts of products and processes throughout their life cycle (Guinée and Heijungs, 2017). It entails the collection of so-called inventory data on mass and energy flows, from raw material extraction to process emissions and wastes. In many applications, however, these data may be lacking because of confidentiality or low technology readiness (Skoronski et al., 2020).

Various approaches have been proposed to bridge the gap in inventory data. One such approach entails the development of detailed process models in order to predict the performance at scale of processes at a low technology-readiness level (TRL), for which industrial process data are yet unavailable (Hetherington et al., 2014). However, these process models can themselves be subject to large uncertainty (Van der Spek et al., 2017). It is important, therefore, to quantify these uncertainties and propagate them into the predicted inventories, and ultimately the predicted impacts. Despite this, most LCA studies based on process simulation simply omit the effect of uncertainty, or consider only the uncertainty in the background data which comprise energy and material flows

delivered to the foreground system as aggregated data sets whereby individual plants and operations are not identified (Baaqel et al., 2020; Llantoy et al., 2020). A more rigorous uncertainty analysis calls for including uncertainty on the foreground data alongside background uncertainty, that is, uncertainty from the system of primary concern to the analysis including process operating conditions and thermophysical properties. This is especially important when comparing processes with similar performance indicators where the corresponding uncertainty ranges could indeed overlap significantly.

This paper presents a methodology for combining foreground and background uncertainty in the LCA of such processes and products at a low TRL. The approach is illustrated with the case study of two dialkylimidazolium ionic liquids (ILs), 1-butyl-3-methylimidazolium tetrafluoroborate [bmim][BF₄] and 1-butyl-3-methylimidazolium-hexafluorophosphate [bmim][PF₆]. Both ILs find many applications, for instance as physical separation media (Riisagera et al., 2006; Shiflett and Yokozeki, 2010) where their separation solvating capacity and negligible vapor pressure are advantageous. The current low TRL of ILs justifies the use of process flowsheeting to simulate their production at scale and, therefore, the consideration of foreground data uncertainty. The methodology is described next (Section 2), before introducing the case study (Section 3) and presenting some results (Section 4) and conclusions (Section 5).

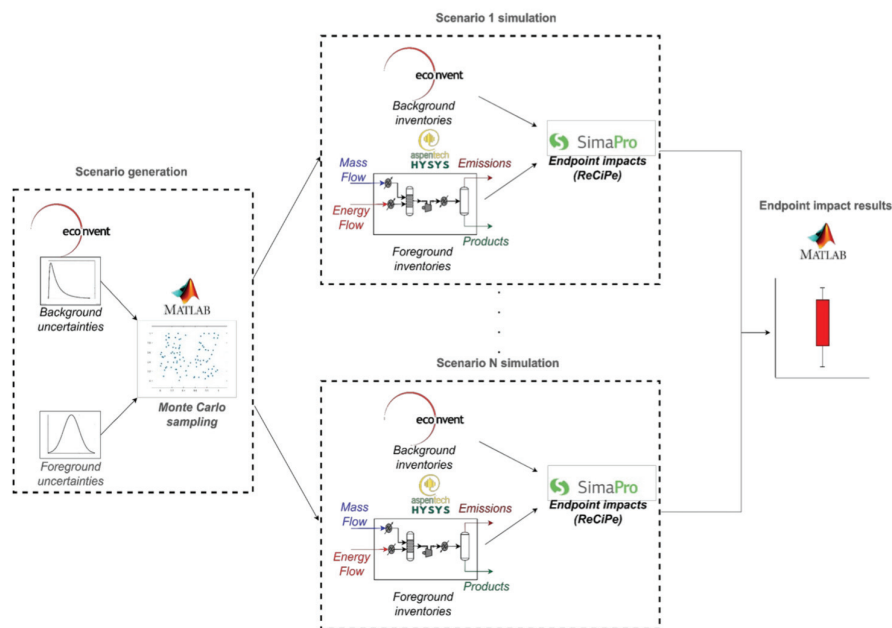


Figure 1: Methodological framework for uncertainty analysis in LCA.

2. Methodology

The proposed methodological framework (Figure 1) combines uncertainty modeling, Monte Carlo sampling, process simulation, and environmental assessment. For each uncertainty realization, the uncertainty is propagated through both the background and

foreground inventories, and ultimately to the environmental damages. These activities are conducted with the state-of-the-art tools Aspen-HYSYS and SimaPro and coordinated using Matlab. They are further detailed below.

2.1. Environmental assessment

The LCA follows the four phases defined in the ISO 14040 standards: (i) goal and scope, (ii) inventory analysis, (iii) impact assessment, and (iv) interpretation. Choices about the system, including the scope, the boundaries of the foreground system and the functional unit, are made in the first phase. Environmental flows corresponding to the system inputs and outputs are collected during the inventory phase, including raw materials, energy streams, emissions and wastes. Inventories for the background processes can rely on databases, such as ecoinvent or GaBi. In contrast, the foreground processes may not be found in such databases due to confidentiality or low technology readiness, but they may instead be predicted using detailed process flowsheeting, such as Aspen-HYSYS or gPROMS. Both the foreground and background inventories are finally translated into environmental impacts during the impact assessment phase through a characterization method.

2.2. Background and foreground uncertainty modeling

A common approach to quantifying the uncertainty in background inventories involves the so-called Pedigree matrix (Weidema and Wesnæs, 1996). Scores between 1–5 are assigned to the inventory data based on five criteria: reliability, completeness, temporal, geographical and technological differences. These scores are combined with a basic uncertainty factor to determine the standard deviation of a log-normal distribution for each environmental flow. In contrast, the foreground uncertainty corresponds to uncertain process parameters, including operating conditions such as pressures and temperatures, reaction rates and conversions, and thermophysical properties such as boiling point and density. These uncertain process parameters can be propagated to the mass and energy flows using the process flowsheet model.

Unlike the background uncertainty which can be retrieved from LCA databases, the main sources of uncertainty in the foreground processes need to be characterized. These may be related to experimental errors in lab-scale procedures, or else inferred from expert opinions when process scale-up is involved. Such uncertainty may furthermore be expressed as parameter ranges or probability distributions.

2.3. Uncertainty propagation

Monte Carlo sampling is used to construct a set of scenarios from the combined foreground and background uncertainty. The number of scenarios is determined automatically based on a user-defined error tolerance between the sample mean and the actual uncertainty mean value (Law and Kelton, 2000). The foreground and background inventory flows for each uncertainty realization are computed using the process model and the inventory database, respectively. In turn, the inventory flows for each scenario are converted to environmental impacts, after normalization by the functional unit and application of a characterization method:

$$C_j = \sum_{i=1}^N f_i b_{ij}$$

where C_j is the predicted endpoint impact j of the scenario in units of environmental damage; f_i is the foreground output in the form of mass and energy flow i under a given uncertainty scenario in units of mass or energy; and b_{ij} is the characterized normalized

background output under a given uncertainty scenario in units of environmental damage of endpoint impact j per unit of mass or energy of flow i with N the number of flows into and out of the process. The scenario generation and evaluation, and the final aggregation into impact uncertainty ranges, are orchestrated in Matlab.

3. Case Study Definition

3.1. Synthesis of Ionic Liquids

Our case study compares the performance of two dialkyl-imidazolium ionic liquids: 1-butyl-3-methylimidazolium tetrafluoroborate [bmim][BF₄] and 1-butyl-3-methylimidazolium hexafluorophosphate [bmim][PF₆]. Both ILs are produced using the same process, for which we developed process models using Aspen-HYSYS to scale-up experimental synthesis procedures.

The synthesis proceeds via the exchange of ions between 1-butyl-3-methylimidazoliumchloride [bmim][Cl] and the corresponding salt. The salt used for [bmim][BF₄] is sodium tetrafluoroborate and for [bmim][PF₆], lithium hexafluorophosphate. The products are the ionic liquid and the chloride salt. The scaled-up process starts by reacting a solution of [bmim][Cl] with excess salt under atmospheric conditions. The mixture is then separated into (i) an upper phase containing the separated IL with some impurities and (ii) a lower phase containing solid chloride salt. The upper phase is washed using the corresponding fluorinated salt to remove impurities before removing excess water from the IL in a vacuum distillation column.

3.2. Life-Cycle Assessment and Uncertainty Analysis

We adopt a cradle-to-gate scope that encompasses all stages from raw material extraction to the final product synthesis. The geographic location for the assessment is Europe. Since chemicals are commonly sold by weight and because the scope of our work is limited to the production phase only, the functional unit is set to “1 kg of ionic liquid”. Background inventory data are collected from the ecoinvent v3.5 database (Wernet et al., 2016) interfaced through SimaPro v9.0. The foreground production processes for both ILs and their precursors are unavailable in ecoinvent and thus modeled using Aspen-HYSYS v9. Thermophysical properties such as density and critical properties are obtained from experiments or from the literature. Other properties such as heat of formation are estimated using group contribution methods (Valderrama and Rojas, 2009). The rest of the properties are estimated from the molecular structure using the Property Constant Estimation System (PCES) in Aspen-HYSYS. The predicted background and foreground inventories in each uncertainty scenario are converted into 17 mid-point impacts using the ReCiPe 2016 characterization method (Huijbregts et al., 2016). Then, they are further combined into 3 damage categories (end-points): human health, ecosystems quality and resources.

We consider a total of 9 uncertain parameters in the (foreground) process model, 5 of which correspond to unit operating conditions (e.g., temperature, pressure) and the other 4 are thermophysical properties (e.g., density, heat of formation). All of these parameters are assumed to follow a normal distribution, with standard deviations estimated based on literature data or from experience. The number of foreground uncertainty scenarios was in the range between 3,000–5,000 for the various runs.

4. Results and Discussions

Figure 2 shows the box plots for each end-point indicator, where [bmim][BF₄] and [bmim][PF₆] are compared under the following scenarios: (A) combined foreground and

background uncertainty, (B) foreground uncertainty only, and (C) background uncertainty only. The central line inside each box represents the median scenario; the upper and lower ends of the box represent the 75% and 25% quartiles, respectively; and the lower and upper extended lines represent the minimum and maximum values, respectively.

Comparing both ILs in terms of their median scenarios only, one could conclude that the production of [bmim][BF₄] presents lower environmental impacts than [bmim][PF₆] in all damage areas. But the box plots show a different reality, whereby the range of impacts of both ILs overlap significantly. All uncertainty considered (scenario A), the damages caused by [bmim][BF₄] on human health (left plot), ecosystems quality (middle plot), and resources (right plot) are higher than those caused by [bmim][PF₆] in 20%, 15% and 30% of the uncertainty scenarios, respectively. This overlap is significantly larger than under the traditional approach of considering solely the background uncertainty (scenario C), where the damages caused by [bmim][BF₄] on human health, ecosystems quality and resources are higher than those of [bmim][PF₆] in 8%, 5% and 20% of the scenarios. Therefore, adding the foreground uncertainty to the background uncertainty (scenario A) is necessary for a more reliable comparative assessment of these two ionic liquids.

Notice that [bmim][BF₄] presents lower impacts on human health, ecosystems quality and resources in nearly all of the scenarios when considering the foreground uncertainty alone (scenario B). But even though these contributions could appear low in comparison to their background uncertainty counterparts, they result in 2 to 3-times larger interquartile ranges for the environmental impacts under combined foreground/background uncertainty propagation. This is mainly due to the multiplicative effect of the foreground and background uncertainty which are linked via the mass and energy flow inventories.

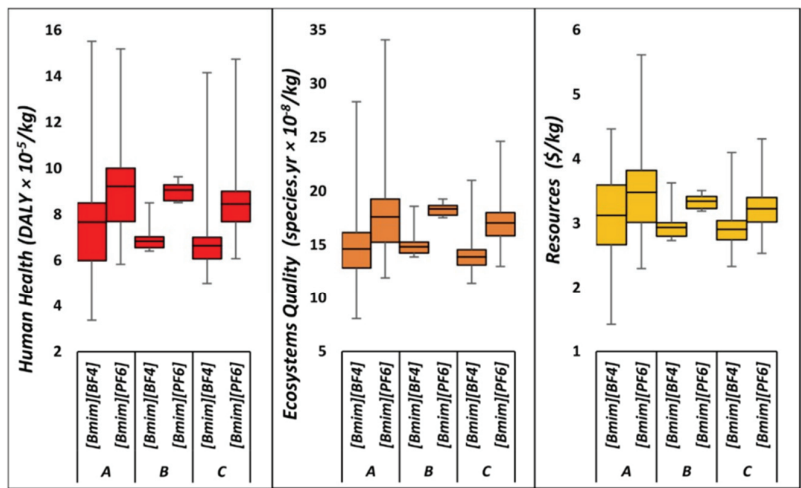


Figure 2: Comparison of end-point indicators under combined foreground/background uncertainty (A), foreground uncertainty only (B), and background uncertainty only (C).

5. Conclusions

A methodology was proposed for the environmental assessment of emerging technologies that combines process flowsheeting with LCA and propagates the uncertainty in both

background and foreground processes to the predicted environmental impacts. This methodology was illustrated with the case study of two dialkyl-imidazolium ILs, [bmim][BF₄] and [bmim][PF₆]. Accounting for the foreground uncertainty was found to increase the uncertainty ranges around all three end-point impact indicators significantly, in comparison with a classical assessment that would only consider the background uncertainty. Future research should, therefore, consider the potential effect of foreground uncertainty more carefully, especially in the comparative LCA of early-stage processes and products, in order to enable more robust decision making.

Acknowledgements: HB is grateful to Saudi Aramco for awarding him a PhD Scholarship. JHP would like to acknowledge the Supergen Bioenergy Hub of the Engineering and Physical Science Research Council (EP/S000771/1).

References

- Baaqel, H., Díaz, I., Tulus, V., Chachuat, B., Guillén-Gosálbez, G., Hallett, J.P., 2020. Role of life-cycle externalities in the valuation of protic ionic liquids – a case study in biomass pretreatment solvents. *Green Chemistry* 22, 3132–3140.
- Guinée, J., Heijungs, R., 2017. Introduction to Life Cycle Assessment, in: *Sustainable Supply Chains: A Research-Based Textbook on Operations and Strategy*. pp. 15–41.
- Hetherington, A.C., Borrión, A.L., Griffiths, O.G., McManus, M.C., 2014. Use of LCA as a development tool within early research: challenges and issues across different sectors. *The International Journal of Life Cycle Assessment* 19, 130–143.
- Huijbregts, M.A.J., Steinmann, Z.J.N., Elshout, P.M.F., Stam, G., Verones, F., Vieira, M.D.M., Hollander, A., Zijp, M., Van Zelm, R., 2016. ReCiPe 2016: A harmonized life cycle impact assessment method at midpoint and endpoint level. *The International Journal of Life Cycle Assessment*.
- Law, A.M., Kelton, W.D., 2000. *Simulation Modeling and Analysis*. McGraw-Hill.
- Llantoy, N., Chàfer, M., Cabeza, L.F., 2020. A comparative life cycle assessment (LCA) of different insulation materials for buildings in the continental Mediterranean climate. *Energy and Buildings* 225, 110323.
- Riisager, A., Fehrmanna, R., Haumann, M., Wasserscheid, P., 2006. Supported ionic liquids: versatile reaction and separation media. *Topics in Catalysis* 40, 91–102.
- Shiflett, M.B., Yokozeki, A., 2010. Separation of CO₂ and H₂S using room-temperature ionic liquid [bmim][PF₆]. *Fluid Phase Equilibria, Ionic Liquids Special Issue* 294, 105–113.
- Skoronski, E., Fernandes, M., Malaret, F.J., Hallett, J.P., 2020. Use of phosphonium ionic liquids for highly efficient extraction of phenolic compounds from water. *Separation and Purification Technology* 248, 117069.
- Valderrama, J.O., Rojas, R.E., 2009. Critical properties of ionic liquids. Revisited. *Industrial & Engineering Chemistry Research* 48, 6890–6900.
- Van der Spek, M., Sanchez Fernandez, E., Eldrup, N.H., Skagestad, R., Ramirez, A., Faaij, A., 2017. Unravelling uncertainty and variability in early stage techno-economic assessments of carbon capture technologies. *International Journal of Greenhouse Gas Control* 56, 221–236.
- Weidema, B.P., Wesnæs, M.S., 1996. Data quality management for life cycle inventories—an example of using data quality indicators. *Journal of Cleaner Production* 4, 167–174.
- Wernet, G., Bauer, C., Steubing, B., Reinhard, J., Moreno-Ruiz, E., Weidema, B., 2016. The ecoinvent database version 3 (part I): overview and methodology. *The International Journal of Life Cycle Assessment* 21, 1218–1230.

Interval-Valued Data Learning for Robustness of Energy Recovery Systems

Rachid Ouaret^{a,*}, Pascal Floquet^a, Jean-Pierre Belaud^a and Stéphane Negny^a

^a*Chemical Engineering Laboratory, Université de Toulouse, CNRS, Toulouse, France, LGC UMR 5503, 4 allée Emile Monso, 31030 Toulouse*

rachid.ouaret@toulouse-inp.fr

Abstract

Modeling of energy recovery systems, Heat Exchanger Networks (HEN) as an example, are complex processes due to many parameters involved and the lack of knowledge on the impact of operational variables on the response. These variables are often affected by different types of uncertainty as to measurement errors, computation errors, or imprecision related to the underlying method. The uncertainty in the data may be treated by considering, rather than a single value for each variable, the interval of values in which it may fall, histograms, or other multivalued data: symbolic data. This work aims to use, when possible, the symbolic data analysis to adapt the classical mathematical HEN models. It deals with the study of continuous interval data through suitable Principal Component Analyses and Regression for two purposes: clustering exchanges (i) classification of exchangers to detect those impacted by uncertainty factors and (ii) evaluation of the relationship between the different process parameters (inlet temperature, heat transfer coefficient, etc.) on interval data. The new method has been tested on a real data set and the numerical results are reported. The symbolic approach provides a simple way to study a great number of scenarios.

Keywords: Machine Learning for Symbolic Data, Interval-valued data, Flexibility and Robustness, Heat Exchanger Networks (HEN)

1. Introduction

Energy saving is an important issue for both industries and society. In the industrial chemical process, Heat Exchanger Networks (HEN) are widely used techniques for reducing external heating and cooling utilities. Data generated by those complex systems has increased drastically over the past few years. Suppose we have 20 Heat Exchangers (Hx) on which 4 variables are measured (2 input and 2 output hot and cold input streams). If we assume that we have hourly values of these variables for 1 year, then each exchanger is described by 35040 values. If, on the other hand, certain characteristics of the exchangers (let's say 3) vary over time, then the 3D array representation conventionally used in data analysis is then $8760(\text{hours}) \times 20(\text{Hx}) \times 7$ (parameters). This is not very large compared to what could be provided by the industry in real-time.

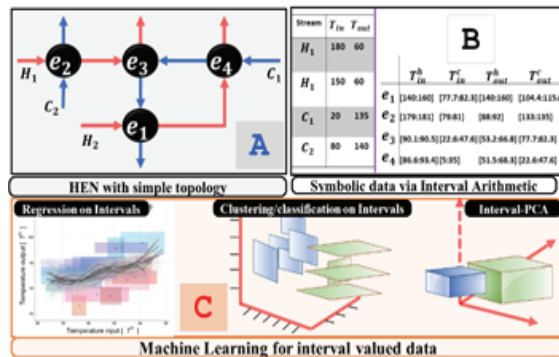
Even such data is ubiquitous for a larger scale of HEN systems, data-oriented based approaches are often analyzed with simplified models by aggregating data. During this operation, some information on the variability aspects is lost. The robustness assessment of HEN is therefore affected since its flexibility is conditioned by the variability of the uncertain parameters. When the size n of entries (exchangers) and p number of variables (features) are very large, classical analysis can be problematic.

To address this problem, we use a more complex object description to capture the variability of measured parameters on each exchanger. When dealing with quantitative variables, complete information can be achieved by describing a set of statistical units in terms of interval data, histogram, rather than a single-valued variable. Mathematically, interval-valued data with measurements on p random variables, are p -dimensional hyperrectangles in R^p . Such data need to be visualized, synthesized, and compared on factor spaces.

This paper is in a complementary perspective to Floquet et al's work (Floquet et al., 2016). They have initiated the robustness analysis of a simple exchange networks using interval arithmetic. They pointed out the "butterfly effect" of the alteration of characteristics of some Hx on the operation of the HEN. Indeed, a maintenance operation of an exchanger, alter the variability of the other parameters of the network. Reducing fouling would imply variations in pressure and shift flows between parallel branches, all changing over time in a way that is difficult to predict (Macchietto et al., 2018). The purpose of this paper is to check whether there are groups of exchangers characterized by the same properties in terms of their responses to external fluctuations. Following the same idea (interval-valued data), the Symbolic Data Analysis (SDA) (Billard & Diday, 2006; Bock & Diday, 2012) is used to study how uncertainty in the output of a model can be apportioned to different sources of uncertainty.

2. Machine Learning for symbolic data

Figure 1 outlines the design and methodological scheme of the proposed method. This methodology draws on two main components: (B) Symbolic data through interval arithmetic, and (C) machine learning for symbolic data. Part A corresponds to the HEN model to be studied or simulated. The interval-valued data is then constructed in part (B). The core function of the



proposed method included in this paper is part (C). The originality of this research lies in the combination of a traditional robustness analysis of HEN with SDA.

2.1. Symbolic Data and statistics of interval-valued variables

In classical statistics, $n \times p$ data matrix $X = X_{ij}$ is defined between n individuals and p variables, where each

cell (i, j) contains a unique value x_{ij} . The symbolic objects are more complex than a simple valued variable description, symbolic data can contain internal variation of the features representing *imprecise knowledge* and can be structured. The symbolic analysis generalizes the classical data analysis, e.g. $x = c$, $c \in R$ is equivalent to the symbolic interval $\xi = [c, c]$. A full conceptualization of symbolic objects can be found in (Bock & Diday, 2012). Let Ω be a set of individuals, D containing the descriptions of individuals and the descriptions of classes of individuals, a , a mapping defined from Ω into D which associates to each $\omega \in \Omega$ a description $d \in D$ by using intervals, histograms, etc. More formally, symbolic object is a triplet $s = (a, R, d)$ where R is a relation between

descriptions, d is a description, and a is a mapping defined from Ω in L depending on R and d . For instance $L = true, false$ or $L = [0,1]$ and R may be one of the relations in $=, \equiv, \Rightarrow, \leq, \subseteq, \in$. To illustrate this point with a simple HEN example, if the outlet temperature of two exchangers ω_1, ω_2 is given by $T_{out}(\omega_1) = 90^\circ C, T_{out}(\omega_2) = 120^\circ C$ the description of the class $D = (\omega_1, \omega_2)$ obtained by a generalization process can be $[90,120]$. The symbolic object s is defined by a triple $s = (a, R, d)$ where $d = [90,120]$, $R = ' \in '$ and a is the mapping: $\Omega \rightarrow true, false$ such that a is the true value of $T_{out}(\omega)Rd$ written $a(\omega) = T_{out}(\omega) \in [90,120]$. An individual ω is in the extent of s if and only if $a(\omega) = true$. Simple statistical descriptions (mean, variance, ...) for interval-valued variables have been defined in (Bertrand & Goupil, 2000). Let consider $Y_j \equiv Z$ be the j^{th} interval-valued random variable, and $Z(\omega_u) = [a_u, b_u]$ is a realization of Z for the observation ω_u over the observed interval $[a_u, b_u]$. The empirical distribution function, $F_Z(\xi)$, is the distribution function of a mixture of m distributions $Z(\omega_u), u = 1, 2, \dots, m$. The central and dispersion parameters of a variable all derived from a strong assumption: the inherent fluctuation within random intervals and rectangles is uniformly distributed: $f(\xi) = \frac{1}{m} \sum_{u: \xi \in Z(u)} \left(\frac{1}{(b_u - a_u)} \right)$. The symbolic sample mean for interval-valued data is given by:

$$\underline{z} = \frac{1}{m} \sum_u \left(\frac{1}{(b_u - a_u)} \right) \quad (1)$$

and the sample variance is given by

$$S^2 = \frac{1}{3m} \sum_u (b_u^2 + a_u^2 + b_u a_u) \frac{1}{4m} \left[\sum_u (b_u + a_u) \right]^2 \quad (2)$$

2.2. PCA for interval-valued data

A principal component analysis is designed to reduce p -dimensional observations into s -dimensional components (where $s \ll p$) in an interpretable way, such that most of the information in the data is preserved. Let $i = 1, 2, \dots, m$ denote n objects (exchangers) described by p features (or variable) Y_1, Y_2, \dots, Y_p (temperatures, ...). The symbolic data matrix used for interval PCA is given by $\underline{X} = (\xi_{11} \dots \xi_{1p} \vdots \dots \xi_{n1} \dots \xi_{np})$ where $\xi_{ij} = [x_{ij}, x_{ij}]$ is the interval of possible values of variable j for the exchanger i , and the symbolic data vector can be denoted by $x_i = (\xi_{i1}, \dots, \xi_{ip}) = ([x_{i1}, x_{i1}], \dots, [x_{ip}, x_{ip}])$. The data point is represented in R^p space by hyperrectangles R_i with 2^p vertices. There are mainly two methods to solve the algebraic mapping to lower dimension: Vertices and Centers methods. To find the factorial axes for Centers method, a classical PCA is applied to the centers $c_i \in R^p$ of the n hyperrectangles R_i . The coordinates of i^{th} center c_i is denoted by x_{ij}^c , where x_{ij}^c is computed for $i = 1, \dots, n$ and $j = 1, \dots, p$. From that case, the centers of hyperrectangles in its rows is an $n \times p$ data matrix, denoted by X and the j^{th} column of X is denoted by the feature Y_j^c . The interval principal components values are obtained by computing first the v^{th} principal component of the center c_i is given by:

$$\psi_{iv}^c = \sum_{j=1}^p (x_{ij}^c - \underline{x}_{ij}^c) \cdot \eta_{iv} \quad (3)$$

where $\eta_{jv} = (\eta_{1v}, \dots, \eta_{pv})$ is the v^{th} of S (sample covariance matrix associated with the dataset). It is possible to find an $[\underline{\psi}_{iv}, \overline{\psi}_{iv}]$ of the possible values of v^{th} principal component ψ_{iv}^c of c_i (Cazes et al., 1997).

For the object i

$$\underline{\psi}_{iv} = \sum_{j=1}^p \max(x_{ij}^r - x_j^c) \cdot \eta_{jv} \quad \text{for} \quad \underline{x}_{ij} \leq x_{ij}^r \leq \overline{x}_{ij} \quad (4)$$

$$\underline{\psi}_{iv} = \sum_{j=1}^p \max(x_{ij}^r - x_j^c) \cdot \eta_{iv} \quad \text{for} \quad \underline{x}_{ij} \leq x_{ij}^r \leq \overline{x}_{ij} \quad (5)$$

3. Application to HEN data

3.1. Data simulation design and preliminary analysis

The starting point for this application is the data table from step B in Figure 1, which is the result of HEN model analysis by interval arithmetic. In order to evaluate the behavior of the exchanges on a more complex network, simulated data on the initial table were undertaken and the simulation procedure is as follows:

- For each variable $(T_{in}^h[\min_{in}^h, \max_{in}^h], T_{in}^c[\min_{in}^c, \max_{in}^c], T_{out}^h[\min_{out}^h, \max_{out}^h], T_{out}^c[\min_{out}^c, \max_{out}^c])$, and for all exchangers, we look for the minimum and the maximum.
- Simulate 40 points (exchanges) using random variable from the uniform distribution. For example, for the exchanger No. 30 (E30), the minimum input temperature for the cold stream is obtained by the following probabilistic simulation scheme: $T_{in}^c(E30) \sim U(\min(\min_{in}^h), \max(\max_{in}^h))$. Retrieve the interval matrix of the 4 variables: $T_{in}^h, T_{in}^c, T_{out}^h$, and T_{out}^c .

Statistical description for the interval-valued data of the initial matrix and the simulated one, using equations 1 and 2, is presented in the following Table:

Tableau 1 Descriptive statistics (mean and standard deviation s:d) for the interval-valued

		T_{in}^h	T_{in}^c	T_{out}^h	T_{out}^c
Initial data	Mean	[124.20:131.42]	[46.32:61.30]	[70.32:84.30]	[84.68:94.95]
	sd	[44.47:45.42]	[38.25:23.67]	[20.81:12.03]	[47.34:38.20]
Simulation	Mean	[112.14:134.02]	[45.34:42.93]	[68.95:81.42]	[76.98:97.71]
	sd	[20.77:28.33]	[22.91:23.85]	[11.38:9.02]	[33.34:25.32]

3.2. Principal Component Analysis for interval-valued data

The interpretation of the position of the interval-valued data in the principal plane is the same as in the classical principal component analysis situation. In Figure 2, we show the results with respect to the first three axes, achieved by the Symbolic PCA using 4 and 5 (centers method). Notice that the 61% of the total inertia is explained by the first two axes in the case of simulation (40 exchangers), and 98.5% of in the case of initial data (4 exchangers). In Figure 2, closeness among clusters exchanger mainly influenced by the

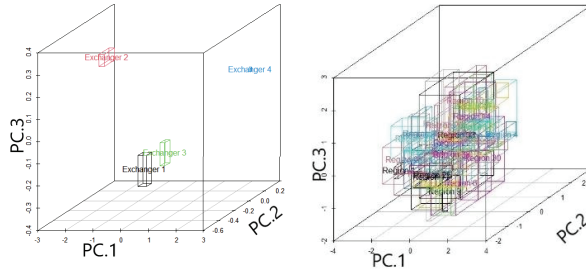


Figure 2 Principal 3D-space with data of interval type of HEN. Factorial for 4 Exchangers (left) and for 40 Exchangers (right).

same descriptors. Only one group consisting of exchangers 1 and 3 can be identified. The constituent elements of this cluster are influenced by the same main factors. Exchangers 2 and 4 are detached from the cluster from a 3D perspective. This observation is in line with the conclusions of the initial study (Floquet et al., 2016). For simulated data, the same analysis can be achieved using additional information, and at this stage, it is also difficult to give an interpretation of the similarity in size and shape among exchangers. In order to shed new light on the variability of outlet temperatures as a function of input streams, we propose a linear regression on interval values.

3.3. Regression and prediction for interval-valued data

Tableau 2 Results of interval-valued data regression

	Dependent variable	
	$T_{out}^h [min_{out}^h, max_{out}^h]$	$T_{out}^c [min_{out}^c, max_{out}^c]$
$\widehat{\beta}_0$	63.22	37.46
$\widehat{\beta}_1$	0.0046	0.17
$\widehat{\beta}_2$	0.244	0.6

The model proposed here is based on the classical formation of HEN models. However, a notable difference lies in the integration of symbolic objects in the model. To understand how the input (cold and hot) temperature stream affects the output temperature, we propose a regression model for interval-valued temperature. For a hot stream:

$$T_{out}^h [min_{out}^h, max_{out}^h] = \beta_0 + \beta_1 \cdot T_{in}^h [min_{in}^h, max_{in}^h] + \beta_2 T_{in}^c [min_{in}^c, max_{in}^c] + \epsilon \quad (6)$$

The following results are obtained using the simulated data in the Table. To evaluate this approach, two datasets were used: (i) *training dataset*, which is a set of the first 35 exchangers used to fit the parameters of the model 6 and (ii) *test dataset*, which is a set of the rest 5 exchangers used to provide an unbiased evaluation of the

Table 3 The prediction results

Hx	Predictions		Test	
	$[\widehat{T}_{out}^h]$	$[\widehat{T}_{out}^c]$	$[T_{out}^h]$	$[T_{out}^c]$
E36	[64.21:75.01]	[61.41:80.81]	[73.61:67.64]	[39.89:119.97]
E37	[73.25:82.21]	[76.12:104.68]	[69.72:89.61]	[115.31:89.80]
E38	[68.06:76.77]	[63.75:87.63]	[57.47:91.65]	[46.96:104.48]
E39	[64.73:77.73]	[69.71:80.77]	[88.25:85.72]	[113.54:66.64]
E40	[60.93:72.54]	[52.76:75.85]	[52.70:88.04]	[83.85:100.02]

estimated model. The Table 3 shows the results of the predictions using the fitted model and corresponding data test. The size of the coefficient for each independent variable $T_{in}^h [min_{in}^h, max_{in}^h]$ and $T_{in}^c [min_{in}^c, max_{in}^c]$

gives the size of the effect that variable is having on the dependent variable T_{out}^h and T_{out}^c . The estimated coefficients tell how much the output stream is expected to increase when that inputs streams (hot and cold) increases by one. For example the coefficients for the model 6, T_{in}^h differed by 1°C (and T_{in}^c did not differ) T_{out}^h will differ by 0.17°C units, on average. The estimated intercept $\widehat{\beta}_0$, is the expected mean value of T_{out}^h and T_{out}^c when all inputs are 0. In our experiments, the T_{in}^h and T_{out}^c never comes close to 0, then intercept has no meaningful interpretation.

4. Conclusion

Symbolic Data Analysis (SDA) extends statistics and multivariate data analysis to deal with data structured in a distributional form with complex internal variations. In this paper, comprehensive modeling via SDA has been presented that moves substantially beyond the traditional modeling in HEN robustness analysis. It proposed some new approaches which intended to redefine the robustness of HEN based on interval data. First, the detection of Exchangers cluster which would be affected by common factors has been modeled by Symbolic Principal Component Analysis. Second, the relation between inherent variation, expressed by intervals, of input and output temperature has been modeled using the linear regression method for interval-valued variables. Future research should include histogram valued-data.

References

- Bertrand, P., & Goupil, F. (2000). Descriptive statistics for symbolic data. In *Analysis of symbolic data* (pp. 106–124). Springer.
- Billard, L., & Diday, E. (2006). *Symbolic Data Analysis: Conceptual Statistics and Data Mining* John Wiley. Chichester.
- Bock, H.-H., & Diday, E. (2012). *Analysis of symbolic data: Exploratory methods for extracting statistical information from complex data*. Springer Science & Business Media.

- Cazes, P., Chouakria, A., Diday, E., & Schektman, Y. (1997). Extension de l'analyse en composantes principales à des données de type intervalle. *Revue de Statistique Appliquée*, 45(3), 5–24.
- Floquet, P., Hétreux, G., Hétreux, R., & Payet, L. (2016). Analysis of Operational Heat Exchanger Network Robustness via Interval Arithmetic. In *Computer Aided Chemical Engineering* (Vol. 38, pp. 1401–1406). Elsevier.
- Macchietto, S., Coletti, F., & Bejarano, E. D. (2018). Energy Recovery in Heat Exchanger Networks in a Dynamic, Big-data World: Design, Monitoring, Diagnosis and Operation. In *Computer Aided Chemical Engineering* (Vol. 44, pp. 1147–1152). Elsevier.

Rules for Predicting Benefits in Simultaneous versus Sequential Heat Integration

Hendrik Pötting^a, Avinash S. R. Subramanian^a and Truls Gundersen^{a*}

^a*Department of Energy and Process Engineering, Norwegian University of Science and Technology (NTNU), Kolbjørn Hejes vei 1B, NO-7491, Trondheim, Norway*
truls.gundersen@ntnu.no

Abstract

Applying simultaneous process optimization and heat integration (SIM) on a process can lead to an increased Net Present Value (NPV) compared to applying a sequential approach (SEQ) to the same process. These benefits do not always arise, but the computational effort is significantly higher. In this paper, three heuristic rules are proposed that predict whether applying SIM will improve a specific process over applying SEQ. These rules are to be applied once the results of SEQ are available and can be used to decide if SIM should be applied or not. The proposed rules are subsequently verified with two test cases.

Keywords: Simultaneous Heat Integration, Sequential Heat Integration, Process Optimization, Non-Linear Programs (NLPs), Trade-Off-Section (TOS)

1. Introduction

Multiple optimization strategies can be applied to ensure that a newly developed chemical process is energy efficient. The traditional way to perform energy optimization is sequential process design and heat integration (SEQ), which consists of two steps. In the first step, the operating conditions of the background process are optimized, assuming that all heating and cooling demands are satisfied by external utilities. In the second step, the heat integration scheme is optimized to derive the lowest possible external utility demand for this process. It follows that the sequential approach leads to a process where the operating conditions are determined without considering the scope for improved heat recovery. An alternative to SEQ is simultaneous process optimization and heat integration (SIM), where the process and the heat integration are optimized in one step. Duran and Grossmann (1986) proposed an NLP formulation for applying SIM to a process. This formulation was used in several publications as a basis for further research. Examples are Dowling and Biegler (2015), who modified the formulation for large scale flowsheet optimization, and Kamath et al. (2012), who proposed a modification that allows for taking phase changes into account. An alternative MINLP formulation was proposed by Kong and Maravelias (2018). However, to avoid the higher computational costs with MINLP problems, this paper focuses on the NLP formulation by Duran and Grossmann (1986).

The benefit of applying SIM instead of SEQ is that the external utility costs can be decreased. In addition, better process parameters can be found than the ones obtained by

applying SEQ. One potential downside of SIM is that the computational costs are significantly higher, because the size of the resulting NLP is larger with increased degree of non-linearity. It follows that the complexity of the optimization model increases, which can lead to reduced robustness, i.e. the optimizer may fail to find the global optimum or even a feasible solution. Hence, the resulting higher effort is justifiable only if applying SIM leads to significant improvements compared to applying SEQ. However, these benefits do not occur for every process; a counter example can be found in Pötting (2020).

In this paper, three heuristic rules are proposed. These heuristics allow for predicting whether applying SIM instead of SEQ can lead to benefits for a specific process. As the rules are to be applied after performing SEQ, they allow the engineer to decide whether SIM should be applied or if an application of SIM would lead to similar results as already obtained from the SEQ approach. The proposed rules are applied to two test cases and optimization with both SEQ and SIM are performed to verify the predictions made.

2. Methodology

The proposed heuristics are derived based on the following reasoning: The process optimization performed by applying SEQ leads to the highest possible net present value (NPV) for the background process. By definition of optimality, any change in operating conditions for applying SIM from the optimum given by applying SEQ would lead to the same or a lower NPV for the background process. Thus, the lower NPV has to be compensated for by providing more heating capacity. The rules proposed in this paper are explained for the case of an increase in the utilized heating capacities. An analogous set of rules can be derived for the case of increasing utilized cooling capacities. The utilized heating capacity of a process can be increased if three different conditions are met:

- Heuristic 1: There needs to exist a Trade-Off-Section in the process which is described in more detail in Section 2.1.
- Heuristic 2: The process needs to be able to utilize the additional heat. This can be done by either reducing the usage of external heating utilities or by using the heat to produce additional products, such as steam or electricity. The use of additional heat to reduce the feedstock consumption is beyond the scope of this paper. Heuristic 2 is explained in Section 2.2.
- Heuristic 3: The creation and utilization of additional heating capacity needs to be cost efficient. Heuristic 3 is discussed in Section 2.3.

2.1. *Heuristic 1: Existence of a Trade-Off-Section*

To be able to increase the amount of provided heating capacity by applying SIM, a process needs to have at least one Trade-Off-Section (TOS). A Trade-Off-Section can be defined as a section of the process which allows for adjusting the operating conditions such that the profitability of the background process is lowered in order to generate additional heating capacity. A section is a TOS if it satisfies the following two conditions:

- At least one of the section's operating parameters has to influence both the heat exchange with the rest of the process and at least one outlet flow of the section.

This influenced effluent heat or material flow is denoted $f_{Infl}^{TOS,out}$ and the operating parameter is referred to as decision variable x . The identified section can have several additional effluent flows that may be influenced by the independent variable as well. If the considered process section has several independent variables or influenced flows, each combination of process section, independent variable and influenced flow has to be evaluated separately. For the process section to be a TOS, only one of these combinations needs to meet the second condition.

- The second condition is that perturbing the decision variable x from its optimal value determined after applying SEQ has an opposite effect on the NPV of the background process compared to the quantity of additional heating capacity generated. This opposite effect allows for compensating the lower NPV of the background process by additional heating capacity. However, the NPV is determined by the entire background process and not only the particular TOS studied. To relate the local behavior of the TOS to the plantwide performance that determines NPV, it is assumed that the influenced effluent flow $f_{Infl}^{TOS,out}$ correlates positively with the plant wide NPV.

Heuristic 1 is formalized in Equation (1). $Q^{TOS,out}$ denotes the heating capacity that is provided by the TOS, and x_{SEQ} is the optimal value of x after applying SEQ.

$$\left. \frac{\partial f_{Infl}^{TOS,out}}{\partial x} \right|_{x=x_{SEQ}} * \left. \frac{\partial Q^{TOS,out}}{\partial x} \right|_{x=x_{SEQ}} \leq 0 \quad (1)$$

2.2. Heuristic 2: Existence of a Suitable Heat Sink

To improve a process by applying SIM instead of SEQ, the additional heating capacity provided by the TOS needs to be utilized by a Heat Sink. Here, it is essential to define two types of Heat Sinks: Natural and Created Heat Sinks. A Natural Heat Sink is a process section that still has a demand for external heating utilities after applying SEQ. It follows that the use of the additional heating capacity to satisfy Natural Heat Sinks leads to a decrease in the cost of external utilities. A Created Heat Sink arises if certain background process parameters change compared to the sequentially optimized process, with two consequences: First, the heating demand of the process increases and second, this additional heat results in an increase in the yield of a product. The process section with the increased heating demand is denoted as Created Heat Sink. An example is a steam turbine that can be increased in size compared to the results of SEQ to use the additional heat for increased power production.

A Heat Sink is denoted as a Suitable Heat Sink if the additional heat can be transferred to the Heat Sink with respect to the minimum temperature difference. To determine whether heat transfer to a Natural Heat Sink is possible, the additional heat source (obtained from changing the operating parameters of the TOS) can be included in the Grand Composite Curve (GCC) of the process. The Natural Heat Sink is suitable if the additional heat can be used in the process. To evaluate if a Created Heat Sink is suitable, only the hot and cold streams forming the Created Heat Sink are of interest. Using these streams only, a reduced GCC diagram can be derived. The Created Heat Sink is suitable

if an increase in the heating demand of the section allows more heat to be transferred from the TOS.

2.3. Heuristic 3: Economic Profitability

The availability of both a TOS and a Suitable Heat Sink is a requirement for potential benefits from applying SIM instead of SEQ. In addition, it must be profitable to accept a sub-optimal operating point of the background process to improve the heat integration. The evaluation of this profitability needs to be done separately for processes with a Natural Heat Sink and those with a Created Heat Sink as they have different types of revenues: Created Heat Sinks are only used if there are no suitable Natural Heat Sinks. The reason is that in the sequentially optimized process, external utilities would have been used to produce the additional product, if that would have been more profitable. As this is not the case, it follows that the earnings from selling the additional product are lower than the costs for the additional external utilities. Accordingly, it is more cost efficient to reduce the amount of external heating utilities than to increase the amount of a product.

For a process with a Suitable Natural Heat Sink, the reduced use of external utilities has to lead to larger savings than the losses from sub-optimal operation of the background process. The rules are meant to exclude processes from being subject to SIM. Therefore, a sufficient but not necessary assumption is that all costs that are not related to revenues or utility costs do not decrease when applying SIM instead of SEQ. In addition, the earnings from products whose flow rates are independent of $f_{Infl}^{TOS,out}$ are assumed not to decrease. With these assumptions, there needs to be at least one $x = x_{SIM}$ for which the process satisfies the condition stated in Equation 2.

$$\left(Q^{TOS,out} \Big|_{x=x_{SIM}} - Q^{TOS,out} \Big|_{x=x_{SEQ}} \right) * c_{heating} + \left(f_{Infl}^{TOS,out} \Big|_{x=x_{SIM}} - f_{Infl}^{TOS,out} \Big|_{x=x_{SEQ}} \right) * \quad (2)$$

Equation 2 consists of two terms: The first term stands for the reduced costs for external heating utilities due to the satisfaction of heating demand with the additional heat. The second term denotes the losses in the production due to the changes in the background process. Pr is the set of all sold products. The fraction $f_p / f_{Infl}^{TOS,out}$ stands for the ratio between the flow rate of the product f_p and the flow rate of the influenced effluent material or heat flow of the TOS, $f_{Infl}^{TOS,out}$, which is assumed to be constant. This assumption holds if the downstream process units always operate in the same optimal operating point and a change in $f_{Infl}^{TOS,out}$ just leads to a linear rescaling of the downstream process units. As this paper deals with grassroots design, it is justifiable to assume that rescaling would be preferred over a non-optimal operating point. The unit cost for the heating utility is denoted as $c_{heating}$ and the specific earnings of a product p is c_p .

The condition stated in Equation (2) is also a conservative estimate for the profitability of a process with a Created Heat Sink. The reason is that satisfying a Created Heat Sink is always less cost-efficient than satisfying a Natural Heat Sink, as explained in Section 2.2. As a consequence, if the condition for a Natural Heat Sink to be cost-efficient does not hold, a process with only Created Heat Sinks is not cost-efficient either.

3. Case Studies

The application of the proposed rules is illustrated with two case studies. The first case study is the example process discussed by Duran and Grossmann (1986). The flowsheet is shown in Figure 1. Following the proposed rules, the first step is to identify a process section that can be a TOS. This process section is chosen to be the steam generation part of the reactor and the following heat exchanger, excluding the reactor itself. The effluent steam flow $f_{Steam}^{TOS,out}$ is both the decision variable x and the influenced outlet flow $f_{Infl}^{TOS,out}$. A decrease in the steam production leads to a higher temperature of the stream entering heat exchanger H1 and therefore to more available heat. As this satisfies the condition stated in Equation 1, the steam generation part is confirmed to be a TOS.

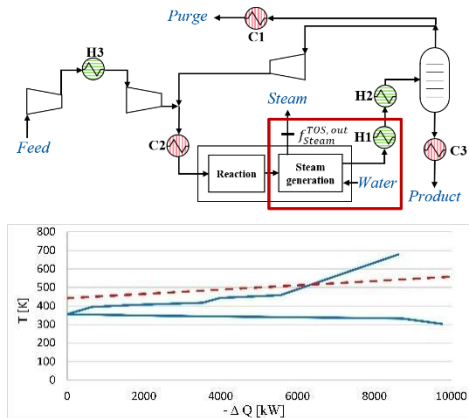


Figure 1: Flowsheet and GCC with additional heating capacity of the example process by Duran and Grossmann (1986). The red box marks the TOS.

The availability of a Suitable Heat Sink is evaluated with the GCC of the process, which is shown in Figure 1 as the blue solid line. The additional heating capacity provided by the TOS is shown as the red dashed line. As heat transfer from the TOS to the process is thermodynamically feasible, it follows that the process has a Suitable Natural Heat Sink. Next, the profitability of applying SIM has to be evaluated. With steam as the sold product, the technical and economic parameters used by Duran and Grossmann (1986) lead to $c_p < c_{heating}$. As it is assumed that applying SEQ always lead to maximum production, it follows that $Q_{Steam,SIM} < Q_{Steam,SEQ}$. Also, the steam flow $f_{Steam}^{TOS,out} = f_{Infl}^{TOS,out}$ does not pass any downstream unit after the TOS. It follows that $f_p = f_{Infl}^{TOS,out}$ and therefore $f_p / f_{Infl}^{TOS,out}$ is constant. Accordingly, Equation 2 holds and the application of SIM is therefore recommended. This recommendation is confirmed

by the results of Duran and Grossmann (1986) as they achieve a 90 % higher annual profit by applying SIM instead of SEQ.

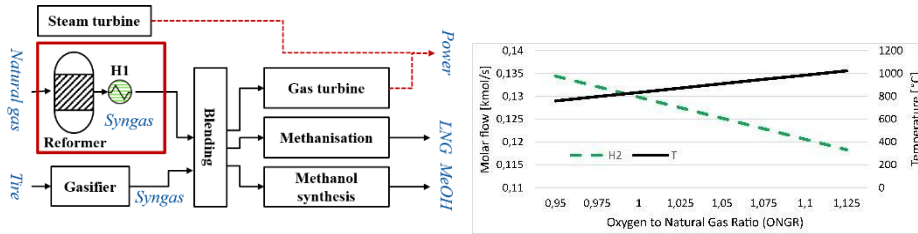


Figure 2: Flowsheet and Reformer behavior of the example process by Pötting (2020). The red box marks the TOS.

The second case study is the process studied by Pötting (2020), which is a simplification of the process proposed by Subramanian et al. (2021). The flowsheet of this process is shown in Figure 2. The reformer and the heat exchanger after the reformer are identified as a TOS candidate. The decision variable x is the oxygen-to-natural-gas ratio $ONGR$ at the reformer inlet. The influenced outlet flow $f_{Infl}^{TOS,out}$ is the molar flow of H_2 $n_{H_2}^{TOS,out}$.

The right part of Figure 2 shows the temperature and the molar flow of H_2 at the reformer outlet, which are modeled to be linear functions of the $ONGR$. With $n_{H_2}^{TOS,out} = n_{H_2}^{TOS,out,max}$ at $x = x_{SEQ}$ and the assumption of constant specific heat capacities, the condition stated in Equation 1 holds. Accordingly, this section is proven to be a TOS.

As the SEQ approach to process optimization results in a threshold problem, there is no Natural Heat Sink in the process. It follows that a Heat Sink needs to be created. This can be done by increasing the size of the steam turbines, which leads to an increased power production. Figure 3 shows the streams of the steam turbine section after applying SEQ as the blue solid line. The red dashed line indicates the additional heating capacity. This dashed line starts to the right of the solid line because it is not replacing already available heat sources, but it can be used as a heat source for the additional Created Heat Sink. An increase of the steam mass flow leads to a horizontal stretch of the solid line. The result is shown by the grey dotted line. As this dotted line allows more heat transfer from the dashed line than to the solid line, the Heat Sink is Suitable.

To evaluate the profitability, it is assumed that the blended syngas flow always contains enough CO to run a stoichiometric methanization and that the specific heat capacity of the effluent stream of the TOS is constant. As $ONGR_{SEQ}$ is $ONGR_{min}$, it follows that $ONGR_{SIM} > ONGR_{SEQ}$. Using the parameters of Pötting (2020), Equation 2 can be shown not to hold. Accordingly,

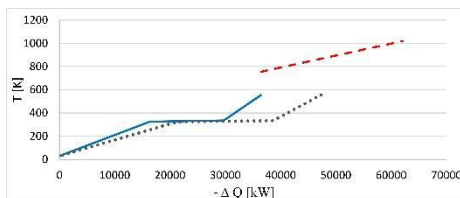


Figure 3: Evaluation of the usability of a Created Heat Sink, Pötting (2020)

applying SIM will not lead to an improvement compared to SEQ. Pötting (2020) also obtained this result.

4. Conclusion and future work

This paper proposes three heuristic rules for evaluating the benefits of applying simultaneous over sequential process optimization and heat integration. The first rule determines if the background process contains a Trade-Off-Section that can provide additional heating capacity by sacrificing product flow rate. The second rule evaluates if this additional heat capacity can be feasibly transferred and used by the rest of the process. The third rule focuses on the economics of generating this additional heating capacity at the expense of profitability of the background process. Two example processes verified these rules. In future work, it is possible to derive a similar set of rules for benefits due to a reduced feedstock consumption.

References

- A. W. Dowling, L. T. Biegler, 2015. A framework for efficient large scale equation-oriented flowsheet optimization. *Computers & Chemical Engineering* 72, 3–20.
- M. A. Duran, I. Grossmann, 1986. Simultaneous optimization and heat integration of chemical processes. *AIChE Journal* 32, 123–138.
- R. S. Kamath, L. T. Biegler, I. E. Grossmann, 2012. Modeling multistream heat exchangers with and without phase changes for simultaneous optimization and heat integration. *AIChE Journal* 58, 190–204.
- L. Kong, C. T. Maravelias, 2018. An Optimization-Based Approach for Simultaneous Chemical Process and Heat Exchanger Network Synthesis. *Industrial & Engineering Chemistry Research* 57, 6330–6343.
- H. Pötting, 2020. Optimization of Energy Systems for Polygeneration Plants. Master Thesis, RWTH Aachen University, Aachen, Germany, 1-109, unpublished, but available on request from the first author.
- A. S. R. Subramanian, T. Gundersen, P. I. Barton, T. A. Adams II, 2021. Optimal Design and Operation of a Hybrid Waste Tire and Natural Gas Feedstock Polygeneration System, In Preparation.

Processing systems design considering resilience

Ákos Orosz^a, Jean Pimentel^b, Ferenc Friedler^{b,*}

^a*University of Pannonia, H-8200 Veszprém, Egyetem u. 10. Hungary*

^b*Széchenyi István University, H-9026 Győr, Egyetem tér 1, Hungary*

f.friedler@ga.sze.hu

Abstract

The resilience of a system is defined as the system's capability of recovering from failures. Traditionally, only predictable aspects are considered when designing processing systems. Evaluation of these aspects is performed via assessment of exact indicators and enumeration of all cause-effect options. However, such evaluation is not appropriate for determining the resilience of processing systems, since resilience is based on unexpected events in addition to the expected ones. Consequently, the cause part of the cause-effect relation is not known or not effective. In the current work, the general formula for determining resilience of a system is embedded into a P-graph based process synthesis algorithm. Thus, the resilience can be considered when selecting the most preferred process during its synthesis. The result is illustrated by synthesizing a process of adipic acid production by nitric acid oxidation of KA oil.

Keywords: Resilience, Process design, P-graph.

1. Introduction

A system is generally evaluated considering several aspects, e.g., profitability, reliability, and sustainability. In addition, since there are a lot of interconnections between the operating units that comprise engineering systems, structural indicators are also important. Agility, resilience, and supportability are just three examples for such indicators (Bernus et al., 2020), they are all highly dependent of the system's structure. The current work focuses on resilience, which expresses the persistence of the system, and also its capability to absorb change, and recover from it (Holling, 1973).

In most cases, resilience is derived from other metrics of the system, more than 20 alternatives for such derivation were listed by Trimintzios (2011). For engineering systems, the main considerations for resilience are controllability (Morari, 1983) and flexibility (Grossmann et al, 2014). Another possible metric for resilience is the system performance level after a disturbance (Bhamra et al., 2011). The repair or replacement times of the system's components can also be taken into account (Gong and You, 2018). Since systems resilience has no general definition, there are a large number of different specific definitions for the different areas of application. A thorough review of these was given by Hosseini et al. (2016), whereas a general overview of resilience theory has been provided by Linkov and Trump (2019).

In a processing system, the disturbances or changes, referred by resilience definitions, mostly involve failures of equipment units. The failures in a processing system can be predictable or unpredictable/unexpected, as well as internal or external. Traditionally, in designing processing systems, the predictable, known aspects are taken into account. On the basis of that, exact indicators, such as cost or systems reliability, can be determined.

However, resilience also needs to contemplate unexpected events. The prediction of these events is either impossible, or would require enumerating an enormous amount of possibilities, requiring a significant computational effort. To overcome this philosophical issue, one solution is to enumerate all possible “effects”. Since the enumeration can conveniently be based on the P-graph framework, it is a firm foundation of a new formulation of resilience.

Formally, a processing system is considered resilient if after any failures it is

- able to perform its designated job in full or on a pre-specified partial level;
- able to return to its original state.

Designing a resilient processing system requires either, or both of these goals to be considered. The first goal, i.e., maintaining operation after failures, can be achieved via redundancies in the system. Two different types of redundancies can be distinguished. *Unit-level* redundancy incorporates identical or similar equipment units (or simply: units) to be built in parallel to critical units in the system. When a unit fails, one of its redundant counterparts can immediately, or with some delay take over the operation. On the other hand, *process-level* redundancy contains multiple separate production paths for one of more materials in the process. In case of a failure, a part of the process (containing multiple units) is turned off and a different production path is utilized. This redundant path may even contain different technologies which are more convenient for the short time interval while the main production path is repaired. Naturally, the two redundancy strategies can be combined in a processing system for the best result. The second goal when designing resilient systems, i.e., returning to the original state, is related to maintenance and repair. Units with shorter repair time can be preferred to reduce the off-time of the system.

The current work proposes a method to perform synthesis of processing systems with redundancies, considering resilience based on the above-mentioned enumeration of “effects”. The resilience indicator incorporates both the redundancies in the system, and the repair times of the failed units. The goal is to generate all processing systems that satisfy the requirement on resilience, and rank them according to their cost of operation (i.e. annualized capital cost and operating cost of units). Generating a ranked list is more practical than generating just one option, as it provides the opportunity for selecting a more convenient solution, in light of additional constraints.

The generation of multiple solutions, and the enumeration of the effects are both combinatorial in nature, and cannot be properly incorporated into standard mathematical optimization; i.e. an appropriate tool is required. For this reason, the proposed method is based on the P-graph framework (Friedler et. al., 1992a), which is an axiom-based combinatorial tool for process network synthesis (PNS). The framework represents a PNS problem as a bipartite graph of materials and operating units, and provides several algorithms, such as algorithm MSG (Friedler et. al., 1993) for generating the maximal structure (superstructure), and algorithm SSG (Friedler et. al., 1992b) for generating all potential solution structures. In addition, any mathematical model can be integrated to ensure that only feasible processing systems are generated.

2. Synthesis of processing systems considering resilience

The method described here is a combination of two algorithms: generating processing systems for process synthesis and determining the resilience of the generated processing systems. The generation of the processing systems is performed by the combinatorial

algorithms provided by the P-graph framework. To determine the resilience of a processing system based on the enumeration of the effects, a proper formula is needed.

The number of products of a processing system is not limited, however, first, systems of single product will be examined. This product has a minimum accepted production flowrate (γ), which is specified in the problem definition. This is the flowrate that the production must reach to be considered as fully operational, anything less is considered as partial operation. It is common in processing systems to halt the operation if the production goes below a certain level (because of technical and/or economic reasons), this minimum production flowrate (δ) is also defined.

In a processing system with redundancies, it is important to distinguish the normal operation, i.e., the operating units that are active when the system functions as planned. The remaining operating units are the backup units; these are employed if some failures happen in the normal operation.

Let K denote the number of units in the normal operation, it is assumed that these units are subject to failures. Let S_n^f denote the set of all combinations of at most failures among the units of the normal operation. Let N_n^f be the cardinality of set S_n^f . To uniformly measure the capability of recovering through repairs, time horizon is introduced as a constant value throughout the synthesis.

Let s be an element of S_n^f . The production flowrate over time of the system as a consequence of failures in s is denoted by function $F(s, t)$, where $t \in [0, h]$. To plot the resilience uniformly on the $[0, 1]$ scale, this production flow rate is normalized to $p(s, t)$ as given in Eq. (1).

$$p(s, t) = \begin{cases} 1, & \gamma < F(s, t) \\ \frac{F(s, t)}{\gamma}, & \delta \leq F(s, t) \end{cases} \quad (1)$$

Eq. (2) defines the relative production of the system during time horizon h after the failures defined by s occur (at time 0).

$$q(s) = \frac{\int_0^h p(s, t) dt}{h} \quad (2)$$

The resilience of the system to at most n simultaneous failures is given by enumerating over all possible sets of failures in S_n^f , and taking the average of the relative production rates, see Eq. (3). If the number of failures is not limited (or limited to the number of units in the normal operation, i.e. K), the result given by the formula is the resilience of the system to any set of failures (Eq. (4)).

$$R_n = \frac{\sum_{s \in S_n^f} q(s)}{N_n^f} \quad (3)$$

$$R = R_K \quad (4)$$

By combining the described resilience analysis method with the synthesis tools of the P-graph framework, the following algorithm is proposed to perform PNS with resilience consideration.

Input of the algorithm:

- Information needed for the PNS problem
- γ and δ flowrates
- The repair times of the units
- h time horizon
- Lower bound for resilience (0 if not specified)

Main steps:

- Algorithm SSG of the P-graph framework generates all potentially useful processing systems
 - Only feasible processing systems are generated, as described by the mathematical model given in the problem formulation
- For each generated processing system
 - Normal operation of the process is determined so that the operating cost is minimal
 - Resilience analysis is performed with the proposed formulas
 - If the resilience of the processing system is below the lower bound, it is discarded
- The list of potential processing systems is ranked based on their cost of operation, it is the output of the algorithm

Selecting the desired processing system from the ranked list is up to the designer or decision maker. A simple choice is to select the one with the lowest cost of operation. Another option is to plot the processing systems on a chart showing the cost of operation against the resilience, i.e., to determine the Pareto front, and make a decision based on that.

3. Case study

This work contemplates process synthesis for a continuous process for adipic acid from KA oil (mixture cyclohexanol/cyclohexanone) and HNO_3 . The problem structure is depicted in Figure 1. Low temperature oxidation of KA oil can be performed in two different reactors. Reactor A handles heat with recirculation, whereas reactor B employs a coolant. Oxidation is enhanced in a second reactor at high temperature. Nitrogen oxides can be recovered as nitric acid by means of a bleacher, where they are oxidized. Alternatively, they can be to recover in a membrane process and oxidized in a separated reactor. Oxidized gases are absorbed in water and returned into the process. Separation of water is performed by evaporation in one or more stages. Adipic acid is recovered by crystallization followed by a solid separation stage. Different sequences of these operations are considered for the synthesis problem. Crystals of adipic acid are recrystallized to improve purity of the product. The liquid effluents from the separation stages are recycled into the process, nonetheless a fraction is purged out from the system to avoid accumulation of by-products. These byproducts are employed in another process in the plant; therefore, catalyst is recovered from the purged stream, and sent to a treatment process previous its final destiny.

The case study was solved to generate all possible designs for the processing system providing that the cost of operation of the system does not exceed 2,560,000 USD/y, and

the resilience is at least 0.4. All 4068 generated processing systems are represented in Figure 2 as cost against resilience. The Pareto front of the solutions is indicated by the red dots. .

One of the processing systems in Pareto front is selected as an example (indicated with a larger green circle on Figure 2) and is presented on Figure 4. Its cost of operation is 1,728,408 USD/y, and its resilience is 0.57. The designed or normal operation is shown with bold lines on Figure 4, the redundant units are shown with gray lines.

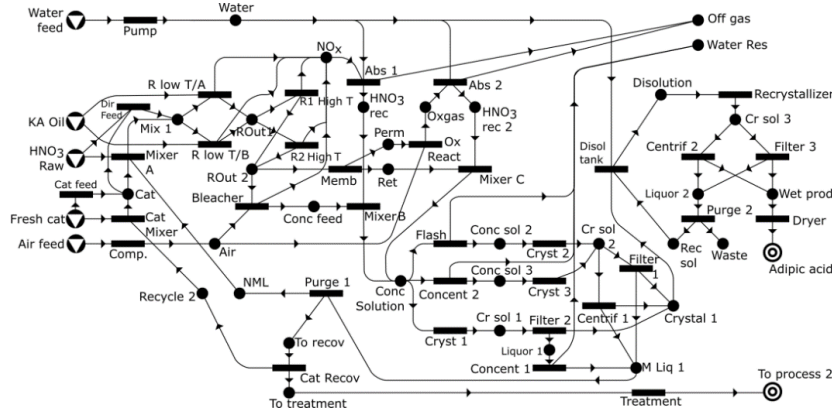


Figure 1. P-graph representation of the superstructure of the case study

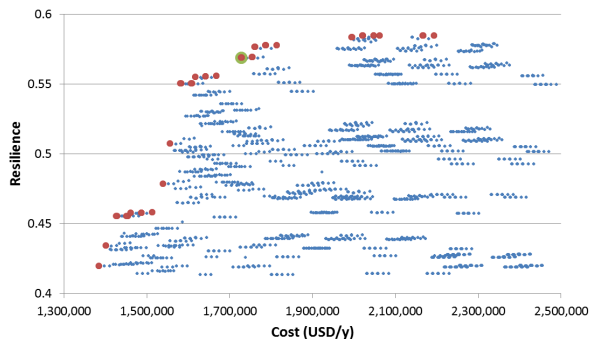
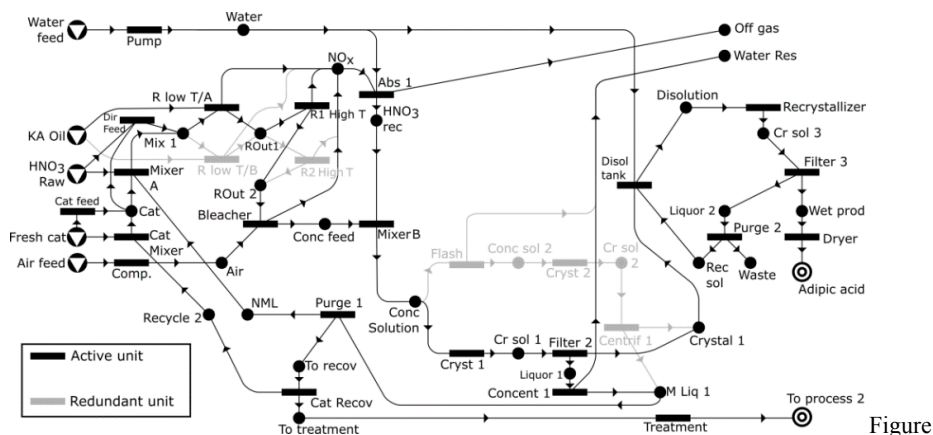


Figure 2. Generated processing systems for the case study, red dots indicate the Pareto front, green circle identifies the process shown on Figure

4. Conclusions

A process synthesis method has been proposed that considers resilience to unpredictable events by combinatorial enumeration of the effects. The method generates all, or n-best potential processes that satisfy the requirement for the resilience, and rank them according to their cost of operation. The capabilities of the method were illustrated through a case study of adipic acid production, where the algorithm generated 4068 potential solutions; 25 of them formed the Pareto front.



3. One of the generated processing systems of the Pareto front

References

- P. Bernus, O. Noran, H.T. Goranson, 2020, Toward a Science of Resilience, Supportability 4.0 and Agility, In Proc. 21st World Congress of the International Federation of Automatic Control (IFAC 2020) Berlin, 13-17 July 2020. IFAC Papers Online. 11350-11357. ISSN 2405-8963
- R. Bhamra, S. Dani, K. Burnard, 2011, Resilience: the concept, a literature review and future directions, *International Journal of Production Research*, 49, 18, 5375-5393
- F. Friedler, K. Tarján, Y.W. Huang, L.T. Fan, 1992a, Graph-theoretic approach to process synthesis: axioms and theorems, *Chemical Engineering Science*, 47, 1973-1988
- F. Friedler, K. Tarján, Y.W. Huang, L.T. Fan, 1992b, Combinatorial algorithms for process synthesis. *Comput Chem Eng*, 16, S313-S320
- F. Friedler, K. Tarján, Y.W. Huang, L.T. Fan, 1993, Graph-theoretic approach to process synthesis: Polynomial algorithm for maximal structure generation, *Computers & Chemical Engineering*, 17, 929-942
- J. Gong, F. You, 2018, Resilient design and operations of process systems: Nonlinear adaptive robust optimization model and algorithm for resilience analysis and enhancement, *Computers & Chemical Engineering*, 116, 231-252
- I.E. Grossmann, B.A. Calfa, P. Garcia-Herreros, 2014, Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes, *Computers & chemical engineering*, 70, 22-34
- C.S. Holling, 1973, Resilience and stability of ecological systems, *Annual review of ecology and systematics*, 4, 1, 1-23
- S. Hosseini, K. Barker, J.E. Ramirez-Marquez, 2016, A review of definitions and measures of system resilience, *Reliability Engineering & System Safety*, 145, 47-61
- I. Linkov, B.D. Trump, 2019, *The science and practice of resilience*, Springer International Publishing
- M. Morari, 1983, Design of resilient processing plants—III: A general framework for the assessment of dynamic resilience, *Chemical Engineering Science*, 38, 11, 1881-1891
- P. Trimintzios, 2011, Measurement frameworks and metrics for resilient networks and services: Technical report, European Network and Information Security Agency (ENISA), Tech. Rep., February

Applying quality assurance concepts from software development to simulation model assessment in smart equipment

Jonathan Mädler,^a Isabell Viedt,^a Leon Urbas^a

^aTechnische Universität Dresden, 01062 Dresden, Germany

jonathan.maedler@tu-dresden.de

Abstract

Smart equipment as a cyber-physical system can be envisioned to support owner/operators with model-based assistance systems. In fluctuating process scenarios, which are expected within modular plants, frequent adaptation of the required simulation models to changing conditions will be necessary. Therefore, automated functional quality assessment for simulation models would provide a significant added value for the owner/operator. Currently, quality assessment of simulation models is dominated by the verification and validation methods developed by Sargent and co-authors (Sargent and Balci, 2017). Although these methods provide a wide spectrum of assessment methods, they base on a rather narrow view on quality of simulation models. In addition, quality and quality advancement are typically not measured continuously. To overcome these obstacles, the authors discuss the applicability of central quality assurance methods from the field of software development to simulation model assessment. These are the principle of test-driven development and the application of quality models to measure quality and quality advancement. A framework and a workflow for test-driven development of simulation models are proposed and evaluated in a case study on a soft sensor for a fermentation process. It was found, that in principle, the framework provides transparent and continuous quality assessment in a semi-automated manner and lays the foundation for guidance strategies during model development.

Keywords: smart equipment, quality assessment, validation, soft sensor

1. Introduction

As a cyber-physical system, smart equipment can be envisioned to support owner/operators with model-based assistance systems. The importance of concepts and technologies like design space, soft sensors or model predictive controllers in plant operation is rapidly growing. Unfortunately, the setup of these systems is highly dependent on the concrete process running within the plant. This is especially a challenge for modular plants (c.f. VDI, 2020), which are to be used within fluctuating process scenarios for various chemical or biochemical applications. Therefore, simulation models require frequent adaptation to the changing conditions. Since these simulation models become core part of the functionality of smart equipment and the modelling process will become an important part of the user experience, an automated quality assessment methodology must capture a broad view on functional quality and quality attributes. Quality assessment should promote a transparent and continuous view on functional quality and quality advancement during the modelling process. To optimally support the

owner/operator, the quality assessment must be as efficient as possible and run automated. Since several model-based assistance systems applying different modelling approaches should be supported, the methodology has to be flexibly applicable to different model classes and allow the simple integration of additional requirements.

In this paper, the authors discuss the applicability of quality assurance concepts from software development to simulation models. A framework for automated functional quality assessment for simulation models is presented. This lays the foundation for further development of a fully guided assessment framework for smart equipment. The remainder of this paper is organized as follows: section 2 introduces and discusses standard methods for quality assessment of simulation models and highlights the gap in research. Following, in section 3 the authors discuss quality models and methodologies from the area of software development and present a concept to apply these methodologies to the area of simulation models. In section 4 a case study is conducted to evaluate the suggested methodology. Lastly, section 5 provides a conclusion and suggests further research.

2. State of the art in quality assessment of simulation models

Today's quality assessment of simulation models is dominated by the verification and validation (V&V) methods (Sargent and Balci, 2017). Sargent (2013) describes four distinct steps in the model development process. Firstly, in the center of the modelling process is the data validation, which represents the process of ensuring the correctness and appropriateness of the data for model building, evaluation, testing and simulation experiments. Conceptual model validation is utilized in determining the correctness of the underlying assumptions of the conceptual model and the applicability of the model for the intended purpose. Computerized model verification describes the process of assuring the correctness of the implementation of the conceptual model and the computer programming. Lastly and most importantly, Sargent (2013) defines the operational validation as the assurance of model accuracy over a domain of validity. Even though a substantial number of V&V methods exists in literature, the most commonly used ones rely on graphical inspection (Roungas et al., 2018), which is not suitable for automated model quality assessment. Roungas et al. (2018) propose a framework that classifies existing methods according to their suitability for certain simulation models and their characteristics. One example is the use of face validation during operational validation for simulation models without access to the source code but available data from the real system. Sargent and Balci (2017) find that many studies do not apply V&V consequently or even neglect it.

In conclusion, the recent methods are not able to meet the requirements for functional quality assessment of simulation models in smart equipment according to the introduction. These cyber-physical systems will rely heavily on continuous integration and continuous deployment strategies (CI/CD) (c.f. Bruckner et al., 2020). Although the V&V methods provide a wide spectrum of assessment methods, they mainly focus on model accuracy as quality property. Other quality attributes like reliability or efficiency are considered as boundary conditions only. This approach reduces the overall transparency of simulation model quality. Another important drawback of the current approach is that usually quality specification and testing is exclusively addressed at the end of each model development phase like conceptual modelling or implementation. This limits the transparency of quality advancement within the modelling process. In addition, many V&V methods rely on expert evaluation of graphical representations. This is not appropriate for automated quality assessment.

Therefore, for this paper the authors aim to resolve the issues with the state of the art in quality assessment for simulation models by transferring quality assurance concepts from software development to simulation models. The next section outlines the concepts of test-driven development and quality models and presents a new framework for quality assessment of simulation models.

3. State of the art in quality assessment of simulation models

3.1. Test-driven software development

According to the test-driven development concept (c.f. Beck, 2003), requirements for software parts have to be translated into tests. Developers are forced to formulate the requirements and implement the tests before starting the actual development of the software. Afterwards, these tests are used to check the progress of software development and the advancement of software quality continuously. This approach increases the transparency of the overall quality of software.

3.2. Quality models in software development

For an assessment of software quality, attributes that characterize good software quality must be determined (Wagner, 2013). The abstract term of quality needs to be made tangible. To accurately assess the quality of software, one needs to differentiate between structural and functional quality (Balci, 1998). Structural quality relates to the source code and architecture of software. Functional quality describes the execution of software (Balci, 1998). Quality models are an approach to defining requirements for perceived quality. A widespread approach for structuring quality models in software development are so-called FCM models (Cavano and McCall, 1978). Factors (F) describe the desired attributes of the software. Criteria (C) are allocated to those factors and are tested by metrics (M). Figure 1 depicts the hierarchal structure of FCM models. The metrics provide a quantitative measure that weighs attributes against requirements (Cavano and McCall, 1978). Quality models exist for different types of software applications and aspects. The QUAMOCO framework or the ISO/IEC 9126 are used to assess software product quality (Wagner, 2013). A widely used quality model for software product quality is the ISO/IEC 25010. This standard is derived from the software quality model of McCall and Cavano, which was one of the first approaches to assess software quality (Cavano and McCall, 1978, ISO/IEC, 2011). The quality model of ISO/IEC 25010 for product quality consists of eight quality factors: functional suitability, reliability, performance efficiency, operability, security, compatibility, portability and maintainability.

3.3. A test-driven approach to quality assurance for simulation models

In summary, some important lessons can be learned from software quality assurance. Firstly, the concept of test-driven development provides an approach to increasing the transparency of quality advancement. Secondly, quality models define what good quality

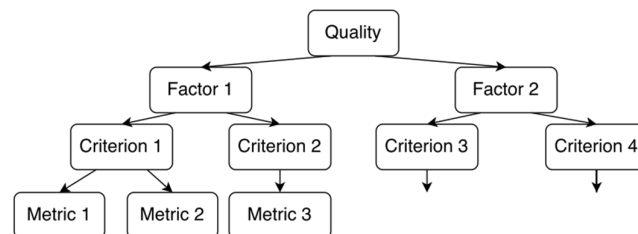


Figure 1: Taxonomy of FCM (Factor-Criteria-Metrics) models

attributes are. Those can be represented by quality factors and criteria and become measurable by metrics. The definition of quality needs to be expandable because quality expectations evolve with time.

Based on these lessons and the requirements in the introduction, a new framework for the assessment of functional quality in smart equipment is developed. Figure 2 presents the suggested workflow. The specification of the purpose of the simulation model should always be the first step within model development. Afterwards, a generic FCM model must be adapted based on the specified purpose. The factors and criteria within ISO/IEC 25010 provide the frame for a set of metrics to achieve transparent quality assessment. The user decides which factors, criteria and metrics will be included. The next step is to determine the target range or target value for each metric that can be used to calculate a quality rating. Afterwards, a standard modelling approach applying first principles modelling, data driven modelling or a combination of both can be taken. The advancement is then tested using the specified FCM model continuously. If the simulation model sufficiently fulfills the target, the user can exit model development successfully. If this is not the case, the smart equipment can provide recommendations of action to the user for further development based on the various criteria analyzed applying the FCM model.

4. Case Study

4.1. Study design

The authors conduct an initial evaluation of the proposed new framework for functional quality assessment using a case study. The use case for the quality assessment is a soft sensor for process monitoring. A simulated first principles model for yeast fermentation provides the raw data (Nagy, 2007). The soft sensor is used for online estimation of biomass concentration during the fermentation. The data-driven soft sensor is a feed forward neural network (FFNN) and is trained through Levenberg-Marquardt backpropagation. Inputs for the soft sensor are reactor temperature, dissolved oxygen concentration, yeast concentration and ethanol concentration. A quality model based on ISO/IEC 25010 was adapted to the use case. The quality factors functional suitability, reliability, performance efficiency and maintainability were considered. The quality factors performance efficiency and maintainability were not separately examined in the evaluation due to the study design without a real process plant. Metrics were matched to the quality factors functional suitability and reliability for the quality assessment. The equations for those metrics are shown in Table 1. Important metrics for soft sensors are the difference between estimated and measured values, which covers the criterion functional appropriateness, and the coverage of the targeted area of validity of the soft

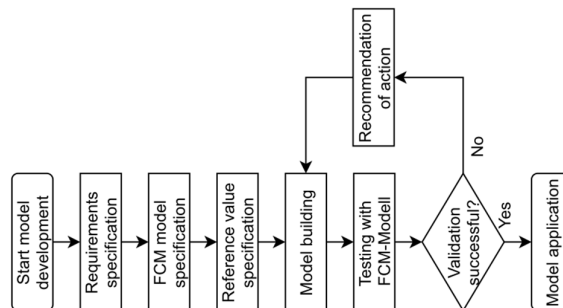


Figure 2: Workflow for functional quality assessment for simulation models in smart equipment

Table 1: Metrics overview for quality assessment

Metric	Equation	Source
Coverage	$\frac{\text{NumberOfInputs} \sqrt{V_{\text{FittingData}}}}{\sqrt{V_{\text{Specified}}}}$	(Marquardt, 2007)
R ²	$1 - \frac{SSE}{SST}$	(Roungas et al., 2018)
RMSE _{test}	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	(Montgomery & Runger, 2010)
Violation _{output}	$\frac{\sum \text{Violation}_{\text{Boundary,Output}}}{\sum \text{Data}_{\text{Output}}}$ with process limit violation	(Marquardt, 2007)
Outlier _{frequency}	$\frac{\sum \text{Outlier}_{\text{Output}}}{\sum \text{Data}_{\text{Output}}}$ with MAD filter	(Liu et al., 2004, Roungas et al., 2008)

sensor with model fitting data, which relates to the quality criterion of functional completeness. The case study consists of seven cases and is designed to emulate quality differences in the soft sensor model to evaluate the method. The cases differentiate between model fitting data (Case 1 – 3), model structure (Case 4) and input data for online estimation (Case 5 – 7). Soft sensors for online prediction use measured online data for the estimation of the output.

4.2. Study results

The quality assessment is able to distinguish between the designed quality differences in the case study. For some cases, like Case 7, almost every metric requires an action to further develop the simulation model, while in other cases, like Case 1, only certain metrics require an action because the targets are not met. For example, if the target range for the outlier frequency is not met, the quality assessment recommends adding a pre-processing layer into the soft sensor to filter out incorrect data points. In case of output violations of the process limits, the quality assessment recommends different actions depending on the severity of the error. Actions range from simple retraining of the model to replacing the entire modelling approach. The high coverage for Case 3 shows that the

Table 2: Results for the different quality metrics

Factor	Functional Suitability			Reliability	
	Correctness	Appropriateness	Completeness	Fault tolerance (Robustness)	
Metric	R ²	RMSE _{test}	Coverage	Violation _{Output}	Outlier _{frequency}
Case 1	0.82	0.0893	0.3904	0.0488	0.0
Case 2	0.77	0.0891	0.3451	0.0487	0.0
Case 3	0.37	0.0888	0.9347	0.0488	0.0
Case 4	0.85	0.0896	0.3904	0.0491	0.0
Case 5	0.55	0.0893	0.3904	0.0976	0.0741
Case 6	0.53	0.0826	0.3904	0.1707	0.0667
Case 7	0.53	0.1343	0.3904	0.5366	0.0167

coverage metric cannot be regarded in isolation, as the high coverage is due to noise increase in the model fitting data. Simply comparing the volumes of the convex hulls does not entirely capture the correctness of the data. One future approach will be to also assess the position of the volumes in space.

5. Conclusion

In this paper, a new approach to quality assessment of simulation models based on the test-driven development concept and FCM models from software development was presented. The framework allows transparent and continuous quality assessment of simulation models within the modelling process. The determination of requirements before the start of model building allows an early linkage of requirements and tests during the planning phase of the tests. This framework lays the foundation for guidance strategies during model development. The structure of the quality model enables only a semi-automated quality assessment because the generic quality model must be adapted manually beforehand but, therefore, allows easy adjustment and extension of the quality model. Future work will explore the applicability of the method to other types of soft sensors and use cases of simulation models like design space models. Additional factors, criteria and metrics should be introduced into the framework. Furthermore, new metrics must be developed if no universally accepted ones are available. Finally, the method should be tested in real world applications.

References

- O. Balci, 1998, Verification, validation, and testing, Handbook of simulation, 10, 8, 335-393
- K. Beck, 2003, Test-driven Development: By Example, Addison-Wesley
- L. Bruckner, M Oppelt, L. Urbas, M. Barth, 2020, The current and future use of simulation in discrete and process industries (white paper)
- J. Cavano , J. McCall , 1978, A framework for the measurement of software quality, Proceedings of the software quality assurance workshop on Functional an performance issues, 133-139
- ISO/IEC, 2011, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models, Standard ISO/IEC 25010:2011, International Organization for Standardization, Geneva, CH
- O. Kahrs, W. Marquardt, 2007, The validity domain of hybrid models and its application in process optimization, Chem Eng Process, 46, 11, 1054-1066
- H. Liu, S. Shah and W. Jiang, 2004, On-line outlier detection and data cleaning, Computer & chemical engineering, 28, 9, 1635-1647
- D. Montgomery and G. Runger, 2010, Applied statistics and probability for engineers. John Wiley & Sons
- Z. Nagy, 2007, Model based control of a yeast fermentation bioreactor using optimally designed artificial neural networks, Chemical engineering journal, 127, 1-3, 95-109
- B. Roungas, S. Meijer, A. Verbraeck, 2018, A framework for optimizing simulation model validation & verification, Int. Journal On Advances In Systems and Measurements, 11, 1&2
- R. G. Sargent, 2013, Verification and validation of simulation models, J. Simul., 7, 1, 12–24
- R. G. Sargent, O. Balci, 2017, History of verification and validation of simulation models, 2017 Winter Simulation Conference, IEEE, 292-307
- VDI, 2020, Modular plants – Fundamentals and planning modular plants – Part 1, Standard VDI 2776:2020-11, 2020, Beuth Verlag
- S. Wagner, 2013, Software Product Quality Control

Studying the Synergy between Dimension Reduction and Clustering Methods to Facilitate Fault Classification

Melisa Mollaian^{a,*}, Gyula Dörgö^b, Ahmet Palazoglu^a

^a*Department of Chemical Engineering, University of California, Davis, 95616 CA, USA*

^b*MTA-PE Lendület Complex Systems Monitoring Research Group, Department of Process Engineering, University of Pannonia, Egyetem u. 10, Veszprém, H-8200, Hungary*
mmollaian@ucdavis.edu

Abstract

Detection of abnormal behaviors (faults) in chemical process systems is a challenging task due to the high amount of data collected through measurements associated with distributed sensor networks. In these multivariate systems, gaining insight into the inner structure of collected high-dimensional data is essential for exploring and discovering paths to assist big data technologies. This exploration is facilitated by the use of many tools, such as dimension reduction techniques and clustering methods. In the present paper, the synergy between dimension reduction techniques and clustering methods has been studied to help discovering methods which complement each other in the field of fault detection and to demonstrate the performance of a fault detection strategy which is tested on the data from the well-known Tennessee Eastman Process simulator. The outcomes can be substantially improved by judiciously pairing the dimension reduction approach with a method from the appropriate group of clustering methods.

Keywords: Data mining, Dimension reduction, Data clustering, Fault detection, High-dimensional data

1. Introduction

Control systems of chemical processes require constant monitoring of different variables throughout the system. The historical data obtained from different plant sections and system variables are high-dimensional and complex, yet they are an extremely useful resource for discovering patterns and behaviors, thereby facilitating the detection of faults and other anomalies. Many fault detection methods utilizing high-dimensional historical data have been developed throughout the years and benefit from the use of data science tools such as dimension reduction techniques and clustering methods, or both. Each of these tools have separately been studied in the context of various applications and evaluated based on their success in class representation. Applying a dimension reduction technique helps to extract the most important features of a high-dimensional dataset and represent them using lower dimensions, and a clustering method is used to identify groups and classes containing those features. It should be pointed out that the attributes of dimension reduction and clustering methods have a direct effect on the outcomes. Therefore, a study on such attributes will help discovering how dimension reduction techniques and a common unsupervised learning tool, clustering, can complement each other in the field of fault detection. Utilizing combinations which complement each other

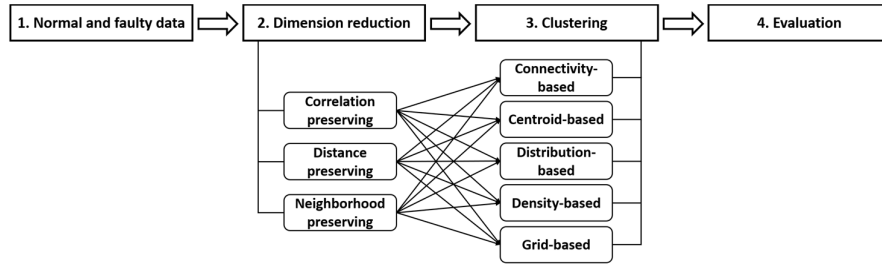


Figure 1: A step-wise summary of the presented approach

improves information extraction from the data and the fault detection process. There have been previous studies such as the work done by Thomas et al. (2018) demonstrating some methods are more compatible with each other. The reason of such observations is explored and discussed in the current paper by studying a wider range of methods.

The dimension reduction techniques in this work are classified as correlation-preserving methods, distance-preserving methods and neighborhood-preserving methods while the clustering methods are categorized as connectivity-based, centroid-based, distribution-based, density-based and grid-based techniques. As shown in Figure 1, the proposed approach consists of a first step where the data containing different states are normalized, followed by a dimension reduction step then a clustering step, ending with the evaluation of the results. Different combinations of categories are tested and characteristics of each group of methods is taken into account to match each class of dimension reduction technique with one or more groups of clustering methods to achieve the best overall functionality of the fault detection process. Each of the methods are briefly explained in Sections 2 and 3 with a focus on their definitive attributes. Evaluation criteria is explained in Section 3. The proposed approach is tested on a dataset containing three different types of faults from the Tennessee Eastman Process simulator mentioned in Section 4. The discussion of the results is available in Section 5.

2. Dimension reduction techniques

The techniques utilized in this study for dimension reduction and visualization purposes are categorized into three groups based on their approach to map the original high-dimensional data onto the low-dimensional space. The considered attributes leading to choose these categories are the input type of the methods, meaning whether they use a distance matrix or the data points for projection, and their examined neighborhood, meaning whether local points are taken into account or they are examined globally (Espadoto et al., 2019).

In the first category, correlation-preserving methods include the ones that consider and preserve the correlation between the samples, whether in a global scale or a local scale. Examples of this category are Principal Component Analysis (PCA) and its variants. PCA is a linear transformation for dimension reduction with the objective of preserving the maximum variance within the data which is performed on a global scale.

Techniques in the second category, distance-preserving methods, preserve distances between samples in a global scale to map the data onto a lower number of dimensions. The methods chosen for this category are Multidimensional Scaling (MDS) and Isometric mapping (Isomap). MDS is a nonlinear method which represents the distances of the data points as similarities between them and Isomap is a nonlinear embedding method which

constructs a neighborhood graph of the samples to find the shortest path between two nodes.

Third category of dimension reduction techniques is neighborhood-preserving methods. Dimension reduction techniques such as t-Stochastic Neighborhood Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) preserve the distance between data at a local scale. t-SNE is a nonlinear technique which represents local structures of a dataset by considering the affinities of its data points as probabilities. UMAP is also a nonlinear method similar to t-SNE with assumption of uniformly distributed data.

3. Clustering methods

Clustering methods differ in defining the similarity criteria for identifying classes (clusters). Therefore, considered methods in this study fall into one of the five aforementioned categories. The approach towards clustering in each category along with a representative example is briefly explained in this section (Han et al., 2012).

3.1. Clustering categories

The methods in the connectivity-based category take the distance between data points into account to assign cluster memberships. Closer data points are assumed to be in similar clusters as opposed to further data points. The most well-known method in this category is agglomerative hierarchical clustering, using a number of possible closeness measures (linkages).

Methods in the centroid-based category find centroids in order to partition the data into a specific number of clusters. This is done by minimizing the distance of points from their closest centroid. The examples chosen for this category are *k*-Means and *k*-Medoids. The main difference between these two methods is their selection of the centroid.

The main idea exploited in the third category, distribution-based clustering, is the assumption that members of a cluster most likely belong to the same distribution. Points are assigned to clusters based on their probability of belonging to a distribution; Gaussian mixture models (GMM) is an example that assumes that the data is constructed of multiple Gaussian distributions (Reynolds, 2015).

In the fourth category, which includes density-based methods, clusters are defined as areas of high density separated by areas of low density. Higher density is defined as smaller regions with higher number of samples. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Ordering Points To Identify the Clustering Structure (OPTICS) are two examples in this category. Both methods assign points to clusters based on a density criterion, while their difference can be simply expressed that OPTICS orders the points to prioritize the memberships.

Methods in the last category divide the space into grids and calculate the density of each grid to find the clusters. These methods are called grid-based clustering methods. An example is CLustering In QUest (CLIQUE). This method partitions high-dimensional data into non-overlapping subspaces and uses these subspaces to identify clusters.

3.2. Cluster evaluation criteria

Multiple metrics have been defined to assess the performance of a clustering algorithm. Metrics used in this study utilize the ground truth class assignments of the data points for evaluation. *Adjusted Rand Index* (ARI) (Rand, 1971) and *Adjusted Mutual Information* (AMI) (Vinh et al., 2009) measure the similarity of the true labelling and the clustering labelling, while ignoring the permutations and with chance normalization, meaning random assignments will have a score close to zero. ARI is bounded between -1 and 1, and AMI has an upper bound of 1, while 1 being the perfect score for both. *Homogeneity*

is a measure of clusters being pure, and *Completeness* measures if all members of a class have been assigned to the same cluster. *V-measure* is their harmonic mean, and all three scores are bounded between 0 and 1, 1 being the perfect score (Rosenberg and Hirschberg, 2007).

4. Case study

In order to study the synergy between categories of dimension reduction techniques and clustering methods, all the methods mentioned above were tested on a dataset generated using the Tennessee Eastman Process (TEP) simulator in MATLAB (Bathelt et al., 2015). This benchmark chemical plant consists of five major units (a reactor, a product condenser, a recycle compressor, a vapor-liquid separator, and a product stripper) and eight components in total and has been the reference for testing many data-driven fault detection methods in chemical engineering. The simulation has 41 measured variables and 12 manipulated variables. The original simulation has 20 predefined faults of different types that can be introduced to the system. For this study, three different faults of varying types were selected; Fault 2 which is a step change in a component's composition, Fault 13 which is a slow drift in the reaction kinetics, and Fault 14 which is the sticking of reactor cooling water valve. Each of these faults were separately active for a constant 20 minutes, and then turned off right before the next period started.

The generated dataset consisting of these three faults and 20 minutes of normal operation (without any faults) was first preprocessed by removing the constant variables and normalizing to a zero mean and unit variance. A dimension reduction technique was applied to the dataset followed by a clustering method. All the dimension reduction techniques were used to obtain two-dimensional data, and the parameters for each clustering method were searched for based on the scores described in Section 3.2. They were selected such that the clustering would result in the highest possible scores for maximum number of metrics; In other words, the clustering assignments would match the true labels of the four different states as much as possible. Hence, better scores (closer to 1) do not necessarily correspond to correct number of clusters, but they may contain some extra clusters or some combined clusters. Every combination of mentioned dimension reduction techniques and clustering methods were tested as mentioned to find the categories that complement each other more successfully in isolating the different faults.

5. Results and discussion

The ability of a clustering method to accurately detect different states and faults as separate clusters in this case is highly dependent on the ability of the dimension reduction technique to extract the inner structure and information from the original high-dimensional data. Therefore, if some states are not distinguished from each other after the first step of dimension reduction, complete fault isolation in clustering is almost impossible. For example, in the visualization of the two-dimensional data, it was observed that data points of Fault 14 were close to those of the normal state or even were indistinguishable in most of the cases.

The results of the case study are summarized in Table 1. For ease of comprehension and to satisfy the space restrictions, only the AMI scores for different cluster assignments are shown for each dimension reduction technique. The best score for each dimensional reduction technique is highlighted.

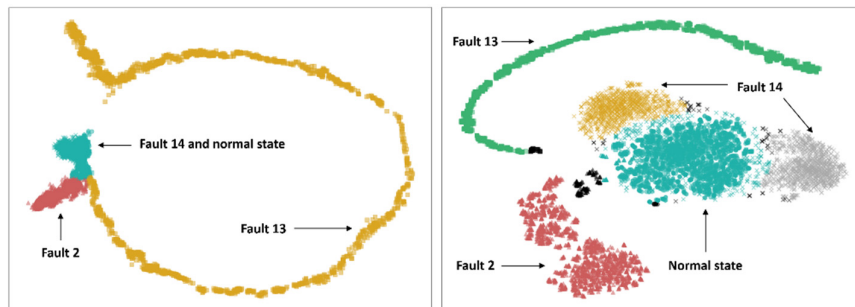
As presented in Table 1, for the correlation-preserving category (e.g., PCA), a density-based clustering had the best performance, and the distribution-based clustering was the

Table 1: AMI Scores for Cluster Assignments After Dimension Reduction

	Hierarchical	k-Means	k-Medoids	GMM	DBSCAN	OPTICS	CLIQUE
PCA	0.511	0.513	0.455	0.630	0.645	0.683	0.373
MDS	0.539	0.510	0.427	0.553	0.600	0.501	0.433
ISO	0.655	0.589	0.573	0.440	0.707	0.700	0.426
t-SNE	0.766	0.610	0.716	0.642	0.785	0.668	0.766
UMAP	0.685	0.595	0.575	0.628	0.704	0.673	0.433

method with the second-best performance. For the distance-preserving category (e.g., MDS and Isomap), DBSCAN achieved the highest AMI score for both techniques. The best clustering result for the neighborhood-preserving category was also DBSCAN. For both methods in this category, hierarchical clustering also produced results which were relatively close to the highest scores. Other evaluation scores demonstrated same results, confirming the most compatible clustering methods and dimension reduction techniques represented in Table 1 for each category. Also, it can be seen the lowest performance for PCA, Isomap and UMAP was obtained with the grid-based clustering method (e.g., CLIQUE), and the lowest performance for MDS and t-SNE were from the centroid-based clustering methods. These two categories of clustering methods do not work well with elongated clusters and tend to separate a long cluster into several smaller ones and as demonstrated later in this section, Fault 13 has such a feature. This fault is a slow drift in the reaction kinetics, and this is well-reflected in the elongated shape of its cluster in the low-dimensional space.

In total, for all five dimension reduction techniques, a density-based clustering method achieved the best performance in accurately isolating the faults, and distance-based clustering was amongst one of the methods with a relatively high score for two of the three categories of dimension reduction techniques. Overall consideration of the results demonstrates that density-based clustering methods lead to higher AMI scores in general, therefore more accurate cluster assignments. Most of the tested dimension reduction techniques represent similarity of data points in terms of their closeness in the lower dimensions, therefore the clusters can be found using a method which searches for areas with high density or clusters them based on the relative distance of data points more easily. An additional observation can be made by comparison of dimension reduction techniques, stating that the clustering results obtained after t-SNE demonstrate better performance compared to all other techniques for the dataset at hand.



(a) Isomap Dimension Reduction

(b) t-SNE Dimension Reduction

Figure 2: DBSCAN clustering applied on (a) Isomap and (b) t-SNE results

Highest AMI scores were achieved using t-SNE followed by DBSCAN clustering and using Isomap followed by DBSCAN. These results are presented in Figure 2. Different colors represent different clusters found through the mentioned steps and the arrows pointing to each cluster show the true categories of the present states on the figures. Three obtained clusters can be seen in Figure 2(a). Two faults (2 and 13) were detected successfully and two of the states (Fault 14 and no fault) were almost overlapping and combined into a single cluster in color blue. Figure 2(b) shows five different clusters in addition to outliers (in black) found by DBSCAN. All four different states were detected almost perfectly, with the exception of the yellow and gray clusters which are labeled separately but both are Fault 14. It can be seen that Fault 14 is mapped very closely to the normal state with no faults.

6. Conclusions

This presented work focused on studying the performance of categories of clustering methods when used in conjunction with categories of dimension reduction techniques to enhance the fault detection process. The tests were carried out on a case study from the TEP simulator to find the dimension reduction techniques and clustering methods that complement each other. Different combinations of dimension reduction techniques followed by clustering methods were applied to the dataset to obtain the highest similarity in cluster assignments to true data labels. The study shows that different categories of dimension reduction tend to produce results with specific characteristics leading to their higher compatibility with some categories of clustering methods. More specifically, the results show that the density-based and distance-based clustering categories have a promising performance in identifying isolated states and possible faults used along with all dimension reduction categories.

References

- A. Bathelt, N. L. Ricker, M. Jelali, 2015. Revision of the Tennessee eastman process model. *IFAC-PapersOnLine* 48:309–14.
- M. Espadoto, R. M. Martins, A. Kerren, N. S. T. Hirata, A. C. Telea, 2019. Towards a Quantitative Survey of Dimension Reduction Techniques. *IEEE Transactions on Visualization and Computer Graphics*, 1–1.
- J. Han, M. Kamber, J. Pei, 2012. *Data mining: Concepts and techniques*, Third Edition. Waltham, Mass.:Morgan Kaufmann Publishers.
- W. M. Rand, 1971. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* 66, 846–850.
- D. Reynolds, 2015. *Gaussian Mixture Models*. Springer, Boston, MA.
- A. Rosenberg, J. Hirschberg, 2007. V-Measure: A conditional entropy-based external cluster evaluation measure. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 410–420.
- M. Thomas, W. Zhu, J. Romagnoli, 2018. Data mining and clustering in chemical process databases for monitoring and knowledge discovery. *Journal of Process Control* 67, 160–175.
- N. X. Vinh, J. Epps, J. Bailey, 2009. Information theoretic measures for clusterings comparison. *Proceedings of the 26th Annual International Conference on Machine Learning*, 1073–1080.

Computational Fluid Dynamics Modelling of Phenol Oxidation in a Trickle-Bed Reactor using 3D Eulerian Model

Tladi J. Makatsa ^{a,b}, Siwela J. Baloyi ^c, Cornelius M. Masuku ^{d*}

^a Department of Civil and Chemical Engineering, University of South Africa, Private Bag X6, Florida, 1710, South Africa

^b Advanced Materials Division, Mintek, Private Bag X3015, Randburg, 2125, South Africa

^c Council for Scientific and Industrial Research (CSIR), Built Environment (BE), Hydraulic Infrastructure Engineering (HIE), P.O Box 395, Pretoria, 0001, South Africa

^d Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA

cmasuku@purdue.edu

Abstract

Catalytic wet-air oxidation is an attractive solution to treating refractory wastewater. The process has been demonstrated at laboratory scale over a low-cost pillared clay catalyst in a trickle-bed reactor. However, complex interaction of fluid dynamics and reaction kinetics makes scaling up of laboratory reactors to industrial reactors very difficult. Changes in hydrodynamic parameters are significant when laboratory reactors are scaled up to commercial reactors. To understand the behaviour of fluids inside a trickle-bed reactor, a computational fluid dynamics model was developed from experimental data using an Euler–Euler model. A commercial software Fluent was used to study hydrodynamic behaviour, temperature distribution and oxidation process. The model indicated that a hot spot was formed near the centre of the reactor due to liquid mal-distribution. Moreover, incorporating monolithic structure in a reactor packing material helped to lower pressure drop due to low velocities inside monolith channels. Furthermore, when the reactor was modelled at 160 °C and 10 bar, phenol was completely oxidized to CO₂.

Keywords: Advanced Oxidation Processes, CFD modeling, Reactor Scale-Up, Eulerian Model.

1. Introduction

Phenolic wastewaters originate from several industrial processes making them a model pollutant for bio-toxic and non-biodegradable organic pollutants. Moreover, phenol is an intermediate product found in the oxidation of aromatic hydrocarbons (Makatsa et al, 2020). It is also listed as a priority pollutant by the United States Environmental

Protection Agency (Lal and Garg, 2014). The use of phenol in petroleum, petrochemical, pharmaceutical, paint, pulp and paper, plastic and refinery industries is common (Zhang et al, 2019). There are different wastewater treatment methods available and catalytic wet-air oxidation (CWAO) is gaining a lot of interest due to developments in heterogeneous catalysis (Baloyi et al, 2019). The reactor choice is very important in this process (CWAO) since it can affect the reaction mechanism, which might result in the formation of undesired products like polymers. These by-products are known to be responsible for catalyst deactivation by blocking access to active sites (Pintar and Levec, 1992).

Trickle-bed reactor (TBR) is widely used in many industrial processes. However, there are few challenges associated with this type of reactors such as measuring the interaction between gas, liquid, and solid flows including high-pressure losses. The reactor flow regime is dependent on superficial mass velocity, fluid properties (density, viscosity, etc.), and design parameters. In most instances, the reactor is operated between a low interaction regime (trickle flow regime) and a high interaction regime (pulse, spray, and bubble regime).

Ranade et al (2011) suggested that the scale of the reactor affects the performance of a TBR. Also, these authors listed several factors that are directly affected during reactor scale-up as follows; reactor to particle diameter ratio, reactor volume, bed porosity, wetting, channeling, liquid mal-distribution, dispersion and reactor operating mode (isothermal/adiabatic). After thorough investigation, these researchers concluded that the wall effect is predominant in laboratory TBR whereas flow mal-distribution is common in industrial TBR due to large bed diameter. On the other hand, when CFD model is developed correctly, the scale of the reactor can be taken into consideration. CFD models are based on conservation of mass, energy and momentum; and provide a time saving and cost-effective approach in the reactor design (Kapfunde et al, 2018).

The purpose of this study is to validate the results obtained in our previous work (Makatsa et al, 2019) using a 3D multiphase Eulerian model of phenol oxidation in a trickle bed reactor. To the best of our knowledge, this is the first study to simulate CWAO of phenol using a novel Al/Zr-PILC catalyst supported on a cordierite monolith.

2. Governing equations

A multiphase Eulerian CFD model of phenol oxidation in a TBR was developed using commercial software ANSYS Fluent 2019R2 and 2020R2. The following sets of mathematical equations are incorporated into a CFD code solver.

Mass conservation equation:

$$\frac{\partial (\varepsilon_k \rho_k)}{\partial t} + \nabla \cdot (\varepsilon_k \rho_k U_k) = 0 \quad (1)$$

Momentum conservation equation:

$$\frac{\partial (\varepsilon_k \rho_k U_k)}{\partial t} + \nabla \cdot (\varepsilon_k \rho_k U_k U_k) = -\varepsilon_k \nabla P_k + \nabla \cdot (\varepsilon_k \mu \nabla U) + \varepsilon_k \rho_k g + F_{K,R} (U_k - U_r) \quad (2)$$

Where ε_k is volume fraction for each phase, ρ_k is the density of the k-th phase, U_k is the cell velocity of the k-th phase and $F_{K,R}$ is an interphase momentum exchange (Ranade et al, 2011). The interface coupling term $F_{K,R}$ can be expressed as follow

$$F_{GL} = \varepsilon_G \left(\frac{E_1 \mu_G (1 - \varepsilon_G)^2}{\varepsilon_G^2 d_p^2} \left[\frac{\varepsilon_s}{(1 - \varepsilon_G)} \right]^{0.667} + \frac{E_2 \rho_G (U_G - U_L) (1 - \varepsilon_G)}{\varepsilon_G d_p} \left[\frac{\varepsilon_s}{1 - \varepsilon_G} \right]^{0.333} \right) \quad (3)$$

$$F_{GL} = \varepsilon_G \left(\frac{E_1 \mu_G (1 - \varepsilon_G)^2}{\varepsilon_G^2 d_p^2} \left[\frac{\varepsilon_s}{(1 - \varepsilon_G)} \right]^{0.667} + \frac{E_2 \rho_G U_G (1 - \varepsilon_G)}{\varepsilon_G d_p} \left[\frac{\varepsilon_s}{1 - \varepsilon_G} \right]^{0.333} \right) \quad (4)$$

$$F_{LS} = \varepsilon_L \left(\frac{E_1 \mu_G \varepsilon_S^2}{\varepsilon_G^2 d_p^2} + \frac{E_2 \rho_L U_G \varepsilon_S}{\varepsilon_L d_p} \right) \quad (5)$$

Where F_{GL} , F_{GS} , F_{LS} are gas-liquid, gas-solid and liquid-solid momentum exchange terms. To understand turbulence inside the reactor, a standard k- ε model was chosen and the software solved the following mathematical equations (Lopes and Quinta-Ferreira, 2010);

$$\mu_{t,L} = \rho_L C_\mu \frac{k_L^2}{\varepsilon_L} \quad (6)$$

The liquid viscosity turbulence $\mu_{t,L}$ is calculated from the transport equations by determining kinetic (k_L) and dissipation energy (ε_L) from the following equations (Lopes and Quinta-Ferreira, 2007)

$$\frac{\partial}{\partial t} (\rho_L \alpha_L k_L) + \nabla \cdot (\rho_L \alpha_L \vec{u}_L k_L) = \nabla \cdot \left(\alpha_L \frac{u_{t,L}}{\sigma_k} \nabla k_L \right) + \alpha_L G_{k,L} - \alpha_L \rho_L \varepsilon_L + \alpha_L \rho_L \Pi_{k,L}$$

$$\frac{\partial(\rho_L \alpha_L k_L)}{\partial t} + \nabla \cdot (\rho_L \alpha_L \bar{u}_L k_L) = \nabla \cdot \left(\alpha_L \frac{u_{t,L} \nabla k_L}{\sigma_k} \right) + \alpha_L G_{k,L} - \alpha_L \rho_L \Pi_{k,L} \quad (7)$$

$$\frac{\partial(\rho_L \alpha_L \varepsilon_L)}{\partial t} + \nabla \cdot (\rho_L \alpha_L \bar{u}_L \varepsilon_L) = \nabla \cdot \left(\alpha_L \frac{u_{t,L} \nabla \varepsilon_L}{\sigma_k} \right) + \alpha_L \frac{\varepsilon_L}{k_L} \times (C_{1\varepsilon} G_{k,L} + C_{2\varepsilon} \rho_L \varepsilon_L) + \alpha_L \rho_L \Pi_{\varepsilon,L} \quad (8)$$

The following parameters were taken as constants C_μ , $C_{1\varepsilon}$, $C_{2\varepsilon}$, σ_k , σ_ε and assigned the following values 0.09, 1.44, 1.92, 1.0 and 1.3, respectively. Enthalpy was calculated from the conservation of energy in a multiphase Eulerian model as follows (Manoharan and Buwa, 2019; Lopes and Quinta-Ferreira, 2007).

$$\frac{\partial(\rho_q \alpha_q h_q)}{\partial t} + \nabla \cdot (\rho_q \alpha_q \bar{u}_q h_q) = -\alpha_q \frac{\partial p_q}{\partial t} + \bar{\tau}_q : \nabla \bar{u}_q - \nabla \cdot \bar{q}_q + S_q + \sum_{p=1}^n (\bar{Q}_{p,q} + \dot{m}_{pq} h_{pq} - \dot{m}_{qp} h_{qp}) \quad (9)$$

The specific enthalpy of phase q is represented by h_q and \bar{q}_q is a heat flux. The heat exchange intensity between the q and p phases is represented by \bar{Q}_{pq} whereas interphase enthalpy is represented by h_{pq} and S_q is the source term. By activating species transport the solver modeled volumetric reaction using the following equation (Lopes and Quinta-Ferreira, 2010):

$$\frac{\partial \alpha_q \rho_q C_{q,i}}{\partial t} + \nabla \cdot (\alpha_q \rho_q u_q C_{q,i}) = \nabla \cdot (\alpha_q \rho_q D_{q,i} \nabla C_{q,i}) + \alpha_q \rho_q S_{q,i} \quad (10)$$

The kinetic parameters for Al/Zr-PILC catalyst were derived from our previous work (Makatsa et al, 2019). The term in eq. (10) includes the reaction rate as;

$$-r_{ph} = k_{ob} \cdot C_{ph}^\alpha \quad (11)$$

And K_{ob} can be expressed as follow,

$$k_{ob} = k_o \cdot \exp\left(\frac{-E_{ob}}{RT}\right) \cdot P_{O_2}^\beta \quad (12)$$

By assuming ideal plug flow and first order with respect to phenol, K_{ob} can be expressed as follow,

$$k_{ob} = -\left(\frac{1}{\tau}\right) \ln(1 - x_{ph}) \quad (13)$$

3. Results

A mixture of phenol (C_6H_5OH) and oxygen (O_2) was fed to the isothermal-isobaric reactor operated at 160 °C and 10 bar. The gas and liquid inlet velocities were kept constant at 0.012 and 0.00007 m/s, respectively. As shown in Figure 1, the concentration profile of phenol inside the reactor is close to zero along the reactor bed. This is due to increased residence time and improved kinetics. These results are supported by Figure 2(a), which shows the contours of phenol mass fraction inside the reactor. The results show that the pollutant is highly concentrated at the top half of the reactor. However, the concentration is sharply decreased as the stream moves through the reactor, and phenol is completely oxidized to form carbon dioxide (CO_2) as shown in Figure 2(b). From these findings, it can be concluded that C_6H_5OH was completely mineralized to CO_2 and water. Our results are consistent with the findings of (Lopes and Quinta-Ferreira, 2010), in their study phenolic acid was oxidized in a TBR and simulated using the Euler-Euler method at 160 and 200 °C. They concluded that 82% of total organic carbon (TOC) was converted at 160 °C whereas only 84.8% was converted at 200 °C.

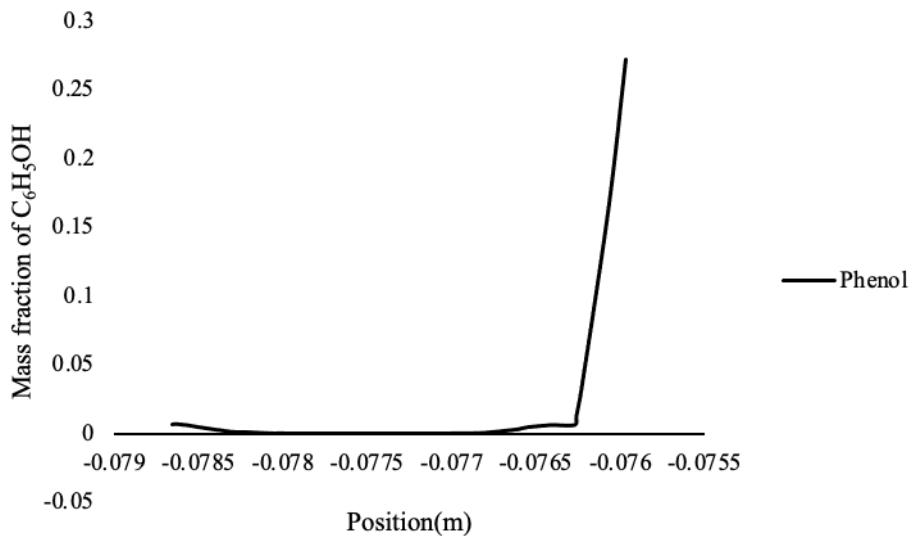


Figure 1: Mass fraction of phenol inside the reactor bed.

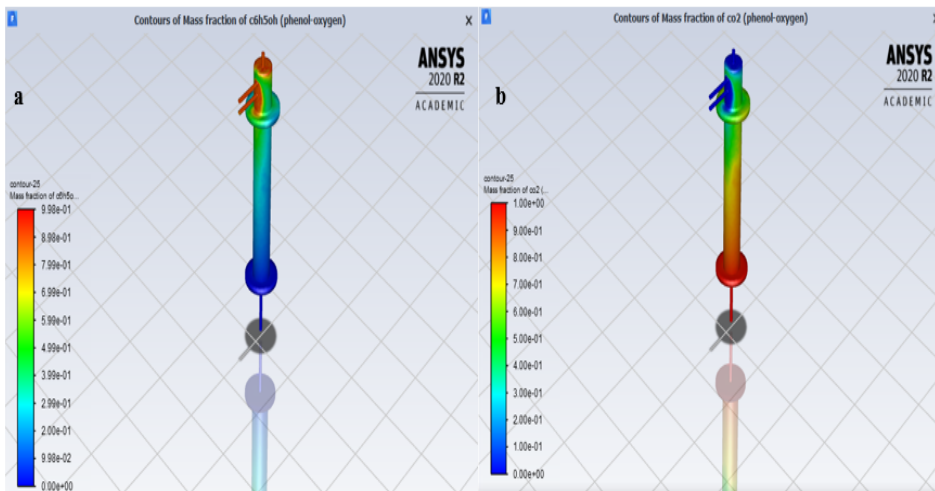


Figure 2: Contours of phenol mass fraction (a) and CO₂ profile inside the reactor (b).

4. Conclusions

A computational fluid dynamics model of a trickle-bed reactor was developed to simulate catalytic wet air oxidation of phenol and experimental results were used to validate the model. A multiphase Eulerian model coupled with energy and species transport equations was used to model the phases. The reactor model was simulated at 160 °C, 10 bar and CFD results showed that phenol was completely removed leaving a significant amount of CO₂.

References

- J. Baloyi, T. Ntho, J. Moma, 2019. Synthesis of highly active and stable Al/Zr pillared clay as catalyst for catalytic wet oxidation of phenol. *Journal of Porous Materials*, 26, 583–597. doi.org/10.1007/s10934-018-0667-3.
- N. Kapfunde, C.M. Masuku, D. Hildebrandt, 2018. Optimization of the thermal efficiency of a fixed-bed gasifier using computational fluid dynamics. *Computer Aided Chemical Engineering*, 44, 1747–1752. doi.org/10.1016/B978-0-444-64241-7.50286-X.
- K. Lal, A. Garg, 2014. Catalytic wet oxidation of phenol under mild operating conditions : development of reaction pathway and sludge characterization. *Clean Technologies and Environmental Policy*, 17, 199–210. doi.org/10.1007/s10098-014-0777-9.
- R.J.G. Lopes , R.M. Quinta-Ferreira, 2007. Trickle-bed CFD studies in the catalytic wet oxidation of phenolic acids. *Chemical Engineering Science*, 62, 24, 7045–7052. doi.org/10.1016/j.ces.2007.08.085.
- R.J.G. Lopes, R.M. Quinta-Ferreira, 2010. Assessment of CFD Euler-Euler method for trickle-bed reactor modelling in the catalytic wet oxidation of phenolic wastewaters.

- Chemical Engineering Journal*, 160, 1, 293–301. doi.org/10.1016/j.cej.2010.03.024.
- T.J. Makatsa, S.J. Baloyi, T.A. Ntho, C.M. Masuku, 2019. Kinetic study of phenol oxidation in a trickle bed reactor over Al/Zr-pillared clay catalyst. *IOP Conference Series: Materials Science and Engineering*, 655, 012050. doi.org/10.1088/1757-899x/655/1/012050.
- T.J. Makatsa, J. Baloyi, T. Ntho, C.M. Masuku, 2020. Catalytic wet air oxidation of phenol: Review of the reaction mechanism, kinetics, and CFD modeling. *Critical Reviews in Environmental Science and Technology*. doi.org/10.1080/10643389.2020.1771886.
- K.G. Manoharan, V.V. Buwa, 2019. Structure-resolved CFD simulations of different catalytic structures in a packed bed. *Industrial & Engineering Chemistry Research*, 58, 49, 22363–22375. doi.org/10.1021/acs.iecr.9b03537.
- A. Pintar and J. Levec, 1992. Catalytic Oxidation of Organics in Aqueous Solutions: I. Kinetics of phenol oxidation. *Journal of Catalysis*, 135, 2, 345–357. doi.org/10.1016/0021-9517(92)90038-J.
- V.V. Ranade, R.V. Chaudhari, P.R. Gunjal, 2011. *Trickle bed reactors*. 1st Ed., Elsevier eBook.
- Y. Zhang, C.M. Masuku, L.T. Biegler, 2019. An MPCC reactive distillation optimization model for multi-objective Fischer–Tropsch synthesis. *Computer Aided Chemical Engineering*, 46, 451–456. doi.org/10.1016/B978-0-12-818634-3.50076-X.

Enhanced deterministic approach for heat exchanger network synthesis

Zekun Yang,^a Nan Zhang,^{a*} and Robin Smith^a

^a *Centre for Process Integration, Department of Chemical Engineering and Analytical Science, The University of Manchester, Manchester M13 9PL, UK*

*nan.zhang@manchester.ac.uk

Abstract

Stage-wise superstructure (SWS) is one of the most widely used methods for Heat Exchanger Network (HEN) synthesis. Approaches to obtain optimal HEN solution can be based on the deterministic algorithm, but it is difficult for solving a large-scale problem since the complexities arisen from nonlinearities of SWS. Several methods which employed stochastic algorithms and meta-heuristic approaches have been proposed to tackle the problem. However, it determines a near-optimal HEN configuration from a series of stochastic solutions which are obtained by the execution of many computational operations, thus it is time consuming and a global optimum will not be achieved from only randomly generated results. In this study, an enhanced SWS is presented, in which new constraints and variables are added to avoid conflicted calculation of non-isothermal mixing energy balance and reduce the redundant combinations. Moreover, the present model is extended to allow a flexible requirement of stream splitting for practical application. Then, a deterministic-based global solver (GAMS/BARON) is applied in solving three case studies, operated on computer: I7-8565U. The results showed that the proposed approach can provide a cost-efficient HEN solution with a lower TAC than that obtained from existing stochastic and deterministic algorithms.

Keywords: Heat exchanger network synthesis, Deterministic approach, Optimization, Process synthesis, Mathematical programming

1. Introduction

Heat exchanger networks are important for the process industries, since it can reduce energy consumption and pollutant emissions by heat integration of process streams. The benefit can be attributed into HENs for achieving energy saving, but heat exchangers bring significant capital investment at the same time. To increase economic benefits, HEN optimization has been widely studied. The approaches used in HEN synthesis can be divided into pinch method and automated method. Pinch Technology has been widely applied and can provide a cost-efficient solution, but it requires experienced designers and may lead to missing promising solutions.

Mathematical programming has been developed to achieve automated HEN synthesis. It generally formulates the design problem through a superstructure. One of the widely used superstructures has been proposed by [Yee and Grossmann \(1990\)](#). Based on the use of stage-wise superstructure (SWS), the HEN synthesis is formulated as an MINLP problem, as targeting the minimum total cost. Due to non-linearity, non-convexity and raised problem sizes for large-scale HEN optimization problems, it is a challenge to find an

optimal HEN solution at accepted time. Several approaches have been reported to solve the HEN synthesis problem by using different algorithms.

Stochastic algorithms treat the HEN design with randomization, not limited by model non-convexity and discontinuity. Some papers have applied these methods in solving HEN synthesis, such as: Genetic Algorithms (GA) (Holland, 1992), Simulated Annealing (SA) (Kirkpatrick, 1983), Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995). Based on these, a number of papers have been published to use hybrid-approach with developed stochastic algorithms. Recently, GA-PSO and SA-RFO methods have been applied to solve the models (Pavão et al., 2016; Pavão et al., 2018) respectively, in which parallel processing techniques and streams with phase change were considered.

However, for employing stochastic algorithms, extensive computational efforts are required to determine near-optimal solutions. Deterministic approaches have been widely employed to solve HEN synthesis problems, such as Branch and Bound (BB) and Outer Approximation (OA) (Zamora and Grossmann, 1997). In recent years, study (Huang et al., 2012) reported a modified superstructure to consider non-isothermal mixing, and solved by commercial solver BARON/GAMS. Faria et al. (2015) developed the Bound Contraction method to solve the HEN synthesis problem, in which variables are partitioned to construct a linear or convex lower bound. Kim et al. (2017) presented a deterministic-based global optimization, which extended the works (Faria et al., 2015). However, large time-resource is necessary according to their approach, and industrial-scale problems have not been investigated by the use their method.

In this work, based on the SWS developed by Huang et al. (2012) and Pavão et al. (2017), an enhanced non-isothermal mixing superstructure is proposed. A global solver, BARON/GAMS is employed as the deterministic approach. Additional variables and constraints are defined to improve the model performance and achieve flexible splitting requirement.

2. Mathematical Model

The mathematical formulation is based on the well-known stage-wise superstructure (SWS) proposed by Yee and Grossmann (1990), and stages are introduced in which all possible matches between hot streams and cold streams are optimized. In this formulation, the iso-thermal mixing assumed in the original SWS method is removed, and some logical constraints are adopted to improve the non-isothermal mixing model. Basically, the set of hot process streams ($i; i \in HP$), cold process streams ($j; j \in CP$), and stages ($k; k \in ST$) are considered. The number of stages is assumed as $ST = \max[i, j]$. Some following assumptions are described below:

- a. Constant stream flow rate, heat transfer coefficients and thermodynamic properties.
- b. Split stream pass exchangers in series and by-pass are not considered.
- c. Utility inlet and outlet temperatures are given.

2.1. Objective function, stage and overall energy balances

The objective function, stage and overall heat balances can be founded in the literature (Yee and Grossmann, 1990). The objective function is to minimize the total annual cost (TAC), which consists of the total capital cost related to the heat exchangers and utility cost. Stage and overall energy balances are required to achieve feasible heat transfer in each stage and each stream with heaters or coolers.

2.2. Non-isothermal mixing (NIM)

Different from the iso-thermal mixing, exchanger energy balance is applied to express the NIM heat transfer. For a hot stream (cold stream), the inlet temperature of a split stream enters the heat exchanger is the same as the stage inlet temperature $Th_{i,k}$ ($Tc_{j,k+1}$). The outlet temperature for each split stream can be different according to the varying heat capacities and heat duties. Continue variables $fh_{i,j,k}$ and $fc_{i,j,k}$ ($0 \leq fh_{i,j,k} \leq 1$; $0 \leq fc_{i,j,k} \leq 1$) are defined as split fractions of hot stream i and cold stream j . The true outlet temperatures of exchangers can be calculated by exchanger energy balance, as formulated:

$$(Th_{i,k} - Thout_{i,j,k+1})Cp_h fh_{i,j,k} \geq q_{i,j,k} \quad i \in HP, j \in CP, k \in ST \quad (1)$$

$$(Tcout_{i,j,k} - Tc_{j,k+1})Cp_c fc_{i,j,k} \geq q_{i,j,k} \quad i \in HP, j \in CP, k \in ST \quad (2)$$

Where $Thout_{i,j,k}$ and $Tcout_{i,j,k}$ are the actual outlet temperatures of exchangers. In addition, the model allows the stream to fully bypass a stage if no exchangers are required. Two continuous variables are defined as:

$$yh_{i,k} \begin{cases} 1 & \text{if hot stream } i \text{ bypass the stage } k \\ 0 & \text{Otherwise} \end{cases} \quad i \in HP, k \in ST \quad (3)$$

$$yc_{j,k} \begin{cases} 1 & \text{If cold stream bypass the stage } k \\ 0 & \text{Otherwise} \end{cases} \quad j \in CP, k \in ST \quad (4)$$

Based on Eq. (3) and Eq. (4), the constraints to direct feasible stream split fraction have been reported by Huang et al. (2012).

2.3. Feasibility constraints

To conduct a feasible HEN configuration, some process constraints reported by Yee and Grossmann (1990), and Huang et al. (2012) are used in this model, which targets the basic feasibilities of temperatures, assessments of temperatures, and feasible minimum approach temperature difference (MATD). Furthermore, the stage inlet temperature of hot stream i must be no less than the split stream temperature that enters the next stage $k + 1$. Similarly, the split stream temperatures of cold stream j as it leaves the stage k must be equal to or larger than the stage inlet temperature. Hence, some new assessments and constraints of temperatures are presented as follows:

$$Thin_i = Thout_{i,j,firstk} \quad i \in HP, j \in CP, k \in ST \quad (5)$$

$$Tcin_j = Tcout_{i,j,lastk} \quad i \in HP, j \in CP, k \in ST \quad (6)$$

$$Th_{i,k} \geq Thout_{i,j,k+1} \quad i \in HP, j \in CP, k \in ST \quad (7)$$

$$Tc_{j,k+1} \leq Tcout_{i,j,k} \quad i \in HP, j \in CP, k \in ST \quad (8)$$

For heat transfer in split streams, the temperature change must be 0 if the match does not exist. Upper bound of temperature change is required to formulate these constraints. The relevant correlation has been proposed by Huang et al. (2012).

3. Optimization

In this study, the proposed model is formulated as a non-convex MINLP problem, as applying global solver BARON to target a near global optimal solution under specified iteration or time-resource. Figure 1 illustrates the optimization procedures. If a feasible upper bound solution cannot be found within acceptable time, the required stage number k can be reduced gradually to decrease the number of binary variables. Additionally, LMTD equation is employed at the last step to avoid overestimated exchanger area obtained by using LMTD approximation.

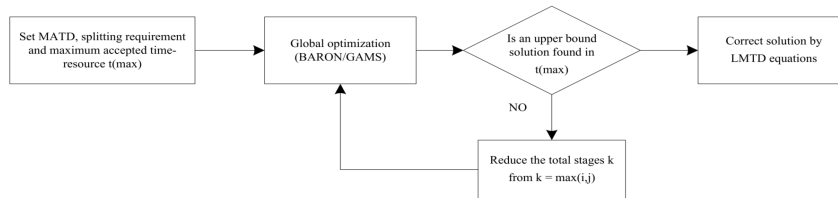


Figure 1. Block diagram of optimization approach

4. Case study

In this section, this case is a crude oil fractionation process originally studied by Kim et al. (2017) by a deterministic-based bounded method. In order to satisfy the practicality with industrial crude oil units, the total split numbers for crude oil stream, C1 and C2 should not exceed four respectively. Pavão et al. (2018) applied the SA-RFO method to solve this problem, and they noted that the solution obtained by the SA-RFO method is better than that reported by Kim et al. (2017). Nevertheless, in their work (Pavão et al., 2018), they ignored the split limitation and the split streams of C2 are larger than four, which becomes impractical for industrial crude oil systems. Hence, in order to evaluate the performance of the proposed method, two cases are conducted and compared with the results of Kim et al. (2017) and Pavão et al. (2018), respectively. The first case is based on limited splitting streams while introducing new constraints. The second case is obtained through the present method directly. To conduct HEN with limited branch number, Eqs (9) and (10) are formulated to achieve required splitting number (splitting number ≤ 4 for C1 and C2 respectively):

$$\sum_{i \in H P} z_{i,j,k} \leq 4 \quad j \in C P, k \in S T \quad (9)$$

$$\sum_{i \in H P} \sum_{k \in S T} z_{i,j,k} - \left(\text{stage number} - \sum_{k \in S T} y_{j,k} \right) + 1 \leq 4 \quad j \in C P \quad (10)$$

This problem is then solved with 11 stages and the MATD is set to 10 °C to be consistent with the literature (Kim et al., 2017). To find a high-quality solution at accepted computational time, maximum computation time is set (4 h in the case study). Furthermore, iteration numbers are recorded with the maximum value of 2500. For computational operations, GAMS 24.6 with BARON is used as the global MINLP solver. The computer model is a Dell Intel ® Core™, I7-8565U CPU, 1.99 GHz with 8 GB RAM.

Figure 2 and Table 1 show the optimal HEN configuration and results comparison with limited branch number, respectively. From the results, a better HEN solution is obtained as a lower TAC (3,455,358 \$/yr) and time-resource (98 % time-resource saving) than previous paper (Kim et al., 2017). Next, Figure 3 presents the optimal HEN solution without split limitation. According to Table 1, the total 3,379,651 \$/y TAC is obtained, which is lower than the best solution so far (3,391,066 \$/y), with 79 % computation time reduction.

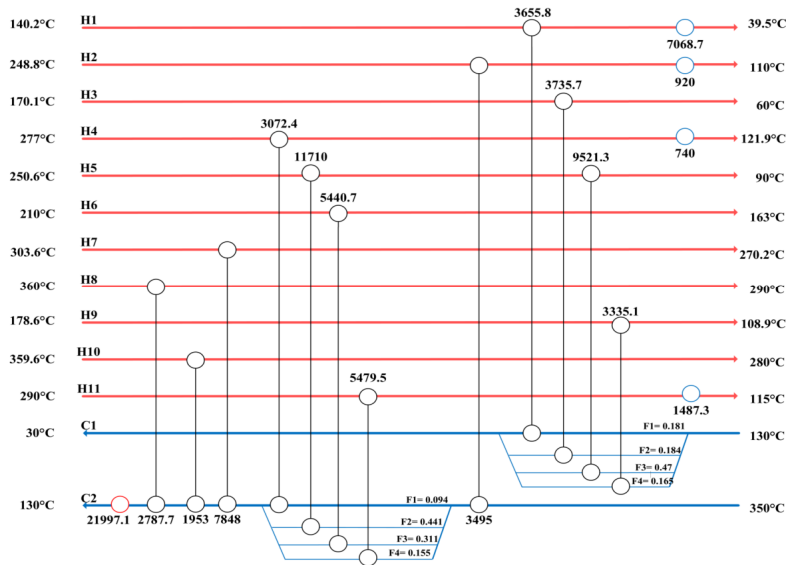


Figure 2. Optimal HEN solution while limiting split streams

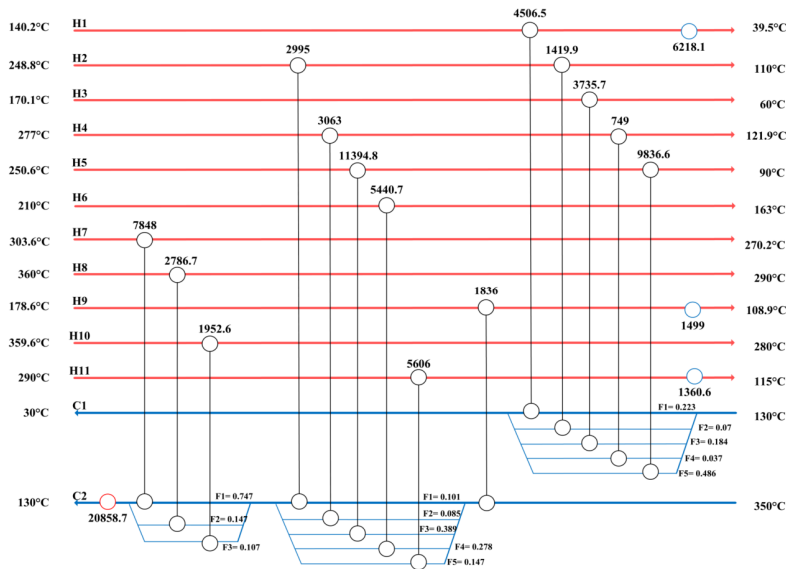


Figure 3. Optimal HEN solution without limiting split streams

Table 1: Results comparison

Reference	TAC (\$/y)	CPU (s)	Units	QH (kW)	QC (kW)	Total Area (m ²)
Kim et al.	3,456,649	47,783	17	23,566	11,783	10131.1
This work	3,455,358	534	17	21,997	10,216	13245.2
Pavao et al.	3,391,066	22,630	18	20,891	9108	13835.1
This work	3,379,651	4564	18	20,858	9077	13690.9

5. Conclusions

Consequently, the proposed method shows better performance than the previous approaches. A cost-efficient HEN solution can be obtained by the enhanced deterministic approach with acceptable time. Different from the existing stochastic algorithms that may have inconsistent solutions in different executions, the deterministic approach applied in proposed method can give better solutions, shown with cost-efficient HEN solutions with a lower TAC and computation time than that obtained from the existing stochastic (0.34 % TAC saving, 79 % computation time reduction) and deterministic algorithms (0.04 % TAC saving, 98 % computation time reduction).

References

- Faria, D. C., Kim, S. Y., & Bagajewicz, M. J. (2015). Global Optimization of the Stage-wise Superstructure Model for Heat Exchanger Networks. *Industrial & Engineering Chemistry Research*, 54, 1595-1604.
- Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*: MIT press.
- Huang, K. F., Al-mutairi, E. M., & Karimi, I. A. (2012). Heat exchanger network synthesis using a stagewise superstructure with non-isothermal mixing. *Chemical Engineering Science*, 73, 30-43.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks (Vol. 4, pp. 1942-1948 vol.1944)*.
- Kim, S. Y., Jongsuwat, P., Suriyaphadilok, U., & Bagajewicz, M. (2017). Global Optimization of Heat Exchanger Networks. Part 1: Stages/Substages Superstructure. *Industrial & Engineering Chemistry Research*, 56, 5944-5957.
- Kirkpatrick, S. (1983). DG Jr., and MP Vecchi. Optimization by simulated annealing. *science*, 220, 671-680.
- Pavão, L. V., Costa, C. B. B., & Ravagnani, M. A. d. S. S. (2016). Automated heat exchanger network synthesis by using hybrid natural algorithms and parallel processing. *Computers & Chemical Engineering*, 94, 370-386.
- Pavão, L. V., Costa, C. B. B., Ravagnani, M. A. d. S. S., & Jiménez, L. (2017). Large-scale heat exchanger networks synthesis using simulated annealing and the novel rocket fireworks optimization. *AIChE Journal*, 63, 1582-1601.
- Pavão, L. V., Costa, C. B. B., & Ravagnani, M. A. S. S. (2018). A new stage-wise superstructure for heat exchanger network synthesis considering substages, sub-splits and cross flows. *Applied Thermal Engineering*, 143, 719-735.
- Yee, T. F., & Grossmann, I. E. (1990). Simultaneous optimization models for heat integration—II. Heat exchanger network synthesis. *Computers & Chemical Engineering*, 14, 1165-1184.
- Zamora, J. M., & Grossmann, I. E. (1997). A comprehensive global optimization approach for the synthesis of heat exchanger networks with no stream splits. *Computers & Chemical Engineering*, 21, S65-S70

Modelling Biogas and H₂S in Anaerobic Digestion

Bozzano Giulia Luisa, Moretta Federico, Rizzo Eleonora and Manenti Flavio
Politecnico di Milano, piazza Leonardo da Vinci, 32, Milan (MI) – 20133, Italy
giulia.bozzano@polimi.it

Abstract

Green Economy is the key word of the last two decades. In this context, an important role is covered by anaerobic digestion (AD), treating organic waste fractions to recover both fertilizing elements and energy. Countless biochemical reactions lead to substrate degradation and biogas development. Despite AD has been used for many years, its study from an engineering point of view is quite recent. The reference model ADM1 (Anaerobic Digestion Model No.1), proposed in the late 90's by IWA (Batstone et al., 2002) to describe and simulate the steady-state digestion of sewage sludges from Wastewater Treatment (WWT) plants, has gone through several adjustments during time. The aim of this work is to improve and extend ADM1 by means of coefficients and equations optimization to better fit various contexts such co-digestion processes. Moreover, provide a description of the complete biokinetic of H₂S, main biogas impurity, through the homologation of Fedorovich and Kalyuzhnyi studies (1998), which perfectly address its production from the organic feedstock reduction by means of Sulphur Reducing Bacteria (SRB) bacteria. Introducing the novelty of H₂S biological abatement process, whose mechanism spontaneously occurs under microaeration condition, also requires the implementation of an inhibition function since O₂ will affect methane production. Python™ is the coding language chosen in order to solve a highly stiff ODE system and earlier literatures are the roots of this new model, that has been validated on real data resulting in 0.2%vol. H₂S and 2%vol. unreacted O₂ release.

Keywords: AD Process, H₂S Biokinetics, Python™, Mathematical Modelling.

1. Introduction

Anaerobic digestion consists in the molecular cleavage of organic matter, carried out by many different families of microorganisms, living in an oxygen free environment defined as anaerobic condition. The final products of this biological transformations are biogas, a gaseous mixture mainly composed by CH₄ (50-60 %vol.), CO₂ (25-50 %vol.), H₂ (0-1 %vol.) and digestate, the semi-solid residue that can be converted into fertilizer and used for rural applications. The feedstock, characterized in terms of carbohydrates, lipids, proteins, and inert materials, undergoes four reaction stages in which the substrate is exploited by bacteria as source of nutrients. The hydrolysis reaction converts carbohydrates, lipids and proteins into their respective monomers i.e. sugars, fatty acids and amino acids by adding a molecule of water to break their polymeric chains. This reaction represents the Rate Determining Step of the entire process. Then, Acidogenesis follows. This is responsible for the production of several short acid molecules (e.g. valeric, butyric and propionic acid and CO₂) accompanied by high pH reduction to around 4.5~5.5. These acids are converted into CH₃COOH and H₂ during Acetogenesis, but however the latter has a strong inhibition effect over some microorganisms, it's quickly consumed during the last step of Methanogenesis. In this way they can grow, duplicate,

and generate other chemicals as their intake wastes. Hydrogenotrophic bacteria convert CO_2 into CH_4 consuming H_2 (Eq. 1) while Acetoclastic bacteria produce CH_4 from CH_3COOH (Eq. 2). So, the overall reactive step stabilizes the final pH level inside the digester at a value about 6–8 (this strictly relies upon feedstock characteristics).



In order to simply represent the trends of these elements, the used kinetic model is a non-structured and non-segregated one, meaning that all the bacterial cells are equivalent in terms of age, size and metabolism abilities, so the reactions are not influenced by the intracellular structure itself but only depends on operative conditions. All these concepts were applied for the construction of Anaerobic Digestion Model No.1 (ADM1), that is a complex Monod-type kinetic created by IWA in 1997. The process described is a wet AD (total solid content < 10 %wt.) performed in a CSTR constant volume in mesophilic conditions (i.e., 35 °C). It consists in an ensemble of 35 differential equations representing the mass balances for all the substrates and bacteria families involved into the process (Batstone et al., 2002). Then, eight additional algebraic equations are needed for pH computation, balance closure and biogas partial pressure evaluations. General equations are functions of ρ , the production term summarizing the reaction rates expressed through 1st order or Monod kinetics, accounting also for different inhibition functions (e.g., low pH value, H_2 , NH_3), moreover acid-base ions dissolution equilibrium and liquid-gas mass transfer expressions ρT_i appear in ADM1 equations.

2. Model improvements

As reported in the previous paragraph, ADM1 was firstly developed for digestion processes which treat principally wastewater. In the last decades, different authors proposed extensions and modifications to the model in order to adapt it to the digestion of different substrates. In H. Li et. al (2019) a mathematical model is proposed to simulate manure mono-digestion in order to get the dynamics of carbon, phosphorus and nitrogen components in wet condition and high total-solid content and it was found valid to describe a high total-solid inhibition. Rivas-Garcia et al. (2020), in their studies, have tried to extend ADM1 to co-digestion proposing different values for kinetic parameters through regression of experimental data. In summary, to properly run a real-case simulation, ADM1 has to be adapted to the anaerobic digestion of various feedstocks, other than sewage sludges. Being the model based on default parameters, these have been investigated to better fit meaningful different dynamics behavior of all the species through iterative approximations starting from previous studies (Rivas-Garcia et al., 2020; Heng Li et al., 2019; F. Blumensaat et al., 2003). New values for parameters are reported in Table 1 as optimized parameters. This allows the adaptation of ADM1 to process different types of substrates, becoming more suitable in industrial applications where not only wastewater AD is performed. So, the ADM1 extension proposed in this paper makes it more suitable for a wider range of substrates than the previous thanks to the new found kinetic constant values, which open the doors to the possibility to describe different feedstock AD with great results. This aspect is further confirmed directly with real data comparison (section 3) and by reproducing the results of the works aforementioned, meaning that the new model is so capable to describe also specific situation with good approximation. A second contribution addresses equations, CSTR mass balances become different for macromolecules and intermediate molecules. For the former, the normal shape of the equation is used, for the latter the inlet term $S_{in,i}$ has to be removed since

only macromolecules, such as carbohydrates, lipids, proteins, and inert materials, are directly introduced with feed stream (Eq. 3-4). Then, a clear distinction between influent initial conditions and starter initial conditions is required: the former is defined by the inlet concentration of the macromolecules derived by their influent flowrate value, the latter is referred to the bacteria amount initially present inside the reactor necessary to start the digestion process (Eq. 5). To further specify, the starter composition can be defined as inoculum, if only bacteria are present, or as fresh digestate if it consists in an already treated digestate recirculated to the feed stream or taken from another reactor. Another novelty brought by this work addresses hydrogen sulfide release and its biological abatement together inside the digestion chamber. Developed from sulfur rich substrate by means of sulfur reducing bacteria (SRB), H₂S, a toxic acid responsible for serious units' corrosion, has to be reduced to few ppm to be harmless. Its typical value ranges between 0.2~2 %vol. of the final biogas amount. This small percentage makes the acid treatment suitable with Sulfur Oxidizing Bacteria (SOB), that being aerobic bacteria, have to be activated by low amount of oxygen. As normal industrial practice, the oxygen is injected in the headspace, in order to directly encounter the SOB and reducing the inhibition effect on methanogens.

Table 1. Optimized kinetic constants and misc. coefficients.

Constant type	Reference compound	Default value	Optimized value
Hydrolysis $k_{hyd,i}$ [d ⁻¹]	Carbohydrates	10	1.25
	Proteins	10	0.525
	Lipids	10	0.8
Semi-Saturation $K_{S,i}$ [kgCOD m ⁻³]	Sugar	0.5	0.3
	Fats	0.4	0.8
	Propionate	0.1	0.2
	Acetic Acid	0.15	0.2
	Hydrogen	0.000007	0.0000099
Michealis-Menten $k_{m,i}$ [kgCOD kgCOD ⁻¹ d ⁻¹]	Sugar	30	20
	Amino Acids	50	45
	Propionate	13	15
	Hydrogen	35	55
Degradation $k_{dec,i}$ [kgCOD kgCOD ⁻¹ d ⁻¹]	Sugar	0.02	0.05
	Fats	0.02	0.05
	Acetic Acid	0.02	0.05
	Hydrogen	0.02	0.09
Inhibition $K_{I,i}$ [kmol m ⁻³]	Ammonia Inhibition	0.0018	0.003

In order to treat sulfur compounds in the reactive mixture the model proposed by Flores-Alsina (2016) properly works but requires several input information concerning feedstock characterization that is quite impossible to have without expensive and time-consuming laboratory analysis, because its aim concerns modelling also phosphorous and iron ions interaction in AD processes. Treating claimed micronutrients such as P, Fe and Ca makes the model too complex for industrial purposes, so the homologation of Kalyuzhnyi and Fedorovich (1998), a simpler but complete work about these topics, allows to mathematically account for H₂S development during digestion. Reaction rates for SO₄²⁻ and H₂S (Eq. 6) follow Monod kinetics and three more bacteria families are added (i.e., Acetogenic SRB, Acetoclastic SRB and Hydrogenotrophic SRB). On the other hand, the chemical oxidation of the acid to solid elemental S_x (that typically accumulates in the reactor ceil) is described trough the incorporation of Sharma et al. (2014) kinetic expression adapted to anaerobic conditions.

$$dS_i/dt = HRT(S_i^{in} - S_i) + \sum_j^{NR} \rho_i v_{i,j} \quad (3)$$

$$dS_i/dt = HRT(-S_i) + \sum_j^{NR} \rho_i v_{i,j} \quad (4)$$

$$dX_i/dt = HRT(-X_i) + \sum_j^{NR} \rho_i v_{i,j} - k_{dec} X_i \quad (5)$$

$$\rho_i = \frac{k_{m_i} (S_i S_{SO_4^{2-}}) (1 - H_2 S_{dis}/K_{I_i})}{(K_{S_i} + S_i) (K_n + S_{SO_4^{2-}})} X_i \quad (6)$$

Where for the symbology of Eq. 3-5 refer to Batstone et al. (2002), and for Eq. 6 refer to Kalyuzhnyi and Fedorovich (1998). As introduced, this new model for AD process description, brings dissolved oxygen reactant S_{O_2} (since methanogenesis happens in liquid phase) and its mass balance (Eq. 7) for the first time, whose principal task is the biological abatement of H_2S . However, it is also responsible for the inhibition of the Methanogenic bacteria (MPB). Because these latter work in liquid phase, it is crucial to evaluate the right amount of oxygen to insert, since it will penetrate in the substrate by natural diffusion. To reduce inhibition, SOBs are often placed on wood frameworks installed in the headspace, so all the O_2 injected will be captured from them as nutrient. To be complete and reliable, the model has to account also for these phenomena introducing a new function that relates the content of oxygen dissolved S_{O_2} into liquid phase to the methane production balance, actually, multiplying it in dissolved methane mass balance. The inhibition function (Eq. 8) is then tuned properly by a constant whose value was set be $K_{I,O_2} = 0.00035 \text{ kgCOD/m}^3$, in order to fit a case study, later presented. The shape of the function is kept similar to the other inhibition functions, already present in the original ADM1 model, not to impact on the stiffness of the model itself. I_{O_2} , then, is present as multiplication factor in liquid phase methane mass balance (Eq.9). Again, for the symbology of Eq. 9 refer to Batstone et al. (2002).

$$dS_{O_2}/dt = HRT(S_{O_2}^{in} - S_{O_2}) - \rho T_{O_2} - k_{m_{O_2}} \frac{S_{O_2}}{K_{S_{O_2}} + S_{O_2}} S_{H_2S}^{\alpha} \quad (7)$$

$$I_{O_2} = \frac{1}{1 + S_{O_2} / K_{I,O_2}} \quad (8)$$

$$dS_{CH_4}/dt = HRT(-S_{CH_4}) + (1 + Y_{ac})\rho_{11}I_{O_2} + (1 - Y_{H_2})\rho_{12}I_{O_2} - \rho T_{CH_4} \quad (9)$$

Another improvement concerns the characterization of the feedstock. Specifically, the introduction of a pseudo-component covering all the charged substrates. This is defined as a weight average of the bio-composition of all the inlets and result as percentages of the macromolecules in the feedstock. These values are then converted in kgCOD/m^3 in order to fit the model's unit of measures. Specific analysis for compounds such as phosphorus, calcium or metals are not necessary. Finally, the model so defined can be easily supported by the generation of a detailed database collecting all the aforementioned values of a huge variety of possible substrates, simplifying the usability of the model itself and spreading its simulation range to different industrial applications.

3. Case Study

To validate the developed model, the simulation of a local CSTR, with a Hydraulic Retention Time (HRT) of 35 days, is employed to test the new mathematical model. The plant is designed to discharge around $28 \text{ m}^3/\text{d}$ of digestate and produces $3 \text{ m}^3/\text{d}$ of biogas

from the co-digestion of different origins feedstock (Table 2). The amount of air used for microaeration treatment corresponds to the 5 %vol. of the headspace volume (i.e. 0.0315 m³/d of O₂ injected). The code implementation in Python™ requires a time span of 35 days divided into 10,000 subintervals and an absolute and relative tolerance of 10⁻¹⁰ and 10⁻¹¹ respectively. The effluent biogas is not only composed of methane, carbon dioxide and hydrogen, but contains also small amounts of hydrogen sulfide that has been previously developed from sulfur containing compounds and in part biologically oxidized. Table 3 shows the final composition of the model simulation in terms of volumetric percentages of biogas main components and allows to quickly compare them with the real plant outlet biogas composition (Unit effluent) provided. Oxygen is responsible for H₂S abatement and its excess leaves the headspace unreacted. From the trends reported below it is possible to see the reaching of a stationary phase within the canonical 30 days for AD processes. Picks in the biogas production (Fig. 1a) refer to different degradation times for the hydrolysis products from the substrate. It is also possible to notice as the trend showed by the original ADM1 does not depict properly the biogas trend. This due to the lack of evaluation of H₂S and other aforesaid compounds present in the feedstock and especially the methanogenesis oxygen induced inhibition in liquid phase. However, both the trends seem to be really similar at the steady-states. It's necessary to say that these trends represent the instantaneous value of biogas produced, so at the end the total biogas production refers to the area under these curves. Finally, it is clear from the Figure 1a that this kind of substrate, in this condition, is not suitable to be simulated from the original model. Furthermore, the dynamic trend of the sulfur compounds is analyzed (Fig 1b) from the new balances added, showing the sulfur present in the feed quickly reduced to H₂S before undergoing to degradation due to SOB_s.

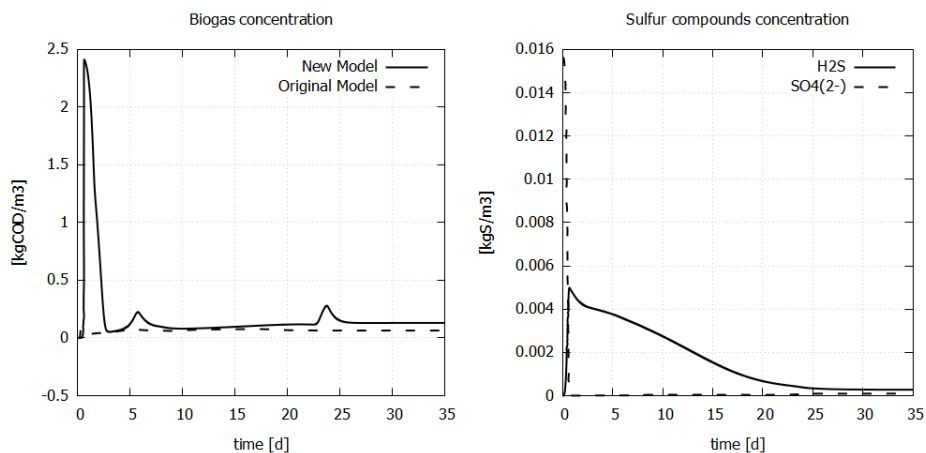


Figure 1. Trends of exiting biogas (a-left) and of dissolved H₂S and SO₄²⁻ (b-right).

Table 2. Feedstock composition.

Feed	m ³ /d
Cattle slurry	19
Bovine manure	7.4
Sorghum and Triticale silage	4.4

Table 3. Exiting composition of biogas compounds.

Component	Simulation [%vol.]	Unit effluent [%vol.]
CH ₄	70.7	50-75
CO ₂	27.06	25-45
H ₂	0.011	< 1
H ₂ S	0.2	0.02-2
O ₂	2	< 2

4. Conclusions

The presented model accounts for 43 differential equations, 35 original ones completed with the mass balances for SO_4^{2-} , H_2S and O_2 for both the liquid and gas phase and three new bacterial mass balances dedicated to SRBs. SOB's balances, instead, are not considered having a negligible contribution in the microbial environment. New kinetic constants have been proposed. Moreover, the new model greatly simplifies the simulation itself. In Flores-Alsina et al. studies (2016) the possibility to compute sulfur dynamics was presented with a mathematical model which needed several input information concerning feedstock characterization, which increases both the difficult availability of those data and the model complexity. The model here presented, requires a reduced number of input data that are summarized in the definition of a pseudo-component, which characteristics are defined by its percentages of macromolecules (carbohydrates, lipids, proteins), the amount of sulfur and inert materials coming from the feedstock charged to the digestion chamber and the air/oxygen feed to the headspace for biological H_2S abatement. A data base of biomasses and their characteristics has been also built for the pseudo-component definition. Original ADM1 has also other gaps that the presented model tries to fill with easy co-digestion simulation and complete H_2S dynamics. Case study results not only show realistic outlet values, provided in range percentages by the company, but also correct time depending on trends. Although this is a new work, it seems promising given the first results, and further studies, experiments and optimization processes will make the model complete and more effective.

References

- D.J. Batstone, J. Keller, I. Angelidaki, S.V. Kalyuzhnyi, S.G. Pavlostathis, A. Rozzi, W.T.M. Sanders, H. Siegrist and V.A. Vavilin, 2002, Anaerobic digestion model No 1 (ADM1), *Water Science & Technology* 45(10):65-73.
- F. Blumensaat, J. Keller, 2003, Modelling of two-stage anaerobic digestion using the IWA Anaerobic Digestion Model No. 1 (ADM1), *Water Research* 39 (2005) 171-183.
- H. Li, Z. Chen, D. Fu, Y. Wang, Y. Zheng, Q. Li, 2020. Improved ADM1 for modelling C, N, P fates in anaerobic digestion process of pig manure and optimization approaches to biogas production; *Renewable Energy* 146:2330-2336.
- K. Sharma, N. Derlon, S. Hu, Z. Yuan, 2014, Modeling the pH effect on sulfidogenesis in anaerobic sewer biofilm, *Water Res* 49:175-185
- L. Krayazelova, J. Bartacek, I. Diaz, D. Jeison, E. Volcke, P. Jenicek, 2015, Microaeration for hydrogen sulfide removal during anaerobic treatment: a review, *Rev Environ Sci Biotechnol*, 14:703-72
- P. Rivas-Garcia, J.E. Botello-Alvarez, L.R. Miramontes-Martinez, J.J. Cano-Gomez, R. Rico-Martinez, 2020, New model of hydrolysis in the anaerobic co-digestion of bovine manure with vegetable waste: Modification of anaerobic digestion model No.1; *Revista Mexicana de Ingeniería Química* Vol.19, No.1 109-122.
- V.V. Fedorovich and S.V. Kalyuzhnyi, 1998, Mathematical modelling of competition between sulphate reduction and methanogenesis in anaerobic reactors, *Bioresource Technology*, 65:227-242.
- X. Flores-Alsina, K. Solon, K. C. Mbamba, S. Tait, V. K. Gerney, U. Jeppsson, D. J. Batstone, 2016, Modelling phosphorus, sulfur and iron interaction for dynamic simulations of anaerobic digestion processes, *Water Research* 65:370-382

Implementation of first-principles surface interactions in a hybrid machine learning assisted modelling of flocculation

Nima Nazemzadeh^a, Rasmus Fjordbak Nielsen^a, Krist V. Gernaey^a, Seyed Soheil Mansouri^a, Martin P. Andersson^{b,*}

^a *Process and Systems Engineering Centre (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søtofts Plads, Building 228A, 2800 Kgs. Lyngby, Denmark*

^b *Combustion and Harmful Emission Control Centre (CHEC), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søtofts Plads, Building 228A, 2800 Kgs. Lyngby, Denmark*

* *martan@kt.dtu.dk*

Abstract

Machine learning algorithms are drawing attention for modelling processes in the chemical and biochemical industries. Due to a lack of fundamental understanding of complex processes and a lack of reliable real-time measurement methods in bio-based manufacturing, machine learning approaches have become more important. Hybrid modelling approaches that combine detailed process understanding with machine learning can provide an opportunity to integrate prior process knowledge with various measurement data for efficient modelling of the (bio)chemical processes. In this study, the application of a hybrid modelling framework that combines various first-principles models with machine learning algorithms is demonstrated through a laboratory-scale case of flocculation of silica particles in water. Since flocculation is a process that occurs across length- and time scales, an integrated hybrid multi-scale modelling framework can improve the phenomenological understanding of the process. The first-principles models utilized in this study are molecular scale particle surface interaction models such as Derjaguin-Landau-Verwey-Overbeek (DLVO) theory and interfacial tension predictions combined with a larger-scale population balance model.

Keywords: Hybrid modelling, Flocculation, Surface interactions, Interfacial tension energy

1. Introduction

The integration of machine learning and first-principles model in hybrid form can be achieved in different ways. For instance, a general hybrid model is proposed by Oliveira (2004) to model a stirred tank bioreactor. Another way of applying hybrid modelling approaches is represented in Psychogios and Ungar (1992), which provides interaction between a machine learning algorithm and a first-principles model of a fed-batch bioreactor. Based on the availability of first-principles models and prior process knowledge, one can apply a variety of different hybrid models. The hybrid modeling approach implemented in this study is represented in Figure 1. The model inputs are the data collected from experiments, which includes the pH measurements throughout each batch operation and the particle size distribution. According to the availability of first-principle models for the flocculation process, such hybrid modelling scheme (see Figure

1) is used, since a first-principles model does not exist to fully describe the process. Hence, a machine learning model (here artificial neural network) can be trained based on the model inputs and the first-principles models use the output of the machine learning algorithm and model inputs all together to predict a future particle size distribution.

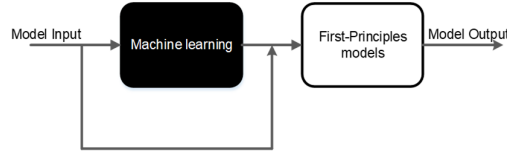


Figure 1: Hybrid modelling schemes according to the availability of first-principles models

Various approaches have been used to model particle processes, including flocculation, by using a population balance model (PBM). In this study, a discretized PBM is implemented using the work of Nazemzadeh et al. (2019) to model the particle size distribution over time. The discretized PBM is represented in Eq. (1).

$$\frac{dN_i}{dt} = \sum_{\substack{j \geq k \\ v_{i-1} \leq (v_j + v_k) \leq v_{i+1}}} \left(1 - \frac{1}{2} \delta_{j,k}\right) \eta_i \alpha_{j,k} \beta_{j,k} N_j N_k - N_i \sum_k \alpha_{i,k} \beta_{i,k} N_k + \sum_k \gamma_{j,i} S_j N_j - S_i N_i \quad (1)$$

Where N_i is the number of particles in a size bin of the discretization scheme, $\delta_{j,k}$ is the Dirac delta function, η_i is the proportionality coefficient related to the discretization scheme, $\alpha_{j,k}$ is the collision efficiency of two colliding particles, $\beta_{j,k}$ denotes the collision frequency of particles, γ is the breakage distribution function, and S denotes the breakage rate of a particle. Collision efficiency of the two colliding particles is calculated by using DLVO theory (Yotsumoto and Yoon, 1993) as a first-principles model. This theory in combination with Fuch's stability ratio (W) (Ahmad et al. 2008) allows us to determine the collision efficiency from the interaction energies between two particles in the system (see Eq. (2)). The remaining parameters of the PBM are defined by the semi-empirical equations found in the literature (Li et al. 2019).

$$\frac{1}{\alpha_{i,j}} = W_{i,j} = \int_{r_i+r_j}^{+\infty} \frac{\exp(V_T/K_b T)}{s^2} ds \quad (2)$$

Where V_T is the total interaction energy between two particles with a surface-surface separation distance and radii r_i and r_j , K_b is the Boltzmann constant and T is temperature.

Moreover, a solid-liquid interfacial tension (IFT) prediction model by Andersson et al. (2020) is implemented to determine the surface interaction energy of two particles in contact with each other. In future studies, we plan to use the predicted IFT values as an additional soft sensor to the hybrid model to evaluate the unknown parameters in the first-principles model.

2. Modelling framework

In this study, the hybrid model uses the population balance model as the main first-principles model to represent the particle size evolution as a function of time. The prior knowledge of flocculation kinetic parameters alongside with DLVO theory is used as additional first-principles models on top of the PBM. The kinetic model contains parameters such as a velocity gradient, which is not measured during the process

operation. Moreover, the breakage term in the PBM contains parameters that cannot be correlated with the process variables. A modelling workflow related to the selected modelling approach is represented in Figure 2. This workflow is structured by extending the modelling framework developed by Nielsen et al. (2020). The difference between the framework in the present study and the one developed in Nielsen et al. (2020) is the utilization of first-principles models for the kinetic parameters in a PBM (i.e. DLVO theory and semi-empirical equations of other kinetic parameters). The machine learning algorithm (black box in Figure 1) is an artificial neural network consisting of three hidden layers with 34, 40, and 50 neurons respectively. The output of the neural network is the unknown/unmeasured parameters of the population balance model.

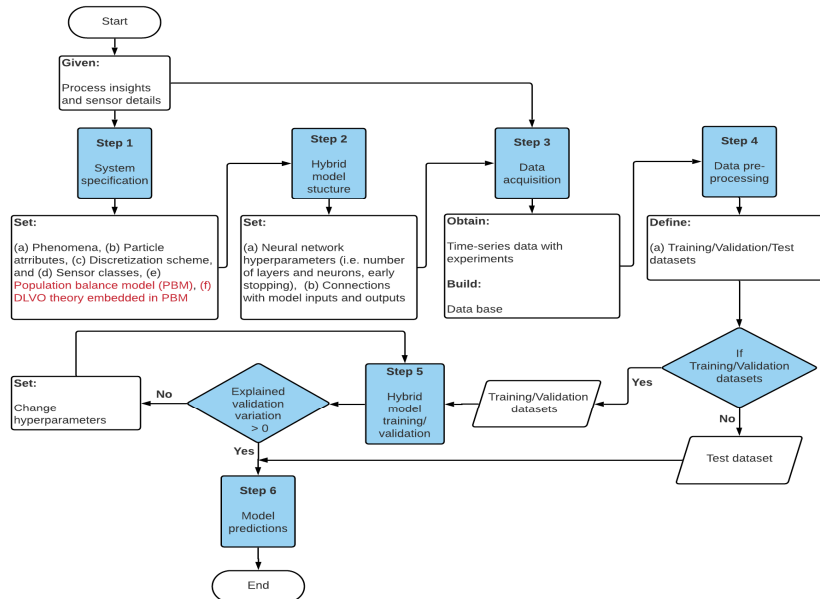


Figure 2: Overview of the hybrid multi-scale modelling workflow

3. Application example

The application of the proposed modelling framework is demonstrated through a laboratory-scale silica particle flocculation. To this end, an experimental setup is prepared that uses a dynamic image analysis sensor (Nielsen et al. 2019) to monitor the particle size distribution. The flocculation is stirred by an impeller to increase the chance of occurrence of particle collisions in the system. Moreover, an in-line pH probe is installed in the tank to measure the pH continuously. Several batches of experiments are carried out at different pH values. Variation of pH imposes a different surface charge density on the particles by protonation/deprotonation of the surface silanol groups on the silica. The experiments show that for the batches, in which pH is in the range of 2 to 3, agglomeration of particles takes place as shown in Figure 3 (a). However, at a higher pH value, the process is dominated by breakage phenomena and successful collisions of particles are less likely to happen (see Figure 3 (b)).

Silica nanoparticles have a point of zero charge (PZC) in the range of pH 2 to 3 (Cloarec et al. 2016). For a pH in the vicinity of the PZC, silica nanoparticles are neutral. Therefore, when two particles approach each other the attraction forces can overcome the repulsion forces and the particles collision will be more likely successful. However, for a pH

higher/lower than the PZC, the particles become deprotonated/protonated and negatively/positively charged. Hence, the repulsion forces of surfaces with the same charge can easily overcome the attraction forces and no agglomeration takes place.

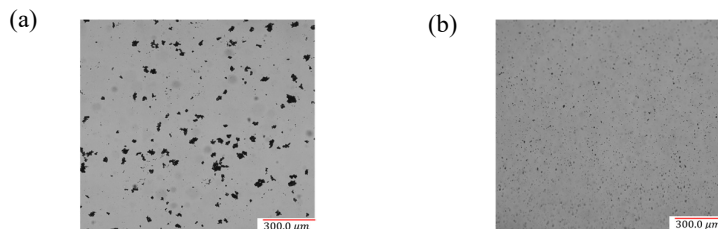


Figure 3: End-of-batch representation of particles: (a) at $pH = 2.45$, (b) at $pH = 7.61$

To determine the interfacial tension energies of silica particles, a structure of a silica nanoparticle containing 54 Hydrogen, 121 Oxygen, and 47 Silicon atoms was constructed and then optimized by using density functional theory (DFT) calculations. The geometry optimization was carried out for a neutral and singly deprotonated molecular structure by using the BP functional (Becke 1988), the TZVP basis set (Schäfer et al. 1992) and the COSMO implicit solvent model (Klamt and Schüürmann, 1993) with infinite dielectric constant using TURBOMOLE, v 7.4 (Ahlrichs et al. 1989). The IFT calculations were carried out using a solid-liquid interfacial tension model in COSMO-RS (Klamt et al., 2010) by using the BP_TZVP_C30_1601 parameterization in COSMOtherm (Eckert and Klamt 2010). The IFT of singly deprotonated silica nanoparticle and water was approximately -16 mN/m . The negative value of IFT means that a deprotonated silica nanoparticle has a favorable interaction with water. In this study, the IFT energies are only used to characterize the process. However, in future studies these interaction energies are planned to be utilized directly in the hybrid model.

The hybrid multi-scale framework in §2 is applied to the experimental data collected from the laboratory-scale flocculation of silica particles in water. Figure 4 represents the end-of-batch predictions for two batches of experiments: (a,c) dominated by breakage and (b,d) dominated by agglomeration. In both cases, the hybrid multi-scale model quite accurately predicts the particle size distribution of the end time-point. The predictions are improved compared to the similar case study in Nielsen et al. (2020). In that study, PBM was the only first-principles model and all the kinetic parameters of the model were estimated by means of a neural network. The agglomeration end-of-batch predictions in that study showed that the relative volume density of large particles were not predicted quite accurately due to either lack of data for training, or the fact that a pH sensor was not sufficient to describe the process. However, for the agglomeration batch in the present study, the model predicted the relative volume density of particles more accurately even for larger size-bins. A pure machine learning approach is also utilized to model the flocculation process to compare the performance of a black-box model and the hybrid model. The black-box model uses Eq. (3) to predict the particle number concentration for each size-bin in a future time-point with the rate that is determined by the machine learning algorithm. Hence, no phenomenon is defined for the process and the future particle number concentration is predicted by a single rate at each time-point for each size-bin. The rest of the modelling components are chosen to be the same as the ones in §2 (see Figure 2). The end-of-batch predictions for the black-box model are represented in Figure 4 (c,d). The predictions show that for the breakage batch, the performance of

the hybrid and black-box models are reasonably comparable. We believe that the reason is the fact that the variations of relative volume densities in that experimental batch were not very high and both models can predict the end-of-batch distribution in a similar way. However, in the agglomeration batch, the hybrid model outperformed the machine learning model in end-of-batch predictions.

$$\frac{dN_i}{dt} = rate_{i,t} \quad (3)$$

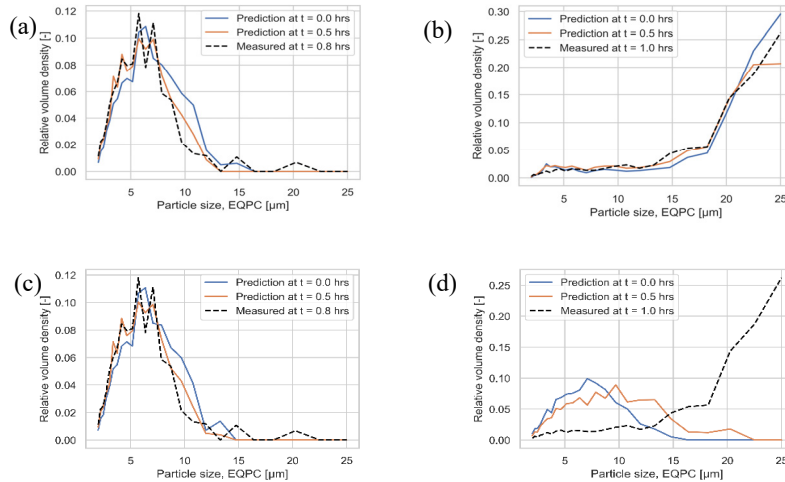


Figure 4: End-of-batch predictions of (a) agglomeration batch, (b) breakage batch for the hybrid model and (c) breakage batch, (d) agglomeration batch for the black-box model

4. Conclusion and future developments

In this study, the application of a hybrid multi-scale framework was demonstrated on a laboratory-scale case of silica particles flocculation by using a population balance model, DLVO theory, and semi-empirical equations of flocculation kinetics to model the process. The main reason that we used machine learning for modelling the flocculation process is that a prior process knowledge does not exist to fully describe the process. The end-of-batch predictions of two experimental batches have shown that such modelling approach can accurately estimate the evolution of particle sizes in the system in a future time horizon for both phenomena (agglomeration and breakage) involved in the flocculation process. Moreover, a pure black-box model was also used to perform the same end-of-batch predictions. Both modelling approaches showed similar predictions for the batch dominated by breakage due to the very small variations of the particle size distributions in the breakage batch. Nonetheless, the hybrid model outperformed the black-box model by accurately predicting the end-of-batch particle size distribution. We believe that the integration of machine learning algorithms and first-principles models helps us to model complex processes such as flocculation more accurately and predict the dynamics of the system in a future unknown time-horizon.

5. Acknowledgment

We would like to thank Novozymes A/S and Greater Copenhagen Food Innovation program (CPH-FOOD) for co-financing this research and also for their support during this project.

References

- R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, 1989, Electronic structure calculations on workstation computers: The program system turbomole, *Chemical Physics Letters*, 162, 165-169.
- A.L. Ahmad, M.F. Chong, S. Bhatia, 2008, Population Balance Model (PBM) for flocculation process: Simulation and experimental studies of palm oil mill effluent (POME) pretreatment. *Chemical Engineering Journal*. 140, 1-3, 86-100.
- M.P. Andersson, T. Hassenkam, J. Matthiesen, L.V. Nikolajsen, D.V. Okhrimenko, S. Dobberschütz, S.L.S. Stipp, 2020, First-principles prediction of surface wetting, *Langmuir*, 36, 42, 12451-12459.
- A.D. Becke, 1988, Density-functional exchange-energy approximation with correct asymptotic behavior, *Physical review A*, 38, 6, 3098-3100.
- J.P. Cloarec, C. Chevalier, J. Genest, J. Beauvais, H. Chamas, Y. Chevolot, T. Baron, A. Souifi, 2016, pH driven addressing of silicon nanowires onto Si₃N₄/SiO₂ micro-patterned surfaces, *Nanotechnology*, 27, 29, 295602.
- F. Eckert, A. Klamt, 2013, COSMOtherm Version C3. 0, Release 13.01. Leverkusen (Germany): COSMOlogic GmbH & Co KG.
- A. Klamt, F. Eckert, W. Arlt, 2010, COSMO-RS: an alternative to simulation for calculating thermodynamic properties of liquid mixtures. *Annual review of chemical and biomolecular engineering*, 1, 1, 101-122.
- A. Klamt, G.J. Schüürmann, 1993, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *Journal of the Chemical Society, Perkin Transactions 2*, 5, 799-805.
- Z. Li, P. Lu, D. Zhang, F. Song, 2019, Simulation of Floc Size Distribution in Flocculation of Activated Sludge Using Population Balance Model with Modified Expressions for the Aggregation and Breakage, *Mathematical Problems in Engineering*, 2019, 5243860.
- N. Nazemzadeh, L.W. Sillesen, R.F. Nielsen, M.N. Jones, K.V. Gernaey, M.P. Andersson, S.S. Mansouri, 2020, Integration of Computational Chemistry and Artificial Intelligence for Multi-scale Modeling of Bioprocesses, *Computer Aided Chemical Engineering*, 48, 295-300.
- R.F. Nielsen, N.A. Kermani, L. la Cour Freiesleben, K.V. Gernaey, S.S. Mansouri, 2019, Novel strategies for predictive particle monitoring and control using advanced image analysis, *Computer Aided Chemical Engineering*, 46, 1435-1440.
- R.F. Nielsen, N. Nazemzadeh, L.W. Sillesen, M.P. Andersson, K.V. Gernaey, S.S. Mansouri, 2020, Hybrid machine learning assisted modelling framework for particle processes, *Computers & Chemical Engineering*, 140, 106916.
- R. Oliveira, 2004, Combining first principles modelling and artificial neural networks: a general framework, *Computers & Chemical Engineering*, 28, 5, 755-766.
- D.C. Psychogios, L.H. Ungar, 1992, A hybrid neural network-first principles approach to process modeling, *AIChE Journal*, 38, 10, 1499-1511.
- A. Schäfer, H. Horn, R. Ahlrichs, 1992, Fully optimized contracted Gaussian basis sets for atoms Li to Kr, *The Journal of Chemical Physics*, 97, 4, 2571-2577.
- H. Yotsumoto, R.H. Yoon, 1993, Application of extended DLVO theory: II. Stability of silica suspensions, *Journal of Colloid and Interface Science*, 157, 2, 434-441.

Optimal physical property data for process simulations by optimal experimental design

Lorenz Fleitmann^{a,b}, Jan Pyschik^a, Ludger Wolff^a and André Bardow^{a,b,c,*}

^a*Institute for Technical Thermodynamics, RWTH Aachen University, Schinkelstrasse 8, 52062 Aachen, Germany*

^b*Energy and Process Systems Engineering, Department of Mechanical and Process Engineering, ETH Zurich, Tannenstrasse 3, 8092 Zurich, Switzerland*

^c*Institute of Energy and Climate Research (IEK-10), Forschungszentrum Jülich, Wilhelm-Johnen-Strasse, 52425 Jülich, Germany*
abardow@ethz.ch

Abstract

Simulations of chemical processes require accurate physical properties, which are usually estimated from experiments. Experimental effort can be minimised by optimal experimental design (OED). OED tailors experimental measurements to minimise the expected uncertainty of the fitted parameters. However, in process design and optimisation, the primary purpose of an experiment is usually not to determine property parameters as accurately as possible but to enable most accurate process simulations. Therefore, in this work, we present OED of physical property measurements leading to optimal predictions of process simulations by the so-called *c*-optimal experimental design. A *c*-optimal design minimises the uncertainty of the simulation results by weighting parameters by their influence on the process model. The *c*-optimal OED is employed to design liquid-liquid equilibrium measurements as a basis for simulations of an extraction and a hybrid extraction-distillation process. For the same simulation accuracy, the *c*-optimal design can almost halve the number of experiments compared to state-of-the-art OED that neglects the process. Our work shows that *c*-optimal design can reduce experimental effort in chemical engineering successfully by tailoring experiments to their process application.

Keywords: model-based experimental design, parameter precision, *c*-optimal design, liquid-liquid-equilibrium, non-random two-liquid model, extraction-distillation.

1. Introduction

Accurate estimation of property parameters is crucial for the simulation of chemical processes (Mitsos et al., 2018). The basis for accurate parameter estimation is experimental data. If such data is not available, new experiments are required, usually consuming time and large volumes of material samples, leading to high costs. Therefore, experimental effort should be minimised by careful design of experiments. Optimal experiments minimise the uncertainty in the experimentally determined parameters. The systematic selection of optimal experiments is enabled by model-based optimal experimental design (OED) (Atkinson et al., 2006). OED analyses the

propagation of uncertainty from inaccuracies in experimental measures to estimated parameters. The propagation is used to optimise the experimental design, i.e., the settings and inputs of the experiments (Franceschini and Macchietto, 2008). However, for process systems engineering, the primary purpose of experiments is rarely the knowledge of parameters itself, but the use of these parameters in process simulation. An optimal experimental design minimising parameter variances does not necessarily lead to the most accurate process simulations. OED for process simulations needs to consider the use of the fitted parameters within the process models.

Recently, OED was applied to design plant experiments using a flowsheet simulator (Asprion et al., 2019). The authors designed experiments at the plant to measure property data capturing the parameter use in the process simulation. However, for this approach, challenging plant experiments need to be performed instead of lab-scale experiments. Recker et al. (2013) presented a pioneering approach considering the process model within OED of lab-scale experiments. In their work, state-of-the-art OED was extended by scaling the well-known A -criterion heuristically by process sensitivities to parameters. A similar OED objective was proposed by Lucia and Paulen (2014) for robust model predictive control. The authors tailored the modified E -criterion by the sensitivities of the optimal robust economic objective value with respect to the range of parametric uncertainty.

Although these approaches successfully integrated process sensitivities into OED, the OED objectives rely on heuristic scaling. However, full consideration of process information by systematic uncertainty propagation is enabled by established theory:

c -optimal experimental design. In c -optimal design, the objective is to minimise the variance of a linear combination of the parameters (Atkinson et al., 2006). By linear approximation of the variance propagation from the parameters through the process model, c -optimal design can be employed to minimise the uncertainty of process simulations.

So far, c -optimal OED has not been applied to chemical engineering problems. Therefore,

we adapt c -optimal design to experiments for the measurement of liquid-liquid-equilibria

to simulate extraction and extraction-distillation processes. We compare the c -optimal objective to state-of-the-art OED methods using the D -optimality criterion and an A -optimality criterion that is heuristically scaled by process sensitivities.

2. Method: Calculation of c -optimal experimental designs

To design optimal experiments, OED needs to link the uncertainties of experimental measures to the uncertainties of the model parameters. Mathematically, the uncertainties of the model parameters are expressed in the parameter variance-covariance matrix V_{θ} .

The parameter variance-covariance matrix is approximated by the product of the Fisher information matrix $F(\hat{\theta}, \xi)$ and the number of experiments N_{exp} .

$$V_{\theta} \approx \left[N_{exp} F(\hat{\theta}, \xi) \right]^{-1} \quad (1)$$

The Fisher information matrix $F(\hat{\theta}, \xi)$ depends only on the chosen experimental design ξ and an initial parameter guess $\hat{\theta}$. As the parameter variance-covariance matrix

is proportional to the inverse of $F(\hat{\theta}, \xi)$, OED mainly focuses on designing an optimal Fisher information matrix by selecting an optimal design ξ^* . The Fisher information matrix $F(\hat{\theta}, \xi)$ is obtained from the model $g(w, \theta)$. The model $g(w, \theta)$ describes the experiments by linking the uncertainties of experimental measurements V_w to the parameters θ . The Fisher information matrix $F(\hat{\theta}, \xi)$ is calculated for an experimental design ξ by taking the model sensitivity to experimental measures A_μ and the model sensitivity to parameters B_μ for each experiment μ of the experimental design.

$$F(\hat{\theta}, \xi) = \sum_{\mu=1}^{N_{exp}} v_\mu B_\mu^T \left(A_\mu V_w A_\mu^T \right)^{-1} B_\mu \quad \text{with} \quad A_\mu = \frac{\partial g}{\partial w} \Big|_{w_\mu, \hat{\theta}} \quad \text{and} \quad B_\mu = \frac{\partial g}{\partial \theta} \Big|_{w_\mu, \hat{\theta}} \quad (2)$$

State-of-the-art OED objective functions are based on the Fisher information matrix and aim to optimise a scalar measure of the Fisher information matrix. The c -optimal design additionally considers the use of the estimated parameters in the process model. Variance

propagation through the process model weights the parameter variances by the sensitivities of the process model result $h(\theta)$ to the parameters θ in the c -optimal design.

$$c(\hat{\theta})^T F(\hat{\theta}, \xi)^{-1} c(\hat{\theta}) \quad \text{with} \quad c(\hat{\theta}) = \frac{\partial h}{\partial \theta} \Big|_{\hat{\theta}} \quad (3)$$

We compare the c -optimal design with state-of-the-art OED objectives. The most popular

objective is D -optimal design. D -optimal design maximises the determinant of $F(\hat{\theta}, \xi)$ to minimise the volume of the confidence ellipsoid spanned by all parameters.

$$\log(\det F(\hat{\theta}, \xi)) \quad (4)$$

Inspired by Recker et al. (2013) and Lucia and Paulen (2014), we also investigate a heuristically scaled Fisher information matrix together with A -optimal experimental design. The A -optimal design maximises the trace of the Fisher information matrix. We scale the Fisher information matrix with a diagonal weighting matrix W . The weighting matrix W includes the sensitivities of the process model output $h(\theta)$ to the parameters θ .

$$\text{tr}\left(F_{heur}(\hat{\theta}, \xi)\right) = \text{tr}\left(W F(\hat{\theta}, \xi) W\right) \quad \text{with} \quad W = \text{diag}\left(\frac{\partial h}{\partial \theta} \Big|_{\hat{\theta}}\right) \quad (5)$$

All experimental designs are computed by a solution algorithm with monotonic global convergence from literature (Yu, 2010). Derivatives are obtained by numerical differentiation. The algorithm provides experimental designs with a continuous distribution of experimental effort, i.e., which share of the total experimental effort should be spent on which measurements. The results serve as targets for experiments in the laboratory. The identified experiments are optimal for the initial parameter set; thus, an iterative procedure might be required in practice. For comparison of the OED objectives, each experimental design ξ can be assessed by OED-efficiencies, e.g., c -efficiency $\zeta_c(\xi)$. The c -efficiency is the ratio of the objective values (Eq. (3)) of a

c -optimal design and of the design under consideration. The inverse of $\zeta_c(\xi)$ shows how many more experiments are needed for the same accuracy of the process simulation compared to a c -optimal design.

3. Case Study and Results

For two solvent-based processes, we apply c -optimal experimental design to tailor liquid-liquid-equilibrium (LLE) measurements: (1) the extraction of acetone from aqueous solution with toluene and (2) the extraction-distillation process for the same mixture including a distillation column to separate the extract. The extraction and distillation column are modelled using pinch-based process models (Bausa et al., 1998; Redepenning et al., 2017) employing the non-random two liquid (NRTL) model as an activity coefficient model. The liquid-liquid-equilibrium experiments are designed to estimate isothermal NRTL- t -parameters for the extraction and linearly temperature-dependent NRTL- t -parameters for the extraction-distillation process. We challenge the results of the c -optimal design by the D -optimal design and the heuristically scaled A -optimal design.

3.1. OED for the simulation of a pinch-based extraction column

For the extraction process, we determine at which concentrations which share of experimental effort should be invested in LLE-experiments for minimum uncertainty in the solvent demand of the process. For simplicity, we assume that LLE-experiments are performed with the overall composition of the components that corresponds to the centre of the tie lines. Therefore, an experiment is fully characterised by a scalar measure α which runs from 0 at the binary subsystem to 1 at the critical point of the two-phase region.

For each optimality criterion used, three distinct locations for measurements are sufficient

for an optimal design (Figure 1). As commonly found in OED, the optimal design suggests repeating measurements at these locations rather than to explore other locations. The locations and, in particular, the share of the total experimental effort differ for the three objectives: The c -optimal design shifts 80% of the experimental effort towards the operating range of the extraction column resulting in a c -efficiency ζ_c about 3 times higher than the D -optimal design (Table 1). Consequently, 3 times more D -optimal experiments would be required to obtain the same accuracy as the c -optimal experiments. Interestingly, the c -optimal design does not place any experiments in the operating range of the extraction column and still places about 20% of the experimental effort in the high-curvature region of the binodal curve.

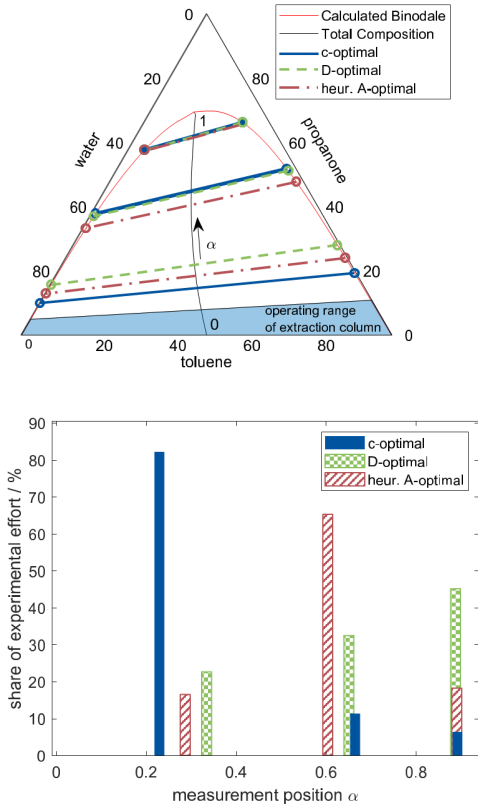
The D -optimal design places 75% of the experiments in the high-curvature region of the binodal curve, as already found by Dechambre et al. (2014). In the high-curvature region of the binodal curve, the phase equilibrium model is susceptible to the property parameters leading to low uncertainty in the parameter estimation.

The c -optimal design balances the parameter requirements: Experiments need to be linked

to the operating range of the process, but the exclusive focus on the process neglects that

accuracy can be increased by including experiments for high parameter precision. The heuristic A -optimal design achieves a c -efficiency ζ_c of only 0.28. The experimental design from the heuristic A -optimal design is closer to the c -optimal design. However,

the c-efficiency is lower than the c-efficiency of the D-optimal design, which shows that the similarity of designs is not necessarily correlated to the efficiency.



(a) Location of LLE experiment

(b) Share of experimental effort

Figure 1: Optimal designs for LLE experiments for the extraction process.

Table 1: c-efficiency ζ_C for the c- and D-optimal experimental designs ξ_C^* and ξ_D^* and the heuristically scaled A-optimal design $\xi_{A,heur}^*$ for the extraction and the extraction-distillation process

Experimental design ξ	c-efficiency ζ_C	
	Extraction	Extraction-Distillation
c-optimal ξ_C^*	1	1
D-optimal ξ_D^*	0.36	0.61
heuristically scaled A-optimal $\xi_{A,heur}^*$	0.28	0.57

3.2. OED for the simulation of an extraction-distillation process

For the extraction-distillation process, we design the location of the LLE-experiments within the miscibility gap as before and additionally, the temperature for each LLE-experiment to minimise the uncertainty of the reboiler energy demand. Other key performance indicators of the process could serve as targets as well by computing the corresponding sensitivities $c(\theta)$. For each design objective, the design mainly focuses on the boundaries of the design space (Figure 2). On the boundaries, the parameter sensitivity is highest leading to low parameter uncertainty. For the extraction-distillation process, parameter precision is more important for accurate process simulation than for the extraction only: The D -optimal design yields a c -efficiency ζ_c of about 0.6 (Table 1).

Still, the c -optimal design increases accuracy by 64%. The heuristic A -optimal design is comparable to the D -optimal design. However, the heuristic A -optimal design does not improve in c -efficiency ζ_c compared to the D -optimal design, despite the consideration of process sensitivities. In this case study, the heuristics are not only exceeded by the c -optimal design, but also by the conventional design that does not consider the process. Therefore, the selection of the right optimality criterion is crucial for process engineering problems. For the extraction and extraction-distillation process, only c -optimal design tailors experiments for the process simulations.

4. Conclusion

In this work, we investigate a method for optimal experimental design (OED) of lab-scale experiments that consider the use of the estimated parameters in a subsequent process simulation: c -optimal experimental design. The c -optimal design includes the propagation of uncertainties from measurements to the outputs of the process model. We compare

c -optimal design with state-of-the-art OED, i.e., D -optimal design and a heuristically scaled A -optimal design. The c -optimal design yields non-trivial experimental designs that outperform state-of-the-art OED by providing more accurate results of process simulation. The c -optimal design does not simply perform experiments in the operating region of the process but balances parameter precision with uncertainty propagation to the process model. Thereby, the experimental effort can be significantly reduced while maintaining the same accuracy.

Acknowledgement

The authors gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - Cluster of Excellence 2186 "The Fuel Science Center" ID: 390919832.

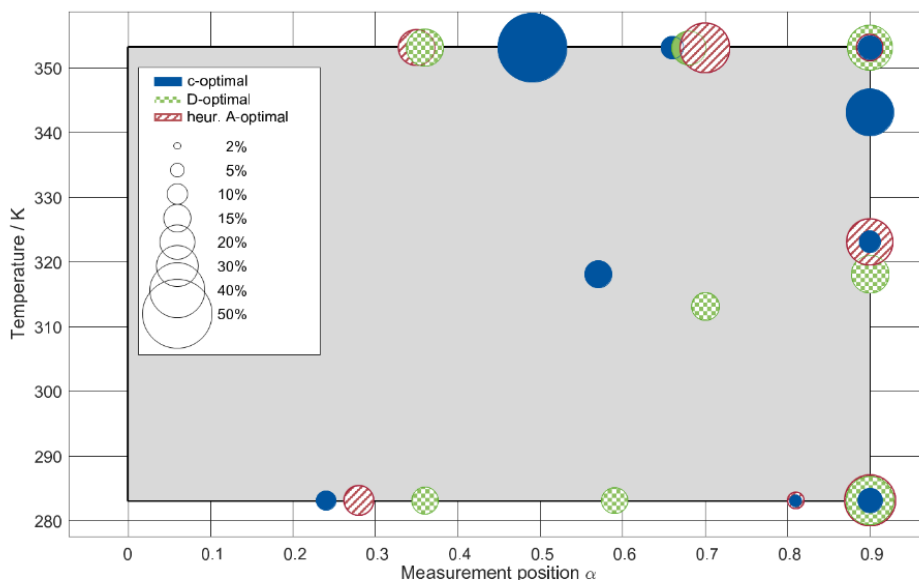


Figure 2: Optimal designs for LLE experiments for the extraction-distillation process. The size of the circles corresponds to the share of the experimental effort. The grey box indicates the design space.

References

- N. Asprion, R. Böttcher, J. Mairhofer, M. Yliruka, J. Höller, J. Schwientek, C. Vanaret, M. Bortz, 2019. Implementation and Application of Model-Based Design of Experiments in a Flowsheet Simulator. *Journal of Chemical & Engineering Data* 65 (3), 1135–1145.
- A. C. Atkinson, A. N. Donev, R. Tobias, 2006. Optimum experimental designs, with SAS. Vol. 34 of Oxford statistical science series. Oxford University Press, Oxford and New York.
- J. Bausa, R. v. Watzdorf, W. Marquardt, 1998. Shortcut methods for nonideal multicomponent distillation: I. Simple columns. *AIChE Journal* 44 (10), 2181–2198.
- D. Dechambre, L. Wolff, C. Pauls, A. Bardow, 2014. Optimal Experimental Design for the Characterisation of Liquid–Liquid Equilibria. *Industrial & Engineering Chemistry Research* 53 (50), 19620–19627.
- G. Franceschini, S. Macchietto, 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science* 63 (19), 4846–4872.
- S. Lucia, R. Paulen, 2014. Robust Nonlinear Model Predictive Control with Reduction of Uncertainty via Robust Optimal Experiment Design. *IFAC Proceedings Volumes* 47 (3), 1904–1909.
- A. Mitsos, N. Asprion, C. A. Floudas, M. Bortz, M. Baldea, D. Bonvin, A. Caspari, P. Schäfer, 2018. Challenges in process optimisation for new feedstocks and energy sources. *Computers & Chemical Engineering* 113, 209–221.
- S. Recker, N. Kerimoglu, A. Harwardt, O. Tkacheva, W. Marquardt, 2013. On the integration of model identification and process optimisation. In: 23rd European Symposium on Computer Aided Process Engineering. Vol. 32 of Computer Aided Chemical Engineering. Elsevier, pp. 1021–1026.
- C. Redepenning, S. Recker, W. Marquardt, 2017. Pinch-based shortcut method for the conceptual design of isothermal extraction columns. *AIChE Journal* 63 (4), 1236–1245.
- Y. Yu, 2010. Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics* 38 (3), 1593–1606.

On the practical identifiability of kinetic models of hydroxymethylfurfural hydrogenation in batch reaction systems

Philipp Deussen^a, Federico Galvanin^{a,*}

^a*Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom*

f.galvanin@ucl.ac.uk

Abstract

Hydroxymethylfurfural (HMF) is an organic compound that occurs in numerous foods and is used as feedstock in a variety of chemical processes. HMF can be hydrogenated to form 2,5-Dimethylfuran (DMF), which is an important component in biofuel production. Several kinetic models have been proposed in literature for this hydrogenation reaction, including power law and Langmuir-Hinshelwood-Hougen-Watson (LHHW) models based on reaction species. A critical aspect to address is the parametric identifiability in these models, i.e. the correct estimation of model parameters from experimental data. In this paper, a two-step identifiability approach is proposed exploiting model-based design of experiments (MBD_{oE}) techniques to assess the identifiability of candidate kinetic models of HMF hydrogenation at variable experimental conditions in a batch reaction system. Information maps indicate the most informative regions of the experimental design space to be used to precisely estimate the kinetic model parameters.

Keywords: identifiability analysis, model-based design of experiments, kinetics of HMF hydrogenation

1. Introduction

HMF is an organic compound that can be found naturally in many foods and is used industrially as a platform chemical in plastics, pharmaceuticals and agrochemicals. It can be hydrogenated to DMF - an alternative biofuel and additive for diesel or jet fuel. It significantly lowers pollution versus its fossil equivalents (van Putten et al., 2013). Proposed mechanisms for this hydrogenation reaction include power-laws (Grilic et al., 2014, Luo et al., 2015, Gyngazova et al., 2015), single-site LHHW (Jain and Vaidya, 2016) and dual-site LHHW (Gawade et al., 2016) kinetics. However, these models lack an explicit temperature dependence. Thus, an expansion of the power-law model to include temperature dependency is proposed and the estimability of the new set of kinetic parameters from experimental data (practical identifiability) assessed (Chis et al., 2016).

MBD_{oE} techniques (Franceschini and Macchietto, 2008) have been proposed in literature to design sets of experiments yielding the most informative data to be used for model identification. Experiments can be designed for discriminating between candidate kinetic models or, given a suitable model structure, to improve precision in parameter estimation

(Galvanin et al., 2016). A two-step identifiability approach is proposed in this paper using MBDoe techniques to assess the identifiability of three new candidate kinetic models of HMF hydrogenation in a batch reactor. Temperature, duration and initial concentrations of reagents are the decision variables defining the design space of the models. First, the design space is screened using Latin Hypercube Sampling (LHS) to generate conditions for running in-silico experiments under variable modelling/noise assumptions. Second, experiments are mapped using the Fisher information matrix (FIM) as information metric (Galvanin et al., 2016). Combined with a-posteriori statistics from parameter estimations using in-silico data, the analysis indicates the most informative regions for precise estimations of model parameters for experimentation without numerical optimisation.

2. Methodology

The proposed kinetic model (1) is written in terms of control variables, \mathbf{w} , state variables $\mathbf{x}(t)$ and parameters $\boldsymbol{\vartheta}$ to be determined. The function $\mathbf{h}(\mathbf{x}(t))$ (2) relates state variables and measured responses $\hat{\mathbf{y}}(t)$. The vector for the design of experiments $\boldsymbol{\varphi}$ (3) lies in the design space, ϕ , and contains \mathbf{w} , initial conditions \mathbf{y}^0 , duration τ and sampling times \mathbf{t}_{sp} .

$$\mathbf{f}(\mathbf{x}(t), \dot{\mathbf{x}}(t), \mathbf{w}, \boldsymbol{\vartheta}, \tau) = 0 \quad (1)$$

$$\hat{\mathbf{y}}(t) = \mathbf{h}(\mathbf{x}(t)) \quad (2)$$

$$\boldsymbol{\varphi} = (\mathbf{y}^0, \tau, \mathbf{w}, \mathbf{t}_{\text{sp}}) \quad (3)$$

In any experiment ($1, \dots, n_{\text{exp}}$) the sensitivities of the measurable outputs \hat{y}_i ($1, \dots, n_m$) to perturbations of the parameters $\boldsymbol{\vartheta}$ ($1, \dots, n_\theta$) are expressed in the sensitivity matrix \mathbf{Q} (4) and computed at sampling points \mathbf{t}_{sp} . The $n_\theta \times n_\theta$ -dimensional variance-covariance matrix \mathbf{V}^θ (5) is obtained from the standard deviation of measurement errors (σ) and \mathbf{Q} . Each kl -th element in the parameter correlation matrix \mathbf{C} (6) can be computed from \mathbf{V}^θ elements. Correlations scale from -1 (total anticorrelation), to 0 (no correlation) to 1 (total correlation). Critical correlation begins at 0.95. The Fisher information matrix (FIM) (7) approximates the Hessian matrix of the log-likelihood function of parameters $\boldsymbol{\vartheta}$ using the predicted outputs \hat{y}_i (Franceschini and Macchietto, 2008).

$$\mathbf{Q} = \begin{bmatrix} \frac{\partial \hat{y}_1}{\partial \vartheta_1} & \dots & \frac{\partial \hat{y}_1}{\partial \vartheta_{n_\theta}} \\ \dots & \dots & \dots \\ \frac{\partial \hat{y}_{n_m}}{\partial \vartheta_1} & \dots & \frac{\partial \hat{y}_{n_m}}{\partial \vartheta_{n_\theta}} \end{bmatrix} \quad (4)$$

$$\mathbf{V}^\theta = [\mathbf{Q}^T \sigma^{-2} \mathbf{Q}]^{-1} \quad (5)$$

$$C_{kl} = \frac{V_{kl}^\theta}{\sqrt{V_{kk}^\theta V_{ll}^\theta}} \quad (6)$$

$$\mathbf{H} = \sum_{i=1}^{n_{\text{exp}}} \sum_{j=1}^{n_m} \begin{bmatrix} 1 \\ \sigma_{ij}^{-2} q_{ij}^T q_{ij} \end{bmatrix} \quad (7)$$

Organic systems are often sloppy. Sloppiness (Chis et al., 2016) is identified if the eigenvalues of the FIMs of different parameters vary by three orders of magnitude or

more and is described by the ratio of the axes of the confidence area. Sloppiness occurs when under specific experimental conditions some parameters have very low impacts on the outputs (i.e. negligible sensitivity), but a set of experimental conditions to identify the full set of model parameters in ϕ does exist.

A two-step procedure is proposed to test the identifiability of selected models in ϕ :

Step 1: The sensitivity and information metrics rely on ϕ and \mathfrak{D} and are thus only locally valid. Since local optima are undesirable, the design space is screened by LHS. LHS is more efficient than randomly selecting points as the position of already generated points is considered before generating subsequent ones, avoiding local clusters (Montgomery, 2012). In-silico experiments are run using these points and the correlation matrix (6) is computed. Models are checked for sloppiness by testing whether critical correlations persist throughout ϕ even when a large number of samples is taken. If correlations persist, these models are practically unidentifiable and must be rejected or reparametrised.

• **Step 2:** A parameter estimation is carried out using a Student t-test to assess the precision of the estimates and a χ^2 test to assess the adequacy of model predictions. The model passes when the parameter t -values (8) are larger than the 95% confidence ($\alpha = 0.05$) reference t -value and the sum of weighted residuals (9) is less than a reference χ^2 value.

$$t_i = \frac{\hat{\vartheta}_i}{t \left(\frac{1-\alpha}{2} + \frac{1-\alpha}{2} n_{exp} n_m - n_{\theta} \right) \sigma_i} \quad i = 1, \dots, n_{\theta} \quad (8)$$

$$\chi^2 = \sum_{i=1}^{n_m} \sum_{j=1}^{n_{sp}} \frac{(y_{ij} - \hat{y}_{ij})^2}{\sigma_i^2} \quad (9)$$

The trace of the FIM (7) of the experiments is used as a scalar measure of information in ϕ . If plotted on a graph against conditions of ϕ it allows to identify the most informative regions of the design space to provide a ranking of experiments (Galvanin et al., 2016).

3. Case Study

3.1. Definition of experimental design space

The model by Gyngazova et al. (2017) (10-15) was developed for HMF hydrogenation in a lab-scale batch reactor (Figure 1) and was used as a basis for candidate kinetic models.

$$dC_{HMF}/dt = -k_1 C_{HMF} \quad (10)$$

$$dC_{BHMF}/dt = k_1 C_{HMF} - k_2 C_{BHMF} - k_5 C_{BHMF} \quad (11)$$

$$dC_{MFA}/dt = k_2 C_{BHMF} - k_3 C_{MFA} \quad (12)$$

$$dC_{DMF}/dt = k_3 C_{MFA} - k_4 C_{DMF} \quad (13)$$

$$dC_{DMTHF}/dt = k_4 C_{DMF} \quad (14)$$

$$dC_{DHMTHF}/dt = k_5 C_{BHMF} \quad (15)$$

$$\phi = [C_{HMF}(0), C_{DMF}(0), T, \tau] \quad (16)$$

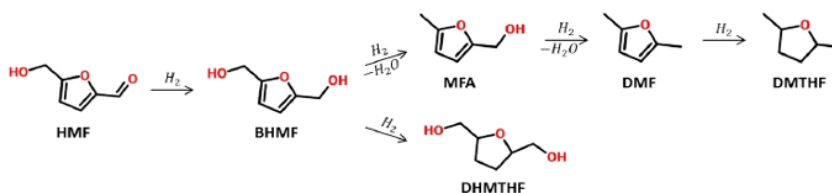


Figure 1 – Proposed HMF hydrogenation mechanism in THF solvent using Ni/C catalyst

The design vector ϕ (16) is defined in the design space ϕ (Table 1). The experimental setup bounds temperature and duration. Ranges for reagents were obtained from literature (Gyngazova et al., 2017). In ϕ , 30 in-silico experiments were run assuming an even t_{sp} [min] distribution. Gaussian noise with $\sigma = 0.03$ M was used to perturb simulated concentrations; this value reflects the expected error in actual experiments (Bindwal and Vaidya, 2014). Matlab v.9.5 was used to generate in-silico measurements and LHS points (Khaled, 2020). gPROMS ProcessBuilder v.1.4.0 was used for MBD_{oE} (supporting A-, E- and D- optimality) and nonlinear parameter estimation using maximum likelihood.

Table 1 – Definition of design space ϕ

	Lower	Upper	Unit
$C_{HMF}(0)$	0.01	0.2	M
$C_{DMF}(0)$	0.01	0.2	M
T	300	600	K
τ	60	360	min

3.2. Candidate kinetic models

• **M1:** A temperature dependency is introduced into the original model (10-16) using Arrhenius equation (17). A_i [min^{-1}] and E_{Ai} [J mol^{-1}] are estimated instead of k_i [min^{-1}]. The new parameters are computed using experimental data from Gyngazova et al. (2017), which fit the reformulation in (18) with $R^2 > 0.91$.

$$k_i = A_i e^{-\frac{E_{Ai}}{RT}} \quad (17)$$

$$\ln(k_i) = -\frac{E_{Ai}}{RT} + \ln(A_i) \quad (18)$$

• **M2:** Buzzi-Ferraris and Manenti (2009) outline a reparametrisation (19) to reduce correlation. B_i [min^{-1}] and C_i [K] are new, re-balanced parameters to be estimated. Initial guesses are obtained by rearranging equations (20 & 21). The mean temperature, T_m [K], is fixed at a suitable average of experimental temperatures (Schwaab et al., 2008).

$$k_i = B_i e^{C_i \left(\frac{1}{T} - \frac{1}{T_m} \right)} \quad (19)$$

$$B_i = \exp \left(\ln(A_i) + \frac{C_i}{T_m} \right) \quad (20)$$

$$C_i = -E_{Ai}/R \quad (21)$$

• **M3:** this is a reduced model based on M1 that estimates parameters E_{Ai} and keeps A_i constant at the values obtained from fitting experimental data in (18); these are the values that would otherwise be used as an initial guess in the parameter estimation.

Table 2 - Correlation matrix for M1 obtained from 30 experiments generated using LHS

θ	A_1	A_2	A_3	A_4	A_5	E_{A1}	E_{A2}	E_{A3}	E_{A4}	E_{A5}
A_1	1.00									
A_2	-0.18	1.00								
A_3	-0.23	-0.35	1.00							
A_4	-0.03	0.00	-0.06	1.00						
A_5	-0.17	0.95*	-0.34	0.04	1.00					
E_{A1}	1.00*	-0.18	-0.22	-0.03	-0.17	1.00				
E_{A2}	-0.19	1.00*	-0.35	0.00	0.95*	-0.19	1.00			
E_{A3}	-0.22	-0.35	1.00*	-0.06	-0.34	-0.22	-0.35	1.00		
E_{A4}	-0.03	0.00	-0.05	1.00*	0.04	-0.03	0.00	-0.05	1.00	
E_{A5}	-0.17	0.94	-0.33	0.04	1.00*	-0.18	0.94	-0.33	0.04	1.00

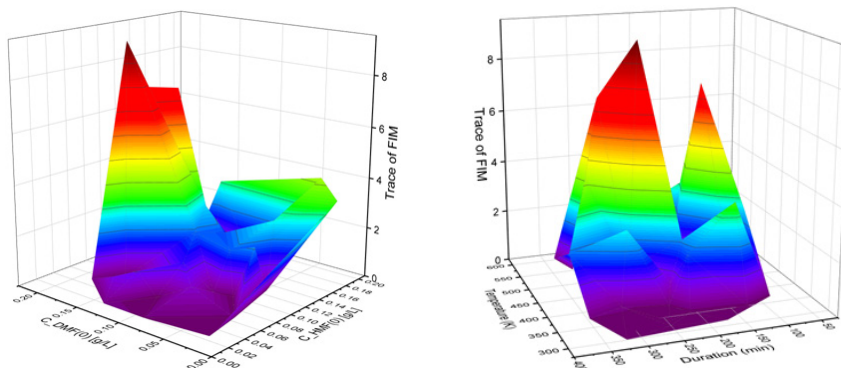
3.3. Simulation results for candidate kinetic models

- **M1:** Critical correlations (Table 2) exist between parameters E_{A_i} and A_i . The model passes the χ^2 test despite large residuals affecting individual experiments. The t-test was passed, however, the t-values for A_i were orders of magnitude higher than the ones for E_{A_i} . This was due to unrealistically large estimated values for A_i , passing the t-test only because, according to (8), the corresponding standard deviation σ was small in comparison. This makes the t-test untrustworthy and explains the good results obtained despite high correlations. The contribution to the overall information content (FIM) is also much higher for E_{A_i} than for A_i , underlining a low sensitivity to A_i in ϕ for M1.

- **M2:** First, the value of T_m was fixed at 450K (centre of the temperature design range). Then, using this value of T_m , the new parameters B_i and C_i were calculated from data points generated using M1, which reliably describes the original system at these conditions. The conditions suggested by the LHS were then used with these estimates for B_i and C_i . Correlations were significantly reduced and critical correlations appeared only between C_2 and C_5 . However, the deviation between M1 and M2 was so large that M2 no longer accurately represented the original system. The above estimation procedure was repeated fixing T_m at different values clustered around the centre of the design space to see if the choice of T_m affected the lack of fit. Parameter correlations were reduced but still the model was not adequate to represent the system in a large set of conditions in ϕ .

- **M3:** Using 30 experiments generated using LHS, there were no critical correlations between the estimated parameters for this reduced model. All parameters were estimated satisfactorily except E_{A3} (Table 3), which depends heavily on the MFA concentration (12-13). The χ^2 test was passed overall, but some experiments with low MFA concentrations deviated more in their predictions. MFA concentration may practically be too low to be detected at some experimental conditions, and this affects parameter estimation.

A FIM-based information analysis was then conducted on M3 to determine the most informative regions of the design space to explore in future experiments for a precise estimation of model parameters (Figure 2). The analysis identified two potential regions: 1) $C_{DMF}(0) = 0.19$ M and $C_{HMF}(0) = 0.10$ M in a $\tau = 230$ min long batch experiment at $T = 510$ K; 2) $C_{DMF}(0) = 0.09$ M and $C_{HMF}(0) = 0.17$ M in a $\tau = 280$ min long batch experiment at $T = 530$ K. The temperature in the reactor could thus be bounded between 450-550 K to improve parameter estimability.



(a) Initial HMF and DMF concentrations (b) Temperature and experiment duration
 Figure 2 – Information maps in terms of design variables obtained by plotting the trace of the FIM of all experiments using an information surface in model M3.

Table 3 – M3: confidence intervals after parameter estimation using LHS conditions

θ [kJ mol ⁻¹]	Final Value	Initial Guess	95% t-value (reference value = 1.65)
E_{A1}	66.99	58.79	1939.57
E_{A2}	66.49	64.48	676.13
E_{A3}^{**}	6.57**	42.00**	0.001**
E_{A4}	76.36	83.61	3772.45
E_{A5}	63.86	62.17	492.01

4. Conclusions

A two-step identifiability study was applied to three candidate kinetic models for HMF hydrogenation. M1 and M2 have been found unsatisfactory due to high correlation and low fidelity, respectively. The reduced model M3 showed satisfactory fitting performance but limitations on representing E_{A3} resulting from low and potentially undetectable MFA concentrations. Applying a FIM-based information analysis allowed to define the most informative regions of the experimental design space for a precise estimation of kinetic parameters. The most informative was at high DMF and lower HMF concentrations in a batch experiment lasting 230 min at 510 K. Future work will aim to confirm these results using robust MBDoE techniques to precisely estimate the full set of model parameters.

References

- Bindwal, A. B. and Vaidya, P. D., 2014, Reaction Kinetics of Vanillin Hydrogenation in Aqueous Solutions Using a RuC Catalyst. *Energy & Fuels*, 28(5), 3357–3362.
- Buzzi-Ferraris, G. and Manenti, F., 2009, Kinetic models analysis. *Chem Eng Sc*, 64, 1061-1074.
- Chis, O., Villaverde, A., Banga, J. and Balsa-Canto, E., 2016, On the relationship between sloppiness and identifiability. *Math Biosc*, 282, 147-161.
- Franceschini, G. and Macchietto, S., 2008, Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sc*, 63, 4846-4872.
- Galvanin, F., Enhong, C., Al-Rifai, N., Gavriilidis, A. and Dua, V., 2016, A joint model-based design approach for the identification of kinetic models in continuous flow laboratory systems. *Comp Chem Eng*, 95, 202-215.
- Gawade, A.B., Tiwari, M.S. and Yadav, G.D., 2016, Bio-based green process: Selective hydrogenation of 5-hydroxymethylfurfural to 2,5-dimethylfuran under mild conditions using Pd-s2.5h0.5pw12o40/K-10 Clay. *ACS Sus Chem Eng*, 4(8), 4113-4123.

- gPROMS Process Builder version 1.4.0. Process Systems Enterprise, London, England.
- Grilic, M., Likozar, B. and Levec, J., 2014, Hydrodeoxygenation and hydrocracking of solvolysed lignocellulosic biomass by oxide, reduced and sulphide form of NiMo, Ni, Mo and Pd catalysts. *App Cat B: Env.*, 150-151, 275-287.
- Gyngazova, M.S., Negahdar, L., Blumenthal, L.C and Palkovits, R., 2017, Experimental and kinetic analysis of the liquid phase hydrodeoxygenation of 5-hydroxymethylfurfural to 2,5-dimethylfuran over carbon-supported nickel catalysts. *Chem Eng Sc*, 173, 455-464.
- Jain, A.B. and Vaidya, P.D., 2016, Kinetics of catalytic hydrogenation of 5-hydroxymethylfurfural to 2,5-bis-hydroxymethylfuran in aqueous solution over RuC. *Int J Chem Kin*, 48(6), 318–328.
- Khaled, N., 2020, Latin Hypercube, MATLAB Central File Exchange.
- Luo, J., Arroyo-Ramirez, L., Wei, J. and Yun, H., 2015, Comparison of HMF hydrodeoxygenation over different metal catalysts in a continuous flow reactor. *App Cat A: Gen*, 508, 86–93.
- Montgomery, D.C., 2012, *Design and Analysis of Experiments*, 8th edition. Wiley.
- Schwaab, M., Lemos, L., Pinto, J., 2008, Optimum reference temperature for reparametrization of the Arrhenius equation Part 2 Problems involving multiple reparametrizations. *Ch Eng Sc*, 63.
- van Putten, R., Waal, J. C. and de Jong, E., 2013, Hydroxymethylfurfural, A Versatile Platform Chemical Made from Renewable Resources. *Chem Rev*, 113.

Continuous-Effort Approach to Model-Based Experimental Designs

Kennedy Putra Kusumo^a, Kamal Kuriyan^a, Salvador García-Muñoz^b, Nilay Shah^a, and Benoît Chachuat^{a,*}

^a *Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom*

^b *Small Molecule Design and Development, Lilly Research Laboratories, Eli Lilly & Company, Indianapolis, Indiana 46285, United States*

Abstract

Model-based design of experiments is a technique for accelerating the development of mathematical models. Through maximally informative experiments, time and resources for estimating uncertain model parameters are minimized. This article presents a method for computing effort-based experimental designs, whereby designs are akin to experimental recipes. As well as identifying which experiments are the most informative, the optimal experimental effort to dedicate to each experiment is also optimized. Upon discretizing the experimental design space and treating the efforts as continuous decision variables, this method leads to convex optimization problems regardless of the model structure, which is ideal for large, parallel experimental campaigns. The case study of a batch reactor model with four parameters is presented to illustrate the methodology.

Keywords: optimal experiment design, model-based design of experiments, experimental effort, information, parametric uncertainty

1. Introduction

Optimal experiment design (OED) has shown great success in accelerating the development of mathematical models through providing maximally informative experimental campaigns. The seminal example is its application to construct *statistical* models such as response surface models (RSM), where OED led to the discovery of the celebrated factorial designs and their derivatives (Box and Wilson, 1951; Box and Hunter, 1961). Although highly effective, this approach can still require significant resources and time to develop for problems with many experimental degrees of freedom. In addition, its applicability to dynamic systems remains limited (Georgakis, 2013).

When applied to a *mechanistic* modelling framework, this approach leads to a systematic, optimization-based methodology, for which the name model-based design of experiments (MBDoE) was coined (Franceschini and Macchietto, 2008). MBDoE is often applied to expedite the development of knowledge-driven or hybrid (static or dynamic) models, which have superior extrapolation capability compared to statistical models and may be transferable to other similar systems. Such system-tailored strategies can significantly reduce the experimental effort for large-scale design problems, but come at the price of a much higher complexity and computational burden.

Consequently, MBDoe has become almost synonymous with sequential design and experimentation, that is, one experiment designed and conducted at a time with parameter update in between each experiment. This seems to have caused the concept of *effort*-based experimental designs, which is widely used in classical OED for the simultaneous design of experiments (Nguyen and Miller, 1992; Fedorov and Leonov, 2014), to be somewhat forgotten in MBDoe.

This paper aims to reintroduce the concept of effort within MBDoe. The focus is on methods that (i) discretize the continuous experimental spaces into a finite set of experimental candidates then (ii) seek to determine optimal effort to dedicate to each candidate. Section 2 states the problem of interest, reviews some background material on OED, and presents the methodology. Then, Section 3 illustrates the methodology with the case study of a batch reactor with two experimental degrees-of-freedom and four model parameters.

2. Methodology

2.1. Experimental Designs

Consider an experimental system, whereby $x \in X \subset R^{n_x}$ denotes the vector of experimental variables or controls. An experimental campaign design, or simply experiment design, Ξ is often made up of repeated combinations of $N_c > 0$ unique candidate experiments. It may be formally defined as

$$\Xi := \left\{ x_1 \dots x_{N_c} p_1 \dots p_{N_c} \right\} \quad (1)$$

where $x_i \in X$ and $p_i \in [0, 1]$ refer to the experimental variables and effort associated with the i th candidate experiment, respectively. If r_i is the number of times the experiment x_i is repeated, and $N_t \geq N_c$ is the total number of experiments, then $p_i = r_i/N_t$.

Designs where p_i are only allowed to be multiples of the fraction $1/N_t$ are called discrete (effort) designs and require for the N_t to be specified a priori. In contrast, continuous (effort) designs allow p_i to vary continuously, subject to the convex combination constraints

$$\sum_{i=1}^{N_c} p_i = 1 \quad \text{and} \quad p_i \geq 0 \quad \forall i \quad (2)$$

thereby making them independent of N_t . For theoretical considerations, continuous designs may be regarded as discrete designs with $N_t = \infty$, justifying that the distinction between continuous and discrete designs become less relevant as N_t gets larger.

2.2. Optimal Experimental Design Formulations

Classical OED formulations are concerned with linear response models of the form

$$y = F(x) \theta \quad (3)$$

where $y \in R^{n_y}$ denotes the vector of responses, and $\theta \in R^{n_\theta}$ the vector of model parameters. The $n_y \times n_\theta$ matrix F is called the regressor matrix. It is comprised of

vertically stacked regression vectors f_j^T , which are user-chosen functions of x that should differ between responses.

Such mathematical models can be built using a data set $(x_1, \dots, y_1), \dots, (x_{N_t}, \dots, y_{N_t})$ collected during an experimental campaign. The root problem of OED, therefore, is to determine an experiment design that is maximally informative. One way of measuring such information content is by means of the (normalized) Fisher Information Matrix (FIM) which, for the experimental campaign (1) and the linear model (3), can be written as

Optimality Criterion	ϕ	
	Standard	Eigen-form
Determinant	$\log \log (\det \det (M))$	$\log \log \left(\prod_k \lambda_k\right)$
Average of trace	$-\left(M^{-1}\right)$	$-\sum_k \lambda_k^{-1}$
Extreme	$\lambda_{\min } (M)$	λ_k

Table 1: standard information criteria for parameter precision.

$$M(\Xi) = \sum_{i=1}^{N_c} p_i A_i, \text{ with } A_i = F\left(x_i\right)^T F\left(x_i\right), \tag{4}$$

where $A_i \in R^{n_\theta \times n_\theta}$ is the atomic information matrix associated to the i th candidate experiment. This expression clearly shows the FIM's dependence on Ξ , and takes advantage of the FIM's additivity property for independent experiments.

An *optimal* experiment design Ξ^* is a design that maximizes some scalar criterion ϕ of the FIM M

$$\Xi^* \in \arg \arg \phi(M(\Xi)) \tag{5}$$

Notice that this formulation is applicable to both discrete and continuous (effort) designs. Three classical scalar criteria of the FIM are reported in Table 1 – the focus in the remainder of this short paper is on D-optimal designs.

Now turning to generic nonlinear process models of the form

$$y = f(x, \theta) \tag{6}$$

Where the mapping f needs not be available in closed form, e.g. could be defined implicitly via the solution of algebraic and/or differential equations. The main challenge for computing optimal designs for nonlinear models is that the FIM for a nonlinear model becomes a function of the model parameters θ , making the optimal designs also a function of θ . A popular approach in MBDoe entails linearizing the model (6) around a nominal model parameter value θ^n ,

$$y \approx f\left(x, \theta^n\right) + \frac{\partial f}{\partial \theta}\left(x, \theta^n\right)\left[\theta - \theta^n\right] \tag{7}$$

After rearrangement, the linearized model (7) may be recast into the form of (3), with a linearized response $\tilde{y}:=y - f\left(\cdot, \theta^n\right)$, model parameter $\tilde{\theta}:=\theta - \theta^n$, and regressor matrix

$\tilde{F} := \frac{\partial f}{\partial \theta}(\cdot, \theta^n)$. In turn, the latter can be substituted in Equation (4) to compute a linearized FIM, so standard methods can then be applied to compute an optimal design for the linearized model. Such a design is said to be *locally-optimal* around θ^n for the original nonlinear model (6).

2.3. Continuous-Effort Approach to MBDoe

Searching over Ξ in Problem (5) is particularly challenging as it entails determining the optimal number of candidates $N_c \in Z^+$, the optimal experimental variables $x_i \in R^{n \times N_c}$, and the optimal experimental efforts $p \in [0, 1]^c$, all together (see, e.g., De Castro et al., 2019). Consequently, existing MBDoe techniques rely on approximations.

In the traditional MBDoe setting (Franceschini and Macchietto, 2008), the number of experiments and associated efforts are usually fixed so the search proceeds over the experimental variables only. Yet a limitation is that these problems are typically riddled with local optima. In contrast, we propose to discretize the continuous experimental variable space X into a sampling set X_s comprising $N_s \gg N_c$ experimental candidates. This reduces the problem (5) to searching over the experimental efforts for the N_s candidates,

$$\left(p_1^*, \dots, p_{N_s}^* \right) \in \arg \arg \left\{ \phi \left(\sum_{i=1}^{N_s} p_i A_i \right) : \sum_{i=1}^{N_s} p_i = 1, p_i \geq 0 \forall i \right\} \quad (8)$$

When computing continuous-effort designs, $p_i \in [0, 1]$, the problem formulation (8) is convex insofar as the function ϕ is concave. And all three standard criteria listed in Table 1 are indeed concave functions. In addition to precluding locally suboptimal designs, this property thereby enables the use of powerful convex optimization solvers to solve large-scale experimental design problems efficiently—the size of the problem being directly proportional to the number of sample design candidates N_s but independent of the total number of experiments N_t to be designed. Notice also that the complexity associated with the chosen model structure is moved away to a prior step, which involves precomputing the atomic information matrices A_i , $i = 1 \dots N_s$ of each design candidate. Although costly, these atomic information matrices are independent of the experimental efforts and could thus be precomputed in parallel to save time.

Lastly, it is worth mentioning that the methodology is by no means restricted to a specific discretization strategy. Though without an effective discretization scheme, computation of the atomic matrices could quickly become intractable as the number of experimental variables n_x increases. We thus see much potential in developing smart discretization strategies tailored to continuous-effort MBDoe approaches, as part of future work.

3. Case Study

Consider an experimenter developing a model of a batch reactor running the reaction $A \rightarrow \nu B$. The proposed model is given by

$$\frac{dc_A}{dt} = -r, \quad \frac{dc_B}{dt} = \nu r, \quad r = kc_A^\alpha, \quad k = \exp \exp \left(\theta_0 + \theta_1 \left(\frac{T - T_{ref}}{T} \right) \right) \quad (9)$$

where the order of the reaction α , the stoichiometric ratio ν , the (transformed) pre-exponential constant θ_0 , and the (transformed) activation energy θ_1 of the reaction are the unknown model parameters to be estimated from experimental data. The reaction rate r is modelled to be temperature-dependent, with a reaction constant k that follows Arrhenius-type kinetics. The experimental variables of the system are the sampling times $t_{sp} \in [0, 200]$ in minutes, the initial A concentration $c_A^0 \in [1, 5]$ in moles/L, and temperature $T \in [273.15, 323.15]$ in K. The nominal model parameter values are $\theta_0 = -4.50$, $\theta_1 = 2.20$, $\alpha = 1$, and $\nu = 0.5$. The setup measures c_A and c_B at a unique sampling time t_{sp} during each batch. It is decided to compute a D-optimal experimental design, and all the results are obtained using the open-source python library Pydex (<https://github.com/KennedyPutraKusumo/pydex/>).

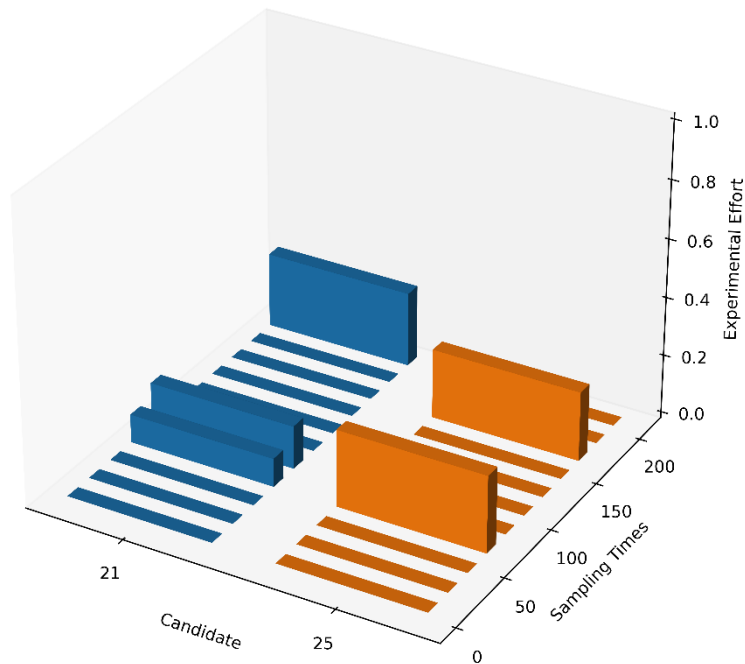


Figure 1: D-optimal continuous-effort experimental design. The effort of each candidate equals the sum of efforts allocated to each sampling time variant.

Using the discretization approach, 25 candidate experiments are enumerated from a 5×5 mesh-centered grid of c_A^0 and T . For each candidate, 11 sampling times are enumerated from an equally-spaced time grid. The atomic information matrices are computed, with the linearization via finite differences using numdifftools (Brodtkorb and D’Errico, 2015) taking a total of 2.40 CPU-sec, followed by solution of (8), taking 0.25 CPU-sec using MOSEK (MOSEK ApS, 2020) through CVXPY (Diamond and Boyd, 2016).

The computed optimal non-zero efforts are presented in Figure 1. Out of the 25 candidate experiments, only two are part of the optimal design: Candidates 21 and 25.

They correspond to experiments with $\left[c_A^0, T \right]$ of [5.00, 273.15] and [5.00, 323.15], respectively. The optimal effort for Candidate 21 is 49.85%, with three sampling time variants. 9.71% samples at $t_{sp} = 60$, 14.67% samples at $t_{sp} = 80$, and the remaining 25.48% samples at $t_{sp} = 200$. Notice that $t_{sp} = 60$ and $t_{sp} = 80$ are adjacent sampling times, which suggests that the "true" optimal sampling time could lie between 60 and 80 minutes, with an optimal effort of $9.71 + 14.67 = 24.38\%$. Furthermore, one may reasonably guess that the true sampling time is closer to 80 minutes as the effort allocated there is larger, although this is difficult to ascertain. Candidate 25 is allocated 50.15%, a near equal split with Candidate 21. There are two different sampling time variants, 26.40% samples at $t_{sp} = 60$, whilst the other 23.75% will sample later at $t_{sp} = 160$. Candidates 21 and 25 tell us that an informative experimental campaign involves running experiments with c_A^0 as high as possible, and alternating between the highest and lowest temperatures. The pattern followed by the optimal sampling times comprises a measurement early during the batch and a second measurement near the end of the batch.

In practice, using a continuous design requires a rounding procedure in order to convert the optimal fractional efforts to integers. Out of existing rounding procedures, Adam's method of apportionment is probably superior (Pukelsheim and Rieder, 1992). A multiplier v is set equal to N_t initially. Each effort p_i is multiplied by v , and rounded up to the nearest integer; the result is called the apportionment of each candidate a_i . If $\sum a_i > N_t$, v is reduced using a heuristic, e.g. multiplied by 0.90. If $\sum a_i < N_t$, v is increased, again by a heuristic. This is done iteratively until $\sum a_i = N_t$. The final apportionment a_i 's are the desired rounding.

4. Conclusion

The success of model-based experimental design in process systems engineering is accompanied by challenges that still needs to be addressed. We introduced a methodology that brings the concept of continuous efforts into model-based experimental designs, ideal for designing parallel experimental campaigns. The methodology involves discretizing the continuous experimental variables into a set of experimental candidates. This approximation brings advantages over a standard model-based experimental design, including a guaranteed convex optimization, and a problem size that does not scale with the number of experiments. Although effective, efficiency of the methodology hinges upon effective discretization procedures. Even without an effective discretization procedure (gridding), we demonstrated the tractability and efficiency of the method in designing an experiment for a batch reactor. There is a wealth of unexplored areas that follow up from this work. Studies into the different attitudes an experimenter can have towards sampling times, various robustification strategies against model-parametric uncertainties, and constrained experimental designs are amongst the most promising directions.

Acknowledgements: This work is co-funded by Eli Lilly & Company and the Engineering and Physical Sciences Research Council (EPSRC) as part of its Prosperity Partnership Programme under grant EP/T518207/1. The authors thank Shankar Vaidyaraman for the insightful discussions throughout the preparation of this article.

References

- G. E. Box, J. S. Hunter, 1961. The $2k-p$ fractional factorial designs. *Technometrics* 3 (3), 311–351.
- G. E. Box, K. B. Wilson, 1951. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society: Series B (Methodological)* 13 (1), 1–38.
- P. A. Brodtkorb, J. D’Errico, 2015. numdifftools 0.9.11. <https://github.com/pbrod/numdifftools>.
- Y. De Castro, F. Gamboa, D. Henrion, R. Hess, J. B. Lasserre, 2019. Approximate optimal designs for multi-variate polynomial regression. *Annals of Statistics* 47 (1), 127–155.
- S. Diamond, S. Boyd, 2016. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research* 17 (83), 1–5.
- V. Fedorov, S. L. Leonov, 2014. *Optimal Design for Nonlinear Response Models*. CRC Press, 2013.
- G. Franceschini, S. Macchietto, 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science* 63 (19), 4846–4872.
- C. Georgakis, 2013. Design of dynamic experiments: A data-driven methodology for the optimization of time-varying processes. *Industrial & Engineering Chemistry Research* 52 (35), 12369–12382.
- MOSEK ApS, 2020. MOSEK Optimizer API for Python 9.2.29. URL <https://docs.mosek.com/9.2/pythonapi/index.html>
- N. K. Nguyen, A. J. Miller, 1992. A review of some exchange algorithms for constructing discrete D-optimal designs. *Computational Statistics & Data Analysis* 14 (4), 489–498.
- F. Pukelsheim, S. Rieder, 1992. Efficient rounding of approximate designs. *Biometrika* 79 (4), 763–770.

Application of Machine Learning and Global Sensitivity Analysis for Identification and Visualization of Design Space

Sergei Kucherenko,^{a,*} Oleksiy Klymenko,^b Nilay Shah^a

^a*Imperial College London, United Kingdom*

^b*University of Surrey, United Kingdom*

s.kucherenko@imperial.ac.uk

Abstract

The design space (DS) is defined as the combination of materials and process conditions which provides assurance of quality for a pharmaceutical. A model-based approach to identify a probability-based DS requires costly simulations across the entire process parameter space (certain) and the uncertain model parameter space (e.g. material properties). We demonstrate that application of metamodel-based filters and global sensitivity analysis (GSA) can significantly reduce model complexity and reduce computational time for identifying and quantifying DS. Once DS is identified it is necessary to present it graphically. The output of identification of DS is a multi-dimensional probability map. The projection of the multi-dimensional DS to a 2D representation is still unavoidable irrespectively of the method used to reach such probability mapping. We showed that application of constraint GSA can dramatically reduce the number of required for visualization 2D projections.

Keywords: Design Space, probability map, Sobol' sensitivity indices, global sensitivity analysis

1. Problem Setting

Pharmaceutical products should be approved by a regulatory body to be manufactured on the condition that there is enough evidence to demonstrate the safety and efficacy of the drug product, and that the manufacturing process has “controls” to assure the quality of it. These “controls” represent a set of “admissible operating conditions” that provide an assurance of quality, accounting for variations in the materials and manufacturing processes. They are commonly referred to as the Design Space (DS) (U.S. Department of Health and Human Services, 2006).

Identification of probabilistic DS is a demanding task and for a typical practical problem the traditional approach based on exhausting sampling requires costly computations (García-Muñoz et al, 2015). A novel theoretical and numerical framework for determining probabilistic DS using metamodeling and adaptive sampling was proposed in Kucherenko et al, 2020. It was based on the multi-step adaptive technique using a metamodel for a probability map as an acceptance-rejection criterion to optimize sampling to identify the DS.

Consider a model $f(\vec{x}; \vec{\theta})$, and the vector of constraint functions $\vec{g}(\vec{x}; \vec{\theta})$, where \vec{x} is a vector of process parameters defined in the box domain H^d , $\vec{\theta}$ is a vector of uncertain model parameters defined in m -dimensional real space R^m with a given pdf $\varphi(\vec{\theta})$. Define probability $p(\vec{x})$ of occurrence of an undesirable (reliability estimation analysis) or desirable (pharmaceutical/chemical engineering) event as

$$p(\vec{x}) = P\left(\vec{g}(\vec{x}; \vec{\theta}) \geq \vec{g}^*\right) = \int_{\Omega(\vec{x})} \varphi(\vec{\theta}) d\vec{\theta} \quad (1)$$

Here \vec{g}^* is a vector of given thresholds, domain $\Omega(\vec{x})$ is defined as $\Omega(\vec{x}) = \left\{ \vec{\theta}: \vec{g}(\vec{x}; \vec{\theta}) \geq \vec{g}^* \right\}$. Probability (1) can be presented in a convenient form as:

$$p(\vec{x}) = \int_{R^m} I(\vec{g}(\vec{x}; \vec{\theta}) \geq \vec{g}^*) \varphi(\vec{\theta}) d\vec{\theta} = E_{\vec{\theta}} \left[I(\vec{g}(\vec{x}; \vec{\theta}) \geq \vec{g}^*) \right], \quad (2)$$

where I is an indicator function. Define design space (DS) as $DS(\vec{x}; p(\vec{x}) \geq p^*)$, where p^* is a critical (acceptable) value of probability of meeting constraints. A brute force numerical estimation of $p(\vec{x})$ requires $N_x \times N_{\theta}$ model runs, where N_x is the number of points in the space of the process parameters \vec{x} , N_{θ} is the number of sampled points in the $\vec{\theta}$ - space. For a typical practical problem this number can be very large which results in costly computations. For such models, global sensitivity analysis (GSA) is an efficient practical tool which can dramatically reduce computational costs and provide useful additional information about importance of parameters and model structure.

GSA is a widely used tool for identification of key parameters whose uncertainty most affects model output. In the context of DS identification GSA can be applied to identify uncertain parameters contributing to the acceptance of imposed constraints with potential further model reduction. This problem is linked to the problem of identification of critical regions of inputs $\vec{\theta}$ responsible for "extreme" values (values for which constraints are satisfied) of model response. It is known as a factor mapping setting of GSA. It typically requires rather complex techniques such as Monte Carlo Filtering. We propose an efficient GSA formulation using an indicator function $I(\vec{g}(\vec{x}; \vec{\theta}) \geq \vec{g}^*)$ as a "model output" which allows the application of a well-developed GSA method of Sobol' indices.

In order to illustrate the developed approach a single-stage batch reactor problem was considered, with two "bounded" variables. This system was introduced in Samsatli et al. 1999 and used in Kucherenko et al., 2020. It is assumed that the following reactions take place in a fixed volume vessel with a given initial charge of solution A (1 m³ of solution A with a concentration 2000mol/m³): $2A \xrightarrow{k_1} B \xrightarrow{k_2} C$. The processing time and temperature profiles may be adjusted to maintain the purity and production rate of the desired product B above certain levels (representing constraints of the form

$\vec{g}(\vec{x}; \vec{\theta}) \geq \vec{g}^*$. The kinetic parameters of this model $\vec{\theta}$ (two activation energies E and two pre-exponential factors k) are considered to be uncertain.

2. Machine Learning and Global Sensitivity Analysis

As a first step we applied a framework for determining probabilistic DS using metamodeling and adaptive sampling developed in Kucherenko et al, 2020. The (Q)RS-HDMR method developed in Zuniga et al., 2013 was used to build metamodels. Consider the ANOVA-HDMR decomposition of an integrable function $f(x) \equiv f(x_1, x_2, \dots, x_n)$ defined in the unit hypercube H^n

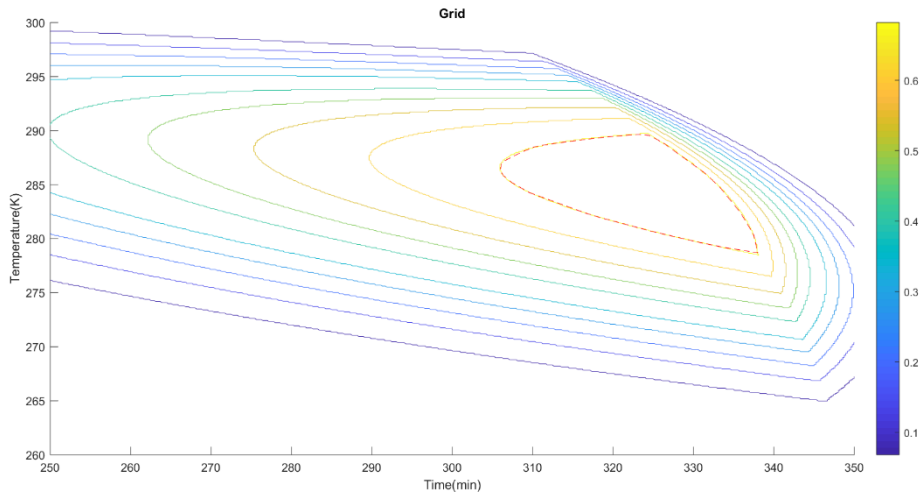


Figure 1. Contour plots of the probability map. DS results with 3 non important uncertainty parameters fixed. Broken line is DS corresponding to $p^*=0.68$.

$$f(x) = f_0 + \sum_i f_i(x_i) + \sum_{i < j} f_{ij}(x_i, x_j) + \dots + f_{12\dots k}(x_1, \dots, x_n) \quad (3)$$

It is known that this decomposition is unique if the mean value of each term with respect to integration from the set of any variable it depends on is equal to zero, in which case pairs of component terms are orthogonal with respect to integration. (Q)RS-HDMR method exploits the fact that for many practical problems only low order interactions of the input variables are important. This assumption can dramatically reduce the computational effort for building metamodels.

Assuming that component functions in (3) are piecewise smooth and continuous, they can be decomposed using a complete basis set of orthonormal polynomials:

$$f_i(x_i) = \sum_{r=1}^{\infty} \alpha_r^i \varphi_r(x_i), \quad f_{ij}(x_i, x_j) = \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} \beta_{pq}^{ij} \varphi_{pq}(x_i, x_j), \quad \dots \quad (4)$$

Here $\varphi_r(x_i)$, $\varphi_{pq}(x_i, x_j)$ are sets of one- and two-dimensional basis functions and α_r^i and β_{pq}^{ij} are the coefficients of decomposition. From the orthogonality of the basis functions it follows that:

$$\alpha_r^i = \int_0^1 f_i(x_i) \varphi_r(x_i) dx_i, \quad r = 1, \dots, k, \quad (5)$$

$$\beta_{pq}^{ij} = \int_0^1 \int_0^1 f_i(x_i) \varphi_p(x_i) \varphi_q(x_j) dx_i dx_j, \quad p = 1, \dots, l, \quad q = 1, \dots, l'.$$

The choice of the basis functions depends on the probability distributions of the inputs.

For practical purposes the summation in (4) and (5) is limited to some maximum orders k, l, l'

$$f_i(x_i) \approx \sum_{r=1}^k \alpha_r^i \varphi_r(x_i), \quad f_{ij}(x_i, x_j) \approx \sum_{p=1}^l \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j) \quad (6)$$

Hence, the HDMR approximation function up to the second order interaction can be constructed as

$$\tilde{f}(x) = f_o + \sum_{r=1}^k \alpha_r^i \varphi_r(x_i) + \sum_{p=1}^l \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j) \quad (7)$$

There is a similarity between this method and the surrogate-based feasibility analysis (Bhosekar. and Ierapetritou, 2018). However, feasibility analysis relates to the ability of a process to satisfy all relevant constraints and it is formulated in deterministic setting, while we are dealing with a probabilistic problem setting (1). In our approach metamodels were built to approximate a probability map and then use them as an acceptance-rejection criterion to optimize sampling to identify the DS, while the surrogate-based feasibility analysis aims at finding a surface defining the boundary of the feasible space within the box bounded design space as accurately as possible using only built surrogate.

At the second step we applied GSA which revealed that only one parameter out of four, namely k_{j2} is important. These results were used for complexity reduction, namely by fixing (E_{j1}, E_{j2}, k_{j1}) parameters to their nominal values and computing the DS. In this case results identical to the benchmark results were obtained with a much lower number of points in the process parameters space (Figure 1). Complexity reduction after application of GSA resulted in a 16 times speed-up. The SobolGSA software (Kucherenko et al., 2019) was used for building metamodels and computing Sobol' sensitivity indices.

3. Visualization of Design Space

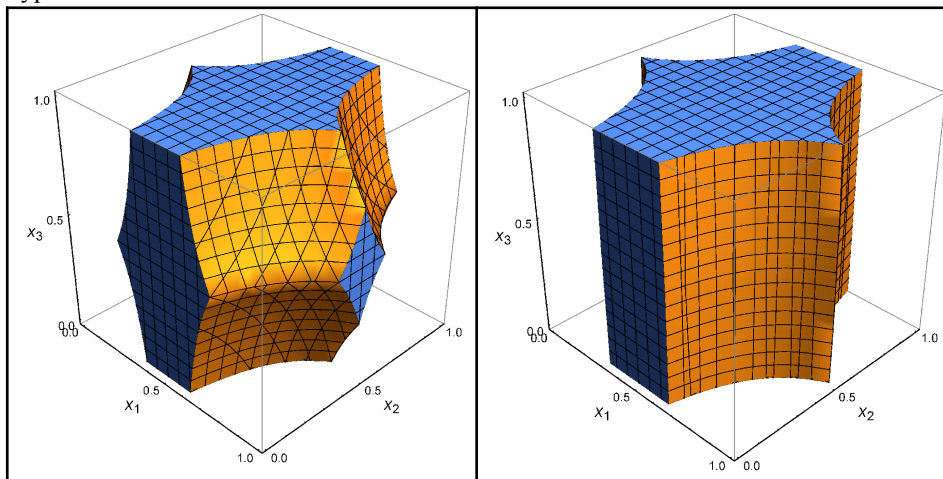
A challenge remains in the visualization of DS needed to communicate higher dimensionality probability maps and the inclusion of this information into the formal documentation required for the approval of DS by a government agency. For more than three parameters process parameters visualization of DS is a challenging task. Guidance for Industry, Q8(R2) Pharmaceutical Development (2009) has the following recommendations: "When multiple parameters are involved, the design space can be presented for two parameters, at different values (e.g., high, middle, low) within the range of the third parameter, the fourth parameter, and so on." Considering DS in d -dimensional process parameters space, a graphical presentation of DS via 2D plots prescribed in [3] would result in $N_p = \frac{d(d-1)}{2} 3^{d-2}$ 2D plots.

GSA can be applied for visualization of DS. This problem belongs to a class of models with constraints. Most existing GSA techniques were designed under the hypothesis that model inputs are independent. However, in many cases there are dependencies among inputs, which may have significant impact on the results. A constraint GSA (cGSA) as an extension of variance based Sobol' sensitivity indices that was developed in Kucherenko et al, 2017.

Consider DS in d -dimensional process parameters space and assume that GSA revealed that parameter x_i is unimportant. It means that any 2D projections corresponding to different values of x_i will be the same. Hence instead of three values ("high", "middle", "low") for this parameter only one is sufficient for the graphical presentation, which would result in the reduction of N_p to $\frac{(d+4)(d-1)}{2} 3^{d-3}$ 2D plots. In the case of s non-important parameters N_p will be reduced to $[(d - s)(d + 5s - 1) + 9s(s - 1)]3^{d-s-2}/2$. For illustration we consider a 3D Sobol' "g-function":

$$f(x) = \prod_{i=1}^3 \frac{|4x_i - 2| + a_i}{1 + a_i} \tag{8}$$

with three different sets of parameters a_i and x_i uniformly distributed in the unit hypercube H^3 .



(a)

(b)

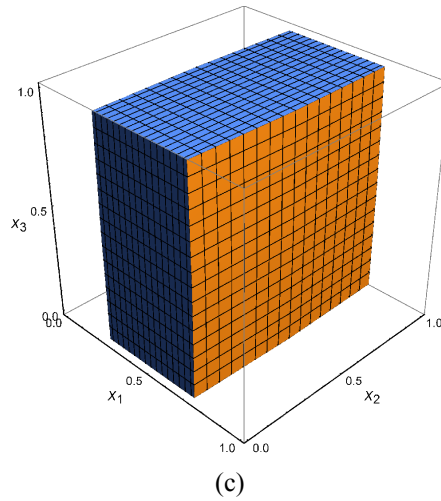


Figure 2. 3D DS shape. (a) $a = \{0; 1; 2\}$; $SI = \{0.706; 0.0818; 0.0234\}$, $SIT = \{0.856; 0.229; 0.104\}$; (b) $a = \{0; 1; 99\}$; $SI = \{0.761; 0.0957; 0.002\}$, $SIT = \{0.887; 0.238; 0.003\}$; (c) $a = \{0; 49; 99\}$; $SI = \{0.996; 0.0004; 0.0005\}$, $SIT = \{0.996; 0.0005; 0.0006\}$. SI and SI^T are the values of the main effect and total Sobol' sensitivity indices respectively.

A synthetic DS is defined as $DS(\vec{x}; f(\vec{x}) \geq 1)$ where f acts as “probability $p(\vec{x})$ ”. Its shape for different sets of parameters a_i is shown in Figure. 1. In the first case (Figure 2, a) all three 2D projections are important and the prescription given above requires plotting $N_p = \frac{d(d-1)}{2} 3^{d-2} = 9$ 2D projections. GSA reveals that in this case all three input parameters are important. In the second case (Figure 2, b) although all 2D projections onto the planes (x_1, x_2) , (x_1, x_3) and (x_2, x_3) are important, 2D projections on the plane (x_1, x_2) are independent of parameter x_3 . Hence only $N_p = \frac{(d+4)(d-1)}{2} 3^{d-3} = 7$ 2D projections are required for the visualization of DS. GSA shows that in this case parameter x_3 is not important. In the third case (Figure 2, c) only 1D projection to the x_1 axis is important while for two other parameters DS shape is defined in the whole domain of their change $[0,1]$ hence no 2D projections are of any interest. GSA shows that in this case only parameter x_1 is important.

4. Conclusions

We showed on a typical problem from the chemical industry that application of metamodel-based filters can significantly reduce model complexity and computational costs for identification of DS. GSA provides important information about the importance of uncertain parameters and can further reduce computational costs. We propose formulations via indicator functions which allow transformation of constrained problem into a class of problems with independent variables and enabled application of a well-developed variance-based method of Sobol' indices. Our approach is applicable to a wide range of process designs where the operational aspects are treated in terms of ranges rather than fixed points and where system models include uncertain parameters.

Graphical presentation of multidimensional DS is challenging task. The projection of the multi-dimensional DS to a 2D representation is required by the industry regulatory bodies. We showed that constraint GSA can be applied to dramatically reduce the number of required for visualization 2D projections.

Acknowledgements

We acknowledge the financial support of Eli Lilly and Company.

References

- Guidance for Industry, Q8(R2) Pharmaceutical Development, 2009, U.S. Department of Health and Human Services
- A. Bhosekar, M. Ierapetritou, 2018. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers & Chemical Engineering*, 108, 250-267.
- S. García-Muñoz, C. V. Luciani, S. Vaidyaraman, K. D. Seibert, 2015, Definition of Design Spaces Using Mechanistic Models and Geometric Projections of Probability Maps. *Organic Process Research and Development*. 19 (8), 1012-1023.
- S. Kucherenko, O. Klymenko, N. Shah, 2017, Sobol' indices for problems defined in non-rectangular domains, *Reliability Engineering and System Safety* 167, 218–231.
- S. Kucherenko S., O. Zacheus, 2019, SobolGSA manual, Imperial College London. <https://www.imperial.ac.uk/process-systems-engineering/research/free-software/sobolgsa-software/>
- S. Kucherenko, D. Giamalakis, N. Shah, S. García-Muñoz, 2020, Computationally efficient identification of probabilistic design spaces through application of metamodeling and adaptive sampling. *Computers & Chemical Engineering*, 132(4), 106608.
- N. J. Samsatli, L.G. Papageorgiou, N. Shah, 1999, Batch process design and operation using operational envelopes. *Computers and Chemical Engineering*. 23, 887-890.
- M.M. Zuniga, S. Kucherenko, N. Shah, 2013, Metamodelling with independent and dependent inputs. *Computer Physics Communications*, 184(6), 1570–1580.
- U.S. Department of Health and Human Services, Food and Drug Administration, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, 2006, Q8 Guidance for industry. <http://www.ich.org/home.html>.

A Hybrid Model Coupling Advanced Oxidation Processes (AOP) and Conventional Bio-processes for the Removal of Recalcitrant Contaminants in Wastewaters

Kourosh Nasr Esfahani, Montserrat Pérez-Moya, Moisès Graells*

*Chemical Engineering Department, Universitat Politècnica de Catalunya, Av. Eduard Maristany, 16, 08019 Barcelona, Spain.
moises.graells@upc.edu*

Abstract

The combined use of Advanced Oxidation Processes (AOP) and conventional bio-processes has been suggested for the efficient treatment of wastewaters with a significant presence of recalcitrant contaminants. However, no models have been proposed to describe such combined processes, which prevents further design and operational optimization. Thus, by combining models previously reported this work contributes a first model integrating AOPs and bio-processes. Mass balances were reformulated by consistently linking the variables used in each model. The model was implemented using Simulink® and it was tested and analyzed using several process configurations (photo-Fenton processes followed by a biological treatment as well as a biological treatment followed by a photo-Fenton process). The outlet concentrations of total organic carbon (TOC) and substrate (S) were monitored for various wastewater compositions, which allowed analyzing the performance of the integrated system and suggesting suitable treatment arrangements. Different treatment options were discussed and the capability of the integrated bio/AOP and AOP/bio models was shown to allow a systematic approach to design, operation, and control of integrated wastewater treatment plants in the future. Further research will be oriented to improve the understanding of the potential key parameters as well as their inclusion in the model.

Keywords: Hybrid Model, AOP, Biological process, Wastewater treatment.

1. Introduction

Wastewaters contain an increasing number of persistent contaminants. Pollutants with high chemical stability and/or low biodegradability (e.g. emerging contaminants: pharmaceuticals, cosmetics, and personal care products) cannot be efficiently addressed by common wastewater treatment processes (WWTP). Conversely, Advanced Oxidation Processes (AOPs), in particular, the photo-Fenton process, can oxidize a broad range of non-biodegradable from wastewaters by means of a catalytic reaction of ferrous iron and hydrogen peroxide (H_2O_2) in an acidic or circumneutral pH under UV-VIS radiation yields highly oxidant hydroxyl radicals. However, such chemical oxidation for complete mineralization is much more expensive, which limits large-scale applications.

Hence, combining AOPs with biological treatments has been reported as an opportunity to reduce design and operating costs (Huang et al., 2017; Oller et al., 2011). Several configurations of hybrid AOPs as pretreatment or post-treatment with bioremediation for wastewater treatment have been experimentally studied (Nidheesh et al., 2021). However, the design and operation of such a hybrid process require suitable models. A great deal of mathematical models for WWTP have been developed and reported, while not many works have addressed the modeling of AOPs. Models combining AOPs and bio-processes have hardly been discussed. The modeling needs, and perspectives of AOPs and WWTP are different, as well as it is the nomenclature and the lumped parameters selected to characterize the system. This may explain the divergence of the research efforts of both areas, as well as the problems that converging them to a unique hybrid model entail. Therefore, a first and key step is consistently linking the variables used in both modeling approaches (biological and chemical) in a pilot coupled model. Consequently, this study consisted on selecting kinetic models for WWTP and AOP, standardizing nomenclature, mapping lumped parameters, and finally extending the models to incorporate those essential aspects considered by only one model. This has been applied to the combination of photo-Fenton processes followed by a biological treatment as well as a biological treatment followed by a photo-Fenton process. Particularly, new equations for non-biodegradable contents in the WWTP needed to be introduced, as well as equations discriminating biodegradable and non-biodegradable matter in the AOP model. Hence, simulation examples are discussed in regard of the expected results, achievements, and limitations of the model.

2. Methodology

The main problem to be addressed is the identification of the significant parameters and reactions in each subsystem and appropriately linking kinetic expressions and material balances in a consistent model of the resulting hybrid system. The integration strategy consists on preparing two configurations: AOPs as a pre-treatment followed by biological treatment (plan AB), also in the opposite direction, considering the post-treatment of AOPs (plan BA). Three case studies will be used to test the model: wastewater with only non-biodegradable contaminants (W1), partially biodegradable wastewater (W2), and completely biodegradable wastewater (W3). The schematic diagram of the proposed approach is illustrated in Figure 1.

Figure 1. Schematic diagram of the proposed integrated modeling

The mathematical models already proposed for the photo-Fenton process and the activated sludge biodegradation (ASM1) were first selected and adapted, and later, the models were coupled by linking of the associated variables. The coupled ordinary differential equations (ODEs) were solved simultaneously concerning the conceptual consistency of the related parameters using Simulink®.

2.1. Biological treatment kinetic model

There is abundant literature developing reliable models for WWTPs with formulations fully elaborated and complicated. However, reduced versions are still required for quickly analyzing design trade-offs and layouts, which is the purpose at this stage.

Furthermore, in the preliminary steps of developing a model, using simplified models may speed up the model evolution with consistent results and later, the model can be enriched with more complex concepts. Thus, the biological wastewater treatment model selected was the Activated Sludge Model No.1, ASM1, consisting of four non-linear differential equations (Vlad et al., 2011). The mass balance equations were adopted for modeling of the active sludge biomass (X) at the level of the aeration tank, the mass balance of the substrate (S), the mass balance of the oxygen in the water mass (DO), and the balance of the recycled biomass (X_r) at the level of the settling tank. The biomass growth rate (μ), was modeled by the Monod law.

2.2. Modeling of the AOP (photo-Fenton process)

Cabrera Reina et al. (2012) proposed and validated a photo-Fenton process model for batch operation and then the model was re-written by Audino et al. (2019) to describe the continuous operation. The model did not consider biodegradability, although it should be taken into account to apply integrated treatment processes including Fenton-based processes and biological treatment (Huang et al., 2017).

Thus, the model by Audino et al. (2019) was adapted by separating the biodegradable (subscripted by b) and non-biodegradable (subscripted by nb) parts of the organic matter to enable biodegradability footprint tracking in the combined model. To do that, the equations and subsequently related mass balances were replicated for non-biodegradable contents. The proposed model includes the two ferric species, hydrogen peroxide, the radicals formed from peroxide (R), the dissolved oxygen, three states accounting for the biodegradable organic matter (two kinds of partially oxidized organics (MX_{1b} , MX_{2b}) plus the parent compound (M_b) present at the beginning of the reaction) which are responsible for the lumped parameter measured as TOC_b .

Three complementary states for the non-biodegradable organic matter were introduced to the model including two kinds of partially non-bio oxidized organics (MX_{1nb} , MX_{2nb}) plus the parent non-bio compound (M_{nb}) present at the beginning of the reaction, aggregated as TOC_{nb} (i.e. $TOC = TOC_b + TOC_{nb}$). The values of the kinetic constants of the new reactions were also replicated at this stage.

3. Development of the integrated models

The two simulation schemes are next presented. The AOP pre-treatment followed by WWTP is designated as AB, while using AOP as a post-treatment is designated as BA.

3.1. Combined modeling: configuration AB

The biodegradability of wastewaters can be improved in a combined AOPs system, which is favorable to WWTPs, while WWTPs in the same combined system may stabilize the waste and reduce the use of chemical reagents (Huang et al., 2017). A partial oxidation treatment may produce intermediates even more recalcitrant, but chemically-oxidized intermediates are often less recalcitrant than the parent compound. Many studies have proved that prior chemical oxidation may cause the biodegradability changes of a waste stream (Mantzavinos and Psillakis, 2004). Therefore, the main role of the chemical pre-treatment may be explained as partial oxidation of the biologically persistent part to produce biodegradable reaction intermediates which remove the pointless expense of chemicals and energy (Oller et al., 2011).

A conceptual challenge for developing a mathematical model for the AB is to find which variables in the AOP model require to be properly linked to which variables in the WWTP model. Moreover, the effect of the chemically-oxidized intermediates from the AOPs on the performance of the WWTP needs to be modeled. In this first step, these intermediates are considered to have no adverse impact (toxicity or inhibitory effect). In this regard, for the integration of the processes using the set of reported reactions, the outlet biodegradable TOC and the oxygen concentration in the AOP were introduced equally as the inlet substrate content and the dissolved oxygen in the biotreatment, respectively. The following linking conditions were associated to couple the photo-Fenton and the WWTP models as continuous stirred tanks reactor (CSTR):

- The outlet flowrate from the AOP reactor was directly entered to the WWTP;
- The biodegradable TOC_b from the AOP was fed as the inlet substrate to the WWTP;
- The oxygen flow from the AOP was connected to dissolved oxygen in the bioreactor;
- The non-biodegradable compounds were assumed to be inert through the WWTP

Figure 2 illustrates the configuration AB.

Figure 2. AB scheme presenting the photo-Fenton process (A) followed by the bio-treatment(B)

3.2. Combined modeling: configuration BA

The alternative post-treatment of AOPs are often suggested for WW containing the recalcitrant organic pollutants with low toxicity concentrations. The transformation products generated during biodegradation would be recalcitrant and even hazardous (Zimbron and Reardon, 2011; Sánchez Pérez et al., 2014). In this study, the combined model represents the biological treatment followed by the photo-Fenton reactor as CSTR using first-order kinetics and mass balances. The same conceptual integration process was applied correspondingly through linking the outlet substrate and the dissolved oxygen in WWTP to inlet biodegradable TOC and the oxygen content in the AOP model as presented in Figure 3.

Figure 3. BA schematic presenting the bio-treatment(B) followed by the photo-Fenton process(A)

4. Simulation results

The preliminary simulations of the outlet organic matter concentration (TOC and/or substrate) in the single-reactor design (A or B) as well as the combined system for both AB and BA configurations were performed according to the data of the kinetic constants provided by Cabrera Reina et al. (2012) and Vlad et al. (2011). The simulation of the hybrid processes indicated promising results compared to each individual process. The integrated model was able to predict and track the biodegradability evolution, which is crucial for the WWTPs. Figure 4 shows the results obtained for the three case studies (non-biodegradable wastewater, W1; partially biodegradable wastewater, W2; and biodegradable wastewater, W3) and the four arrangements (A, B, AB, and BA). In general, the simulations produced the expected profiles as for the higher biodegradability, the lower outlet organic matter concentration was predicted. The outcomes of the integrated models produced similar trends and all of them confirm decreasing outlet organic matter concentrations as the inlet biodegradability increases ($W1 < W2 < W3$). This is not the case for a single AOP (Figure 4a), for which all profiles overlap according to the non-selective oxidizing capacity of the hydroxyl radicals. Figure 4a also shows a conversion limit given by an insufficient supply of hydrogen peroxide. Conversely, the single WWTP (Figure 4b) presents the most disparate outcomes: the total organic matter removal for a complete biodegradable input (W3) and no change at all for a totally non-biodegradable feed (W1).

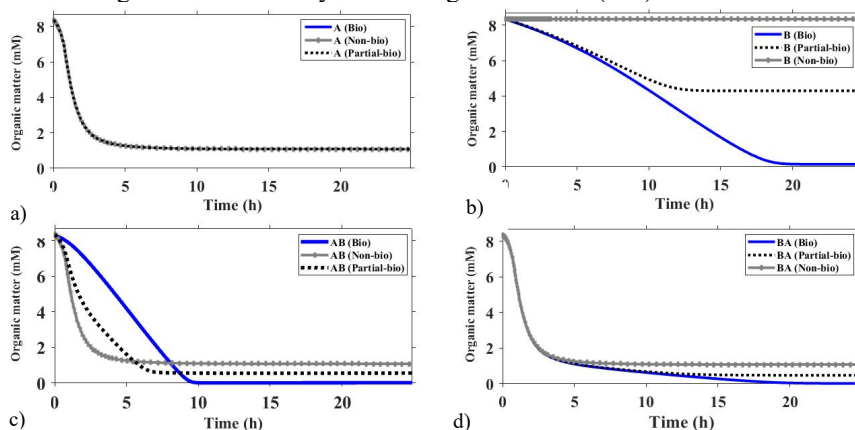


Figure 4. Simulated profiles of the single and integrated models for a) A, b) B, c) AB, and d) BA.

On the other hand, the combined processes AB (Figure 4c) and BA (Figure 4d) resulted in different behavior in the transition period but, they showed the expected higher steady-state performance compared to the single processes A and B. Again, a total conversion for is not attained, this time only for the partially and non-biodegradable cases (W2 and W1) due to the lack of reagents (hydrogen peroxide), which hints at the need to adjust this variable. The different kinetics produce different transient periods in the combined models but show the same long term behavior given by the biodegradability ratio.

These preliminary simulations of the combined model meet the expectations, but reveal the limitations and next steps, mainly the incorporation of new features allowing to discriminate the behavior of processes as a function of the biodegradability and the chemical nature of the contaminants, so that toxicity and inhibitory effects of recalcitrant species in the bio-treatment can be considered and the true system trade-offs could be addressed for decision-making support.

5. Conclusions

This contribution proposes a combined model coupling the photo-Fenton process and WWTP for the removal of recalcitrant contaminants in wastewaters. The lumped parameter TOC in the photo-Fenton process is considered equivalent to the substrate content in WWTP for the model integration of the processes. Additionally, dissolved oxygen is directly linked to the oxygen content in the photo-Fenton process. The simulations were consistent through all case studies designed including biodegradable, partial biodegradable, and non-biodegradable influent. The simulation results from different configurations led to the variation curves of the outlet organic matter in the effluent that may be useful to design and control the integrated wastewater treatment systems. The analysis of the model arose some limitations to be addressed, such as the need, to include the inhibitory effect of recalcitrant species in the bio-treatment. Further research needs to be carried out through the proposed model to validate and improve the understanding of the potential key parameters as well as their inclusion in the model.

Acknowledgments

Financial support received from the Spanish "Ministerio de Economía, Industria y Competitividad" and the European Regional Development Fund, both funding the research Project AIMS (DPI2017-87435-R) is fully acknowledged. Kourosch Nasr Esfahani thankfully acknowledges financial aid received by the Generalitat de Catalunya through the program FI SDUR (Grant ref. BDNS 481561).

References

- F. Audino, G. Companyà, M. Pérez-Moya, A. Espuña, M. Graells, 2019, Systematic Optimization Approach for the Efficient Management of the Photo-Fenton Treatment Process. *Science of The Total Environment*, 646, 902–913.
- A. Cabrera Reina, L. Santos-Juanes Jordá, J. L. García Sánchez, J. L. Casas López, J. A. Sánchez Pérez, 2012, Modeling Photo-Fenton Process for Organic Matter Mineralization, Hydrogen Peroxide Consumption and Dissolved Oxygen Evolution, *Applied Catalysis B: Environmental*, 119-120, 132-138.
- D. Huang, C. Hu, G. Zeng, M. Cheng, P. Xu, X. Gong, W. Xue, 2017, Combination of Fenton Processes and Biotreatment for Wastewater Treatment and Soil Remediation, *Science of The Total Environment*, 574, 1599-1610.

- D. Mantzavinos, E. Psillakis, 2004, Enhancement of Biodegradability of Industrial Wastewaters by Chemical Oxidation Pre-Treatment, *Journal of Chemical Technology & Biotechnology*, 79 (May), 431–54.
- I. Oller, S. Malato, J. A. Sánchez-Pérez, 2011, Combination of Advanced Oxidation Processes and Biological Treatments for Wastewater Decontamination—A Review. *Science of The Total Environment*, 409(20), 4141–4166.
- P.V. Nidheesh, C. Couras, A. V. Karim, H. Nadais, 2021, A Review of Integrated Advanced Oxidation Processes and Biological Processes for Organic Pollutant Removal, *Chemical Engineering Communications*, January, 1–43.
- J. A. Sánchez Pérez, I. Carra, C. Sirtori, A. Agüera, B. Esteban, 2014, Fate of Thiabendazole through the Treatment of a Simulated Agro-Food Industrial Effluent by Combined MBR/Fenton Processes at $\mu\text{g/L}$ Scale. *Water Research*, 51, 55–63.
- G. Vlad, D. N. Robescu, C. R. Mocanu, 2011, Mathematical Modeling of a Biological Wastewater Treatment Process. Case Study: The Wastewater Treatment Station of Românofir S.A. Trading co.-Tâlmăciu. 73, 299–310.
- J. Zimbron, K. Reardon, 2011, Continuous Combined Fenton's Oxidation and Biodegradation for the Treatment of Pentachlorophenol-Contaminated Water. *Water Research*, 45, 5705–5714.

Analysis of the dissolution mechanism in the phosphoric acid manufacturing process: modelling and simulation

Sanae Elmisaoui^{a,b,c}, Abderrazak M. Latifi^{a,b,*}, Lhachmi Khamar^{a,d}, Mohamed Salouhi^c

^a*Mohammed VI Polytechnic University, Benguerir, Morocco.*

^b*Laboratoire Réactions et Génie des Procédés, CNRS-ENSIC, Université de Lorraine, Nancy, France.*

^c*L3GIE, EMI, Mohammed V University, Rabat, Morocco.*

^d*LIPIM, National School of Applied Sciences in Khouribga, Sultan Moulay Slimane University, Beni Mellal, Morocco.*

Abderrazak.latifi@univ-lorraine.fr

Abstract

This paper deals with the modelling and simulation of the dissolution of phosphate ore particles in a dilute phosphoric acid solution in a batch stirred tank reactor. A shrinking core model with elimination of the products is used to describe the reaction and mass transfer phenomena involved in the three phases considered, i.e., liquid bulk, liquid film surrounding the particles and solid phase. The model is based on mass balance equations in the three phases and consist of algebraic equations in the liquid bulk, an ODE in the solid phase, and PDEs in the liquid film. An estimability analysis method based on global sensitivities is used to determine the most estimable unknown parameters involved in the model equations from the available experimental data. The values of the least estimable parameters are fixed from the literature, while gProms environment is used to implement and solve the equations and to identify the most estimable parameters. The results obtained exhibit a good agreement between the model predictions and the measurements with coherent values of the identified parameters. Furthermore, the model is tested on the measurements carried out by van der Sluis et al. (1987) and showed that the dissolution process is mainly controlled by the diffusion step.

Keywords: Phosphate ore, Phosphoric acid, dissolution mechanism, shrinking core model, parameter identification.

1. Introduction

Phosphoric acid is the key element in the phosphate industry, particularly in the fertilizer industry, and even in the pharmaceutical and food industries (Becker, 1983). It is mainly produced by the wet process which consists of reacting the phosphate ore with concentrated sulfuric acid solutions in a digestion tank. Two main phenomena therefore take place in the tank, i.e., the dissolution of the ore and the crystallization of the gypsum. In this work, only phosphoric acid is used in the attack of the phosphate ore in order to decouple the two phenomena and deal only with the dissolution. The latter is

one of the most complex phenomena involved in the wet process of industrial manufacture of phosphoric acid. Its understanding is one of the key issues for the optimal design and operation of the phosphoric acid processes.

The objective of the present paper is to investigate the mechanism of the attack of the phosphate ore by a phosphoric acid solution in the digestion tank. More specifically, the shrinking core model (SCM) with elimination of the products will be used to describe the chemical reactions and mass transfer phenomena that take place in the tank. The resulting model will inevitably involve several unknown parameters that are often deduced from experimental measurements by means of a parameter optimization method. However, it is recognized that the available experimental measurements do not always contain the necessary information to accurately identify all the unknown parameters. A global estimability analysis will therefore be carried out to determine which parameters are estimable from the available measurements. The most estimable parameters will then be identified, and the values of the least estimable parameters will be fixed from the literature or from previous studies. gProms environment will be used to implement and solve the model equations and for parameter identification as well.

2. Model formulation

Tri-calcium phosphate (TCP) is one of the most interesting elements to be recovered from the phosphate ore. The objective of the digestion is to dissolve the TCP in a dilute phosphoric acid solution before its attack with a concentrated sulfuric acid solution to produce concentrated phosphoric acid and gypsum. Two main reactions take place in the digestion tank during the dissolution process (Elmore and Farr, 1940), but in this work we will focus on the most important one which transforms TCP into mono-calcium phosphate (MCP) as:



The conversion rate of this reaction depends on different operating parameters, including temperature, phosphoric acid concentration, porosity and size distribution of phosphate ore particles, residence time, solid rate, and hydrodynamics.

2.1. Dissolution mechanism

The attack of the phosphate ore particles by a phosphoric acid solution is a non-catalytic liquid/solid reaction with a complex kinetic mechanism. We assume that three phases are involved in the mechanism, i.e., liquid bulk, liquid film surrounding the particles, and solid phase (Figure 1).

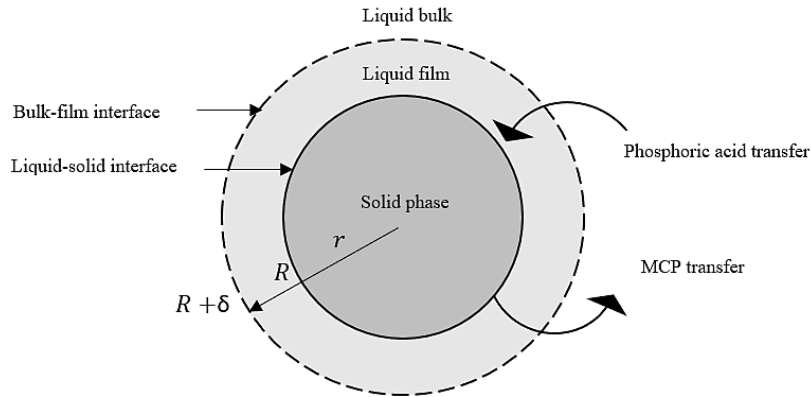


Figure 1: Schematic illustration of the SCM in the dissolution mechanism

The dissolution mechanism considered assumes that the phosphoric acid diffuses through the film towards the solid, adsorbs to the phosphate ore particles, and finally reacts with the TCP reagent at the solid surface. The MCP product takes the opposite path where it first desorbs from the solid surface, then diffuses through the liquid film towards the liquid bulk. It is noteworthy that the chemical reaction and all the mass transport phenomena take place simultaneously in the digestion tank.

3. Phosphate ore dissolution modelling

The dissolution model developed is based on the following assumptions: (i) the particles are well dispersed in the liquid phase, are spherical with the same diameter, and shrink uniformly during the dissolution process, (ii) the digestion tank is perfectly mixed, (iii) the reaction is irreversible and only takes place at the surface of the particles, (iv) the adsorption and desorption steps are assumed to be very fast and are not considered.

Moreover, the shrinking core model (SCM) with elimination of the products used is adapted from the model developed by Salmi et al. (2017) with two main major differences. The first one assumes that the solid reagent is not soluble in the liquid, which is the case of the Moroccan phosphate ore, and the second one considers that the surface of the particles is not saturated with the solid reagent. The model equations are based on the transient mass balances in the solid phase, in the liquid film, and in the liquid bulk. They are presented below.

- In the solid phase, the mass balance can be written as:

$$\frac{x_{TCP}}{M_{TCP}} \frac{dm_s}{dt} = -4\pi R^2 k_{MCP} (C_{MCP}|_{r=R} - C'_{MCP}) \quad (2)$$

where m_s is the mass of the solid (i.e., phosphate ore particles) and R its radius.

x_{TCP} is the mass composition of TCP in the particles, M_{TCP} is the molecular weight of TCP, and $C_{MCP}|_{r=R}$ and C'_{MCP} are the concentrations of MCP at the solid surface and

in the liquid bulk, respectively. k_{MCP} is the liquid/solid mass transfer coefficient of MCP, it is expressed by D_{MCP}/δ using the film theory, where δ is the thickness of the liquid film surrounding the particles and D_{MCP} the diffusion coefficient of MCP. Eq. (2) can be developed in terms of TCP conversion rate X and particle radius R as:

$$\frac{dX}{dt} = \frac{3M_{TCP}D_{MCP}}{x_{TCP}\rho_s R_0\delta} (1-X)^{\frac{2}{3}} (C_{MCP}|_{r=R} - C'_{MCP}) \quad (3)$$

where ρ_s and R_0 are the density and the initial radius of the particles respectively, and

$$X = 1 - \left(\frac{R}{R_0}\right)^3$$

- In the liquid film, the mass balance is expressed by the second law of Fick as:

$$\frac{\partial C_i}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(D_i r^2 \frac{\partial C_i}{\partial r} \right); \quad i = H_3PO_4, MCP \quad (4)$$

with the following boundary conditions:

$$-D_i \frac{\partial C_i}{\partial r} \Big|_{r=R} = \mathfrak{R}_i \quad (5)$$

• at the solid interface:

where the kinetic rate equation is assumed to be of first order with respect to phosphoric acid at the particle surface, and is defined as: $\mathfrak{R}_i = v_i k_r C_{H_3PO_4}|_{r=R}$. k_r is the rate constant, D_i is the diffusion coefficient of component i and v_i is the stoichiometric coefficient of component i in the reaction (1).

$$\bullet \text{ at the liquid film/liquid bulk interface: } C_i|_{r=R+\delta} = C'_i; \quad i = H_3PO_4, MCP \quad (6)$$

- The mass balance in the liquid bulk provides the MCP and acid concentrations as:

$$C'_{MCP} = \frac{3n_{TCP}^0 X}{V_L} \quad \text{and} \quad C'_{H_3PO_4} = C_{H_3PO_4}^0 - \frac{4n_{TCP}^0 X}{V_L} \quad (7)$$

where n_{TCP}^0 is the initial number of moles of TCP in the solid and V_L is the liquid volume in the tank. The initial conditions used are : $n_{TCP}^0 = 0.0322$ mole, $C_{H_3PO_4}^0 = 4760$ mol/m³, $C'_{MCP,0} = 0$, $C_{MCP,0} = C_{H_3PO_4,0} = 0$, and $X = 0$. The liquid film thickness is calculated using the mass transfer correlation of Rakoczy and Masiuk (2011) as:

$$\delta = R \left[1 + \alpha \left(\frac{R}{R_0} \right)^{8/9} D_{MCP}^{-1/3} \right]^{-1} \quad \text{where} \quad (8)$$

$$\alpha = 0.46 \left(\frac{\varepsilon^{2/3} R_0^{8/3}}{2\nu_L} \right)^{1/3}$$

α is a parameter which considers the hydrodynamics through the energy dissipation per unit mass (ε) of reaction mixture in the tank. ν_L is the kinematic viscosity of the mixture.

The dynamic model is described by a set of partial differential equations (PDEs) in the liquid film and an ordinary differential equation (ODE) in the solid phase and algebraic equations in the liquid bulk. The model equations are implemented and solved within gProms software which provides an environment for modeling, simulation, and parameter identification of dynamic systems (Process Systems Enterprise, 1997-2020).

4. Simulation and parameter identification

As can be noticed, the model involves several unknown parameters, i.e., k_r , $D_{H_3PO_4}$ and D_{MCP} , that should be identified from the available experimental measurements by means of a parameter optimization method. To test the model, we used the measurements of the conversion rate of TCP over time and for different temperatures in a stirred batch reactor where the digestion of the phosphate ore particles by a phosphoric acid solution takes place (van der Sluis et al, 1987). In their study, the authors assumed that the dissolution process is controlled by the diffusion of MCP and determined the corresponding values of the mass transfer coefficient of TCP. In this work, the developed model is adapted to fulfill the same conditions assuming that the diffusion of acid is not the rate-limiting step. The influence of the diffusion coefficient $D_{H_3PO_4}$ on the model is therefore neglected. However, the reaction at the solid surface as well as the diffusion of MCP in the film are considered. The competition between these steps will be analyzed by means of Damköhler number $Da_{MCP} = k_r R / D_{MCP}$. However, the question is to know whether these experiments contain the necessary information to identify all the unknown parameters of the model. To answer this question, we used our recently developed estimability analysis method based on global sensitivities (Bouchkira et al, 2021) to rank the unknown parameters from the most to the least estimable.

Table 1 presents the data used in the simulations. The measured values of the kinematic viscosity at different temperatures can be found in (van der Sluis et al., 1987).

Table 1: Model parameter values

Parameter	ρ_s	ε	x_{TCP}	M_{MCP}	R_0	m_s^0	V_L
Value	7.2×10^3	5.0×10^3	51	3.1×10^{-1}	5.45×10^{-4}	2×10^{-2}	4×10^{-4}
Unit	kg.m^{-3}	W.kg^{-1}	% wt	kg.mol^{-1}	m	kg	m^3

5. Results and discussion

The estimability analysis method showed that k_r is the only estimable parameter. It is then identified using a gradient-based optimization method, while D_{MCP} is deduced from the measurements of van der Sluis et al. (1987) by means of the mass transfer correlation of Rakoczy and Masiuk (2011). Table 2 presents the computed results which consist of the values of the hydrodynamic parameter α , the diffusion coefficient D_{MCP} , and the constant rate k_r along with its corresponding 95% confidence intervals, for three different values of temperature. The tight confidence intervals show that k_r is accurately determined. Moreover, the use of an Arrhenius-type of equation allows us to determine the pre-exponential factor (1.18×10^{-4} m/s) and the activation energy (8641 J/mol). The low value of the activation energy confirms that the diffusion of the MCP product is indeed the limiting step in the dissolution of the particles.

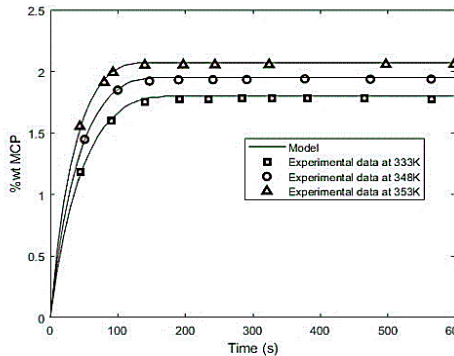


Figure 2: Comparison of simulated and measured values of MCP concentrations

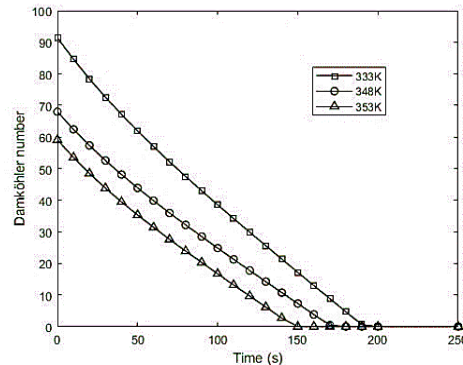


Figure 3: Profiles of Damköhler number for three different values of temperature

Figure 2 exhibits a very good agreement between the model predictions of the MCP concentrations and the measurements carried out at three values of temperature, thus highlighting the quality of the model.

On the other hand, as the dissolution proceeds, the radius of the particles as well as the thickness of the film decrease over time. The resulting Damköhler number also decreases over time and with increasing temperature (Figure 3).

Table 2: Computed values of α , D_{MCP} and k_r at different values of temperature

T(K)	α	D_{MCP} (m ² /s)	k_r (m/s)
333	0.27	3.1×10^{-11}	$5.24 \times 10^{-6} \pm 3.40 \times 10^{-7}$
348	0.29	4.7×10^{-11}	$5.86 \times 10^{-6} \pm 4.29 \times 10^{-7}$
363	0.31	6.3×10^{-11}	$6.78 \times 10^{-6} \pm 3.89 \times 10^{-7}$

In the beginning of the digestion, the dissolution is controlled by the diffusion of MCP in the liquid film for the three values of temperature. This control is reduced over time in favor of the reaction of TCP at the surface of the particles which takes over at the end of the dissolution. The control is actually shared by diffusion and reaction throughout the process. The diffusion prevails at the beginning and the reaction at the very end of dissolution. This phenomenon is very likely to occur in such transient processes.

6. Conclusions

It is interesting to notice that the model developed in this work leads to the same conclusions as in van der Sluis et al. (1987) on the phenomenon that controls the dissolution using the same set of experimental measurements and assuming that the diffusion of the acid in the liquid film is not limiting. Furthermore, the model developed here is more accurate since it is based on more realistic assumptions and predicts a diffusion control which decreases with time. Indeed, we did not assume that the concentration of MCP at the surface of the solid is given by its saturation concentration, it is determined by solving the second Fourier law in transient regime in the liquid film. We also considered the thickness of the liquid film which decreases over time as the radius of the particles. However, the model can be further improved by investigating the phenomena taking place at the surface of the particles and by carrying out additional experiments to better calibrate and validate the model prior to its use in subsequent optimization of the reactor performances.

References

- P. Becker, 1983, Phosphates and phosphoric acid: Raw materials, technology and economics of the wet process, Fertilizer Science and Technology Series. New York.
- I. Bouchkira, A. M. Latifi, L. Khamar and S. Benjelloun, 2021, Global sensitivity based estimability analysis for the parameter identification of Pitzer's thermodynamic model, Reliability Engineering and System Safety.
- K. L. Elmore and T. D. Farr, 1940, Equilibrium in the system calcium oxide-phosphorus pentoxide-water, Industrial Engineering Chemistry, 32, 580-586.
- R. Rakoczy and S. Masiuk, 2011, Forced Convection Mass-Transfer Enhancement in Mixing Systems, Advanced Topics in Mass Transfer, IntechOpen, DOI: 10.5772/14740.
- T. Salmi, V. Russo, C. Carletti, T. Kilpiö, R. Tesser, D. Murzin, T. Westerlund and H. Grénman, 2017, Application of film theory on the reactions of solid particles with liquids: Shrinking particles with changing liquid films, Chemical Engineering Science, 160, 161-170.
- S. Van der Sluis, Y. Meszaros, W. G. J. Marchee, H. A. Wesselingh, and G. M. Van Rosmalen, 1987, The digestion of phosphate ore in phosphoric acid, Industrial & Engineering Chemistry Research, 26, 12, 2501-2505.

Focusing experiments in the early phase process design by process optimization and global sensitivity analysis

Stefanie Kaiser^{a,*}, Tabea Menzel^a, Sebastian Engell^a

^a *Department of Chemical Engineering, TU Dortmund University, Emil-Figge-Str. 70, 44227 Dortmund, Germany*

Abstract

Accurate process models which are the key to a reliable model-based process design usually need to be identified on the basis of expensive laboratory experiments. In this work, we present an integrated methodology which enables to focus these experiments on the most relevant model parameters by combining a global sensitivity analysis and optimal design of experiments. We apply the methodology to the homogeneous catalyzed hydroformylation of 1-dodecene as an example process. The comparison to an ordinary optimal experimental design and a factorial design show that by this approach the experimental effort could be reduced. Furthermore, we compare the use of a local and a global sensitivity analysis; the global sensitivity analysis can indeed enhance the process design.

Keywords: early phase process design, global sensitivity analysis, optimal design of experiments.

1. Introduction

In the early design phase of new chemical processes, the most cost influencing decisions are taken, so this is the key phase of process development. As nowadays product cycles in the chemical industry become shorter, the pressure to reduce the development time is increasing. Sequential steps of laboratory experiments, identification of model parameters and physical properties and process simulation and optimization lead to long development cycles, thus a new methodology that integrates these steps and helps to focus the work is required. Optimal design of experiments (ODoE) is used to reduce the effort of experimental work. An overview of ODoE can be found in Franceschini & Macchietto (2008). Chen & Grossmann (2017) describe the recent developments in optimization-based process synthesis, including the handling of uncertainty in the process parameters. Asprión et al. (2019) developed an interface that enables to integrate optimal experimental design in a flowsheet simulator. The framework can be used for focused model improvement and parameter estimation but does not include a method for process design. Recker et al. (2013) integrated process optimization and optimal design of experiments. In order to focus the experiments on the relevant parameters, they weight the Fischer information matrix in the optimal experimental design. However, in their approach uncertainties in the process optimization are neglected which can lead to suboptimal processes. Commonly, local sensitivity analysis is used for the identification of the cost driving parameters, but this has the drawback

that it is only valid at one point and not for the complete operating range. This issue can be overcome by replacing the local sensitivity analysis with a global sensitivity analysis. In this work, we present a methodology that integrates superstructure optimization under uncertainties, a global sensitivity analysis and optimal design of experiments. For the best design identified by the superstructure optimization the sensitivities of the cost function with respect to the uncertain parameters are computed and used as weights in the optimal design of experiments. The proposed method is applied to a case study of the homogeneously catalyzed hydroformylation of 1-dodecene. The results are compared with the non-weighted optimal experimental design and a factorial design in order to see the effect of applying the weighting on the number of experiments that is required to get a sufficiently good process model.

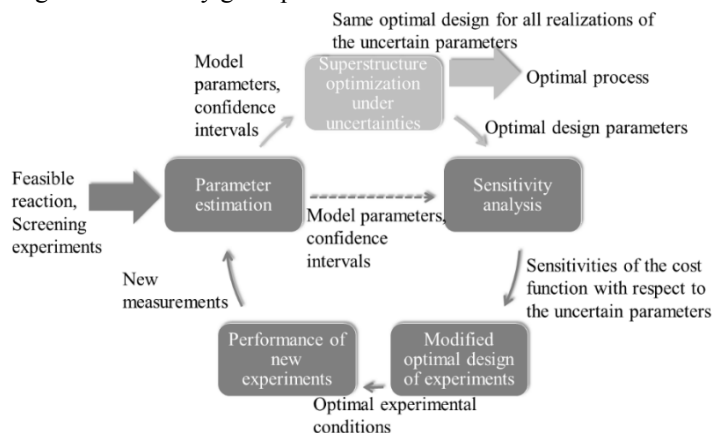


Figure 1: Schematic representation of the proposed methodology for integrated process.

2. Methodology

It is proposed here to use model-based tools already in the early design phase although only little knowledge is available during this phase. It is assumed that first experiments have been performed and the key elements of the process, in particular the chemical reaction system have been identified. The general procedure of the proposed methodology has been described in Kaiser & Engell (2020) and it is depicted in Figure 1. Based on the first screening experiments, kinetic and thermodynamic models with still significant parametric uncertainties are assumed to be available and are used in a subsequent step for superstructure optimization under uncertainties as introduced by Steimel & Engell (2015). For the best design, a sensitivity analysis - either local or global - is performed followed by a subsequent optimal design of experiments (ODoE) which is weighted with the computed sensitivities. The focus of this paper is on the weighted optimal design of experiments using the previously computed sensitivities.

2.1. Sensitivity analysis

Sensitivity analysis is performed in order to obtain information about the influence of each uncertain parameter on the cost function. The local sensitivities are computationally less expensive but are valid only in a small region around the nominal values of the parameters. The global sensitivities are valid in average over the complete region of the uncertain parameters and they give information about nonlinear effects. However, the global sensitivity analysis is computationally more expensive. It is assumed that the operating conditions of a process can be adjusted depending on the realization of the uncertain parameters. Therefore, the cost function is optimized at each

sampling point in the computation of the local and the global sensitivities with respect to the operating conditions (u_{opt}) for given design parameters.

2.1.1. Local sensitivity analysis

The local sensitivities are computed as described by Kaiser & Engell (2020). The uncertain parameters are sampled using 1 % deviation from their nominal values and a linear regression is used to correlate the n uncertain parameters x_i to the regressed objective value \hat{Z}_j :

$$\hat{Z}_j = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (1)$$

The least squares method is used to determine the intercept β_0 and the regression coefficients β_i . The regression coefficients are standardized with zero mean and a standard deviation of one. For N samples the standardized regression coefficients are computed as:

$$SRC_j = \frac{\beta_j \hat{s}_i}{\hat{s}} \quad \text{with} \quad \hat{s} = \left[\sum_j \frac{(Z_j - \bar{Z})^2}{N-1} \right]^{1/2} \quad \text{and} \quad \hat{s}_i = \left[\sum_j \frac{(x_j - \bar{x})^2}{N-1} \right]^{1/2}. \quad (2-4)$$

2.1.2. Global Sensitivity Analysis

The global sensitivities are computed using the Sobol' indices (Sobol, 2001). The effect of a single parameter can be computed by

$$S_i = \frac{var(y|x_i)}{var(y)} \quad (5)$$

where $var(y|x_i)$ describes the conditional variance of the output y with respect to the parameter i and $var(y)$ describes the general variance of y . The total effect S_T also takes non-linearity and interaction effects into account. It can be computed as

$$S_{T,i} = 1 - \frac{var(y|x_{\sim i})}{var(y)} \quad (6)$$

where $var(y|x_{\sim i})$ is the conditional variance of the output y with respect to all the parameter except parameter i . Using a latin-hypercube sampling, N different scenarios of the realization of the uncertain parameters within the 95 % confidence intervals are generated.

2.2. Modified optimal design of experiments

In order to design new experiments to gain information about the parameters that were identified as cost driving, a modified ODoE is used. As in the ordinary experimental design, a metric of the Fisher information matrix (FIM) is minimized. The FIM is defined as

$$FIM = \sum_{\tau=\tau_1}^{\tau_N} Q^T(u_\tau) \text{diag}^{-1}(\sigma_1^2, \dots, \sigma_n^2) Q(u_\tau) \quad (7)$$

with $Q(u_\tau)$ containing the derivatives of the model outputs with respect to the parameters of the experiment with the input u_τ . To drive the designed experiments towards gaining information about the relevant parameters, an intuitive way is weighting the FIM with the previously computed sensitivities (Recker et al., 2013).

$$\tilde{FIM} = W^{1/2} \sum_{\tau=\tau_1}^{\tau_N} Q^T(u_\tau) \text{diag}^{-1}(\sigma_1^2, \dots, \sigma_{n_y}^2) Q(u_\tau) W^{1/2} \quad (8)$$

where the matrix W contains the sensitivities of the process costs with respect to the model parameters θ_i at the current estimated value of the parameters θ_{est} .

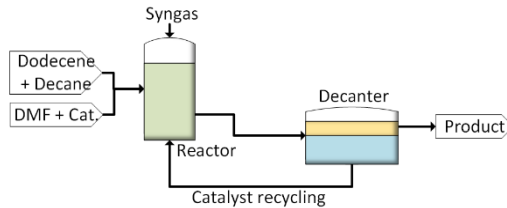


Figure 2: Simplified process flowsheet.

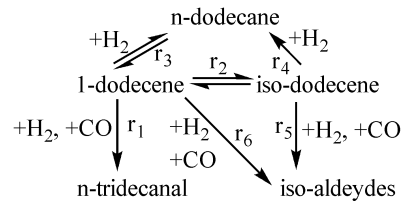


Figure 3: Reaction network of the hydroformylation.

$$W = \begin{bmatrix} \frac{\partial C_{process}}{\partial \theta_1} \Big|_{u_{opt, \theta_{est}}} & \dots & 0 & \ddots & \ddots & 0 & \dots & \frac{\partial C_{process}}{\partial \theta_p} \Big|_{u_{opt, \theta_{est}}} \end{bmatrix} \quad (9)$$

In this work we use an A-optimal DoE which aims at minimizing the trace of the FIM .

3. Case study

The proposed methodology is applied to the case study of the homogeneously catalyzed hydroformylation of 1-dodecene using a thermomorphic solvent system. The process is described in Hernandez et al. (2018). A thermomorphic solvent system consisting of the polar solvent dimethylformamide and a non-polar solvent decane that separates into two phases at low temperatures is used to recycle the expensive rhodium catalyst. The catalyst dissolves in the polar phase and can be recycled into the reactor, while the product tridecanal dissolves in the non-polar phase which can be purified in a further thermal separation step. A simplified process flowsheet is shown in Figure 2. A detailed description of the process model can be found in Hernandez et al. (2018). The kinetic model depicted in Figure 3 was developed by Hentschel et al. (2015). The pre-exponential factors and the activation energies of the hydroformylation of n-dodecene (r_1) and of the isomerization of n-dodecene (r_2) in the kinetic model are considered as uncertain parameters.

The other kinetic parameters have been observed as less influencing and are hence fixed to their pre-defined values. We use the standardized regression coefficients (eq. 2-4) as local sensitivities and the total effects (eq. 6) as global sensitivities. The planned experiments are isothermal batch experiments where the evolution of the concentration is measured over the batch time. The degrees of freedom in the ODoE are the reaction temperature, the initial concentration of n-dodecene, the initial concentration of iso-dodecene, the ratio of H2 to CO and the catalyst concentration. Furthermore, the sampling times of the concentration measurements are included as degrees of freedom. The number of measurements is fixed to 6. The measurements are generated using a simulation model with the true parameters, corrupted by white noise with a standard deviation of 5 %.

The yield of n-undecanal in the product stream with respect to the n-dodecene in the feed is used as the cost function.

3.1. Comparison of different design strategies

In this section we compare the results of different strategies for model refinement: (i) A standard ODeE without weighting the FIM. (ii) A modified ODoE weighting the FIM with local sensitivities. (iii) A modified ODoE weighing the FIM with global sensitivities.

In all cases, the same two experiments performed at two different temperatures are used as the basis for the first parameter estimation and then 25 new experiments are planned in iterative steps of parameter estimation and ODoE according to the applied method. The number of process optimizations that are used for the computation of the weighting matrix W is 10 in case of the local sensitivity analysis and 1200 in case of the global sensitivity analysis. For the standard ODoE, no process optimization is required.

The progress over the number of experiments is compared to the results of a static factorial design, in which experiments are performed at the lower and upper bounds of the five degrees of freedom, leading to 32 experiments. After the parameter estimation, the 95 % confidence intervals are computed. For different scenarios within these confidence intervals the minimal and the maximal yields are computed. As the predicted parameters and hence the predicted yields depend on the random process noise, the procedure is repeated ten times for each case and the mean values are considered. The evolution of the maximum and minimum computed process cost is shown in Figure 4.

The figure shows the change of the maximum and minimum predicted yield for all three cases as well as the true yield and the yield predicted using the factorial sampling. It can be seen that for all cases a better prediction of the yield in comparison to a factorial design can be reached after only 9 iterations for the weighting obtained from the global sensitivity analysis and after 15 iterations using the local sensitivity analysis. The absolute differences in the minimum expected yield between the two approaches is small after 6 iterations. Standard ODoE performs somewhat erratically in comparison.

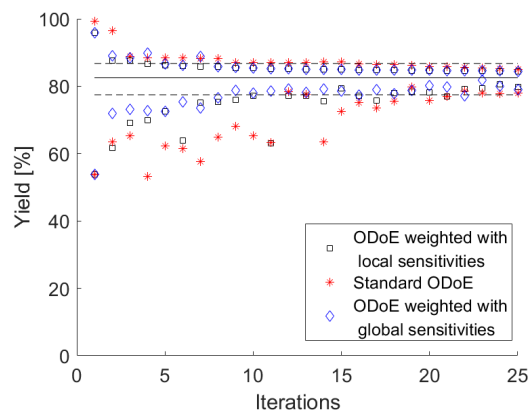


Figure 4: Evolution of the predicted intervals of the yield for the different methods. The true yield is represented by the solid line and the interval of the predicted yield using a factorial design is represented by the dashed line.

Since in the ordinary ODoE equal weights are put on determining the uncertain parameters, it can occur that experiments are planned that improve one parameter that has a small influence on the process yield. By using specific weights on the parameters that have the highest impact on the cost function, more efficient experiments can be planned. The advantage of using global sensitivities can be explained by the fact that the

model is highly nonlinear with respect to the uncertain parameters. By taking these effects into account, the ODoE can further reduce the predicted interval of the cost function.

4. Conclusion and outlook

We presented an integrated methodology that can accelerate early phase process design. Optimal design of experiments can be used to efficiently design experiments to determine model parameters that are needed for a model-based process design. The most relevant parameters that should be determined more precisely by experimental work can be identified reliably by a sensitivity analysis. A modified optimal design of experiments weighted with these sensitivities enables to focus the experimental work on the most relevant model parameters and can hence be a promising tool for process design. In a case study it was shown that in case of nonlinear and interaction effect a global sensitivity analysis is superior compared to a local sensitivity analysis. It is expected that the global sensitivity analysis is even more beneficial in cases with structural uncertainties, because these cannot be quantified using local sensitivities. This will be the subject of further investigations.

This work presented here focused on the experimental design. It will be embedded into a complete process design procedure, including superstructure optimization under uncertainty and the presented modified ODoE.

Acknowledgement

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

References

- Asprion, N., Bö, R., Mairhofer, J., Yliruka, M., Hö, J., Schwientek, J., ... Bortz, M. (2019). Implementation and Application of Model-Based Design of Experiments in a Flowsheet Simulator. *J. Chem. Eng. Data*, 65(3), 1135–1145.
- Chen, Q., & Grossmann, I. E. (2017). Recent Developments and Challenges in Optimization-Based Process Synthesis. *Annu. Rev. Chem. Biomol. Eng.*, 8(1), 249–283.
- Franceschini, G., & Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chem. Eng. Sci.*, 63(19), 4846–4872.
- Hentschel, B., Kiedorf, G., Gerlach, M., Hamel, C., Seidel-Morgenstern, A., Freund, H., & Sundmacher, K. (2015). Model-based identification and experimental validation of the optimal reaction route for the hydroformylation of 1-dodecene. *Ind. Eng. Chem. Res.*, 54(6), 1755–1765.
- Hernandez, R., Dreimann, J., Vorholt, A., Behr, A., & Engell, S. (2018). Iterative Real-Time Optimization Scheme for Optimal Operation of Chemical Processes under Uncertainty: Proof of Concept in a Miniplant. *Ind. Eng. Chem. Res.*, 57(26), 8750–8770.
- Kaiser, S., & Engell, S. (2020). Integrating Superstructure Optimization under Uncertainty and Optimal Experimental Design in early Stage Process Development. In *Computer Aided Chemical Engineering* (Vol. 48).
- Recker, S., Kerimoglu, N., Harwardt, A., Tkacheva, O., & Marquardt, W. (2013). On the integration of model identification and process optimization. In *Computer Aided Chemical Engineering* (Vol. 32, pp. 1021–1026). Elsevier B.V.
- Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.*, 55(1–3), 271–280.
- Steimel, J., & Engell, S. (2015). Conceptual design and optimization of chemical processes under uncertainty by two-stage programming. *Comput. Chem. Eng.*, 81, 200–217.

Design Space Approximation with Gaussian Processes

Panagiotis Demis,^a Sergei Kucherenko,^b Oleksiy V. Klymenko^{a,*}

^a*Department of Chemical and Process Engineering, University of Surrey, Guildford GU2 7XH, United Kingdom*

^b*Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom*

o.klymenko@surrey.ac.uk

Abstract

The design and operation of process systems are required to meet multiple constraints related to production schedules, product quality, safety, economic performance and environmental footprint. These constraints define the set of feasible design and/or operational parameters which is called process Design Space (DS). In most instances, process constraints are defined as functions of state variables of the system in which case the full-scale process model must be solved for their verification which can be a computationally demanding task for large-scale nonlinear models. This a challenge for online applications such as model-predictive control or real-time optimisation.

In this study we present a computationally efficient method of evaluating the feasibility of a set of model parameters using a surrogate indicator function of the DS through Gaussian Process approximation of deterministic inequality model constraints. The method allows finding a compromise between the computational effort required and the level of confidence reflecting the accuracy of DS approximation.

Keywords: design space, constraints, Gaussian process, uncertainty.

1. Introduction

Mathematical models developed for the design and control of manufacturing processes are subject to multiple constraints imposed on a wide range of process parameters (equipment sizes, mass and energy flowrates, controllability, reliability, etc.), product characteristics (quality and yield) as well as economic performance measures and other Key Performance Indicators (KPIs) of the process. Many of these models are highly nonlinear and/or large in scale which makes them computationally expensive to run. This can be problematic, especially when the model needs to be solved repeatedly either in the context of model-predictive control/real-time optimisation or simply to verify the feasibility of a state of the system.

The latter is particularly important in assuring the quality of a product in fine chemicals industries such as (bio)pharmaceuticals where deviations of product characteristics from those approved by a regulator can have drastic safety and financial repercussions. In such cases, the Design Space (DS) defined as “the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have

been demonstrated to provide assurance of quality” (Pharmaceutical Development Q8(R2), 2009) plays a crucial role in process design in accordance with the Quality-by-Design paradigm, control and product approval.

Considering a generalised process model $f: R^n \rightarrow R^m$ that maps an n -dimensional input space into an m -dimensional space of the outputs of interest, let us define a set of C inequality constraints delineating combinations of model inputs leading to feasible outputs (i.e., the DS) as $g(x, f(x)) \geq 0$, $g: R^{n+m} \rightarrow R^c$. In the following, we will refer to such constraints, which depend on model function $f(x)$, as ‘implicit’. Generally, the whole process model must be solved to evaluate implicit constraints and establish whether x lies in the DS. In most cases, this also precludes any possibility of finding an explicit expression or procedure for determining the feasibility of a given set of model inputs.

This fact and the computational complexity of ‘discovering’ the true DS have led to a situation when only DS descriptions in the form of hyperrectangles (i.e., Cartesian products of ranges of model inputs) can be approved by regulators while these are generally only subsets of much larger feasible subspaces. There have been relatively few attempts in the literature to systematically approach the problem of model-based non-convex DS identification (Bano et al., 2019; García-Muñoz et al., 2015; Kotidis et al., 2019; Kucherenko et al., 2020), including finding a hyperrectangle of maximum volume inscribed into the DS (Ochoa et al., 2019). While the cited approaches also considered that the model constraints depend on uncertain parameters, they stress that a high number of process model evaluations is required to obtain a reasonable description of the DS.

Emerging research in DS modelling investigates the construction of a feasibility probability map to obtain an easy-to-evaluate feasibility indicator for each constraint using regression and classification methods (Knudde et al., 2019; Kusumo et al., 2020). To address the computational inefficiencies of DS identification, we adopt the application of Gaussian Processes (GPs) exploiting their intrinsic measures of uncertainty in the feasibility estimates. Our method introduces a confidence threshold to the estimated feasibility measures to control the quality of approximation thus balancing the computational effort of training the GP with the accuracy of DS representation. The proposed framework accelerates the identification of the whole non-convex DS, rather than its orthogonal subspace, tailoring space filling designs to enable extended flexibility of process operation while preserving the feasibility of model constraints.

2. Theory

In the following for simplicity we will assume that the maximum ranges of model input variations have been mapped onto a unit hypercube denoted as $H^n = [0, 1]^n$.

The joint distribution of model inputs x is assumed to be represented by a probability density function $p(x)$. Model constraints will be denoted as $g(x)$ regardless of whether they are explicit (i.e., do not depend on model outputs $f(x)$) or implicit (i.e., dependent on $f(x)$ as in $g(x, f(x))$).

The Design Space denoted as Ω^n is a subspace of H^n such that all of its points are feasible with respect to the constraints. The DS can be defined through its indicator function as

$$\Omega^n = \{x: I(x) = 1\} \subset H^n, \text{ where } I(x) = \begin{cases} 1, & g(x) \geq 0 \\ 0, & g(x) < 0 \end{cases} \quad (1)$$

A single constraint will be considered to develop the methodology below. When multiple inequality constraints determine the feasibility of a point in the input space the indicator function, $I(x)$, can be formulated as the product of C indicators corresponding to individual constraint functions $g_i(x)$ such that $I(x) = \prod_{i=1}^C I_i(x)$.

2.1. Gaussian process interpolation of model constraints

The unit hypercube H^n containing the feasible DS Ω^n is sampled from distribution $p(x)$ and the values of model constraint $g(x)$ are computed at the sample points. A Gaussian Process interpolant is then fitted on the set of sample points, and additional samples can be generated as required by the convergence criteria.

A Gaussian Process (GP) over the constraint function g is a stochastic process comprised by an infinite collection of random variables any finite subset of which, $Y = \{g(x^{(1)}), \dots, g(x^{(N)})\}$, is normally distributed as shown in (Rasmussen, 2004)

$$p(Y) = N(m(x), K(x, x')) \quad (2)$$

where $m(x) = E[x]$ is the mean function, and $k(x, x') = E[(x' - m(x'))(x - m(x))]$ denotes a positive-definite covariance function corresponding to two points $(x, x') \in H^n$.

Bayesian inference of GPs allows using a prior distribution in Eq. 4 to derive a posterior conditional distribution based on a set of model evaluations. The latter is used as a predictive distribution over $g(\cdot)$ at any testing point $x_* \in H^n$. Considering a training set $X = \{x^{(1)}, \dots, x^{(N)}\}$ and a test sample $X_* = \{x_*^{(1)}, \dots, x_*^{(N_*)}\}$ distributed according to $p(x)$, the joint distribution of the corresponding random variables based on a GP is given by

$$p(Y, Y_*) = N\left(\begin{bmatrix} m(X) & m(X_*) \end{bmatrix}, \begin{bmatrix} K_{XX} & K_{XX_*} & K_{X_*X} & K_{X_*X_*} \end{bmatrix}\right) \quad (3)$$

where $K_{XX_*} = k(X, X_*)$. GP interpolation is based on conditioning the probability density function (Eq. 3) for non-parametric estimation of Y_* . Assuming $m(x) = 0$, it follows that

$$p(Y, X, X_*) = N(M_*, \Sigma_*) \quad (4)$$

where $M_* = K_{X_*X} K_{XX}^{-1} Y$ and $\Sigma_* = K_{X_*X_*} - K_{X_*X} K_{XX}^{-1} K_{XX_*}$.

2.2. Confidence level of indicator approximation

Conditioning the GP on a training sample X and a test point $x \in H^n$, the posterior density function of $GP(x)$ is $p(Y, X, x) = N(\mu_x, \Sigma_x)$, where μ_x, Σ_x are given by Eq. 4. Considering $g(x) \geq 0 \Leftrightarrow I(x) = 1$, the conditional posterior distribution is used to express the probability that test point x is feasible according to the approximation of the constraint by the GP:

$$P(I(x) = 1) \approx P_I(x) = P(GP(x) \geq 0) = 1 - P(GP(x) < 0) \tag{5}$$

where $P_I(x) = P(GP(x) \geq 0)$ is computed using the normal cumulative distribution function $F_{GP(x)}(0)$ of the GP-based approximation:

$$P_I(x) = 1 - F_{GP(x)}(0) = 1 - \int_{-\infty}^0 p(Y, X, x) du = \frac{1}{2} \left[1 - \left(\frac{-\mu_x}{\Sigma_x \sqrt{2}} \right) \right] \tag{6}$$

Using the measure of uncertainty around the approximation of the exact $I(x)$ in Eq. 6, a GP-based indicator $\tilde{I}_\varepsilon(x)$ is computed by introducing a threshold $\varepsilon \in (0, 1)$ in the estimated $P_I(x)$. Hence, an approximate DS is described as

$$\tilde{\Omega}^n = \{x: \tilde{I}_\varepsilon(x) = 1\} = \{x: P_I(x) \geq \varepsilon\} \tag{7}$$

The uncertainty threshold ε represents the lowest level of confidence which can be tolerated in the approximation of $I(x)$ by the GP. Therefore, ε serves as a quality metric when evaluating $\tilde{I}_\varepsilon(x)$ at the design point x and as a classifier parameter which is applied on the intrinsic uncertainty of the GP-based estimates rather than their mean values. The method can be extended for multiple inequality constraints, when C GPs must be trained for the identification of the corresponding DS such that

$$\tilde{\Omega}^n = \left\{ x: \prod_{i=1}^C \tilde{I}_\varepsilon^i(x) = 1 \right\}.$$

Convergence of the approximate indicator $\tilde{I}_\varepsilon(x)$ of $\tilde{\Omega}^n$ to the true one $I(x)$ can be assessed using the volume of the approximate DS $V_{\tilde{\Omega}_\varepsilon^n}$. The latter can be computed using a large but computationally inexpensive Monte Carlo (MC) sample of the GP of size $N_* \gg N$:

$$V_{\tilde{\Omega}^n} = \int I(x) dx \approx N_*^{-1} \sum_{i=1}^{N_*} \tilde{I}_\varepsilon(x_i) = V_{\tilde{\Omega}_\varepsilon^n} \tag{8}$$

Convergence of $V_{\tilde{\Omega}_\varepsilon^n}$ to $V_{\tilde{\Omega}^n}$ depends on the size of training sample N , size of N_* and the value of ε . With increasing N , Σ_x in Eq. 6 decreases for all $x \in H^n$ leading to convergence of $\tilde{I}_\varepsilon(\cdot)$ to $I(\cdot)$. It can be shown that for any $\varepsilon \in (0, 1)$ there exists $\Sigma_0(\varepsilon)$ corresponding to a finite number of training samples N_ε such that for each $x \in H^n$

$\Sigma_x \leq \Sigma_0(\varepsilon)$ and $\left| V_{\tilde{\Omega}_\varepsilon^n} - V_{\tilde{\Omega}^n} \right| < \delta_\varepsilon$. Thus, the convergence of $V_{\tilde{\Omega}_\varepsilon^n}$ is uniform in ε over $(0, 1)$ for training sample of size $N > N_\varepsilon$ and $\frac{d}{d\varepsilon} V_{\tilde{\Omega}_\varepsilon^n} \rightarrow 0$ for all ε except $\varepsilon = 0$ or 1 . The latter can be used as a convergence criterion for the approximation of $V_{\tilde{\Omega}^n}$.

3. Results

The proposed framework was tested with an inequality constraint based on Sobol’ product function (Eq. 9) because of its properties including nonlinearity, non-smoothness, symmetry and scalability in higher dimensions.

$$h(x_1, \dots, x_n) = 1 - \prod_{i=1}^n \frac{|4x_i - 2| + a_i}{1 + a_i} \geq 0, \quad a_i = i - 1 \quad (9)$$

Considering that inputs are uniform iid variables, sampling is performed in H^n using the space-filling Sobol’ pseudo-random sequence (Sobol et al., 2011), which accelerates convergence in higher dimensions compared to conventional random number generators.

Without implying any knowledge of $h(\cdot)$, the GP is defined by zero mean and squared exponential covariance function as commonly suggested in most applications. Based on a 95% confidence threshold for the definition of the feasible DS, and $N = 2^{11}$ training samples for the GP, Figure 1 shows the 2-dimensional boundary of Eq. 9. Herein, $h(\cdot)$ represents a non-smooth example of how DS identification based on integration of the Gaussian uncertainty around the interpolant can achieve a robust DS approximation without significant loss of feasible volume. However, in higher dimensions and for highly non-smooth constraint functions the performance of a GP-based surrogate indicator it is expected to deteriorate.

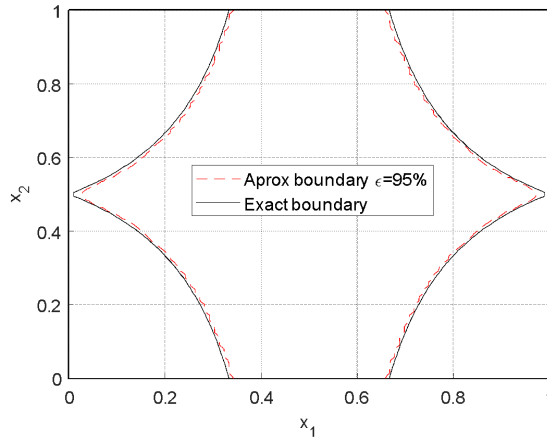


Figure 1: Exact DS boundary and its approximation based on 95% confidence

Figure 2(a) demonstrates how increasing the number of training samples, N , reduces the uncertainty in a 6-dimensional DS volume approximation, by showing that the effect of ε on $V_{\Omega_\varepsilon}^{\sim n}$ becomes less significant as sampling size grows. Based on this, $\frac{d}{d\varepsilon} V_{\Omega_\varepsilon}^{\sim n}$ is suggested as a measure of convergence of $V_{\Omega_\varepsilon}^{\sim n}$ and hence of the approximate indicator

$\tilde{I}_\varepsilon(x)$. Indeed, only if $\left| \frac{d}{d\varepsilon} V_{\Omega_\varepsilon}^{\sim n} \right| < 1$ around $\varepsilon \approx 0.5$ any discrimination between feasible and infeasible subspaces of H^n is possible. In applications where a conservative value of

ε (close to one) is required, because no violation of the constraints can be afforded, the approximated $V_{\sim n}$ is smaller, thus there is a tradeoff between the level of confidence in the approximation and feasible DS volume for any fixed training sample of size N .

The problem of approximating the DS is magnified in higher dimensions due to the curse of dimensionality. As dimensionality n increases, the volume of a subspace with a characteristic linear dimension $L < 1$ decreases exponentially. Thus, the size of the training sample N required to achieve a good approximation of Ω^n increases significantly. Figure 2(b) illustrates the effects of the curse of dimensionality when using a fixed sample size ($N = 2^{10}$) and indicates that the convergence of $V_{\sim n}$ to its true value may be slow in higher dimensions.

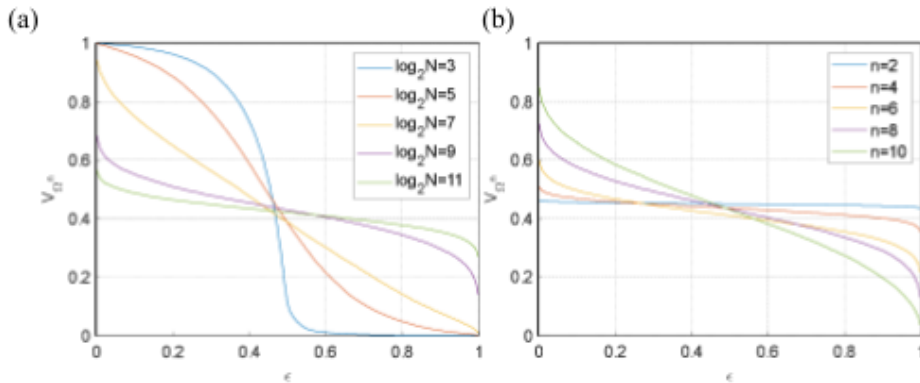


Figure 2: The effect of confidence threshold ε on DS volume approximation when increasing size of training sample (a) and dimensionality (b)

4. Conclusions

A computational framework for efficient approximation of the process Design Space is introduced, based on the use of Gaussian Processes as surrogate models of inequality constraint functions describing process safety, reliability and product quality. Gaussian Processes play a dual role as a flexible and inexpensive interpolation method for a wide class of constraint functions, accelerating sampling-based exploration of the process DS and measuring the uncertainty in the estimated constraints feasibility. In our approach, the quality and robustness of DS approximation can be controlled through the size of a training sample and desired confidence level. However, the method is not immune to the curse of dimensionality with a large number of training points and full model evaluations required in higher dimensions for reliable DS identification.

References

- G. Bano, P. Facco, M. Ierapetritou, F. Bezzo, M. Barolo, 2019, Design space maintenance by online model adaptation in pharmaceutical manufacturing, *Comput. Chem. Eng.*, vol. 127, p. 254–271.
- S. García-Muñoz, C.V. Luciani, S. Vaidyaraman, K.D. Seibert, 2015, Definition of Design Spaces Using Mechanistic Models and Geometric Projections of Probability Maps. *Organ. Process Res. Dev.*, 19(8), p. 1012-1023.

- ICH Harmonized Tripartite Guidelines, 2009, Pharmaceutical Development Q8(R2), <https://database.ich.org/sites/default/files/Q8%28R2%29%20Guideline.pdf>.
- N. Knudde, I. Couckuyt, K. Shintani, T. Dhaene, 2019, Active learning for feasible region discovery, in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, p. 567–572.
- P. Kotidis, P. Demis, C.H. Goey, E. Correa, C. McIntosh, S. Trepekli, N. Shah, O.V. Klymenko, C. Kontoravdi, 2019, Constrained global sensitivity analysis for bioprocess design space identification, 125, p. 558-568.
- S. Kucherenko, D. Giamalakis, N. Shah, S. García-Muñoz, 2020, Computationally efficient identification of probabilistic design spaces through application of metamodeling and adaptive sampling, *Comput. Chem. Eng.*, 132, 106608.
- K.P. Kusumo, L. Gomoescu, R. Paulen, S. García Muñoz, C.C. Pantelides, N. Shah, B. Chachuat, 2020, Bayesian Approach to Probabilistic Design Space Characterization: A Nested Sampling Strategy, *Ind. Eng. Chem. Res.* 59, p. 2396-2408.
- M.P. Ochoa, A. Deshpande, S. García-Muñoz, S. Stamatias, I.E. Grossmann, 2019, Flexibility Analysis For Design Space Definition, *Proceedings of the 9th International Conference on Foundations of Computer-Aided Process Design*.
- C.E. Rasmussen, 2004, Gaussian Processes in Machine Learning. In: Bousquet O., von Luxburg U., Rätsch G. (eds) *Advanced Lectures on Machine Learning. ML 2003. Lecture Notes in Computer Science*, vol 3176. Springer, Berlin, Heidelberg.
- I.M. Sobol', D. Asotsky, A. Kreinin, S. Kucherenko, 2011, Construction and Comparison of High-Dimensional Sobol' Generators, *Wilmott*, p. 64-79.

Employing Adaptive Just-In-Time-Learning in a Transfer Learning Frame for Soft-Sensor Design

Burak Alakent^a

^aDepartment of Chemical Engineering, Bogazici University, Istanbul, Turkey

burak.alakent@boun.edu.tr

Abstract

Online learning methods, such as moving window (MW) and just in time learning (JITL), have been proposed in the literature to remedy concept drift problem deteriorating soft-sensor performance. While these methods are effective against different types of drifts, a single method may not be sufficient in combating against heterogeneous concept drifts. In the current study, we propose combining MW and JITL methods within a transfer learning frame coupled with a relevant instance selection method to improve the prediction accuracy offered by either method. The proposed method involves i) forming a relevant sample of historical observations via backward elimination of the clusters composed of the extended nearest neighbors of the query point, ii) constructing a task transferred JITL model via kernel ridge regression, and iii) using a transductive MW learner. Employing the proposed method on two publicly available real benchmark datasets yields highly accurate predictions, showing convenience for industrial applications.

Keywords: Concept drift, K-means clustering, Kernel regression, Online learning, Relevant sample.

1. Introduction

Soft-sensors are predictive models used to exploit information extracted from the secondary process variable measurements for online prediction of the quality variables in chemical processes [1]. Rise of the machine learning in the last ~20 years has rendered a shift in the design methodology of soft sensors from mechanistic to data-based paradigm. While state-of-the-art learners make it possible to obtain sufficiently accurate predictions for future operations similar to those used in the training set, changes in process conditions, i.e. concept drifts, may easily deteriorate the prediction accuracy of any learner. In “virtual” concept drifts, only the distribution of the secondary variables (features) is perturbed, while the functional relation between the secondary and quality variables changes under “real” concept drifts [2]. Various online learning methods are proposed to reduce the harmful effects of concept drifts, such as moving window (MW) and just-in-time-learning (JITL) models. The former method involves using a sliding window, consisting of the most recent observations for the training set, while the latter involves choosing the most relevant observations, usually obtained by a variation of nearest neighbor (NN) criterion. It is known that MW is affective against real and gradual concept drifts while JITL is affective against abrupt recurrent virtual drifts [3]. Industrial processes, however, encounter heterogeneous concept drifts, hence it is not reasonable to

expect a single MW or JITL model to maintain its prediction accuracy during the whole operation.

Transfer learning offers a new perspective for predicting a target task, for which a small number of labeled observations exist, via utilizing information from a source domain with a larger number of labeled samples [4]. We have recently proposed transductive MW (MW_{tr}) [5], and task transferred JITL coupled MW_{tr} ($JITL_{TT}-MW_{tr}$) learners [6] for combining MW and JITL models in a transfer learning perspective, and shown that high predictive accuracy can be reached against heterogeneous concept drifts. In these studies, however, i) we did not check the relevance of the NN points with the current task, and ii) we used linear learners for the JITL model. In the current study, we improve these issues using an adaptive task transferred JITL coupled transductive MW ($AdJITL_{TT}-MW_{tr}$) learner. Application of $AdJITL_{TT}-MW_{tr}$ on three cases from two publicly available industrial real datasets shows that the proposed online learner has significantly high predictive accuracy, and is robust with respect to learning parameters.

2. Online Learning Methods

In supervised machine learning, a historical dataset of N points $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$ with $\mathbf{x}_n \in \mathbb{R}^p$ and $y_n \in \mathbb{R}$ is used to predict the response (or quality) variable (\hat{y}_{N+1}) of the test (query) point \mathbf{x}_{N+1} . When the mapping function from the features to the quality variables, $f(\mathbf{x}_n)$, changes with respect to the domain of the features and/or time, online learning is more convenient in adapting the predictive function to the current task, contrary to batch learning, which offers a constant predictive function.

2.1. Moving Window (MW) and Just-in-Time-Learning (JITL)

Most of the online learning schemes in soft-sensor design for chemical processes involve choosing relevant training samples. The most recent samples are selected in MW modeling, while JITL is commonly based on the selection of the most similar samples to the test point in the feature space. The rationale behind sample selection methods lies in combating real and virtual concept drifts; the current task is assumed to be more accurately represented by the recent samples, and samples collected in the current domain of the query point. To improve the prediction performance of single MW or JITL models, novel methods using ensemble techniques [7], and combinations of MW and JITL [8] are proposed in the literature. Prediction accuracy of MW and JITL models depends on the window size (W) and the size of NN set (K). While a small W usually yields a more accurate representation of the current task, the resulting training set would only span a limited subspace of the whole domain of operation. In JITL, sample selection based on feature similarity usually leads to a larger training set, comprising samples from various time segments, and spanning a wider subspace. Since linear learners are less prone to overfitting for small multicollinear datasets, we use a linear learner (Lasso) for MW, and a nonlinear learner (Kernel Ridge Regression) for JITL in the current study.

2.2. Least Absolute Shrinkage and Selection Operator (Lasso)

Sum of squares of fitting errors is regularized using the L1 norm of coefficient estimates and parameter λ in Lasso, also yielding an embedded feature selection mechanism [9]:

$$\min_{\boldsymbol{\beta}} \left\{ \lambda \sum_{j=1}^p |\beta_j| + \sum_{n=1}^N (y_n - \mathbf{x}_n^T \boldsymbol{\beta})^2 \right\} \quad (1)$$

2.3. Kernel Ridge Regression (KRR)

In KRR, regression in feature space is regularized via L2 norm of regression coefficients:

$$\min_{\boldsymbol{\beta}} \left\{ \lambda \boldsymbol{\beta}^T \boldsymbol{\beta} + \sum_{n=1}^N (y_n - \Phi(\mathbf{x}_n)^T \boldsymbol{\beta})^2 \right\} \quad (2)$$

Here, $\Phi(\cdot)$ corresponds to the nonlinear mapping of the input variables to the feature space. Using the “kernel trick”, which makes it possible to determine the inner product in Hilbert feature space using a kernel $k(\mathbf{x}_i, \mathbf{x}_j)$ computed in the input space, the prediction for y_{N+1} can be determined as follows:

$$\hat{y}_{N+1} = \sum_{n=1}^N c_n k(\mathbf{x}_{N+1}, \mathbf{x}_n) \quad (3)$$

In the above formulation, $\mathbf{c} = (\mathbf{I}\lambda + \mathbf{G})^{-1} \mathbf{y}$, in which \mathbf{G} is the gram matrix with $G_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, 2, \dots, N$ and $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ [10]. Simply put, using a positive semi definite kernel, representing the similarity between samples in the training set, the query point may be predicted using the similarity of the query point to the training samples. We use exponential kernel in the current study, i.e. $k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma^2\right)$.

3. Transfer Learning

In transfer learning, information in a source domain is used in predicting the task defined in the target domain. It is assumed that either i) the source domain (D_S) and target domain (D_T) differ, or ii) the source task ($f_S(\mathbf{x})$) and target task ($f_T(\mathbf{x})$) are different. The former and the latter scenarios are named transductive and inductive transfer learning, respectively. While there are various information transfer mechanisms, we have employed instance transfer in the current study. In industrial processes under heterogeneous drifts, we assume that the tasks governing the query point and small sized MW samples are identical, but the domains from which these two sets of points are sampled may be different. To solve this issue, we have recently proposed transductive MW learner (see Section 4.3) [5]. We also assume that the query point and JITL samples belong to identical domains, but possibly are governed via different task functions. To solve this problem, we have recently proposed task transferred JITL (JITL_{TT}) learner (see Section 4.2) [6]. In JITL_{TT}, we transfer a single sample from MW to NN set in order to adapt the JITL model to the current task. It should, however, be pointed out that even this mode of information transfer may not be sufficient for realizing a feasible task transfer of the JITL model, since there already may exist samples in the NN set detrimental for query point prediction.

4. The Proposed Method

4.1. Adaptation of the NN Set to the Current Task

The novelty of the current study lies in integrating a relevant data selection method to our recently proposed transfer learning methods. The fundamental motivation here is to select a subset of samples from an extended NN observations set, that “better” represents the current operation conditions. To this aim, first an initial NN set of the query point, based on Euclidean distance similarity of the feature values, is formed; then, this set is extended via including samples in time-proximity. Finally, a subset of the extended NN set that yields the smallest validation error for the MW samples is determined via backward elimination. The algorithm is described in detail as follows:

- i. Determine the time-ordered NN samples set: Obtain $\text{NN} = \{(\mathbf{x}_{n_k}, y_{n_k})\}_{k=1}^K$ from the samples in the historical dataset excluding the MW samples, with the ranked Euclidean distances $\{d_{E(1)}, d_{E(2)}, \dots, d_{E(N)}\}$, such that $d_{E(k)} = \|\mathbf{x}_{n_k} - \mathbf{x}_{N+1}\|_2$.
- ii. Identify different time segments, possibly corresponding to different operating regions in the NN set: Determine successive samples in NN set with L time lags separated, i.e. $n_{k+1} - n_k > L$, for $k = 1, 2, \dots, N - 1$. Let total number of different time segments, from which NN samples are collected, be equal to T , e.g. if there are no two consecutive samples separated with L samples, then $T = 1$.
- iii. Extend each time segment to include the neighboring samples: For each time segment of samples, denoted by \mathbf{n}_t , $t = 1, 2, \dots, T$, all time indices between the minimum and maximum indices, after extending by $\pm L/2$, are included in \mathbf{n}_t , e.g. taking $L = 10$, for a time segment \mathbf{n}_1 initially consisting of samples with indices $\{28, 29, 34, 36, 38\}$, \mathbf{n}_1 is extended so as to include the time indices $\{23, 24, \dots, 43\}$.
- iv. Define the extended NN_E set via concatenating the sample indices from each time segment: $\text{NN}_E = \{(\mathbf{x}_{n_k}, y_{n_k})\}_{k=1}^{K^*}$, where $K^* = \sum_{t=1}^T |\mathbf{n}_t|$.
- v. Determine the principal component scores of the extended NN set: Perform principal components analysis on NN_E , and obtain the first A scores vectors (\mathbf{t}) with eigenvalues greater than the average eigenvalue value.
- vi. Divide NN_E into a predetermined number of clusters (C) using K-means on \mathbf{t} vectors.
- vii. Use backward elimination (BE) to determine the clusters with the smallest validation errors: Remove each cluster one by one, and predict the MW samples using KRR. Stop, when the validation set error starts to increase. Pool all the observations in the remaining clusters into a final ‘‘adapted’’ NN set (NN_{Ad}). Note that the elements and the size of the resulting NN_{Ad} set, which is to be used in JITL in the next section, are likely to be different from those of the initial NN set.

4.2. Adaptive Task Transferred JITL (AdJITL_{TT}) model

JITL_{TT} was proposed in a very recent study [6], in which we transferred the nearest single point from MW samples to adapt the JITL model to the current task using a Lasso estimator. We, here, make two modifications to this method to meet the demands of the current study. First, we use all of the MW samples for regularization, and second, we use a KRR estimator for obtaining a task transferred predictive function $f_{S \rightarrow T}(\cdot)$. We name this method Adaptive Task Transferred JITL learner (AdJITL_{TT}):

$$\min_{\boldsymbol{\beta}} \left\{ \lambda_1 \boldsymbol{\beta}^T \boldsymbol{\beta} + \sum_{k \in \text{NN}_{Ad}} (y_k - \Phi(\mathbf{x}_k)^T \boldsymbol{\beta})^2 + w_1^2 \sum_{m \in \text{MW}} (y_m - \Phi(\mathbf{x}_m)^T \boldsymbol{\beta})^2 \right\} \quad (4)$$

The solution to Eq. 4 can also be easily determined in kernelized formulation shown in Eq. 3 to determine the pseudo-label (y_{N+1}^*) for the $(N+1)^{\text{th}}$ query point.

4.3. Transductive MW Learner (MW_{tr})

In MW_{tr} model, deviations of the fitted value of the query point from its pseudo-label (y_{N+1}^*) is used as an additional penalty in a weighted Lasso formulation [5]:

$$\min_{\boldsymbol{\beta}} \left\{ \lambda_2 \sum_{j=1}^p |\beta_j| + \sum_{n \in \text{MW}} (y_n - \mathbf{x}_n^T \boldsymbol{\beta})^2 + w_2^2 (y_{N+1}^* - \mathbf{x}_{N+1}^T \boldsymbol{\beta})^2 \right\} \quad (5)$$

In this way, a model consistent with both MW samples and the JITL prediction is obtained, completing the formulation of the proposed AdJITL_{TT}-MW_{tr} method. The final prediction of the query point is obtained via $\hat{y}_{N+1} = \mathbf{x}_{N+1}^T \hat{\boldsymbol{\beta}}$, in which $\hat{\boldsymbol{\beta}}$ is obtained from Eq. 5.

5. Results and Discussion

5.1. Datasets Studied

The proposed method was employed on two publicly available industrial benchmark datasets, a debutanizer column (DC) and a sulfur-recovery unit (SRU) [11]. DC is located in a desulfuring and naphtha splitter plant, and a soft sensor is to be designed using seven online measured secondary process variables for predicting the butane content in the bottom product. We divided DC dataset into 1000/1382 samples for training/testing via rolling-origin-recalibration (RORC) evaluation [5,6], respectively. A SRU is used for removing polluting agents from acid gas, and a soft sensor is to be designed using five online measured secondary process variables for predicting H₂S and SO₂ content in the tail gas of Maxisulfur plant. We divided the SRU dataset into 5000/5076 samples for training/testing, respectively. We used feature vectors of lagged measurement for JITL calculations, while we used only the current variable measurements for MW predictions [5,6]. For AdjITL_{TT}-MW_{tr}, RORC evaluation was conducted over a multidimensional grid of parameter values assigned to $\lambda_1, \lambda_2, w_1, w_2$ and σ^2 at fixed values of W, K and C for the training set. The parameter values, which yield the smallest RMSE for the training set, were used for testing. Additionally, we report a posteriori best-case results for MW, JITL and MW_{tr} learners on the same test set data; hence, the current results reflect a conservative prediction performance of AdjITL_{TT}-MW_{tr} compared to other methods.

5.2. Metrics Used for Assessing Predictive Performance

Root mean squared error (RMSE) and R^2_{pred} of test samples are used to assess the prediction accuracy of the learning methods:

$$\text{RMSE} = \sqrt{\frac{1}{N_{test}} \sum_{n=1}^{N_{test}} (y_n - \hat{y}_n)^2} \quad (6)$$

$$R^2_{pred} = 1 - \frac{\sum_{n=1}^{N_{test}} (y_n - \hat{y}_n)^2}{\sum_{n=1}^{N_{test}} (y_n - \bar{y})^2} \quad (7)$$

Above, \bar{y} refers to average of the quality variable measurements.

5.3. Comparisons of Prediction Accuracy of AdjITL_{TT}-MW_{tr} with Other Methods

Table 1 shows that AdjITL_{TT}-MW_{tr} is superior in prediction accuracy compared to other methods for all three quality variables. Low RMSE values obtained from predictions of AdjITL_{TT}-MW_{tr} over a range of K (size of the initial NN) and C (number of clusters) values demonstrate the robustness of the proposed method with respect to its parameters (Fig. 1). Skipping the adaptation step (Section 4.1) is shown to increase the RMSE values significantly, indicating that adapting the NN set to the current task is indeed helpful in decreasing prediction errors. It is also seen that using four clusters yields marginally more accurate predictions compared to other cluster sizes.

6. Conclusion

The current study aims to improve the recently suggested methods on reconciling MW and JITL models in an inductive/transductive transfer learning setting via including an adaptation of nearest neighbors in a kernelized regression frame. The merit of the current method lies in coupling a small size MW learner, aiming to focus only on the most recent task function, with a JITL learner, constructed from a larger number of historical samples, that are consistent both with the current domain and the task. The proposed AdjITL_{TT}-MW_{tr} method is shown to yield significantly high predictive accuracy on benchmark

industrial datasets, and is robust with respect to its learning parameters. Overall, AdJITL_{TT}-MW_{tr} offers a promising soft-sensing methodology for chemical processes.

Table 1. Comparison of prediction accuracy of various learners on all datasets.

Dataset studied/ Online Learning Method	DC		SRU-H ₂ S		SRU-SO ₂	
	RMSE	R^2_{pred}	RMSE	R^2_{pred}	RMSE	R^2_{pred}
MW ⁽ⁱ⁾	0.0288	0.978	0.0205	0.839	0.0164	0.900
JITL	0.0537	0.923	0.0201	0.845	0.0245	0.776
MW _{tr}	0.0149	0.994	0.0153	0.910	0.0135	0.931
AdJITL_{TT}-MW_{tr}	0.0116	0.996	0.0138	0.926	0.0120	0.946

ⁱ: Test results for MW, JITL and MW_{tr} are taken from a previous study [5].

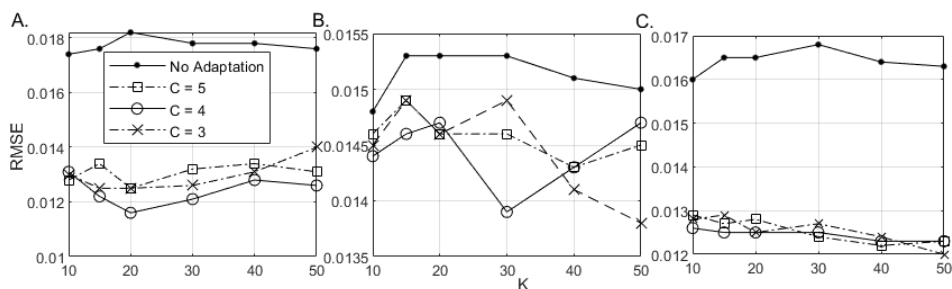


Figure 1. RMSE values for the test set obtained using AdJITL_{TT}-MW_{tr} with $K = \{10, 20, \dots, 50\}$ and $C = \{3, 4, 5\}$ for (A) DC ($W=3$), (B) H₂S in SRU ($W=2$), and (C) SO₂ in SRU ($W=2$) datasets. No adaptation refers to using the NN samples directly, i.e. skipping the adaptation step (Section 4.1), in JITL_{TT}-MW_{tr}.

References

1. F.A.A. Souza, R. Araujo, 2016, Review of Soft Sensors Methods for Regression Applications, *Chemometr. Intell. Lab. Syst.*, 152, 69-79.
2. G. Ditzler, M. Roveri, C. Alippi, R. Polikar, 2015, Learning in Nonstationary Environments: A Survey, *IEEE Comput. Intell. Mag.*, 10, 12-25.
3. H: Kaneko, K. Funatsu, 2013, Classification of the Degradation of Soft Sensor Models and Discussion on Adaptive Models, *AIChE J.*, 59, 2339-2347.
4. K. Weiss, T.M. Khoshgoftaar, D. Wang, 2016, A survey of transfer learning. *J. Big Data*, 3, 9-40.
5. B. Alakent, 2020, Soft sensor design using transductive moving window learner, *Comput. Chem. Eng.*, 140, 106941.
6. B. Alakent, 2021, Soft-Sensor Design via Task Transferred Just-in-Time-Learning coupled Transductive Moving Window Learner, Under revision in *J. Proc. Control*.
7. H. Jin, J. Li, M. Wang, B. Qian, B. Yang, Z. Li, L. Shi, 2020, Ensemble Just-In-Time Learning-Based Soft Sensor for Mooney Viscosity Prediction in an Industrial Rubber Mixing Process, *Adv. Polym. Tech.*, 6575326.
8. A. Urhan, B. Alakent, 2020, Integrating Adaptive Moving Window and Just-in-Time Learning Paradigms for Soft-Sensor Design, *Neurocomputing*, 392, 23-37.
9. R. Tibshirani, 1996, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B.*, 58, 267-288.
10. J. Hainmueller, C. Hazlett, 2014, Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach, *Polit. Anal.*, 22, 143-168.
11. L. Fortuna, S. Graziani, A. Rizzo, M.G. Xibilia, 2007, *Soft Sensors for Monitoring and Control of Industrial Processes*, Springer, England.

AC-Optimal Power Flow Solutions with Security Constraints from Deep Neural Network Models

Zachary Kilwein^a, Fani Boukouvala^a, Carl Laird^b, Anya Castillo^b, Logan Blakely^b, Michael Eydenberg^b, Jordan Jalving^b, Lisa Batsch-Smith^b

^a*Georgia Institute of Technology, North Ave NW, Atlanta, GA, 30332, USA*

^b*Sandia National Labs, 1515 Eubank Blvd SE, Albuquerque, NM, 87123, USA*

zkilwein3@gatech.edu

Abstract

In power grid operation, optimal power flow (OPF) problems are solved several times per day to find economically optimal generator setpoints that balance given load demands. Ideally, we seek an optimal solution that is also “ $N-1$ secure”, meaning the system can absorb contingency events such as transmission line or generator failure without loss of service. Current practice is to solve the OPF problem and then check a subset of contingencies against heuristic values, resulting in, at best, suboptimal solutions. Unfortunately, online solution of the OPF problem including the full $N-1$ contingencies (i.e., two-stage stochastic programming formulation) is intractable for even modest sized electrical grids. To address this challenge, this work presents an efficient method to embed $N-1$ security constraints into the solution of the OPF by using Neural Network (NN) models to represent the security boundary. Our approach introduces a novel sampling technique, as well as a tuneable parameter to allow operators to balance the conservativeness of the security model within the OPF problem. Our results show that we are able to solve contingency formulations of larger size grids than reported in literature using non-linear programming (NLP) formulations with embedded NN models to local optimality. Solutions found with the NN constraint have marginally increased computational time but are more secure to contingency events.

Keywords: Optimization, Machine Learning, Neural Networks, Power Grid Modeling

1. Introduction

1.1 Background

Power grids are very large, complex systems that require operation that is both financially optimal and secure to unforeseen outage events. The US Federal Energy Commission estimates that effective optimization of power grids can save tens of billions of dollars annually (Cain et al. 2012). Furthermore, the need for powerful optimization tools will only increase as more renewables and non-conventional energy sources are added to the grid. This is largely done via optimal power flow (OPF) programs which balance real-time load demands with generator outputs in the most economically optimal way. Some approaches to solving OPF include rigorous formulations with non-linear and non-convex functions (AC-OPF), convex relaxations

of this problem (e.g. Second Order Cone OPF), and linear approximations (e.g. DC-OPF) that are standardly used in industrial applications. In the case of AC-OPF, further distinction can be made between local and global solution algorithms due to non-convexities. This work focuses on the optimization of the rigorous AC-OPF formulation with local solvers, since it can handle large problems without approximating underlying power flow physics and very often achieves solutions close to global optimality (O'Neill et al 2012).

As grids are operated close to their optimal points (i.e., less conservatively), security must be considered as they are more prone to system component failure that may result in serious financial and safety consequences. Operating at suboptimal set points will yield significant cost increases, whereas economically-optimal though insecure set points have potentially catastrophic consequences under a contingency event, as cascading failures can cause blackouts for millions of people. Therefore it is incumbent upon system operators to balance these interests effectively. One way to check system security is to simulate system failures and enforce the resulting security of the system as constraints. If this is done for every component in the system, it is said to be *N-1* secure. While the scalability and accuracy of AC-OPF algorithms have seen major improvements for both local and global techniques, ensuring *N-1* security remains challenging. The current practice is to solve the OPF problem without consideration of contingencies, and then check the solution for *N-1* feasibility. If the solution is not feasible, heuristics are used to add constraints to the OPF problem, and the process is repeated. This approach yields suboptimal results, and since only a subset of the contingencies is typically checked, this may also yield insecure results. While it is possible to formulate a two-stage stochastic programming problem that considers all contingencies, this large scale problem is not tractable for realistic networks. This work aims at solving security constrained (SC) AC-OPF problems with the aid of deep Neural Networks (NN) that can be trained offline to learn the input space of the original optimization problem that is *N-1* secure. The trained NN model can then be used as a single constraint to enforce *N-1* security of any resultant AC-OPF solution across the input space and contingency events.

1.2 Literature Review and Contribution

Capitanescu et al. (2011) provide a thorough review of previous research in the area of SC OPF and future methods that may address some of the challenges detailed above. Heuristic methods are frequently used in practice, but more robust relationships can be derived as algebraic constraints. However, due to the problem size and complexity, fully detailed formulations remain intractable and often reduced or simplified techniques are required. Gutierrez-Martinez et al. (2011) trained non-linear regression models and single layer NN models to find the security boundary of a given load profile for the worst case contingency. Our work builds on this by using deep NN models as a map of the secure and insecure space of all contingencies of interest, variable generation patterns, and higher dimensional representations of the security boundary. Velloso et al. 2020 replace the full SC AC-OPF with a NN model that includes physics-based constraints via a Lagrangian-dual. Venzke et al. (2020) trained a NN security classifier with ReLU activation functions and added the constraints to the AC-OPF problem as a set of linear constraints with binary variables that results in a mixed-integer non-linear program (MINLP). They linearize the power equations to convert this to a MILP. This work instead maintains the non-linearity of the NN based security-constraints and formulates the overall problem instead as a nonlinear program (NLP) that can be solved efficiently via interior-point methods. Some key advantages of our approach include: (a)

computationally challenging security simulations are performed offline, (b) security is represented by a single NN constraint without integer variables, (c) the fully detailed AC-OPF formulation is used, (d) the security constraint scales linearly with grid size, and (e) our method is implemented within the Pyomo platform that enables easy integration with Python-based algorithms for training ML models.

2. Overview of Methods

2.1. AC Optimal Power Flow Problem Formulation

Below, the AC-OPF formulation used in this work is given. Real power, reactive power, voltage magnitude and voltage angle are denoted by p , q , v and θ respectively. N and K denote the sets of all nodes and branches in the system.

$$\text{Min } C(p^g) \quad (1)$$

$$\text{s. t. } \sum_{k(n,m) \in K_n^{\text{out}}} p_{k(n,m)}^f + \sum_{k(n,m) \in K_n^{\text{in}}} p_{k(n,m)}^t - p_n^g + p_n^d = 0, \quad \forall n \in N \quad (2)$$

$$\sum_{k(n,m) \in K_n^{\text{out}}} q_{k(n,m)}^f + \sum_{k(n,m) \in K_n^{\text{in}}} q_{k(n,m)}^t - q_n^g + q_n^d = 0, \quad \forall n \in N \quad (3)$$

$$p_{k(n,m)}^f = g_k v_n^2 - v_n v_m (g_k \cos \theta_{n,m} + b_k \sin \theta_{n,m}), \quad \forall k \in K \quad (4)$$

$$p_{k(n,m)}^t = g_k v_m^2 - v_n v_m (g_k \cos \theta_{n,m} - b_k \sin \theta_{n,m}), \quad \forall k \in K \quad (5)$$

$$q_{k(n,m)}^f = - (b_k + b_k^{\text{sh}}) v_n^2 - v_n v_m (g_k \sin \theta_{n,m} - b_k \cos \theta_{n,m}), \quad \forall k \in K \quad (6)$$

$$q_{k(n,m)}^t = - (b_k + b_k^{\text{sh}}) v_m^2 + v_n v_m (g_k \sin \theta_{n,m} + b_k \cos \theta_{n,m}), \quad \forall k \in K \quad (7)$$

$$v_n^{\text{min}} \leq v_n \leq v_n^{\text{max}}, \quad \forall n \in N \quad (8)$$

$$\theta_{nm}^{\text{min}} \leq \theta_n - \theta_m \leq \theta_{nm}^{\text{max}}, \quad \forall \{n, m\} \in K \quad (10)$$

$$p_n^{\text{min}} \leq p_n \leq p_n^{\text{max}}, \quad \forall n \in N \quad (11)$$

$$q_n^{\text{min}} \leq q_n \leq q_n^{\text{max}}, \quad \forall n \in N \quad (12)$$

$$(p_k^f + q_k^f)^2 \leq (\text{Thermal Limit})^2, \quad \forall k \in K \quad (13)$$

$$(p_k^t + q_k^t)^2 \leq (\text{Thermal Limit})^2, \quad \forall k \in K \quad (14)$$

Eq 1 minimizes the cost of real power at generator buses. Eqs 2-3 represent the nodal power balance derived from Kirchoff's Law. Power flow through each branch is given in Eqs 4-7. Finally, Eqs 8-14 ensure the AC-OPF solution is secure with no contingencies ($N-0$). To check static $N-1$ security of an AC-OPF solution, (v_n, p^g) are set for all generators, (p^d, q^d) for all loads, and (v_n, θ) at the reference bus. Then under each contingency in a given set, the specific element is taken out of the system (e.g., broken line) and Eqs 2-7 are re-solved, without the objective function. If a solution is found that satisfies Eqs 8-14, the solution is secure to that contingency. If this is true for all contingencies, the system is said to be $N-1$ secure. An extensive formulation guaranteeing $N-1$ security requires a scenario for each contingency, leading to a very large non-linear two-step stochastic programming problem. In this work, we train a NN

to capture the $N-1$ security boundary and then embed this NN into the AC-OPF formulation.

2.2. Sampling and Data Set Construction

While one major advantage of this approach is to shift computational costs to offline contingency simulations, it is still necessary to have a methodology that samples data points in an intelligent manner. The input space for the problem is very large with a majority of points far from the security boundary. In order to train a NN model that effectively distinguishes these regions, many points must be generated on the security boundary itself in addition to secure/insecure points. Our sampling algorithm presented below.

Sampling Algorithm for Boundary Points

```

Load_Factors=[LFmin,...,LFmax], Load_Dirs=[LD-0, ..., LD-j]
for LF in Load_Factors:
    sol=Solve_ACOPF( $p^d * LF$ , Power Factor =  $p^d / (q + p^d)^2$ )
    ysec=check_Nmin1(sol) (check N-1 compliance of baseline sol'n)
    for j in Load_Dirs:
        while ysec=True: (continue until no longer N-1 secure)
            sol=solve_acopf( $p^d * LD^{-j} * 1.1$ )
            ysec=check_Nmin1(sol)

```

A set of load factors is constructed between maximum and minimal values that give $N-0$ security. For each of these load factors, an AC-OPF problem is solved given this load profile, and the resulting model is used to check $N-1$ contingencies via the method described in Section 2.1. If the system is $N-1$ secure, the load vector is ramped up by 10% increases until the security check fails. The points directly before and after failure are saved to enhance boundary classification. This is done for all LF's, several load ramping directions (normalized vectors generated via Latin Hypercube sampling), and variable generation patterns (i.e.. variable generator costing parameters) resulting in an increased set of boundary data points. Intermediate solutions that satisfy $N-1$ contingency are saved as secure and points that are infeasible for AC-OPF are saved as insecure to complement the boundary data. This approach provides a set of points that straddle the security boundary, producing a more balanced and targeted training set.

2.3. Neural Network Model Training

Using the constructed data-set, the next stage is to train a NN model that can predict whether a given AC-OPF solution is $N-1$ secure. An important decision is the selection of input variables. The real power setpoints at generators (p^g) are included as they are the ultimate variables in the objective function of AC-OPF, allowing for a direct mapping. Power load (p^d) at all buses are also included since they have a strong correlation with grid security and improve NN accuracy. Thus, the input space of the NN model scales linearly with the number of buses and generators. More input variables may augment NN accuracy in future cases but didn't show strong independent correlation with system security. The Sequential model from Tensorflow's Keras is used as the basis for the feed-forward NN model. It is composed of two hidden layers of 20 nodes each and hyperbolic tangent activation function plus a softmax output layer. Output variable y_{sec} was set to 1 for insecure points and 0 for secure points. ADAM

optimizer was used with a standard cross-entropy loss function. Training continued until validation error plateaued or 4000 epochs were reached.

2.4. Constrained Optimization Formulation

The trained NN can now be thought of as a large non-linear constraint that must be satisfied in order to return an AC-OPF solution that is secure under all contingencies. The NN constraint can be added to the original optimization problem (Eqs 1-14) to encode $N-1$ security.

$$y_{sec} = NN(\{p^d\}, \{p^g\}) \tag{15}$$

$$y_{sec} \leq \alpha \tag{16}$$

Using the trained weights and biases of the NN, Eqs 15-16 are embedded within the AC-OPF model from Egret using Pyomo symbolic equation construction (Knuevan et al 2019). The resulting constraint is fully differentiable and able to be solved with IPOPT. Parameter α is a scalar that controls the constraint conservativeness.

3. Results

Two case studies are used to demonstrate the method: the first (Case 30) contains 30 buses, 6 generators, 21 loads, and 41 branches. The contingency set includes branches {1, 2, 3, 6, and 10}. The next (Case 118) contains 118 buses, 54 generators, 99 loads, and 186 branches. The contingency set includes branches {55,168,169,171,172,173,175}. The contingency set is selected during the sampling process by eliminating contingency events that always make the system insecure or never pose an issue to security. The unconstrained AC-OPF returns a solution that is not $N-1$ secure for 45% of the sampled points for Case 30 and 64% for Case 118. This proves that standard AC-OPF is inadequate for considering grid security. With the NN constraint, the vast majority of insecure operating conditions can be identified and avoided. The control of conservativeness by α can be illustrated through a receiver operator curve (ROC) which shows the false positive and true positive rate of the NN classifier. Figure 1 shows the ROC results for the NN classifier with the full training set and for a model with half the training data.

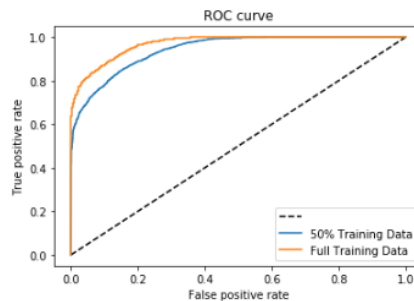


Figure 1. Receiver Operator Curve (ROC) Results (Case 118)

Again, there is an expected trade-off between conservativeness and accuracy. While a high true positive rate allows us the most confidence in grid security, the accuracy trade-off would be important to tailor to the specific grid application. For an α value of 0.5, model accuracy is 89%, much better than base case AC-OPF. The results for Case 30 show nearly identical behavior on the ROC. Another important factor to consider is

computational time of the proposed approach, where it distinguishes itself from the extensive formulation. Literature values for Case 118 take ~400 sec to run SC AC-OPF (Kang et al 2015) on a single processor of Sandia's Red Mesa supercomputer. Our NNSC AC-OPF approach never exceeds **15 sec for all points, with an average of 7 sec.**

One more important thing to consider is the effect of the security constraint on the objective function. We can quantify the cost of incorporating the NNSC constraint by considering the change in objective function value compared to the base case AC-OPF. In order to distinguish meaningful cost increases (to preserve security) from an overly conservative NNSC, models are tested with points that are base case secure and insecure. Increases in the objective function are negligible for over 70% of secure test points and the majority of remaining points constitute increases of less than 5%. This shows that α of 0.5 does a good job of not being overly conservative with respect to security. For insecure inputs, the objective function increases up to 50% in order to meet the NNSC constraint. These solutions come with far greater confidence about security under contingency but also come at an economic cost. Further comparisons with the extensive formulation and cost/risk strategies would be needed to definitively set α in real-world applications.

4. Conclusions

Power grid optimization and security is a vital area of research that can help save money, avoid black-outs, and promote the integration of non-traditional energy sources onto the grid. Using detailed simulations, it is possible to optimize very accurate models of power grids using non-convex/non-linear formulations such as AC-OPF. Adding security constraints for all N-1 contingencies to this problem is computationally challenging and extensive formulations are intractable in large-scale grids. This work addresses this challenge by approximating security constraints with a NN model that can be trained offline and then embedded within a Pyomo-based model as a single non-linear constraint. Results show that this approach can accurately estimate the security boundary and provide a framework to balance optimality and security of system set points. The NLP formulation allows for computationally tractable solutions even for large grids.

Disclaimer: Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

- B. Knueven et al., 2019, Egret v.01(beta), Sandia National Lab (SNL-NM), Python
- M.B. Cain, R.P. Oneill, A. Castillo, 2012, History of Optimal Power Flow and Formulations, Federal Energy Regulatory Commission, p:1-36
- F. Capitanescu, J.M. Ramos, P. Panciatici, D. Kirschen, A.M. Marcolini, L. Platbrood, L. Wehenkel, 2011, State-of-the-art, challenges, and future trends in security constrained optimal power flow, Electric Power Systems Research.
- A. Gulli, S. Pal, 2017, Deep learning with Keras, Packt Publishing Ltd

- V.J. Gutierrez-Martinez, C.A. Canizares, C.R. Fuerte-Esquivel, A. Pizano-Martinez, X. Gu, 2011, Neural-Network Security-Boundary Constrained Optimal Power Flow, IEEE Transactions on Power Systems.
- W.E. Hart, C.D. Laird, J.P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson, J.D. Siirola, 2017, Pyomo – Optimization Modeling in Python. Second Edition. Vol. 67. Springer.
- J. Kang, A. Jayaraman, C. Laird, M. El-Halwagi, V. Sarin, 2015, An Efficient Interior-Point Decomposition for Parallel Solution of Large-Scale Nonlinear Problems with Significant Variable Coupling, Doctoral Thesis, Texas A&M
- R. O'Neill, A. Castillo, M.B. Cain, 2012, The IV Formulation and Linearizations of the AC Optimal Power Flow Problem, FERC staff technical paper
- X. Pan, M. Chen, T. Zhao, S.H. Low, 2020, DeepOPF: A Feasibility-Optimized Deep Neural Network Approach for AC Optimal Power Flow Problems, Under Review.
- A. Velloso, P. Van Hentenryck, 2020, Combining Deep Learning and Optimization for Security-Constrained Optimal Power Flow, Under Review.
- A. Venzke, D. Molzahn, S. Chatzivasileiadis, 2019, Efficient Creation of Datasets for Data-Driven Power System Applications, Electrical Engineering and Systems Science: Systems and Control, Under Review.
- A. Venzke, D.T. Viola, J. Mermet-Guyennet, G.S. Misyris, S. Chatzivasileiadis, 2020, Neural Networks for Encoding Dynamic Security-Constrained Optimal Power Flow to Mixed-Integer Linear Programs, Under Review.

The impact of CO₂ pricing in SC Resilience – An optimisation model

João Pires Ribeiro^{a,*} and Ana Barbosa-Póvoa^a

^a*Centre for Management Studies, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal*
pires.ribeiro@tecnico.pt

Abstract

Environmental and resilience are currently vital concerns when managing Supply Chains (SC). Doing so in a competitive setting, with diverse stakeholders' perspectives and dealing with unforeseen disruptive events is challenging. Today, this is a reality where, with the actual COVID-19 pandemic, supply chains face reduced demand and stoppages at different levels, calling for the urgent need to invest in designing and planning resilient SC. But resilience must not leave apart other vital goals as is the environmental goal, which nowadays requires special attention. This is especially critical in the process industries where environmental concerns are often at stake. We address this challenge in the current work by representing the cost associated with CO₂ emissions, considering the EU emissions trading system (EU ETS). This system makes the cost associated with emissions a variable value attributed by the market. A Mixed Integer Linear Programming model (MILP) is here presented which allows to understand the supply chain resilience of different supply chain structures. This is done with the objective of maximising the Expected Net Present Value (ENPV) while facing disruptions, and the presence of uncertainty in demand is considered. The results show that our model can help decision-makers to create resilient SC with good environmental behaviour and without compromising financial results.

Keywords: Supply Chain Resilience, CO₂ Emissions, Modelling

1. Introduction

Supply chain (SC) Resilience comes as a recent offspring of Supply Chain Management (SCM), with most of the attention only being given in recent years. Even with a multitude of definitions, one can define such behaviour as follows: "a resilient supply chain should be able to prepare, respond and recover from disturbances and afterwards maintain a positive steady-state operation in an acceptable cost and time" (Ribeiro and Barbosa-Póvoa, 2018). It is a field of study with research gaps to explore (Kamalahmadi and Parast, 2016), thus adding knowledge and simultaneously know-how on SCM. One of these research gaps is related to the inclusion of green concerns with economic goals (Fahimnia and Jabbarzadeh, 2016).

The European Union (EU) has set an objective to reduce 80% of CO₂ emissions by 2050 (having 1990 levels as a reference value). To achieve this goal, one of the policy frameworks underway is the European Union Emissions Trading System (EU ETS).

Through this framework, also known as the "cap-and-trade" system, a maximum limit is imposed on the amount of greenhouse gases emitted. This limit is reduced over time so that total emissions are continuously decreasing. Within this set, organisations can buy emission allowances, which, as the name indicates, allow for a certain amount of CO₂ emissions. If the organisation is able to stay under the estimated emissions level, it can sell and trade these allowances or keep them for future needs. Allowances have become commodities, and auctioning is the default method for allocating them. Its value has been rising over the last year, currently reaching +153% of the reference value that existed until mid-2018, showing the utmost importance of considering CO₂ costs in the supply chain decision-making process.

It is then very relevant to explore the coexistence of environmental and resilience concerns when designing and planning a SC (Jabbarzadeh et al., 2018), even more in process supply chains where the environmental impacts are often a considerable concern (Barbosa-Póvoa et al., 2018). This study comes in a time of great pertinence to practitioners that are now making long-standing decisions in their SC, with the changes that arrive from new trends that complement the sustainability as is adaptability to new markets trends at a very fast pace. This type of decisions tends to be complex, with significant investments and implemented in the long term, adding special necessity for decision aiding tools regarding strategic and tactical decisions.

This paper addresses this need and develops a strategic optimisation model that explores the relationship between Resilience and Green goals when designing and planning SC. A relationship that has failed to reach a consensus on its merits and benefits for companies. If on the one hand, sustainability tends to reduce flexibility and redundancy, crucial for SC Resilience, on the other circular economy principles can add responsiveness to the SC. Here we aim at exploring the relationship between the two concepts and test if sustainability and resilience are compatible in SC.

A European SC case study is studied, applying the new model where resilience and environmental concerns are considered, apart from the common SC profit. The strategic decisions regarding SC's design and planning are modelled in three periods of five years, leading the model to a 15-year time-span. Useful managerial insights are obtained to help in the process of acknowledging the full impacts of disruptive events in the fitness of operations under different scenarios and consequently, how much can proactive management contribute to a competitiveness advantage.

2. Problem Characteristics and Model description

The SC to be modelled is generic and involves five different echelons: Raw Materials Suppliers, Plants, Warehouses and Markets and with the possibility to outsource the production of final products and provide direct shipment to warehouses (see Figure 2). It allows for direct flows, transshipment in plants and warehouses, direct flows between warehouses and markets, and it also includes the possibility of reverse flows. Three types of reverse flows are considered: Non-conforming products that go from the markets to warehouses to be repacked and introduced in the forward SC; End-of-life products that are collected, when the expected lifetime of the product is reached, and are refurbished to be introduced in the forward SC; The products that fail to be refurbished are sent for disposal. Two different cases are considered: Case A, with only forward flows (in grey),

and Case E a Closed-Loop SC, with forward and reverse flows (in grey and orange).

When studying resilience, there is the need to define disruptions on which to test the model, an activity that can prove itself challenging. One must define disruptions that represent a broad set of possible disruptions without falling into the error of defining disruptions that are too specific. Disruptive events are associated with some kind of unknowns in terms of risks. By choosing a type of disruptive events, it is possible to generate a representative set of disruptions that meet the objective of providing generic, but useful, results. These disruptions are then studied on each run of the model, as will be detailed below.

The environmental impacts are assessed using the Life Cycle Assessment (LCA) approach. The system boundary defined is cradle to gate and the functional units considered are one kilogram of product and one kilometre travelled. The SimaPro software was used to determine the climate change impact (in CO₂ equivalent) from: truck transportation in forward flows (per kg.km); truck transportation in reverse flows (per kg.km); final product production (per kg of product).

In combination with the cost of 1 Kg of CO₂, these parameters are used in the model to quantify the SC's environmental impact.

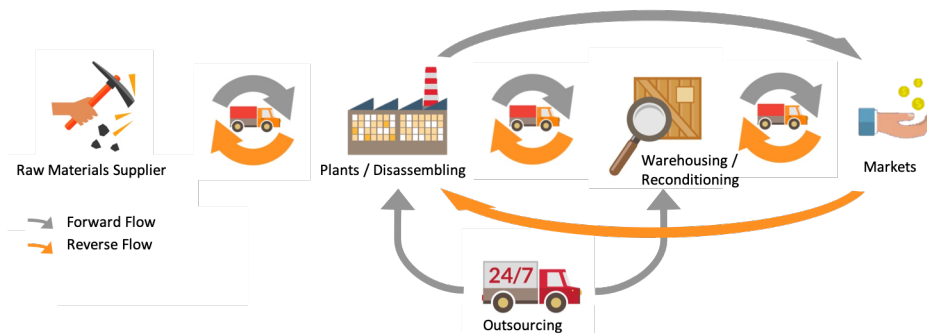


Figure 1: A schematic characterisation of the SC, with its echelons and allowed flows.

Taking as point of origin the work published by Cardoso et al. (2015), the model here presented adapts to create a new model to fulfil the research objectives, described in the introduction, while maintaining the possibility of comparing and discussing the current results with previous results. The model possesses relevant characteristics. It includes a closed-loop SC, with forward logistics and reverse logistics and entails uncertainty, relevant when considering the time span needed for tactical and strategic decisions. Uncertainty is considered in the SC demand and on disruptions. A scenario tree is constructed, combining these two sources of uncertainty. Demand variability is introduced by generating a scenario on each period from a set of three possibilities (Pessimistic, Realistic or Optimistic). The probabilistic nodes are then combined with the variability from the disruption that can only assume two options, or it occurs, or it does not occur in a specific time period to consider. With this, each scenario probability is given by the path's probabilities, with all stages, between the root node and each final leaf node. The proposed model relies on maximising Expected Net Present Value (ENPV), also considering environmental costs, under a set of disruptions and uncertainty.

$$\max \text{ENPV} = \sum pb_s \times \text{NPV}_s \quad (1)$$

Expected Net Present Value (ENPV) is calculated considering each scenario probability (pb_s), in the scenario tree, using cash flows and an interest rate. The net earnings come from the difference between the income (Sales revenue) and the costs (the refunds from non-conforming products, purchases, operational, inventory, transportation, depreciation of invested capital considering taxes, and the environmental costs defined in Equations 2 and 3. Two environmental cost equations are considered, one to account for transportation, Equation 2, and another to account for the production of final products, Equation 3. The costs of the environmental impact of transport are given by the summation of emissions (Kg.km CO₂ eq) associated with forward (*envfactorforward*) and reverse flows (*envfactorreverse*) multiplied by the quantities shipped in forward (*QPL*) and reverse flows (*QNC* for non conform products and *QEL* for end-of-life products) multiplied by the cost of emitting one kg of CO₂ (*CO₂Cost_t*) and the distance between the two locations (*FIPL_{v,w}*), that is given by Equation 2:

$$\begin{aligned} \text{TranspCO}_2 = \sum_t (\text{CO}_2\text{Cost}_t \cdot (& \sum_{s,v,w,p,t} (\text{QPL}_{v,w,p,s,t} \cdot \text{FIPL}_{v,w} \cdot \text{envfactorforward}) + \\ & \sum_{s,v,w,p,t} (\text{QNC}_{v,w,p,s,t} \cdot \text{FIPL}_{v,w} \cdot \text{envfactorreverse}) + \\ & \sum_{s,v,w,p,t} (\text{QEL}_{v,w,p,s,t} \text{FIPL}_{v,w} \cdot \text{envfactorreverse}))) \quad (2) \end{aligned}$$

The cost of production emissions is given by the summation of emissions when a specific product p is produced multiplied by the cost of emitting one kg of CO₂, which is given by Equation 3. Emissions for reverse product flows are expected to be lower than those that are produced directly from raw materials:

$$\text{CO}_2\text{prodtot} = \sum_t (\text{costCO}_2t \cdot \sum_p \text{CO}_2\text{eachp}_p) \quad (3)$$

The opinions diverge on the behaviour of the price of CO₂ emission in the long term since there are two sets of factors that are plausible to happen, and that affects the price of CO₂ in different directions. If in one hand, it is expected a continuous effort to reduce CO₂ emissions and therefore the public efforts to force companies to reduce CO₂ emissions. On the other hand, these efforts can lead to the decarbonisation of the economy that drastically reduces demand for CO₂ allowances. Considering these effects, three scenarios were generated:

Stable scenario - The CO₂ prices are kept constant at 27,1e/ton

Up and Up scenario - The CO₂ prices are expected to rise in each time period.
t1=27.1e/ton t2=40e/ton t3=50e/ton

Up and Down scenario - The CO₂ prices increase in the shorter term but go down again in the longer term. t1=27.1e/ton t2=40e/ton t3=27.1e/ton

Apart from the above objectives function the model involves a set of constraints (entity capacity constraints, transportation constraints, technology constraints) so as to correctly and thoroughly model the SC. This formulation can be explored in further detail in (Cardoso et al., 2013).

3. Case Study

The model is applied to a European SC that is interested in the improvement of its network design. The initial network relies on one plant in Hamburg with the possibility of opening new plants in Bilbao and/or in Milan. Each one of the new plants has a set of suppliers associated. There is also the possibility to open new warehouses or to upgrade the already existing ones. The activities performed by each entity are categorised as production technologies, that result in intermediate products; and assembling technologies, that combine products and result in the final products. Reverse logistics is implemented by disassembling technologies that can use final products (as defined in the Model description) to generate intermediate products.

In this work, the SC will be tested under a reference scenario, where no disruption occurs, and three types of disruptions frequently used and as defined by Rice and Caniato (2003):

Disruption 1: Supply - 100% decrease in the production capacity of the most important plant in time period 2, located in Milan.

Disruption 2: Production - The most important raw material suppliers have their supply suspended in time period 2. These suppliers are located in Birmingham, Badajoz, Marseille, Ljubljana and Lausanne.

Disruption 3: Transportation - The flows that carry the highest quantity of products are stopped in time period 2. In this case study that represents the flows from the factory in Milan to warehouses in Munich, Portsmouth and Bologna.

The disruptions are implemented in the model by not allowing flows between the identified entities, as described above.

4. Results and Discussion

The results of our model related to the ENPV are present in Table 1. Disruption 2 is the one that most affects the SC, as it has the lowest ENPV results (around 0,30E+07€) for both chain configurations as well as for the three cost evolution scenarios of CO₂. On the other hand, disruption 3 is the one that least affects the system, with results very similar to the scenario without disruption in the case of forward SC, Case A. In the closed-loop SC, Case E, there is a decrease in ENPV. However, it ends up having economic results always greater than Case A.

Table 1: Expected Net Present Value (ENPV) for all scenarios and disruptions

ENPV (€)	Case A - Forward Supply Chain			ENPV (€)	Case E - Closed-Loop Supply Chain		
	CO ₂ emissions cost scenarios				CO ₂ emissions cost scenarios		
	Stable	Up & Up	Up & Down		Stable	Up & Up	Up & Down
No disruption	1,74E+07	1,32E+07	1,42E+07	No disruption	1,86E+07	1,51E+07	1,61E+07
Disruption 1	1,49E+07	0,99E+07	1,12E+07	Disruption 1	1,74E+07	1,33E+07	1,42E+07
Disruption 2	1,40E+07	0,89E+07	1,03E+07	Disruption 2	1,72E+07	1,35E+07	1,41E+07
Disruption 3	1,74E+07	1,32E+07	1,42E+07	Disruption 3	1,77E+07	1,43E+07	1,52E+07

From Table 2, a clear conclusion is that Forward SC (Case A) have a higher CO₂ cost than SC with circularity principles (Case E). In cases where there is an increase in cost, this difference is about 1E06€ when there is no disruption and greater than 2E06€ in disruptive scenarios.

Table 2: Total CO_2 cost for all scenarios and disruptions
 Case A - Forward Supply Chain Case E - Closed-Loop Supply Chain
CO₂ emissions cost scenarios *CO₂ emissions cost scenarios*

CO_2 Prod (e)	Stable	Up & Up	Up & Down	CO_2 Prod (e)	Stable	Up & Up	Up & Down
No disruption	1,86E+06	6,15E+06	5,13E+06	No disruption	1,46E+06	5,38E+06	4,50E+06
Disruption 1	1,91E+06	8,07E+06	6,54E+06	Disruption 1	1,54E+06	5,51E+06	4,66E+06
Disruption 2	1,90E+06	8,03E+06	6,51+06	Disruption 2	1,50E+06	5,53E+06	4,86E+06
Disruption 3	1,86E+06	6,16E+06	5,13E+06	Disruption 3	1,41E+06	5,21E+06	4,26E+06

From the interpretation of our results, it can be said that the environmental concern approach compensates, even more, when the price of CO_2 is expected to continue to rise. That is seen for Case E, not only is the return greater (ENPV), but the difference with the scenario without disruption is less. Thus, the variability introduced by disruptions is also reduced.

The best result of Case E is mostly due to the fact that the environmental concern translates into a reduction in long-distance transport and a greater investment in facilities close to each other. This measure will increase resilience in the sense that it increases redundancy and flexibility, with more production facilities and more possible flows. The circular economy principles are also crucial, the production of final products from the reuse of end-of-life products leads to less dependence of factories on raw materials suppliers while reducing the total CO_2 associated with a given product.

Case E brings not only a higher return on ENPV but also a lower CO_2 cost, which provides a fantastic opportunity for companies to take advantage of environmental concerns/obligations to build a more resilient SC.

References

- A. P. Barbosa-Póvoa, C. da Silva, A. Carvalho, jul 2018. Opportunities and challenges in sustainable supply chain: An operations research perspective. *European Journal of Operational Research* 268 (2), 399–431.
- S. R. Cardoso, A. P. Barbosa-Póvoa, S. Relvas, A. Q. Novais, 2015. Resilience metrics in the assessment of complex supply-chains performance operating under demand uncertainty. *Omega* 56, 53–73.
- S. R. Cardoso, A. P. F. D. Barbosa-Póvoa, S. Relvas, 2013. Design and planning of supply chains with integration of reverse logistics activities under demand uncertainty. *European Journal of Operational Research* 226 (3), 436–451.
- B. Fahimnia, A. Jabbarzadeh, 2016. Marrying supply chain sustainability and resilience: A match made in heaven. *Transportation Research Part E: Logistics and Transportation Review* 91, 306–324.
- A. Jabbarzadeh, B. Fahimnia, F. Sabouhi, 2018. Resilient and sustainable supply chain design: sustainability analysis under disruption risks. *International Journal of Production Research* 56 (17), 5945–5968.
- M. Kamalahmadi, M. M. Parast, 2016. A review of the literature on the principles of enterprise and supply chain resilience: Major findings and directions for future research. *International Journal of Production Economics* 171, 116–133.
- J. P. Ribeiro, A. Barbosa-Póvoa, 2018. Supply chain resilience: Definitions and quantitative modelling approaches – a literature review. *Computers & Industrial Engineering* 115, 109 – 122.
- J. B. Rice, F. Caniato, 2003. Building a secure and resilient supply network. *Supply Chain Management Review*, V. 7, No. 5 (Sept./Oct. 2003), P. 22-30: Ill.

Acknowledgements

FCT and P2020 under the project PTDC/EGEOGE/28071/2017, Lisboa -01.0145-Feder-28071 and SFRH/BD/148499/2019

Machine learning-based approach to identify the optimal design and operation condition of organic solvent nanofiltration (OSN)

Changsu Kim, Chanhee You, Do Thai Ngan, Minseong Park, Dongjun Jang, Sungju Lee, and Jiyong Kim*

Department of Energy and Chemical Engineering, Incheon National University (INU), Incheon, Republic of Korea

jkim77@inu.ac.kr

Abstract

Organic solvent nanofiltration (OSN) is one of the most anticipated separation technologies that provides wide-ranged industrial applications such as solvent recovery, solute concentration, and diluent separation. Despite of technical merits of the OSN technology, the numerous characteristics and perplexing nonlinearity on the OSN system have been a critical obstacle for understanding the governing principles, thereby prohibiting practical deployments. Recently, machine learning (ML) based approaches have been widely used for the modelling, discovery and optimization of complex design problems in chemical engineering area such as catalysis, electrochemistry and physicochemical systems. Therefore, this study aims to develop a new ML-based approach for modelling and optimizing the design scheme and operating condition of the OSN system. By collecting commercial OSN data through literatures reviews, the major descriptors for the prediction of the OSN membrane, such as MWCO, solute mole weight, solute concentration, solvent parameter, temperature, pressure, flux, were defined. We then screened noises and outliers of the collected data to ensure a high and consistent density and uniqueness. Support vector machine (SVM) was implemented as a prediction models to simulate the OSN performance and identify the optimal conditions as well as the process scheme. As a result, the optimal operation strategies (i.e., pressure, temperature and solvent and solvent types) were analyzed to meet the targeted specification of the OSN system (mass flux and rejection rate). The proposed ML-based approach can promote a real-world OSN application by reducing a number of time-consuming and expensive experiments for establishing OSN design and operation strategy.

Keywords: Organic Solvent Nanofiltration, Machine learning, Separation, Optimization

1. Introduction

Organic solvent nanofiltration (OSN) is one of promising separation technologies that provide economic and technical benefits compared to existing technologies (e.g., distillation) in petrochemical industries. To accelerate the practical deployment of the OSN technology, A number of studies have been reported in literature from a new material discovery to operation research in real industries.

To facilitate the industrial acceptance and real deployment of novel OSN membrane, precise modeling methodologies should be accompanied, such as conventional the principles of thermodynamics and transport phenomenon. However, due to a huge number of design combinations with different solvents, solutes and membrane materials, predicting performance using first principles of thermodynamics and transport phenomenon requires a number of time-consuming and expensive experiments.

Machine learning (ML) is one of artificial intelligence methods, which have widely used for modeling and predicting multi-dimensional and quite complex systems. Screening out possible metal-organic frameworks (MOF) structure and predicting lifespan of Li-ion battery are salient examples of ML technique employment. Likewise, the emergence of ML techniques on the OSN performance prediction problems has brought about advancements and has become relatively accessible by other disciplines.

In this study, a new machine learning-based approach is developed to identify the optimal design strategy and operation condition of the OSN membrane. 884 OSN membrane datasets are collected and curated to define main descriptors of OSN membrane. Support vector machine (SVM) is then employed to predict the performance of OSN membrane using key descriptors affecting the membrane permeance and rejection, which were obtained through principal component analysis (PCA).

2. Method

2.1. Overview

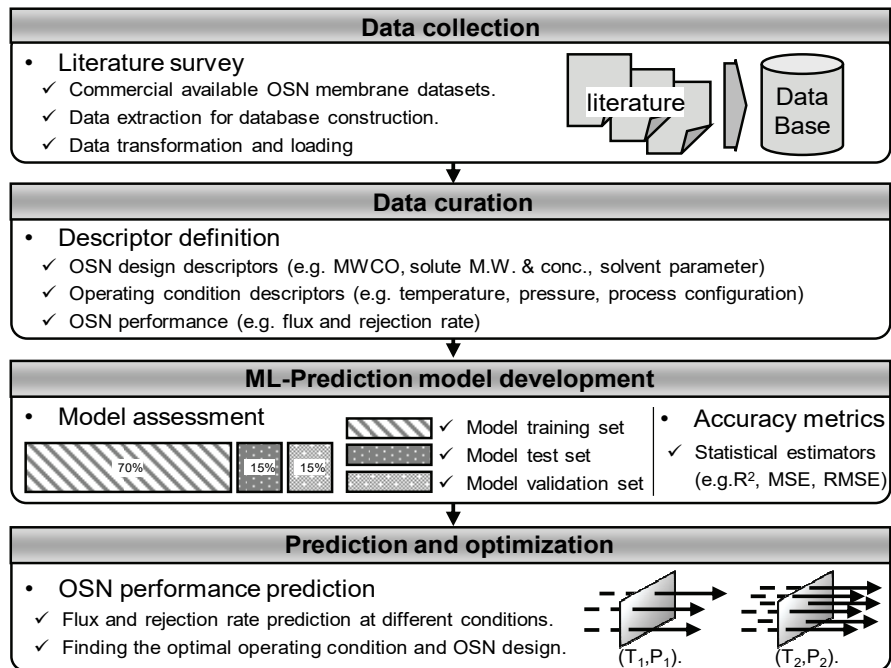


Figure 1: Overview of machine learning-based approaches on the OSN membrane performance prediction and optimization.

The main steps of this study are illustrated in Figure 1. The commercial OSN membrane data is collected from literatures. Datasets include OSN membrane characteristics such as material properties, operating condition, and OSN membrane performances. Among many OSN characteristics, critical OSN descriptors (e.g. molecular weight cut-off, solute mole weight, solute concentration, characteristic solvent parameter, temperature, pressure, process configuration, flux, and rejection) are defined from data curation step. Then, datasets are separated into training sets, validation sets, and test sets in order to verify the capability of prediction model. In this research, prediction model is developed by using the SVM. For prediction capability evaluation, three statistical estimators such as R^2 , MSE, RMSE are used. Finally, from predicting OSN performances (flux and rejection) at different conditions, the optimal operating condition and OSN design are qualitatively analysed by using the PCA.

2.2. Data description

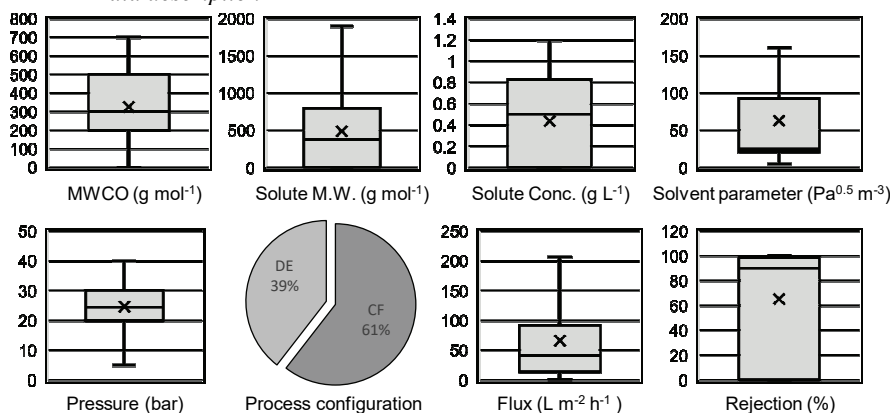


Figure 2: Distribution of OSN descriptors.

The distribution of OSN descriptors is illustrated in Figure 2. Descriptors are varied as followed: MWCO (150-1000 g mol^{-1}), solute mole weight (0-1900 g mol^{-1}), solute concentration (0-1.2 g L^{-1}), characteristic solvent parameter (5-160 $\text{Pa}^{0.5} \text{m}^{-3}$), pressure (20-60 bar), process configuration (dead-end, cross-flow), flux (1-416 $\text{L m}^{-2} \text{h}^{-1}$), rejection (0-100 %). For temperature, most test were experimented at 25°C.

2.3. Support vector machine (SVM)

The SVM is a prominent ML technique for regression and classification analysis. By representing datasets in multidimensional descriptor space, the regression hyperplane is created. The mathematical model and constraint are described in Eqns. 1 and 2, respectively.

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \varepsilon \sum_i (\alpha_i + \alpha_i^*) + \sum_i y_i (\alpha_i + \alpha_i^*) \quad (1)$$

$$0 \leq \alpha_i, \alpha_i^* \leq C, \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, i = 1, \dots, l \quad (2)$$

To ensure a high accuracy of the prediction model, the hyperparameter is optimized by using a grid-search algorithm. Tables 1-3 compare the accuracy of major metrics by the SVM and the HO-SVM (Hyperparameter-optimized SVM).

Table 1: Accuracy comparison between SVM and HO-SVM on training set

Training set	R ²	MSE	RMSE
SVM (Flux)	0.939	155.120	12.454
HO-SVM (Flux)	0.985	69.135	8.314
SVM (Rejection rate)	0.959	77.001	8.775
HO-SVM (Rejection rate)	0.966	58.577	4.671

Table 2: Accuracy comparison between SVM and HO-SVM on validation set

Validation set	R ²	MSE	RMSE
SVM (Flux)	0.934	388.648	19.714
HO-SVM (Flux)	0.928	310.088	17.609
SVM (Rejection rate)	0.903	176.58	13.288
HO-SVM (Rejection rate)	0.940	107.286	10.357

Table 3: Accuracy comparison between SVM and HO-SVM on test set

Test set	R ²	MSE	RMSE
SVM (Flux)	0.914	397.255	19.931
HO-SVM (Flux)	0.923	305.099	17.467
SVM (Rejection rate)	0.906	175.139	13.234
HO-SVM (Rejection rate)	0.960	66.534	8.156

2.4. Principal component analysis (PCA)

PCA is a dimension reduction technique that represents and visualizes multi-dimensional data. From simplifying multi-dimensional descriptor space, the most important information could be extracted by PCA. In this work, the predicted performances of OSN membrane is visualized and the effect of descriptors is analysed by PCA technique.

3. Result

Flux and rejection of the OSN membrane are predicted at conditions of MWCO (150-300 g mol⁻¹), solute mole weight (300 g mol⁻¹), solute concentration (0.1-1.0 g L⁻¹), characteristic solvent parameter (0-100 Pa^{0.5} m⁻³), pressure (10-40 bar), temperature (25°C), and process configuration (cross-flow). In the PCA of both flux and rejection, four dimensions that cover 96% and 87 % data variance are obtained.

Table 4: PCA variance (dimension coverage) of each principal component on flux.

Flux	PC1 (%)	PC2 (%)	PC3 (%)	PC4 (%)
PCA variance	0.3695	0.2	0.2	0.2

Table 5: PCA variance (dimension coverage) of each principal component on rejection.

Rejection rate	PC1 (%)	PC2 (%)	PC3 (%)	PC4 (%)
PCA variance	0.2723	0.2	0.2	0.2

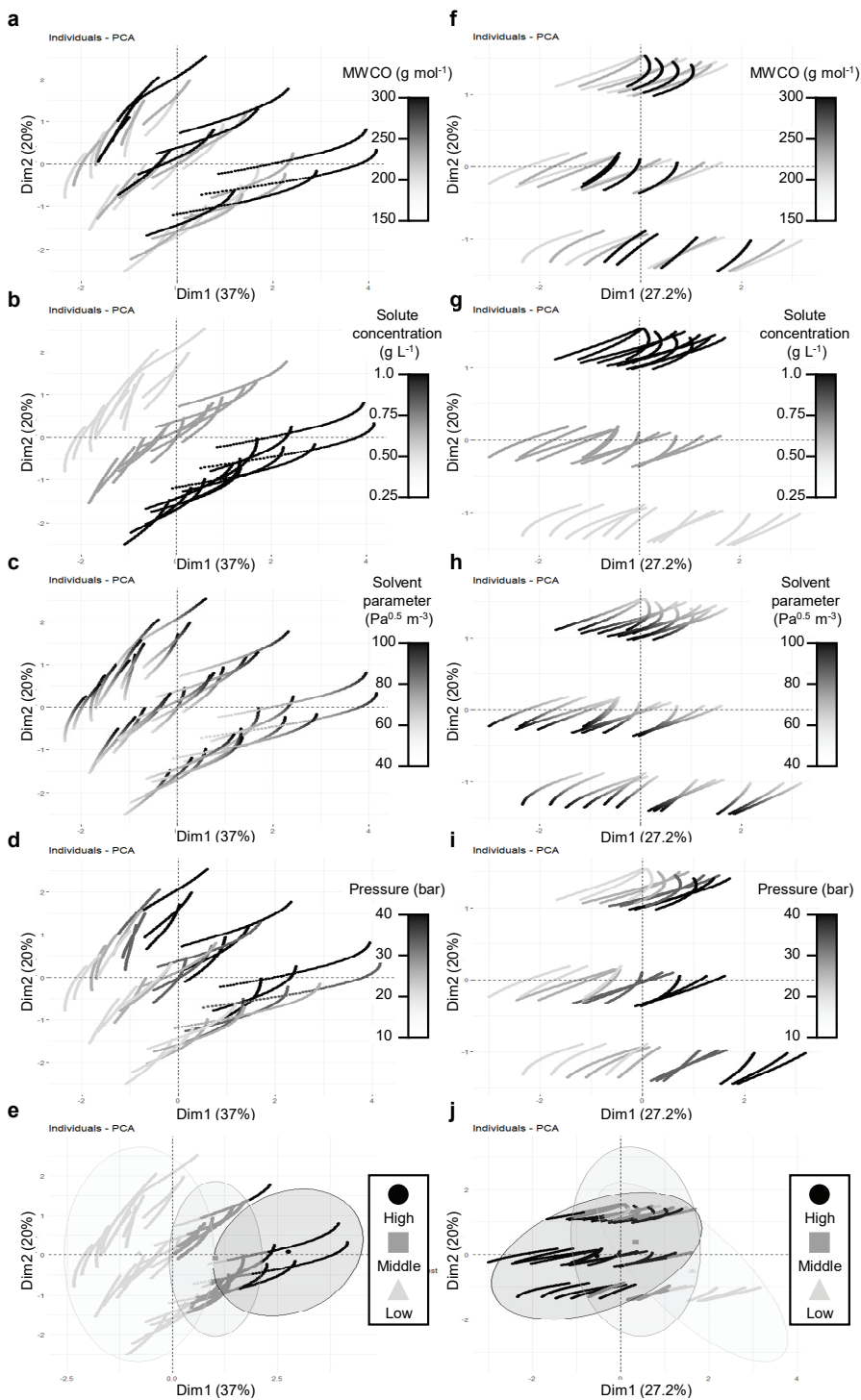


Figure 3: PCA results on OSN membrane descriptors; (a-d) and (f-i) for MWCO, solute concentration, solvent parameter, and pressure, respectively; (e) and (j) for flux and rejection, respectively. (a-d) representing (e) and (f-i) representing (j).

The effect of descriptors on OSN membrane performance is illustrated in Figure 3. In Figure 3 e and j, the flux and rejection rate are described and clustered. The level of OSN membrane descriptors is described in Figure 3 a-d and f-i, which are represented in flux and rejection rate dimension, respectively.

The flux and rejection rate are described in three different levels, which are divided by distribution of datasets. For flux, levels are high (more than $200 \text{ L m}^{-2} \text{ h}^{-1}$), middle ($100\text{--}200 \text{ L m}^{-2} \text{ h}^{-1}$), and low (less than $100 \text{ L m}^{-2} \text{ h}^{-1}$). As well as in rejection rate, levels are high (more than 90%), middle (80-90%), and low (less than 80%).

As shown in Figure 3 e, the clusters representing the flux level show that high flux datasets could be distinguishable from middle to low flux datasets. In case of the high-flux level, OSN membrane tended to have high MWCO (Figure 3 a), high solute concentration (Figure 3 b), and high pressure (Figure 3 d).

For the rejection rate in Figure 3 j, three different clusters are represented to analyse the effect of the major descriptors on the rejection rate. In particular, the high-rejection cluster was highlighted to distinguish the core OSN characteristics leading to the high rejection rate over the middle-and low-rejection rate clusters. On the other hand, it is also found that the OSN system with a low-concentration solute needs to be operated at low pressure to avoid low rejection rate.

4. Conclusions

The performances (flux and rejection rate) of the OSN membrane were predicted with the HO-SVM to investigate and understand the behaviour of the OSN system. In this study, using 9 main descriptors, including material properties and operation conditions, HO-SVM model predicted the OSN performance at different design specifications and operation conditions. As a result, it was found that i) a high-flux OSN could be accomplished by applying high MWCO, high solute concentration, and high pressure and, ii) the operation strategies to ensure high rejection rate in the OSN membrane (e.g., low-pressure operation for low-concentration solute) was analysed. Based on this study, the fundamental knowledge and understanding to design and operate the OSN membrane was preliminary discussed. As a future work, more detailed and practical ML-based guideline for the OSN membrane design and operation will be investigated by pointing out the optimal operation strategies as well as design solutions.

References

- L. S. White, A. R. Nitsch, Solvent recovery from lube oil filtration with a polyimide membrane, *J. Membr. Sci.*, 197 (2000), pp. 267-274
- S. Han, H. T. Wong, A.G. Livingston, Application of organic solvent nanofiltration to separation of ionic liquid and products from ionic liquid mediated reactions., *Chem. Eng. Res. Des.*, 83 (2005), pp. 309-316
- H. Lu, K.N. Plataniotis, A. N. Venetsanopoulos, MPCA: multilinear principal component analysis of tensor object, *IEEE Trans. Neural Networks*. 19 (2008) 18-39.

Development of System Dynamic Model for Sustainability driven Technology Adoption in Indian Transport Sector

Nandita Saraf^a Yogendra Shastri^{a,*}

^aDepartment of Chemical Engineering, Indian Institute of Technology, Bombay, 400076, India

yshastri@che.iitb.ac.in

Abstract

Indian transport sector is the second largest greenhouse gas (GHG) emitter after power sector. For climate change mitigation, transition to more sustainable transport options such as ethanol blended petrol (E-85) and electric vehicle (EV) is necessary. However, to understand how their adoptions evolve among privately owned vehicles, we have developed a system dynamic model. System dynamics captures the underlying causal relation among the variables of the system and hence helps in understanding its behaviour. Causal interactions, feedback loops and delays have been identified in the system. Negative feedback loops connecting fuel supply to its demand, predict their apparent prices. Fuel price adds to the ownership cost of respective vehicle, which is used in logit model to determine respective annual purchase demand. Effect of inconveniences associated with adoption of EV and fluctuation in ethanol supply have been incorporated. Model predicts the penetration of E-85 and EV up-to 32.5 % of total private vehicle stock by 2050. As a consequence, 5.2 billion tons of carbon dioxide emissions can be saved. Market dynamics seems to be insufficient in encouraging high penetration of renewable options. Stronger incentives and development of better infrastructure could improve the situation.

Keywords: System dynamics, logit model, causal-loop diagram

1. Introduction

India is world's third-largest GHG emitter and second most populous country. India's total GHG emission in 2019 were 132 million tons of CO₂e (Joshi& Chen, 2020). Indian government has pledged to reduce emission intensity by 33-35% by 2030 compared to its 2005 level according to the Paris agreement target. Transition to renewable energy sources for transport sector has been recognized as an essential move to meet this pledge. In this regard, government has set a target of achieving 20% blending of ethanol in gasoline and 30% penetration of electric vehicles in new sale by 2030. However, penetration of new transport options among existing once seems difficult owing to large size and complex nature of transport sector. In order to understand how adoption of these options will develop in future, it is necessary to explore the cause-and-effect interaction among variables, which thus defines the dynamics of the transport sector. System dynamics (SD) modelling is a promising approach to perform such studies. This

methodology has been used previously to understand transition from ICE to EV (Pasaoglu et al., 2016; Vilchez et al., 2013). Also it has been used to test the effectiveness of various energy policies on transport sector (Menon & Mahanty, 2015). This has motivated the use of SD modelling in present research work. This research aimed at understanding the adoption of renewable transport options, i.e., E-85 and EV among private vehicles. Model could be used to study the implication of penetration of these options on GHG emissions and fuel prices. Also, this study could help in designing policies in order to meet desirable sustainable development goals.

2. Model development

System dynamics helps in analysing the behaviour of complex socio-economic systems due to underlying interactions which governs the dynamics (Forrester, 1994). To develop a SD model, first step is to develop a causal loop diagram (CLD). It depicts causal relation that exist between variables of the system. This requires a good understanding of how two variables interact. The next step is to formulate mathematical equations which go as input to the model. The specific functional form of equation is decided based on either causal relation or historical trend. Equation parameters are estimated through fitting historical data to the proposed function. SD model developed in this work focuses on private transport vehicles in India. Important assumptions of the model are: -

1. Ownership cost is the only decisive factor in determining the purchase demand for various vehicle options
2. The number of vehicles and distance travelled per vehicle are always related to per capita GDP
3. Fuel price is always related to fuel supply-demand dynamics
4. Electricity price is considered as independent of electricity demand coming from transport sector

The CLD of the model is shown in Fig.1. Population and GDP per capita positively influence the demand for car and two-wheeler, which is divided among various available options. Annual demand of vehicle adds to vehicle stock, which increases the fuel demand. Dynamics of fuel demand and supply impacts the fuel price. Annual fuel expense of vehicle adds to the ownership cost of vehicle, which is used in logit model to determine the annual demand for the vehicle for next time step.

1. Model inputs

Car and two-wheelers ownership and annual distance travelled are the inputs to the model. From annual car and two-wheeler ownership, annual demand for cars and two-wheelers are calculated. Correlation of vehicle ownership and annual distance travelled with per capita income is represented by Gompertz function. S-shape of the function gives more realistic growth of these variables than other logistic functions. Parameter values were estimated by fitting historical data.

2. Gasoline and diesel prices

Gasoline and diesel prices are positively influenced by their respective demand and Brent crude oil price (Fig. 1). Differential equations are formulated to capture the impact of the annual change in fuel demand and Brent crude oil price on annual change in respective fuel prices. Annual change in fuel price for i^{th} time step is added to respective fuel price in $i-1^{\text{th}}$ time step to get the fuel price at i^{th} . The equation coefficients are estimated by minimizing the sum of square of error between actual and predicted

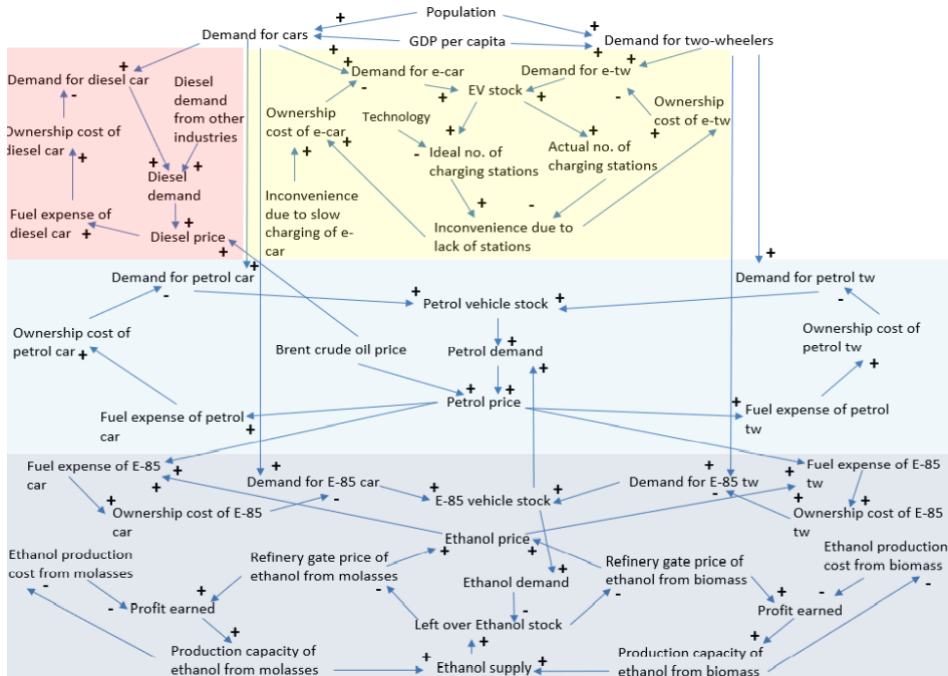


Figure. 1: Causal loop diagram

gasoline and diesel prices between 2001-2018. Calculation of fuel prices are shown in Eq. 1 and 2:

$$P^i = P^{i-1} + 0.7224 \times \frac{dD_p^i}{dt} + 1.1772 \times \frac{dC^i}{dt} \quad (1)$$

$$D^i = D^{i-1} + 0.1702 \times \frac{dD_D^i}{dt} + 0.7712 \times \frac{dC^i}{dt} \quad (2)$$

Where, P^i and D^i are gasoline and diesel prices, $\frac{dC^i}{dt}$, $\frac{dD_p^i}{dt}$ and $\frac{dD_D^i}{dt}$ are annual change in brent crude oil price, gasoline and, diesel demand respectively.

3. Ethanol supply, demand and pricing

Biorefineries producing ethanol from molasses and lignocellulosic biomass have been considered. If biorefinery generates profit, it encourages more investment in the sector, thereby resulting in greater production. Impact is inverse in case sector incur a loss. Since biorefineries require high capital investment and multiple years for erection and commissioning, the model has captured the time-lag in the impact of profit/loss on the actual increase/decrease in the ethanol production using ARX time series model. The equation parameters are estimated by fitting historical data. The ethanol production equations are given by Eq. 3 and 4:

$$EP_M^{i+2} = 0.8967 \times EP_M^i + 0.3224 \times P_M^i + 0.0307 \times P_M^{i-1} \quad (3)$$

$$EP_B^{i+3} = 1.1365 \times EP_B^i + 0.0367 \times P_B^i + 0.0338 \times P_B^{i-1} \quad (4)$$

Where, EP_M^i , EP_B^i , P_M^i and, P_B^i are ethanol production capacity in million litres and profit earned in INR by molasses and biomass based biorefineries respectively.

Literature have reported reduction in production cost of biomass-based ethanol falls within range of 15-25% as capacity doubles (van den Wall Bake et al., 2009). Based on this it is assumed that production cost will reduce by 15% for lignocellulosic ethanol. Production process of ethanol from molasses is fairly matured hence production cost is assumed constant. Ethanol demand comes from E-85 vehicle stock and mandatory blending in gasoline. Refinery gate price of ethanol is the price at which biorefineries sell their ethanol to oil manufacturing companies. If ethanol demand is higher/lower than supply then refinery gate ethanol price will have proportionate influence, as shown by negative causal relation in Fig.1.

4. Electric vehicles and associated inconveniences

Adoption of EVs is associated with many challenges from consumers' side. The model considers two basic inconveniences, i.e., insufficient charging stations and long charging time. Insufficient charging stations lead to limited options for charging. Additionally, slow charging may further affect the desirability of an EV for a consumer. Among these, inconvenience due to lack of charging stations is considered as common for both electric car and two-wheeler as shown in CLD. This is because the charging stations would be shared by both. Inconvenience due to long charging time would be a function of their battery capacity. Hence, this would be vehicle specific. The model converts these inconveniences into monetary values and adds that to the total cost of ownership of the vehicle. Higher inconvenience results in higher ownership cost thereby reducing the demand for EVs. The inconvenience due to lack of charging stations is calculated by comparing existing number of stations with the number of stations ideally required based on electric vehicle stock. Calculation of inconvenience cost due to insufficient charging stations and long charging time is shown in Eq. 5 and 6:

$$I^{CS} = \frac{(CS^i - CS^a) * C^{cs}}{EV^{stock}} \quad (5)$$

$$I^{CT} = \frac{(ST^{car} - FT^{car}) * (C^{SC} - C^{FC})}{E_{car}^{stock}} \quad (6)$$

where, I^{CS} , I^{CT} are inconvenience costs due to insufficient charging stations and long charging time, CS^i , CS^a are ideal and actual number of charging stations, C^S , C^{SC} and, C^{FC} are the setup cost of station, slow and fast charger, EV^{stock} and E_{car}^{stock} are combined EV stock and electric car stock respectively. Inconvenience due to insufficient charging stations is the additional funds required to setup charging stations to meet the ideal requirement. The inconvenience due to long charging time is the fund required to pay in order to save the additional time required to charge electric car by a slow charger.

5. Ownership cost and logit model

Ownership cost of vehicle is the sum of annualised purchase price, fuel expense and, maintenance cost of vehicle. Vehicle purchase price and maintenance cost are inputs to the model. Fuel expense is determined from feedback loops of the model. Ownership cost of vehicle is negatively related to vehicle demand. Literature have used logit model to calculate purchase probability of various vehicle models in competition (Lin & Greene, 2015). Based on this the purchase probabilities n^{th} vehicle option having OC_n as ownership cost is calculated using logit model as shown in Eq. 7.

$$P_n = \frac{e^{\frac{\mu}{OC_n}}}{\sum_{n=1}^m e^{\frac{\mu}{OC_n}}} \quad (7)$$

The parameter μ in the Eq.7 is called scale parameter. It gives the statistical dispersion of probability distribution. Estimation of coefficient μ for car and two-wheelers is performed by comparing the historical sales of petrol and diesel driven vehicles.

3. Results

Based on model formulation and assumptions we have modelled two scenarios: business-as-usual (BAU) and new technology adoption (NTA). BAU scenario assumes no penetration of any renewable transport options. Hence available options in BAU scenario are gasoline and diesel driven cars and gasoline driven two-wheelers. NTA scenario assumes penetration of E-85 and electric cars and two-wheelers. Fig. 2 shows composition of private vehicle stock which includes car and two-wheeler stocks in 2050 for both scenarios. It shows 87% of private vehicle stock will be gasoline and rest will be diesel driven in BAU scenario. There is a reduction in gasoline and diesel driven vehicles in NTA scenario, as it is taken over by E-85 and electric vehicles. As diesel cars only forms the diesel vehicle stock, the reduction is not as significant as gasoline vehicle stock. E-85 vehicles have slightly higher penetration than EV due to associated inconveniences. The overall penetration of E-85 and electric vehicles will be 32% in total private vehicle stock. Due to lower gasoline and diesel vehicle stock, there is a reduction in their respective demands in NTA compared to BAU scenario. Implication of lower demand can be seen in the fuel price plot shown in Fig.3. It can be noticed that gasoline and diesel prices for BAU scenario is higher than NTA scenario due to lower gasoline and diesel vehicle stocks in NTA. Car and two-wheeler ownership saturate towards 2050, therefore new demand for vehicles is very low during this period.

Demand for gasoline vehicles also reduces, thereby stabilizing the gasoline demand. As demand doesn't change much, gasoline price tends to saturate. Diesel demand comes from diesel driven cars, heavy vehicle and other industrial machineries. As the demand for diesel car reduces in NTA scenario, diesel demand coming from heavy vehicles and other industries continues to grow. Therefore, profile of diesel price is increasing throughout the simulation horizon. NTA scenario saves 5.2 billion tons of carbon dioxide than BAU scenario.

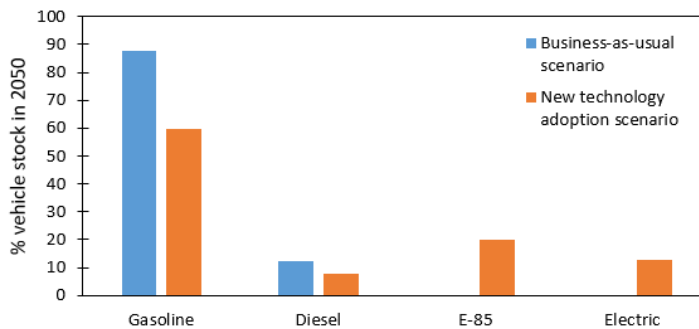


Figure 2: Different vehicle stocks in BAU and NTA scenario

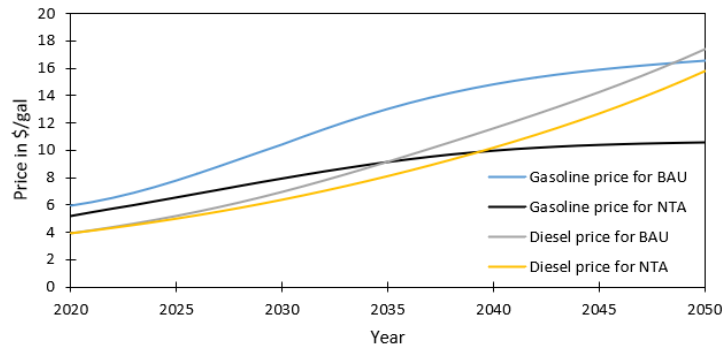


Figure 3: Fuel prices in \$/gallon

4. Conclusion

This work has tried to understand the adoption of renewable travel options, i.e., E-85 and EVs among privately owned vehicles by developing a system dynamics model. Penetration of these options have shown a significant reduction in fuel demand and GHG emissions. However, market dynamics is insufficient to ensure high penetration rate. Favourable scenarios of sufficient charging stations combined with carbon tax and incentives need to be investigated.

References

- Forrester, J. W. (1994). soft OR. *System Dynamics Review*, 10(January), 245–256.
- Lin, Z., & Greene, D. (2015). *Who Will More Likely Buy PHEV: A Detailed Market Segmentation Analysis*
- Menon, B. G., & Mahanty, B. (2015). Assessing the Effectiveness of Alternative Policies in Conjunction with Energy Efficiency Improvement Policy in India. *Environmental*

- Modeling and Assessment*, 20(6), 609–624. <https://doi.org/10.1007/s10666-015-9448-4>
- Pasaoglu, G., Harrison, G., Jones, L., Hill, A., Beaudet, A., & Thiel, C. (2016). A system dynamics based market agent model simulating future powertrain technology transition: Scenarios in the EU light duty vehicle road transport sector. *Technological Forecasting and Social Change*, 104, 133–146. <https://doi.org/10.1016/j.techfore.2015.11.028>
- Joshi M, Chen Han, *The road from paris: india's progress towards its climate pledge. September 2020.*
- van den Wall Bake, J. D., Junginger, M., Faaij, A., Poot, T., & Walter, A. (2009). Explaining the experience curve: Cost reductions of Brazilian ethanol from sugarcane. *Biomass and Bioenergy*, 33(4), 644–658. <https://doi.org/10.1016/j.biombioe.2008.10.006>
- Vilchez, J. J. G., Jochem, P., & Fichtner, W. (2013). EV market development pathways - An application of system dynamics for policy simulation. *World Electric Vehicle Journal*, 6(4), 1030–1038. <https://doi.org/10.3390/wevj6041030>

Multi-Period Design Optimization of Flexible Fixed-Bed Reactors by Stoichiometry-Based Model Reduction

Ronny T. Zimmermann,^a Jens Bremer,^b and Kai Sundmacher^{a,b,*}

^aOtto von Guericke University Magdeburg, Chair for Process Systems Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany

*^bMax Planck Institute Magdeburg, Department Process Systems Engineering, Sandtorstraße 1, 39106 Magdeburg, Germany
sundmacher@mpi-magdeburg.mpg.de*

Abstract

The design of fixed-bed reactors has gained interest in the light of load-flexible operation. This is due to the expectation that some process feeds will be more volatile in the near future, for example, within the framework of the power-to-methane concept. In this regard, the application of an inert shell onto the catalyst particles has proven advantageous. In this work, a multi-period design optimization of a fixed-bed methanation reactor is performed by employing a heterogeneous reactor model, in which the catalyst particle is divided into an active core and an inert shell. Due to the model's computational demand, it is reduced by using stoichiometric relations at the reactor and catalyst particle scale. Subsequently, dynamic transition simulations between the optimized steady states coming from the multi-period design optimization as well as a reactor start-up and shut-down simulation are carried out, and the results of the full model are compared to those of the reduced model. Our results show that the model size can be reduced significantly by using the stoichiometric relations without loss of accuracy in steady state and negligible accuracy loss in transient scenarios. The optimized fixed-bed methanation reactor can be operated over a wide load range by optimally adjusting the operating variables and the dynamic simulations show smooth transitions between the steady states.

Keywords: Multi-Period Design Optimization, Heterogeneous Reactor Model, Stoichiometric Relations, Load-Flexible Reactor, Dynamic Reactor Operation

1. Introduction

Fixed-bed reactors are commonly used in the chemical industry to carry out heterogeneously catalyzed gas-phase reactions. A key design aspect in the presence of exothermic reactions is heat management to avoid temperatures that might damage catalyst and reactor material. For this reason, the fixed-bed is often located in several tubes, which are externally surrounded by a coolant. Despite intensive cooling, critical reactor conditions may arise, which lead to pronounced temperature hot-spots in the reactor, which is known as thermal runaway. The understanding and technical handling of this aspect have been intensively researched for steady-state reactor operation.

However, due to the expectation that certain process feeds will be more volatile in the near future, it must be ensured that the reactors do not only operate safe and economical in a single steady state but in multiple steady states as well as in transitions between them.

It was shown that reactor runaway conditions could be prevented by applying an inert shell on the active catalyst particles. If designed correctly, the effective reaction rate is significantly slowed down at high temperatures, where the mass transport through the inert shell becomes rate-determining. On the other hand, if the shell is sufficiently thin, it does not influence the effective reaction rate at low temperatures. In this way, a safe and economical reactor operation can be ensured, as shown by Zimmermann et al. (2020).

A possible application of such catalyst particles is within the power-to-methane concept (Rönsch et al. (2016)). In this context, hydrogen is produced via water electrolysis in the first part of the process chain. In regions that lack hydrogen infrastructure, hydrogen is subsequently converted with carbon dioxide into methane, which can be fed into the natural gas grid. It is assumed that the employed water electrolyzer operates in a flexible manner, according to the availability of surplus energy. Consequently, to avoid expensive intermediate hydrogen storage, the employed methanation reactor must also be operated flexibly (Bremer and Sundmacher (2019)).

The design of a multi-tubular fixed-bed methanation reactor, which is able to operate with a fluctuating feed by using the described core-shell catalyst particles, is the aim of this work. For this purpose, a multi-period design optimization is performed by employing a heterogeneous reactor model. Since this is a numerically challenging task, the model is reduced by introducing stoichiometric relations on the reactor and particle scale. As the use of the stoichiometric relations is strictly valid only in steady state, dynamic simulations are performed to investigate whether the reduced model can also reflect the behavior of the more complex model in transient scenarios.

2. Heterogeneous Reactor Model

A 1D-1D heterogeneous fixed-bed reactor model is used, which distinguishes mass balance equations for the components $i \in \{\text{CO}_2, \text{H}_2, \text{CH}_4, \text{H}_2\text{O}\}$ and heat balance equations for the gas phase (index G), the inert catalyst particle shell (index S) and the catalytically active catalyst particle core (index C). The balance equations on reactor scale with initial and boundary conditions read as

$$\epsilon_G \frac{\partial c_{G,i}}{\partial t} = - \frac{\partial c_{G,i} u}{\partial z} - \frac{(1-\epsilon)}{d} k_i (c_{G,i} - c_{S,i}(R_S)) \quad (1)$$

$$\epsilon_G (\rho c_p)_G \frac{\partial T_G}{\partial t} + u_{in} (\rho_{in} c_p)_G \frac{\partial T_G}{\partial z} = \frac{\partial}{\partial z} \left[\Lambda_{ax} \frac{\partial T_G}{\partial z} \right] - \frac{4U}{D} (T_G - T_{cool}) - \frac{(1-\epsilon)}{d} h (T_G - T_S(R_S)) \quad (2)$$

$$c_{G,i}(z=0) = c_{G,i,in}, \quad \Lambda_{ax} \frac{\partial T_G}{\partial z} \Big|_{z=0} = u_{in} (\rho_{in} c_p)_G (T_G - T_{in}) \quad (3)$$

$$\Lambda_{ax} \frac{\partial^2 T_G}{\partial z^2} \Big|_{z=L} = 0, \quad c_{G,i}(t=0) = c_{G,i,0}, \quad T_G(t=0) = T_{G,0} \quad (4)$$

and on the catalyst particle scale as

$$\epsilon_C \frac{\partial c_{C,i}}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} \left[r^2 D_{C,i} \frac{\partial c_{C,i}}{\partial r} \right] + \rho_C \nu_i \mathbf{R}, \quad \epsilon_S \frac{\partial c_{S,i}}{\partial t} = \frac{\partial}{\partial r} \left[D_{S,i} \frac{\partial c_{S,i}}{\partial r} \right] \quad (5)$$

$$(\rho c_p)_C \frac{\partial T_C}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} \left[r^2 \Lambda_C \frac{\partial T_C}{\partial r} \right] + \rho_C H_r \mathbf{R}, \quad (\rho c_p)_S \frac{\partial T_S}{\partial t} = \frac{\partial}{\partial r} \left[\Lambda_S \frac{\partial T_S}{\partial r} \right] \quad (6)$$

$$D_{C,i} \frac{\partial c_{C,i}}{\partial r} \Big|_{r=0} = 0, \quad D_{C,i} \frac{\partial c_{C,i}}{\partial r} \Big|_{r=R_C} = D_{S,i} \frac{\partial c_{S,i}}{\partial r} \Big|_{r=R_C} \quad (7)$$

$$\Lambda_C \frac{\partial T_C}{\partial r} \Big|_{r=0} = 0, \quad \Lambda_C \frac{\partial T_C}{\partial r} \Big|_{r=R_C} = \Lambda_S \frac{\partial T_S}{\partial r} \Big|_{r=R_C} \quad (8)$$

$$c_{C,i}(r = R_C) = c_{S,i}(r = R_C), \quad D_{S,i} \frac{\partial c_{S,i}}{\partial r} \Big|_{r=R_S} = k_i (c_{G,i} - c_{S,i}(r = R_S)) \quad (9)$$

$$T_C(r = R_C) = T_S(r = R_C), \quad \Lambda_S \frac{\partial T_S}{\partial r} \Big|_{r=R_S} = h(T_G - T_S(r = R_S)) \quad (10)$$

$$c_{C,i}(t = 0) = c_{S,i}(t = 0) = c_{i,0}, \quad T_C(t = 0) = T_S(t = 0) = T_0 \quad (11)$$

In addition to the radially averaged temperature T_G , the fixed-bed's center temperature \hat{T}_G is approximated by the correlation of Dixon (1996). The reaction kinetic model of Koschany et al. (2016), which considers no side reactions, is employed to describe the methanation reaction. Details about the remaining constitutive equations are given by Zimmermann et al. (2021).

It is known that the n_i - n_r mass balance equations in Eq. 1 can be expressed by n_r mass balance equations of key components, where n_i is the number of components and n_r is the number of linearly independent reactions. The same can be done on the catalyst pellet scale by introducing the stoichiometric relations $N_i/\nu_i = N_j/\nu_j$ as discussed in detail by Burghardt (1986) and Jackson (1977). In the case of piecewise constant diffusion coefficients or Knudsen diffusion, these equations can be integrated analytically. Thus, n_i - n_r concentration profiles can be expressed as

$$c_{G,j} u = c_{G,j,\text{in}} u_{\text{in}} - \nu_j/\nu_i X_i c_{G,i,\text{in}} u_{\text{in}} \quad \text{with} \quad X_i = \frac{c_{G,i,\text{in}} u_{\text{in}} - c_{G,i} u}{c_{G,i,\text{in}} u_{\text{in}}} \quad (12)$$

$$c_{S,j}(r = R_S) = \nu_j/\nu_i k_i/k_j (c_{S,i}(r = R_S) - c_{G,i}) + c_{G,j} \quad (13)$$

$$c_{S,j}(r) = \nu_j/\nu_i \sqrt{M_i/M_j} (c_{S,i}(r) - c_{S,i}(r = R_S)) + c_{S,j}(r = R_S) \quad (14)$$

$$c_{C,j}(r) = \nu_j/\nu_i \sqrt{M_i/M_j} (c_{C,i}(r) - c_{C,i}(r = R_C)) + c_{C,j}(r = R_C) \quad (15)$$

In the dependence of the system, the number of mass balance equations and thus the computational demand can be reduced significantly. However, it has to be kept in mind that the stoichiometric relations are valid only in steady state. Thus, n_i - n_r concentration profiles are assumed as quasi-stationary. Therefore, the degree of approximation in dynamic simulations depends on the choice of the chosen key component.

3. Multi-Period Design Optimization Problem

The discussed model is reduced via the stoichiometric relations. Carbon dioxide is taken as the key component, as it exhibits the highest molar mass of all reactants and, as a result, it exhibits the slowest dynamics at the catalyst particle scale. Consequently, the quasi-

steady-state assumption of the other components should lead to an accurate approximation of the system dynamics, at 40 % of the full model size.

A multi-period design optimization, as discussed by Grossmann and Sargent (1979), is performed by employing the reduced reactor model. Eleven periods of equal temporal length and equally spaced carbon dioxide loads from 500 kg/h to 1500 kg/h are used. As objective function, the mean methane space-time yield is maximized. The operating variables consist of the inlet pressure, velocity, and temperature as well as the coolant temperature and wall heat transfer coefficient. The inlet feed is always a stoichiometric mixture of pure carbon dioxide and hydrogen. The design variables are the reactor's tube length, diameter, and number, together with the catalyst particle diameter, shell porosity, pore diameter, and thickness. In each state, the temperature in the center of the fixed-bed must never exceed 750 K, and the conversion at the outlet of the reactor must be at least 95 %.

The given balance equations are spatially discretized via the Finite Volume Method. In total, 150 logarithmically spaced finite volumes are used at the reactor scale. On the catalyst particle scale, 15 finite volumes with equal volume size are used within the catalyst particle core and five within the catalyst particle shell. Details on the employed software for solving the optimization problem and for performing subsequent dynamic simulations are given by Zimmermann et al. (2021).

4. Results

4.1. Multi-Period Design Optimization

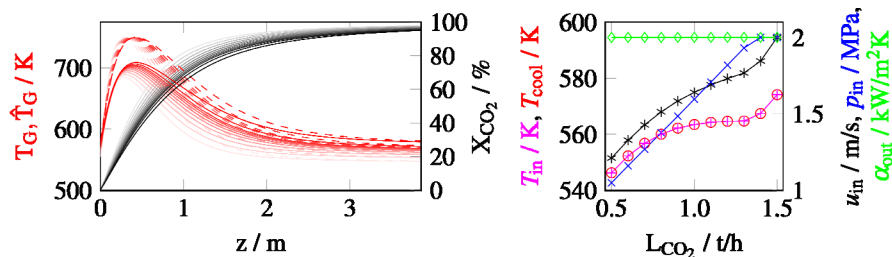


Figure 1: Multi-period design optimization results: The optimal profiles of the radial mean temperature T_G (—), center temperature \hat{T}_G (- - -) and carbon dioxide conversion X_{CO_2} (—) are shown on the left (darker colors correspond to higher loads). The corresponding inlet temperature T_{in} (+), pressure p_{in} (—*), and velocity u_{in} (—*) as well as the coolant temperature T_{cool} (—o) and wall heat transfer coefficient α_{out} (—o) are shown on the right.

The multi-period design optimization with the reduced model is solved within 75 iterations in 118 seconds. Solving the optimization problem of the full model on the employed desktop PC with 32 GB ram is not possible. The optimal temperature and conversion profiles, together with the respective operating variables, are shown in Fig 1. The optimal catalyst particles consist of a shell, where porosity and pore diameter are at their lower bounds of 10 % and 5 nm, respectively. As discussed in the introduction, this allows for a thin shell, which on the one hand, inhibits the effective reaction rate at high temperatures

and, on the other hand, does not significantly influence the effective reaction at low temperatures.

The optimal reactor is then designed such that the maximum heat transfer to the coolant is realized, as the tube diameter is at the lower bound of 2 cm and the coolant wall heat transfer coefficient is at the maximum of 2 kW/m²K for all loads. One hundred eighty tubes are needed to handle the maximum load of 1.5 t_{CO2}/h at the given operating conditions with inlet velocity and pressure at the upper bound. The reactor length of 3.9 m is determined by the residence time of the reactant mixture required to achieve the lower limit of the conversion of 95 % at the outlet at the highest load. To decrease the reactor load, the simultaneous variation of reactor inlet pressure and velocity is optimal, as long as none of the values is at its upper bound. Since lower reactor pressures tend to shift the hot-spot towards the reactor outlet, but lower inlet velocities shift the hot-spot towards the inlet, both effects almost cancel out. Thus, the hot-spot is almost in the same position for each load. Since the decrease of inlet velocity and pressure also decreases the heat transfer to the coolant, inlet and coolant temperature also decrease, such that the hot-spot does not exceed the limit of 750 K. As the reaction is equilibrium-limited, this also allows for outlet conversions of higher than 95 %. Overall, a mean methane space-time yield of 0.44 kg/m³s is achieved.

4.2. Dynamic Transitions

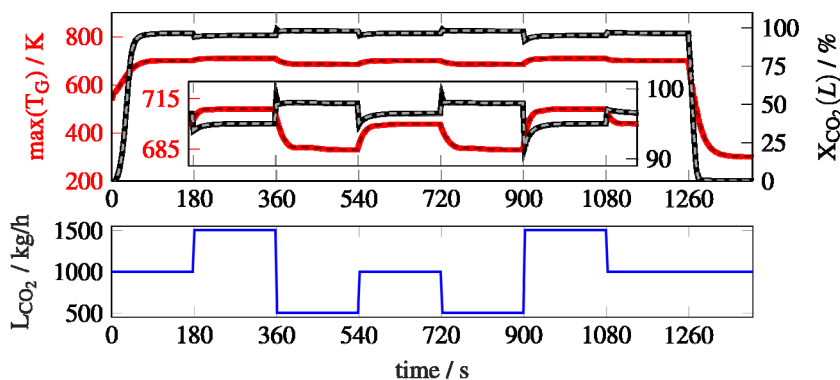


Figure 2: Dynamic simulation results: Maximum radial mean temperature (full model (—), reduced model (----)) and outlet carbon dioxide conversion (full model (—), reduced model (----)) for some load changes as well as a start-up and shut-down.

As the multi-period design optimization neglects the system dynamics, transitions between selected optimal steady states are simulated with both models. For this purpose, the reactor load is changed stepwise, and the operating variables are changed instantaneously according to the results of the multi-period design optimization. Additionally, a reactor start-up and shut-down are investigated. The initial conditions for the start-up are given as $X_{0,CO_2} = 0$ % and $T_0 = 300$ K. For the shut-down, T_{in} and T_{cool} are reduced to 300 K.

The results, shown in Fig. 2, indicate quick transitions in all periods of about one to two minutes. Despite drastic load changes, only slight oscillations in the temperature profile

can be observed during the transitions. Additionally, on positive load changes, a slight undershooting of the conversion is present. It is further evident that the reduced model is suitable to describe the dynamics of the full model. The models predict only a small difference in the outlet conversion in a period of about two seconds after a load change and a maximum hot-spot temperature difference of 0.05 K is observed in the displayed scenario. The total integration times of both models are comparable for the performed simulations (109 s for the reduced model vs. 122 s for the full model).

5. Conclusion

The presented results indicate that a fixed-bed reactor filled with catalyst particles that consist of an active core surrounded by an inert shell can be operated safely and economically within a wide load range. The simultaneous variation of reactor inlet velocity and pressure is favorable in this context, as shown by multi-period design optimization. Subsequent dynamic simulations between the optimal steady states as well as a start-up and shut-down indicate smooth transitions, even at drastic stepwise load changes. It is further concluded that applying stoichiometric relations at the catalyst particle scale gives accurate an approximation of the considered system dynamics.

6. Acknowledgment

This research work was conducted within the DFG Priority Program SPP2080 'Catalysts and reactors under dynamic conditions for energy storage and conversion' and was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 406914011. Ronny Zimmermann and Jens Bremer are also affiliated with the International Max Planck Research School (IMPRS) for Advanced Methods in Process and Systems Engineering, Magdeburg, Germany. (Gefördert durch die Deutsche Forschungsgemeinschaft(DFG)-406914011.)

References

- J. Bremer, K. Sundmacher, 2019. Operation range extension via hot-spot control for catalytic CO₂ methanation reactors. *Reaction Chemistry & Engineering* 4 (6), 1019–1037.
- A. Burghardt, 1986. Transport phenomena and chemical reactions in porous catalysts for multicomponent and multireaction systems. *Chemical Engineering and Processing: Process Intensification* 20 (5), 229–244.
- A. G. Dixon, 1996. An improved equation for the overall heat transfer coefficient in packed beds. *Chemical Engineering and Processing: Process Intensification* 35 (5), 323–331.
- I. E. Grossmann, R. W. H. Sargent, 1979. Optimum design of multipurpose chemical plants. *Industrial & Engineering Chemistry Process Design and Development* 18 (2), 343–348.
- R. Jackson, 1977. *Transport in porous catalysts*. Vol. 4. Elsevier Science & Technology.
- F. Koschany, D. Schlereth, O. Hinrichsen, 2016. On the kinetics of the methanation of carbon dioxide on coprecipitated NiAl(O)_x. *Applied Catalysis B: Environmental* 181, 504–516.
- S. Rönsch, J. Schneider, S. Matthischke, M. Schlüter, M. Götz, J. Lefebvre, P. Prabhakaran, S. Bajohr, 2016. Review on methanation – from fundamentals to current projects. *Fuel* 166, 276–296.
- R. T. Zimmermann, J. Bremer, K. Sundmacher, 2020. Optimal catalyst particle design for flexible fixed-bed CO₂ methanation reactors. *Chemical Engineering Journal* 387, 123704.
- R. T. Zimmermann, J. Bremer, K. Sundmacher, 2021. Load-flexible fixed-bed reactors by multi-period design optimization. *Chemical Engineering Journal*, submitted.

Optimization based sampling for gray-box modeling using a modified upper confidence bound acquisition function

Joschka Winz^{a,*}, Sebastian Engell^a

^aProcess Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, Technische Universität Dortmund, Emil-Figge Str. 70, 44227 Dortmund, Germany

joschka.winz@tu-dortmund.de

Abstract

Chemical process simulations rely on the accurate representation of thermodynamic phenomena. Complex models like the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) provide such accurate descriptions but due to their implicit formulation, process optimization based on such models is computationally very demanding. This issue can be avoided by surrogate modeling, where a data-based model approximates the costly computation. When setting up a surrogate model, the question of which data to collect to fit the surrogate arises. In previous work, methods have been developed to combine sampling with optimization to only collect data in regions of interest for the optimization. These methods however mostly assume that the surrogate model describes the objective function. In this work, an extension to gray-box models is proposed.

Keywords: surrogate modeling, adaptive sampling, Bayesian optimization, chemical processes, gray-box modeling

1. Introduction

Surrogate and gray-box modeling rely on data. When the original relationship that the surrogate should approximate is a simulation or a computer code, the process of acquiring this data is commonly known as sampling. An intuitive sampling method is to sample the space as uniformly as possible with respect to the input variables. This is denoted as space-filling sampling, thoroughly reviewed in Garud et al. (2017). Adaptive sampling methods sample the space non-uniformly in order to cover regions of more complex behavior of the approximated function better. Cozad et al. (2014) use derivative free optimization methods to find points of high deviation of the surrogate from the original function. A similar technique is applied by Garud et al. (2016) who maximize a so-called departure function which describes the sensitivity of the surrogate prediction to samples that are part of the training set.

Kleijnen and Van Beers (2004) do not use optimization of a criterion to find new samples but instead select candidate inputs at which the jackknife variance is maximized. This approach was extended by Eason and Cremaschi (2014) by using a criterion that incorporates both the jackknife variance and a nearest-neighbor distance to address the exploration vs. exploitation trade-off. This has been further refined by Nentwich and Engell (2019) who used surrogate modeling to describe phase-equilibria. Because of the

simultaneous description of phase composition and miscibility gap, both a regression and classification surrogate were combined resulting in a modified sampling criterion.

The goal of these approaches is to obtain a surrogate model that possesses high accuracy over the whole range of input values. This does not reflect the later use of the model which often is an optimization where the surrogate model is incorporated as a submodel. The field of Bayesian optimization deals with combining surrogate modeling, sampling and optimization. Jones et al. (1998) used Kriging surrogates and their description of the prediction uncertainty to statistically describe the expected improvement.

This was extended by Gardner et al. (2014) to deal with constraints for a black-box model for the objective function and by Astudillo and Frazier (2019) to incorporate simple gray-box models with Monte Carlo simulation.

To overcome these assumptions we propose an optimization based sampling algorithm for embedded gray-box models.

2. A novel optimization based sampling approach

The novel approach is based on the upper confidence bound (UCB) acquisition function, which was developed in the field of Bayesian optimization by Cox and John (1992). The idea of UCB-sampling is to evaluate the uncertainty that results from the surrogate prediction in an optimistic manner. This approach is similar to the expected improvement (EI) acquisition function by complementing the surrogate prediction with its standard deviation. The main difference is that using EI the probability distribution is integrated analytically, while using UCB explicit confidence intervals are used. The proposed UCB acquisition function $f^{UCB-GB}(x)$ is optimized for determining a new sample that is added to the training set.

$$f^{UCB-GB}(x) = f^{GB}(x, \hat{y}(x)) + \kappa |g^T| \sqrt{\hat{\sigma}^2(x)} + \Phi(x) \quad (1)$$

Here, the modified acquisition function f^{UCB-GB} is estimated based on a gray-box model f^{GB} , which in turn depends on a surrogate model prediction \hat{y} and decision variables x . The surrogate prediction variance $\hat{\sigma}^2$ is multiplied with the gray-box model sensitivity to the deviation in the surrogate model g and a weighting factor κ . Φ is used for parallelization as described below.

To determine g , the optimization problem can be modified by including additional decision variables z , which are constrained to equal the surrogate prediction, as shown in Eq. (2) for the original (left side) and the UCB objective function (right side).

$$\begin{aligned} \min_{x,z} \quad & f^{GB}(x, z) & \min_{x,z} \quad & f^{GB}(x, z) + \kappa |g^T| \sqrt{\hat{\sigma}^2(x)} + \Phi(x) \\ \text{s. t.} \quad & \hat{y}(x) = z \quad | \lambda^{GB} & \Rightarrow \quad & \hat{y}(x) = z \quad | \lambda^{UCB-GB} \\ & h(x, z) < 0 & & h(x, z) < 0 \end{aligned} \quad (2)$$

From the theory of constrained optimization, at the optimum (denoted by $*$) the sensitivity of the objective function value with respect to surrogate prediction is given by the corresponding Lagrange multiplier $\lambda^{GB,*}$. The problem on the right hand side is dealt with iteratively, g is assumed to be constant during optimization and in the next iteration

updated to $\lambda^{GB,*}$. This is calculated from $\lambda^{UCB-GB,*}$ with a factor $k(x)$ correcting the difference in the gradients of f^{GB} and f^{UCB-GB} , shown in Eq. (3).

$$k(x^*)\lambda^{UCB-GB,*} \approx \lambda^{GB,*} \approx g$$

$$k(x) = \frac{\|\nabla_x f^{GB}(x)\|_2}{\|\nabla_x f^{UCB-GB}(x)\|_2} \quad (3)$$

Parallelization here means techniques to reduce the number of times the surrogate model has to be trained. Common approaches optimize the acquisition function that describes for example the expected improvement for several input combinations at once as in Ginsbourger et al. (2008). New sampling points are computed several times before the surrogate model is retrained. To avoid adding identical samples, $h_{penalty}$ is added to Φ for penalizing sampling near new samples so that different new sampling locations are computed. $h_{penalty}$ is given by

$$h_{penalty}(x_1, x_2, g) = \frac{\kappa |g|^T \sqrt{\hat{\sigma}(x_1)^2}}{\sqrt{2\pi}\sigma_\Phi} \exp\left(-\frac{\|\bar{x}_1 - \bar{x}_2\|_2^2}{2\sigma_\Phi^2}\right). \quad (4)$$

The parameter that determines the broadness of the penalty is σ_Φ . \bar{x} is the normalized value of x to overcome problems with different ranges of the variables. The total amount of penalty is set to the assumed added value in cost function: $\kappa |g|^T \sqrt{\hat{\sigma}(x)^2}$. The exploration vs. exploitation trade-off is addressed because in the early stage of sampling the larger estimated variance $\hat{\sigma}^2$ leads to a higher penalty and more exploration of the whole input space.

Data: Initial training input and output sets X and Y

```

while  $|X| < n_{maxsamples}$  do
    Initialize  $\Phi(x) \leftarrow 0$ ,  $X_{iter} \leftarrow \emptyset$ ,  $Y_{iter} \leftarrow \emptyset$ 
    Train surrogate model  $\hat{y}$  using inputs  $X$  and outputs  $Y$ 
    Calculate jackknife variance  $\hat{\sigma}^2$  using inputs  $X$  and outputs  $Y$ 
    while  $|X_{iter}| < n_{iter}$  do
        Initialize  $x^0$ ,  $z^0$ ,  $g$ 
        while:  $\|x^0 - x^*\|_2 > \epsilon$  do
            Solve Eq. (2), right hand side with initial guesses  $x^0$ ,  $z^0$  to  $x^*$ ,  $z^*$ ,  $\lambda^{UCB-GB,*}$ 
            Update  $x^0 \leftarrow x^*$ ,  $z^0 \leftarrow z^*$ ,  $g \leftarrow k(x^*)\lambda^{UCB-GB,*}$ 
        end
        Update Sets  $X_{iter} \leftarrow \cup(X_{iter}, x^*)$ ,  $Y_{iter} \leftarrow \cup(Y_{iter}, y(x^*))$ 
         $\Phi(x) \leftarrow \Phi(x) + h_{penalty}(x, x^*, g)$ 
    end
    Update Sets  $X \leftarrow \cup(X, X_{iter})$ ,  $Y \leftarrow \cup(Y, Y_{iter})$ 
end

```

Algorithm 1: Sampling method based on the upper confidence bound acquisition function (UCB) with parallelization

With a growing training set exploitation of already explored optima is performed. Based on these definitions Algorithm 1 is proposed, which is the main contribution of this work. $|A|$ denotes the cardinality of a set A . In Algorithm 1 three loops are iterated. The innermost loop adapts the value of g with direct substitution until the change in the optimum regarding the decision variables x is below a threshold ϵ , which leads to $g \approx k(x^*)\lambda^{UCB-GB,*}$. The second loop incorporates the parallelization by filling intermediate sets of samples X_{iter} and Y_{iter} , adjusting the penalty Φ in each step. In the outermost loop X_{iter} , Y_{iter} and Φ are reset after merging the intermediate sets X_{iter} and Y_{iter} with the training sets X and Y . The surrogate model is retrained and the prediction variance $\hat{\sigma}^2$ is calculated by using the jackknife.

3. Case study

The proposed method is applied to the model of the chemical process of the hydroformylation of 1-dodecene in a thermomorphic solvent system, taken from Nentwich et al. (2019). The considered flowsheet can be seen in Figure 1.

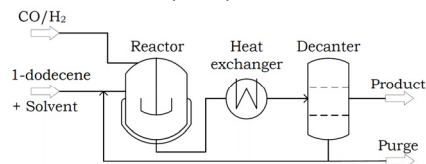


Figure 1: Flowsheet of hydroformylation process

This process consists of three unit operations: reactor, heat exchanger and decanter. In the reactor the catalyzed reaction of n-dodecene with synthesis gas (CO and H₂) to n-tridecanal takes place in a homogeneous liquid phase. The following two unit operations of cooling down the liquid stream coming from the reactor and using a decanter to recycle the catalyst cost efficiently are introduced because a thermomorphic solvent system is used that leads to a temperature dependent miscibility gap. This way, the catalyst can be recycled with the polar phase while the product is removed with the organic phase.

To accurately describe the liquid-liquid equilibrium in the decanter, the equation of state PC-SAFT is used, because it has been shown to provide accurate predictions for the system at hand by Schäfer et al. (2012). Surrogate modeling of the compositions of the two phases from the decanter feed composition and temperature is applied to overcome the issue of computational cost resulting from the implicit formulation. A simplified quaternary system consisting of n-decane, dimethylformamid, n-dodecene and n-tridecanal is used.

The objective function is chosen to be the theoretical cost in \$ per ton of product (CPT) assuming a production of 10,000 tons of n-tridecanal per year. The results of applying the algorithm presented in section 2 to this case study are shown in the next section.

4. Results

To evaluate the performance of Algorithm 1, denoted as UCB, it is compared to the mixed-adaptive approach as developed by Nentwich and Engell (2019). Parameters specific to the UCB-method are ϵ and σ_Φ , which are set to 10^{-5} and 0.05, respectively.

The parameters $n_{max\ samples}$, n_{iter} and the number of jackknifing groups are used in both algorithms. The values are 1000, 20 % of $|X|$ and 5. Additionally, the surrogate structure is the same for all methods.

Artificial neural networks (ANN) with two fully-connected hidden layers of 10 nodes each using the tanh-activation function are used as surrogates as even complex functional relationships can be modeled accurately by this structure.

Additionally, a set of 570 basis functions consisting of polynomial and exponential terms that were determined by using LASSO regression with regularization to minimize Akaike's information criterion (AIC) is applied as a surrogate model. It is expected that the set of considered basis functions and the regularization leads to a response surface which is more convenient for optimization.

The results of the optimization (Eq. (2), left) are shown over the iterations of the sequential sampling. For reference, also results using Latin hypercube sampling (LHS) designs with the same number of samples for each iteration are shown.

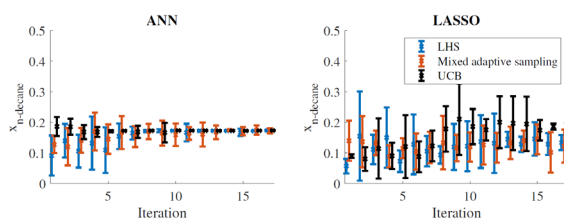


Figure 2: Predicted molar fraction of n-decane at the reactor outlet at the optimum for ANN (left hand side) and LASSO (right hand side) surrogates for different sampling methods; blue: Latin Hypercube sampling (LHS), orange: mixed adaptive sampling, black: upper confidence bound sampling as proposed

In Figure 2 the molar fraction of n-decane at the reactor outlet is shown at the optimum for different stages of sampling, evaluated from six randomly distributed initial points for five runs of each of the methods, because this quantity is sensitive to the surrogate prediction. The number of samples increases by 20 % in each iteration and ranges from initially 60 to more than 1000 samples. It can be seen from the figure that in later iterations the optimum regarding p converges to about 18 % for all considered sampling methods for the ANN surrogates. In the early stages, varying predictions of the optima occur. For LHS and mixed adaptive sampling molar fractions below the final value of convergence can be observed, while using UCB sampling an accurate description of the optimum is obtained already at low numbers of samples.

For LASSO surrogates, convergence to the optimal value is only observed for the UCB sampling. Both mixed adaptive and Latin hypercube sampling lead to an average prediction of around 14 % with a high variance even for large numbers of samples. This can be explained by a too small flexibility of the surrogate model caused by strong regularization, where sampling with the goal to decrease the overall deviation leads to a surrogate with a still mediocre accuracy. When focusing on the region around the optimum with UCB sampling, the accuracy can be increased where it is needed, neglecting deviations elsewhere.

5. Conclusion and outlook

In this work a sampling method for surrogate models is proposed with the goal to focus the improvement in model accuracy on regions that are of interest regarding the

optimization of a nested gray-box model. The main idea is to make use of the upper confidence bound acquisition function developed in the field of Bayesian optimization. Its performance is investigated for the case study of the hydroformylation of 1-dodecene in a thermomorphic solvent system. Using the proposed algorithm, convergence to the optimum is reached with significantly fewer samples compared to conventional methods. In future work, an optimal weighting of prediction and uncertainty will be investigated. Additionally, the combination of the proposed method with other parallelization methods can give further insights into the capability of this method.

6. Acknowledgment

This research has been supported by the project “KI-Inkubator-Labore in der Prozessindustrie - KEEN”, funded by the Bundesministerium für Wirtschaft und Energie (BMWi) under grant number 01MK20014T. This support is gratefully acknowledged.

References

- R. Astudillo, P. I. Frazier, 2019. Bayesian optimization of composite functions. In: 36th International Conference on Machine Learning, ICML 2019. Vol. 2019-June. pp. 547–556.
- D. D. Cox, S. John, 1992. A statistical method for global optimization. In: Conference Proceedings – IEEE International Conference on Systems, Man and Cybernetics. Vol. 1992-Janua. Institute of Electrical and Electronics Engineers Inc., pp. 1241–1246.
- A. Cozad, N. V. Sahinidis, D. C. Miller, 2014. Learning surrogate models for simulation-based optimization. *AIChE Journal* 60 (6), 2211–2227.
- J. Eason, S. Cremaschi, 2014. Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers and Chemical Engineering* 68, 220–232.
- J. R. Gardner, M. J. Kusner, Z. E. Xu, K. Q. Weinberger, J. P. Cunningham, 2014. Bayesian Optimization with Inequality Constraints. In: ICML. pp. 937–945.
- S. S. Garud, I. A. Karimi, M. Kraft, 2016. Smart adaptive sampling for surrogate modelling. In: *Computer Aided Chemical Engineering*. Vol. 38. Elsevier, pp. 631–636.
- S. S. Garud, I. A. Karimi, M. Kraft, 2017. Design of computer experiments: A review. *Computers & Chemical Engineering* 106, 71–95.
- D. Ginsbourger, R. L. Riche, L. Carraro, 2008. A Multi-points Criterion for Deterministic Parallel Global Optimization based on Gaussian Processes. HAL preprint hal00260579 1, 1–30.
- D. R. Jones, M. Schonlau, W. J. Welch, 1998. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* 13 (4), 455–492.
- J. P. Kleijnen, W. C. Van Beers, 2004. Application-driven sequential designs for simulation experiments: Kriging metamodelling. *Journal of the Operational Research Society* 55 (8), 876–883.
- C. Nentwich, S. Engell, 2019. Surrogate modeling of phase equilibrium calculations using adaptive sampling. *Computers and Chemical Engineering* 126, 204–217.
- C. Nentwich, C. Varela, S. Engell, 2019. Optimization of chemical processes applying surrogate models for phase equilibrium calculations. In: *Proceedings of the International Joint Conference on Neural Networks*. IEEE.
- E. Schäfer, Y. Brunsch, G. Sadowski, A. Behr, 2012. Hydroformylation of 1-dodecene in the thermomorphic solvent system dimethylformamide/decane. *Industrial & engineering chemistry research* 51 (31), 10296–10306.

How to design cooperative sustainable industrial systems?

Marianne Boix*, Florent Mousqué, Ludovic Montastruc, Stéphane Négny

*Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, 4 Allée
Emile Monso, 31432 Toulouse, France*

marianne.boix@ensiacet.fr

Abstract

Industrial systems are complex in terms of types of material exchanged (water, electricity, vapour...), types of industries involved, and size of companies, temporality and many others. Regarding this complexity, this paper aims at developing a generic methodology able to design exchanges between companies in an eco-industrial park in terms of electricity, water and several pressures of vapour. The number of companies and their types of processes can also be various. In this sense, a generic model would be able to be adapted to a great diversity of case studies. This multi-objective optimization method can deal with the minimization of total cost, environmental impacts (through a life cycle analysis indicators) or complexity of the network designed (technical aspects linked to interconnections between companies). Four time-periods are studied and the objective functions to minimize are the total cost of the network, its complexity (related to the interdependence between companies) and the energy consumed linked to environmental impact. The procedure developed reaches to design all the pipes constituting the exchanges between companies while minimizing the cost through a multi-objective optimization approach. Additionally, the design of the network takes into account a better dependency of companies what makes the EIP more attractive for them.

Keywords: optimization, industrial ecology, networks, energies, flexibility.

1. Introduction

Regarding the economic, environmental and political challenges, the development of sustainable industrial zones becomes urgent. The competitive advantages offered by the business of industrial clusters need also to be eco-efficient to reduce the environmental impact of economic growth (Kastner et al., 2015). Eco-industrial parks (EIP) constitute a typical illustration of industrial symbiosis and cooperation between different stakeholders in the same area needs to be optimally designed (Boix et al., 2015). Industrial systems are complex in terms of types of material exchanged (water, electricity, vapour...), types of industries involved, size of companies, temporality and many others. Regarding this complexity, this paper aims at developing a generic methodology able to design exchanges between different companies in an EIP. On the other hand, Tian et al. (2014) noted that for different EIPs in China a key measure to support the sustainability of EIPs is the development of the network. Another important factor is the ability of a network to be resilient. In order to reduce the impact of disruptions, resilient network design is a branch of scientific research that aims to develop networks that have a greater propensity to support change. Most particularly, Leo and Engell (2020) took into account

uncertainties where the decisions at one point in time have an impact on the probability of the uncertainty. The general principle of resilience is to maximize the number of interconnections in order to increase the possibilities for distributing flows in the event of a disturbance (Valenzuela-Venegas et al., 2019). As a result, this approach has the bias of seeking to systematically interconnect industries and therefore create dependencies, increase the risk taken, which is a source of reluctance for manufacturers.

In addition, in previous studies, interconnections have been widely studied from the point of view of the complexity of the network but there is an absence of studies regarding resilience. The optimal design of a network by integrating the interdependence generated by the interconnections could be interesting to analyse. Indeed, a powerful obstacle to the development of eco-industrial networks is the trust that companies must have among themselves to accept losing their independence. The work presented in this study provides a new vision since it takes into account the behaviour of industries, wishing to maintain their independence while taking advantage of the advantages they have in coming together on the same site to maximize their benefits (economic, environmental, social). Thus, based on the model developed in a previous study (Mousqué et al., 2020) the methodology proposed may advise industries to integrate the EIP network or to remain autonomous.

The scientific issue here is therefore the development of approaches to design a flexible network that remains optimal in view of the design criteria in a changing operational context. To achieve this, the first objective of this study is to propose an optimal solution for designing a network of exchanges.

2. Methodology

The main novelty of the approach is its ability to consider the great majority of possible exchanges in an EIP:

- Electricity: it can be produced from renewable sources (wind turbines, solar panels) or taken from the external grid network,
- Water: water can be exchanged throughout the network
- Vapour: different levels of pressure are considered through the design of a utility system for each company of the EIP (low, medium and high pressures).

The number of companies and the types of processes can also be various. In this sense, this generic model is able to be adapted to a great diversity of case studies. The multi-objective optimization method can deal with the minimization of total cost, environmental impacts (through a life cycle analysis indicators) or complexity of the network designed (technical aspects linked to interconnections between companies). The aim is to propose simultaneously the optimal design of the Hybrid Renewable Energy System (HRES) feeding the EIP and the exchanges between and within companies while minimizing the objectives previously enounced.

The methodology consists in several steps described in Figure 1:

- Building the superstructure regarding the case study chosen and evaluating the granularity of the model to be developed. It is important to notice that even of a so-called “black box” approach is used; a detailed modelling has to be performed in order to obtain the characteristics of each box of the superstructure.

- Formulation of the problem through a MILP optimization
- Application of the procedure developed to design the network of resources exchanges and evaluation of its flexibility through a multicriteria decision making tool.

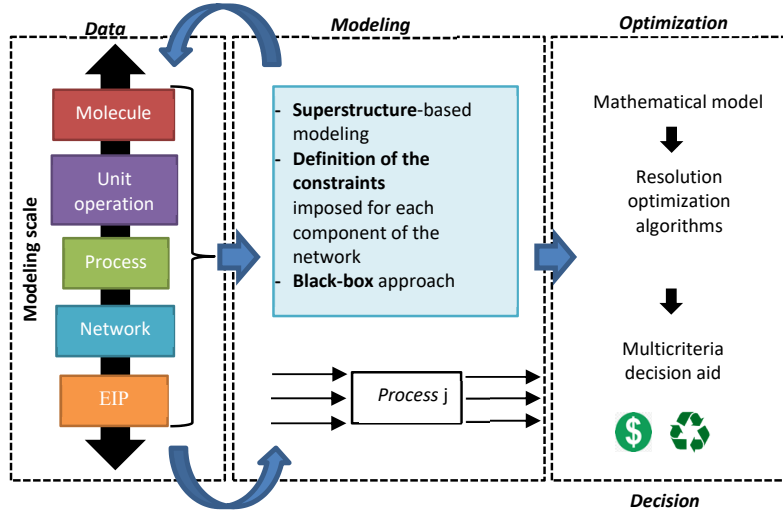


Figure 1. Generic methodology developed

3. Problem statement

Given is a set of industries represented by sinks and demands in terms of vapour, water and electricity throughout four time periods represented by a superstructure previously detailed in Mousqué et al. (2020). The aim is to design the optimal network for making a cooperative symbiosis and to evaluate its flexibility and resilience.

The variables of the optimization model are both continuous and binary: flow-rates (water, energies) exchanged, flow-rates between each box of the superstructure, production of boilers, electricity consumed, existence of each connection and of each component (boilers, turbines). The constraints are represented by all the mass balances, the conservation of materials and minimum and maximum capacity production of each boiler. The MILP model formulation is taken from Mousqué et al. (2020) and solved with ILOG CPLEX solver.

The present paper brings a specific procedure developed and implemented in order to take into account and to evaluate flexibility of the obtained solution (Figure 2). The procedure developed reaches to design all the pipes constituting the exchanges between companies while minimizing the cost through a multi-objective optimization approach. Additionally, the design of the network takes into account a better dependency of companies what makes the EIP more attractive for them.

The objective function to minimize is the total cost of the network calculated through the Net Present Cost (Turkay and Telli, 2011).

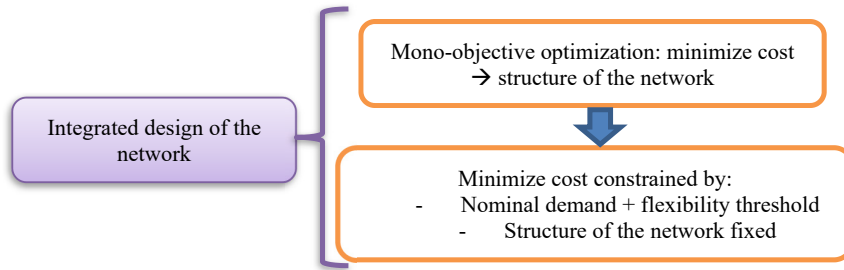


Figure 2. Optimization procedure to consider flexibility of the network

4. Case study

The case study chosen contains 15 different industries grouped in the Yeosu industrial zone in China, this park was previously introduced by Kim et al. (2010). For more details about this particular case study and the mathematical model associated, the reader can refer to Mousqué et al. (2020 b). In this utility network model coupled with a hybrid renewable energy system, steam demand is the key parameter to study the flexibility of the network. Indeed, the capacity to supply the demand of companies with steam is directly linked to the sizing of the production, conversion and transport facilities (boilers, turbines and pipes respectively). On the other hand, the electricity network is connected to the external grid which is considered as an unlimited source, its supply capacity would be unlimited. The flexibility of the connected electricity network being infinite, the study of its flexibility is therefore not relevant.

To detail this case study, nominal demand is discretized over four periods representing the 4 seasons. Each scenario (nominal and disturbances) therefore comprises four periods. In addition, the study parameter taken into account is an industry's vapor demands. This encompasses the demand for very high pressure (VHP), high pressure (HP), medium pressure (MP) and low pressure (LP) steam in an industry. There are therefore 15 input parameters studied given that there are 15 industries. The evaluated parameter therefore corresponds to the smallest maximum deviation among the demands of an industry. To achieve this, the demands of the evaluated industry are variable and the smallest variation in demands is maximized. In fact, during this assessment step, the demands of other industries and all the infrastructures are fixed.

5. Results and analysis

5.1. Flexibility of the nominal solution

The flexibility of the nominal solution has been first evaluated by taking into account the minimal flexibility of each company. Eleven companies among fifteen have a flexibility equal to zero for the nominal solution; that is to say its global flexibility is null. Figure 3 shows the repartition of flexibility for each company of the EIP over the four time periods considered. These results show us that the optimal design in terms of costs leads to a solution that has very little flexibility. This is because the minimization of the cost also entails a minimization of installations which are sized just to the capacity necessary to meet the demand under nominal conditions.

A little flexibility is nevertheless observed in a few industries. This flexibility comes from the turbines installed. In fact, during a disturbance it is possible to stop the condensing turbines, the steam then directly supplies the processes. The additional quantity of steam is then equal to the quantity which would have been lost by condensation in the turbines. More generally, certain elements make it possible to increase the flexibility of a network, such as storage elements or controllable elements that can be stopped. Note that as a function of time, it is over the first period that the model presents the least flexibility, in fact, the flexibility of industries is zero for a large number of industries, nevertheless for industry 2, 4, 12 and 13, this flexibility is respectively 8.6%, 0.7%, 9.3% and 37.7%. The first time step is the period when we observe the least flexibility, this is explained, because it is also the period when the demand is the most important for each process.

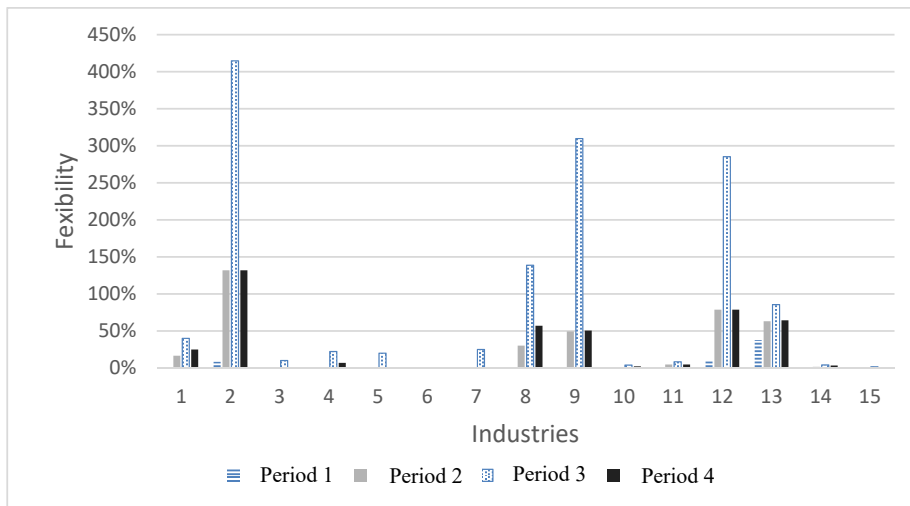


Figure 3. Flexibility of the nominal solution over time periods

5.2. Analyses of the impact of flexibility on the cost

As a reminder, the deviation is then the smallest deviation calculated of the various processes in an industry. The threshold on the flexibility of the optimal solution increases while the cost is minimized. It is important to note that improving flexibility in a whole network can be possible at the design stage by adding some equipment, what leads to additional cost (Figure 4). Regarding the trend of increasing cost versus flexibility, design solutions get results for 20% and 30% flexibility with relatively linear evolution. Indeed, when a flexibility of 30% is taken into account, 7.25% of additional cost is considered to overcome these fluctuations. It implies that, for this solution, all companies will have more than 30% of flexibility over the four time periods.

This overall approach can help to design sustainable industrial systems that can overcome fluctuations in their demands over the time by integrating some additional costs during the conception stage.

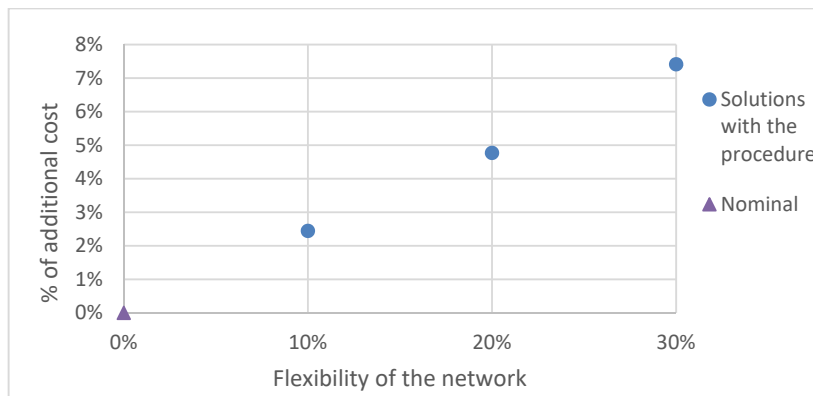


Figure 4. Flexibility of the network regarding cost evaluation

6. Conclusion

A generic procedure to design optimal networks of EIPs has been developed in order to evaluate cooperative exchange network. A generic superstructure that allows to take into account several types of exchanges in an EIP has been developed and the associated MILP model has been adapted to solve large-scale problems. The procedure was validated on a case study involving fifteen industries. Starting from a nominal solution with zero flexibility, one can increase it by adding some cost to overcome potential fluctuations in the network. In addition, in the interest of developing a generic method, this indicator does not require knowing the probabilities of the appearance of the various disturbances.

References

- M. Boix, L. Montastruc, C. Azzaro-Pantel and S. Domenech, 2015, Optimization methods applied to the design of eco-industrial parks: A literature review, *J. Clean. Prod.*, 87, 303–317.
- C.A. Kastner, R. Lau and M. Kraft, 2015, Quantitative tools for cultivating symbiosis in industrial parks; a literature review, *Appl. Energy*, 155, 599-612.
- S.H. Kim, S.G. Yoon, S.H. Chae and S. Park, 2010, Economic and environmental optimization of a multi-site utility network for an industrial complex. *J. Environ. Manage.*, 91, 690–705.
- E. Leo and S. Engell, 2020, A Novel Multi-stage Stochastic Formulation with Decision-dependent Probabilities for Condition-based Maintenance Optimization, *Comp. Aided Chem. Eng.*, 48, 1795-1800.
- F. Mousqué, M. Boix, L. Montastruc, S. Domenech and S. Négny, 2020, Optimal Design of Eco-Industrial Parks with coupled energy networks addressing Complexity bottleneck through an Interdependence analysis, *Comp. Chem. Eng.*, 138, 59.
- F.F. Shein, Z. Liang, D. Wenli, Z. Weimin, Q. Feng, 2020, Large-scale industrial energy systems optimization under uncertainty: a data-driven robust optimization approach, *Appl. Energy*, 259, 114-199.
- J. Tian, W. Liu, B. Lai, X. Li and L. Chen, 2014, Study of the performance of eco-industrial park development in China, *J. Clean. Prod.*, 64, 486-494.
- B.E. Türkay and A.Y. Telli, 2011, Economic analysis of standalone and grid connected hybrid energy systems, *Renew. Energy*, 36, 1931–1943.
- G. Valenzuela-venegas, G. Vera-hofmann and F.A. Díaz-alvarado, 2019, Design of sustainable and resilient eco-industrial parks: Planning the flows integration network through multi-objective optimization, *J. Clean. Prod.*, 10.

Population Balance Modelling of Pan Granulation Processes

Eric Otto^a, Robert Dürr^b, Mateusz Przywara^c, Dorota Antos^c and Achim Kienle^{a,b}

^a*Otto-von-Guericke-University, Universitätsplatz 2, 39106 Magdeburg, Germany*

^b*Max-Planck-Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany*

^c*Rzeszów University of Technology, al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland*

Abstract

Pan granulation is a particle formation process with widespread practical applications such as fertilizer and pharmaceutical manufacturing. The continuous operation mode is especially promising with respect to industrial demands. For process automation and intensification, a suitable dynamical process model is required. Therefore, the focus of this contribution is on identification of agglomeration kernel parameters in a population balance model based on empirical data. To this end, an objective functional, representing the error between model and measurement data, is minimized. It is shown that the steady state particle size distribution of a lab-scale process can be reproduced accurately using the population balance with the identified parameters.

Keywords: Population balance modeling, Pan granulation modeling and simulation, Agglomeration.

1. Introduction

Agglomeration is a particle formation process in which at least two primary particles are combined to form a new one. This principle is often used in many industries, e.g. pharmaceutical manufacturing, food processing and fertilizer production. The properties of the formed agglomerates, e.g. size, shape and porosity, significantly affect certain end-use properties, e.g. dissolubility, processability and storability (Bück and Tsotsas, 2016). In particular for fertilizer production and mineral processing, agglomerates are often formed in drums or pans (Ouchiyama and Tanaka, 1981; Litster and Ennis, 2004) in continuous operation which may provide constant throughput with constant product quality (in terms of specific agglomerate properties) during steady-state operation. Besides sophisticated experiments, modeling those processes in view of emerging agglomerate property distributions supports understanding of the underlying dynamic mechanisms on multiple scales and application of model-based process control and intensification. In this contribution, focus is on modelling of pan granulation, which is also known under the term dish or disc granulation.

The process scheme is shown in Fig 1: a powder is fed constantly to an inclined rotating dish and liquid binder is sprayed on the powder. The wet particles are sticking together, forming larger agglomerates. The flux from the pan represents the product particles. Feed powder as well as product particles vary with respect to individual particle properties. It has been shown that variations of process operation parameters, such as inclination angle, rotational velocity, binder spray rate and composition, significantly affect the product particle properties (Kapur et al., 1993; Kapur and Runkana, 2003; Obraniak and Gluba,

2012). Furthermore, zone formation of different sized agglomerates in the pan is commonly observed (Litster and Ennis, 2004).

It is well-known that for size-enlargement processes, like pan granulation, individual properties, like characteristic volume and porosity, differ from particle to particle. As an alternative to Monte-Carlo modeling approaches (Zhao et al., 2007; Rieck et al., 2018) population balance modeling (PBM) (Ramkrishna, 2000; Ramkrishna and Singh, 2014) represents an established framework to model such processes (Kapur and Runkana, 2003). For standard monivariate PBEs, accounting mostly for characteristic particle size or volume, efficient and accurate numerical solution algorithms have been developed, see e.g. Kumar et al. (2008); Bück et al. (2012) and the references therein.

This contribution aims at modeling of a lab-scale pan granulation setup and model adaption to experimental data in terms of minimization of the error between experimental samples of the particle-volume distribution and the corresponding model predictions in steady state, i.e. for constant feed powder flux and constant product flow.

2. Materials and Methods

2.1. Experimental and Analytical Setup

The experiment was realized in a pilot scale pan granulator shown in Fig. 1 (left). The pan has an inner diameter of 400 mm, and depth of 100 mm, schematically shown in Fig. 1 (right). The primary particles are stored in a tank and then transported into the pan using a vibratory conveyor. The binder is distributed through a nozzle, which is installed in a top-spray configuration at distance of 350 mm above the pan. An external pump and an air compressor ensure the supply of sprayed binder solution. The rotation of the pan is obtained by a motor (Model AR304, ERWEKA GmbH), which allows to adjust the rotational speed in the range of 2 rpm - 80 rpm and the inclination angle of the pan in the range of $0^\circ - 70^\circ$. The control of binder flow, primary particle flow and air flow is obtained using the main control panel (Model CE255, G.U.N.T. Gerätebau GmbH). All components of experimental setup are made of stainless steel and assembled on a metal table.

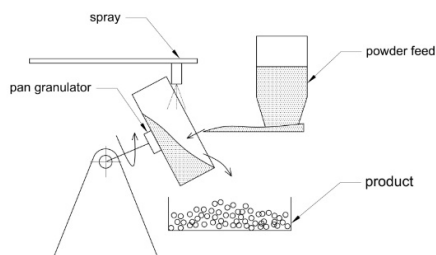


Figure 1: (Left) Experimental pan granulation setup. (Right) Schematic representation of pan granulation process

The particle material is powdered dolomite. It is a grey powder, typically used in the construction industry and as a fertilizer to increase the pH-value of soil. The binder solution contains 1 wt% sucrose and 99 wt% of water. The duration of the experiment

was 75 min with a 50 g/min feed flow rate, and 8,5 g/min binder flow rate. The initial bed mass for the experiment was 400 g. The nozzle air flow pressure was 2 bar. An overview of the process parameters is shown in Table 1. For the offline analysis, 5 particle samples of the pan and 5 samples of the product were taken every 15 minutes after the beginning of the experiment. The particle size distribution (PSD) of each sample was measured offline with a Mastersizer (Malvern Instruments, Malvern, Worcestershire, United Kingdom), which infers particle size via dynamic light scattering. The output data from the analyzer is the normalized volume distribution for each sample and thereby over the process time.

Table 1: Experimental process operation parameters

Parameter	Unit	Value
Initial powder mass	[kg]	0.4
Rotational speed	[rpm]	25
Inclination angle	[°]	35
Binder flow rate	[kg/h]	0.51
Binder composition	[wt%]	1
Primary particle feed rate	[kg/h]	0.3
Average density of initial particles	[kg/m ³]	1300

2.2. Process Modeling

Population balance modeling has proven to be an established concept to describe size enlargement processes like granulation Ramkrishna (2000). Assuming that particle volume is the sole significant particle property, the particle ensemble can be represented by its number density distribution function (NDF) $n(t, v)$ providing information of the number of particles within an infinitesimal volume range. If it is furthermore assumed that size enlargement is dominated by agglomeration, the NDF's dynamics are given by the population balance equation (PBE)

$$\frac{\partial n(t, v)}{\partial t} = \dot{n}_{\text{feed}(t, v)} - \dot{n}_{\text{prod}(t, v)} + \frac{1}{2} \int_0^v \beta(t, u, v - u) n(t, u) n(t, v - u) du - \int_0^\infty \beta(t, u, v) n(t, v) n(t, u) du \quad (1)$$

The corresponding initial NDF is $n(0, v)$. While the left-hand side of Equation 1 describes the dynamics, the first two elements of the right-hand side describe the influx of feed powder and the outflux of product particles from the pan. The feed is further resolved

$$\dot{n}_{\text{feed}}(t, v) = \dot{N}_{\text{feed}}(t) \tilde{n}_{\text{feed}}(v) \quad (2)$$

with $\dot{N}_{\text{feed}}(t) = \text{const.}$ denoting the constant feed rate and $\tilde{n}_{\text{feed}}(v)$ the normalized feed powder NDF. The outflux of product particles $\dot{n}_{\text{out}}(t, v)$ is computed from the measured product mass flow $\dot{m}_{\text{prod}}(t)$ and product volume distribution $q_{3, \text{prod}}(x)$. The last two elements of the right-hand side of Equation (2) denote generation and formation of new particles of volume v by agglomeration of two particles with volumes u and $v - u$. Here,

the agglomeration kernel $\beta(t, u, v)$ contains information about the probability of forming a new agglomerate and is often separated into a volume and time-dependent part

$$\beta(t, u, v) = \beta_0(t)\beta_v(u, v). \quad (3)$$

In general, the volume-dependent part $\beta_v(u, v)$ is a non-negative symmetric function which can be derived either from micro-scale mechanistics, heuristic assumptions or abstract data-driven approaches (see Golovin et al. (2018) and the references therein). The time-dependent coalescence efficiency $\beta_0(t)$ mirrors the effects of the process conditions and operating parameters. In this work, focus is on the steady state operation, thus we assume $d\beta_0/dt = 0$ holds.

3. Model adaption

In the following section the estimation of model parameters is presented. Parameter estimation for (multi-dimensional) population balances has been investigated in great detail in Ramachandran and Barton (2010). For further information regarding techniques and frequently occurring problems the reader is referred to this source.

The unknown kernel parameters \mathbf{p} have to be estimated from experimental data via minimization of a suitable objective function

$$J(\mathbf{p}_{est}) = \|e_u(x, \mathbf{p}_{est})\|_2, \quad (4)$$

where e_u is the error in the weighted particle size distribution (Golovin et al., 2018)

$$e_u(x, \mathbf{p}_{est}) = u_{act}(x) - u_{est}(x, \mathbf{p}_{est}), \quad u(x) = \frac{\pi}{6}x^3n(x), \quad (5)$$

and x represents the characteristic size of the particles. Using local conservation of the particle number, $u_{est}(x, \mathbf{p}_{est})$ is computed from the simulated particle volume distribution. For the simulation the PBE was discretized using the cell-average technique (Kumar et al., 2008). Note that the particles are assumed to be of spherical shape, which is a valid assumption for many particle formation processes. However, for the experiment considered here it has to be checked in future investigations. The experimental distribution $u_{est}(x)$ is computed from the normalized particle size distribution $q_{3,est}(x)$ provided by the Mastersizer measurements and the measured mass in the pan $m_{b,act}$ under the assumption of spherical particle shape and particle material density ρ . In this contribution, only the steady state distribution of the continuous process is considered in the objective function. It is represented by the measurements at $t = 75$ min.

Minimizing the objective function (Eq. 4) using local optimization techniques proved to be difficult for one of the kernel functions since different local minima were found for different initial parameter guesses. Therefore, the optimization was conducted by using a genetic algorithm which is able to provide non-local results (Goldberg, 1989). However, it has to be mentioned that there is no guarantee of finding global minima. The algorithm was implemented in Matlab R2018b using the build-in function `ga`. The parameter search space was restricted to $0 \leq a, b \leq 2$ and $-20 \leq \log_{10}(\beta_0) \leq -4$, based on the author's

experience. The initial population for the genetic algorithm was created randomly within the search space.

4. Results

At first, the proposed model adaption method is used to fit the model with volume independent coalescence kernel $\beta_v(u, v) = 1$. Therefore, only β_0 has to be estimated. The optimal steady state distribution $q_3(x)$ is obtained with $\beta_0 = 2.56 \cdot 10^{-7}$ and depicted on the left-hand side of Fig. 2. The corresponding objective function value amounts to $J = 1.56 \cdot 10^5$. It is seen, that the distribution's mode and the density of large particles are captured, yet significant errors are observed for particles up to 1.5 mm characteristic size. Since this result is unsatisfactory the two-parametric Kapur kernel (Kapur, 1972)

$$\beta_v(u, v) = \frac{(u+v)^a}{(uv)^b} \quad (6)$$

is chosen as coalescence kernel. Due to the free parameters a and b , the kernel possesses the necessary versatility to fit a wide range of agglomeration processes. The optimal steady state volume distribution with the Kapur kernel is presented on the right-hand side of Fig. 2. The objective function value $J = 3.3 \cdot 10^4$ is obtained with the optimal parameters $\beta_0 = 3.76 \cdot 10^{-8}$, $a = 0.92$ and $b = 0.6$. The supremacy of the Kapur-kernel approach, which provides an accurate reproduction of the experimental results, compared to the volume-independent kernel is obvious. In future publications the identified parameters have to be validated with further experiments and thorough analysis of identifiability has to be conducted.

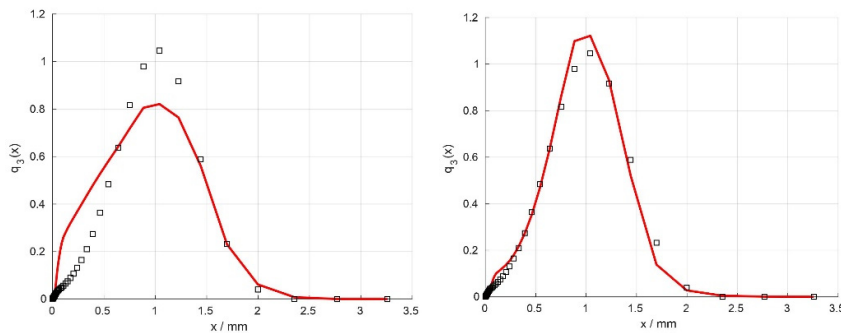


Figure 2: Measured (black) and simulated (red) normalized volume density distributions for the volume-independent kernel (**left**) and the Kapur kernel (**right**).

5. Conclusion

This contribution was concerned with identifying a process model for a continuous pan granulation process. Therefore, agglomeration kernel parameters were identified by fitting a population balance model to lab-scale experimental measurements of particle size distributions by minimizing an error functional. The optimization was implemented as a genetic algorithm. It has been shown that a population balance with a Kapur coalescence

kernel is able to provide very good agreement between model and measurements with respect to the steady state volume distribution of particles. Future work will be concerned with investigating the influence of process conditions on the agglomeration process which is represented by the kernel parameters by adapting the model to additional experiments using the presented methodology (Otto et al., 2020). It is expected, that the Kapur kernel will provide accurate fits. Furthermore, research will include identification of models that also capture the process dynamics. Finally, the identified process models can be used for model-based control and process intensification.

6. Acknowledgements

This work is partly funded by the European Regional Development Fund (ERDF) project "Center of Dynamic Systems". The financial support is hereby gratefully acknowledged.

References

- A. Bück, G. Klaunick, J. Kumar, M. Peglow, E. Tsotsas, 2012. Numerical simulation of particulate processes for control and estimation by spectral methods. *AICHE JOURNAL* 58, 2309–2319.
- A. Bück, E. Tsotsas, 2016. Agglomeration. In: B. Caballero, P. M. Finglas, F. Toldrá (Eds.), *Encyclopedia of Food and Health*. Academic Press, Oxford, pp. 73 – 81.
- D. E. Goldberg, 1989. *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston.
- I. Golovin, G. Strenzke, R. Dürr, S. Palis, A. Bück, E. Tsotsas, A. Kienle, 2018. Parameter identification for continuous fluidized bed spray agglomeration. *Processes* 6 (12).
- P. Kapur, 1972. Kinetics of granulation by non-random coalescence mechanism. *Chemical Engineering Science* 27 (10), 1863 – 1869.
- P. Kapur, P. Kapur, D. Fuerstenau, 1993. An auto-layering model for the granulation of iron ore fines. *International Journal of Mineral Processing* 39 (3), 239 – 250.
- P. Kapur, V. Runkana, 2003. Balling and granulation kinetics revisited. *International Journal of Mineral Processing* 72 (1), 417 – 427.
- J. Kumar, M. Peglow, G. Warnecke, S. Heinrich, 2008. An efficient numerical technique for solving population balance equation involving aggregation, breakage, growth and nucleation. *Powder Technology* 182 (1), 81 – 104.
- J. Litster, B. Ennis, 2004. *The Science and Engineering of Granulation Processes*. Vol. 15. Springer Science & Business Media.
- A. Obraniak, T. Gluba, 2012. A model of agglomerate formation during bed wetting in the process of disc granulation. *Chemical and Process Engineering* 33 (1), 153–165.
- E. Otto, R. Dürr, G. Strenzke, S. Palis, A. Bück, E. Tsotsas, A. Kienle, 2020. Kernel Identification in Continuous Fluidized Bed Spray Agglomeration from Steady State Data (submitted for publication).
- N. Ouchiyama, T. Tanaka, 1981. Kinetic analysis of continuous pan granulation. possible explanations for conflicting experiments and several indications for practice. *Industrial & Engineering Chemistry Process Design and Development* 20 (2), 340–348.
- R. Ramachandran, P. I. Barton, 2010. Effective parameter estimation within a multi-dimensional population balance model framework. *Chemical Engineering Science* 65 (16), 4884–4893.
- D. Ramkrishna, 2000. *Population Balances: Theory and Applications to Particulate Systems in Engineering*. Academic Press, San Diego.
- D. Ramkrishna, M. R. Singh, 2014. Population balance modeling: Current status and future prospects. *Annual Review of Chemical and Biomolecular Engineering* 5, 123–146.
- C. Rieck, M. Schmidt, A. Bück, E. Tsotsas, 2018. Monte carlo modeling of binder-less spray agglomeration in fluidized beds. *AICHe Journal* 64 (10), 3582–3594.
- H. Zhao, A. Maisels, T. Matsoukas, C. Zheng, 2007. Analysis of four monte carlo methods for the solution of population balances in dispersed systems. *Powder Technology* 173 (1), 38 – 50.

Estimation of cutpoint temperature under uncertain feed composition and process conditions using artificial intelligence methods

Junaid Shahzad,^a Iftikhar Ahmad^{a,*}

School of Chemical and Materials Engineering (SCME), National University of Sciences and Technology (NUST) H12, Islamabad, Pakistan
iftikhar.salarzai@scme.nust.edu.pk

Abstract

Crude distillation unit (CDU) is the most energy-consuming and energy deficient process in a refinery, hence saving energy consumption in its operation is the main target. The CDU separates crude oils into light gas streams, naphtha, kerosene, diesel, atmospheric gasoil and residue. The cutpoint temperature management dictates the quantity and quality of the products, therefore a proper model for determination of its causes and effects is fundamental. The cutpoint temperature heavily depends on the composition of crude-oil and process conditions. In this study, the effect of uncertainty in process conditions, such as temperature, pressure, flow rates, etc., and feed composition on cutpoint temperature is researched. A hybrid approach based on Taguchi method and genetic algorithm is used to derive optimum cutpoint temperature for different variants of feed composition and process conditions. Then, an artificial neural networks (ANN) model is developed to predict the optimum cutpoint temperature by eliminating the need for the hybrid approach. The method is validated with two Pakistani crudes oils from the Zamzama field, where the ANN predicted the cutpoint temperature with enough accuracy for industrial application.

Keywords: Deep learning, ANN, Cutpoint temperature optimisation, CDU.

1. Introduction

The efficient design and operation of petroleum refining processes have always been researched due to their energy-intensive nature. The most important refinery process is found in crude distillation units (CDU) that separate the petroleum in its distillates. CDU is one of the most complex process in the refinery, hence there is a huge opportunity for saving energy consumption in its operation (Brueske et al., 2015). It may count on optimisation of scheduling, reduced giveaways in properties of intermediate products, as well as real-time optimisation of the cutpoint temperature of the CDU (Luyben et al., 2013). The cutpoint temperature heavily depends on composition of crude-oils and process conditions. Several methods like categorisation approach, swing-cut modeling, fractionation index model, Taguchi method, weight transfer ratio (WTR) approach, Taguchi and genetic algorithm (GA) hybridization, and monotonic interpolation have been devised for the cutpoint temperature optimisation (Brooks et al., 2018; Trierwiler et al., 2001; Zhang et al., 2001; Li et al., 2007; Alattas et al., 2011; Ali et al., 2012; Menezes et al., 2013; Kelly et al., 2014; Durrani et al., 2018; Franzoi et al., 2020).

In Durrani et al. (2018a), an artificial neural networks (ANN) model capable of multi-outputs is developed from optimised data sets which are generated by the hybrid

framework of the Taguchi method and genetic algorithm (GA) to deal with uncertain feed composition. The current work is an extension of this previous approach, whereby the effect of uncertainty in all process conditions, i.e., temperature, pressure, flow rates, etc., are considered in addition to the uncertainty in feed composition adopted in Durrani et al. (2018b). The application domain is extended to crude-oil assays of the Pakistani oil fields known as Zamzama. Several hundred variations of the Zamzama assay are generated by incorporating artificial uncertainty in the actual crude-oil composition and process conditions. The sample components having variations in crude-oil composition with different cutpoint temperatures and process conditions are derived from the hybrid approach, i.e., Taguchi and GA. The sample components comprised of the crude-oil variants and their cutpoints are being used to train an ANN model. The ANN model has a 0.98 correlation coefficient and required lesser computational power compared to the hybrid method in the prediction of the desired cutpoint temperatures.

In this study, variations will be introduced in 10 process conditions. They are: the standard ideal liquid volumetric flow, temperature and pressure of the Zamzama crude-oil feed to the pre-flashing stage, the pressure of the diesel steam fed to the crude distillation column, the mass flow rate, temperature and pressure of the atmospheric gas oil steam fed to the crude distillation column, the mass flow rate, temperature and pressure of the main steam fed to the crude distillation column, and the operating pressure of the condensing stage in the distillation column, along with the feed compositions. The different sets of data generated by varying the process conditions and feed compositions will be fed to the hybrid model developed in ASPEN HYSYS. The data sets of these uncertain process conditions will be used to generate cutpoints at those different variants. The new sets of optimised cutpoints are used to train and develop an ANN-model.

The process description and process flow along with details of the crude distillation are shown in section 2. The methodology used for generating data from the CDU by the hybrid framework to train and developing the ANN-based model is discussed in section 3. Results for this study are presented in section 4. Section 5 shows the findings from this research.

2. Process description

A representation of the CDU and the process flow can be seen in Figure 1. The selected crude-oils are from Pakistan, from the Zamzama fields, which are sweet, light, and have a specific gravity varying from 0.75 to 0.76. The sulphur and water content are less than 0.05% by weight (for sulphur) and volume (for water). The crude-oil is fed at the temperature of 232.22°C and pressure of 517.107 kPa into the pre-flash column. The pre-flashing stage separates the gases from the liquid phase, reducing the overall duty required in the furnace to promote the separation in the CDU. The furnace pre-heats the crude-oil for the desired fractionation in the distillation tower. The gas stream, which are separated in the pre-flash stage, are mixed with the hot crude-oil stream in a mixer before reaching the a fractionation column. The column has a total of 29 theoretical trays with three strippers to the side and pre-flash condensers along the cutpoint stages and for internal reflux three pumps around installed. The feed is fed at the tray 28 of the column.

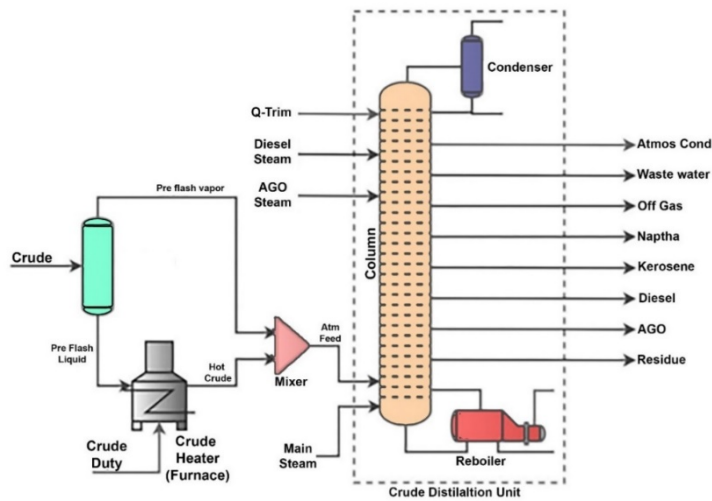


Figure 1. Crude distillation unit.

3. Methodology

The hybrid framework is developed from Taguchi and genetic algorithm in Durrani et al. (2018). The method now is extended to other process conditions of the crude feed and column streams. Variations of Zamzama assay are generated by incorporating artificial uncertainty in the actual crude-oil composition and ten process conditions. The process conditions, which are artificially varied, are listed below:

- Standard ideal liquid volumetric flow, temperature, and pressure of the Zamzama crude-oil feed at the pre-flashing stage.
- The pressure of the diesel steam.
- The mass flow rate, temperature and pressure of the atmospheric gas oil steam.
- The mass flow rate, temperature, and pressure of the main steam.
- The operating pressure of the condensing stage in the CDU.

The main steps for developing and testing are shown in Figure 2 are as follows:

Phase I: Several process data sets are generated at different feed and process conditions by inserting artificial variations which would deviate these feed and operating conditions by a factor of $\pm 1\%$, $\pm 2\%$ and $\pm 3\%$.

Phase II: Cutpoint temperatures are now calculated for these variants through the hybrid framework of Taguchi and GA.

Phase III: ANN model is developed using 75% of the data sets with their associated cutpoints.

Phase IV: The model developed has to be validated now by predicting the cutpoints of the remaining 25% of the data sets with varied feed and process conditions.

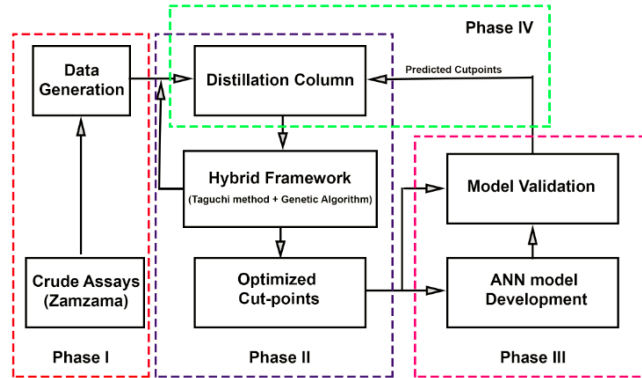


Figure 2. Schematic representation of how the hybrid framework and ANN model are integrated.

4. Result and discussions

In Table 1 is showed the ten different operational variables from 6 different datesets that are varied from their given standard values.

Table 1. Process condition data sets for distillation of Zamzama crude-oils.

Process Condition	Data set 1	Data set 2	Data set 3	Data set 4	Data set 5	Data set 6
Zamzama Mass Flow (KBPD)	97.000	95.060	94.109	95.051	96.952	99.860
Zamzama Temperature (°C)	225.234	220.729	218.522	220.707	225.121	231.875
Zamzama Pressure (kPa)	554.064	542.983	537.553	542.928	553.787	570.401
Condenser Pressure (kPa)	131.752	129.117	127.826	129.104	131.686	135.637
Steam Mass Flow (kg/h)	3299.940	3233.941	3201.602	3233.618	3298.290	3397.239
Steam Temperature (°C)	184.882	181.184	179.373	181.166	184.790	190.333
Steam pressure (kPa)	1002.980	982.920	973.091	982.822	1002.479	1032.553
AGO Steam Mass Flow (kg/h)	1099.980	1077.980	1067.201	1077.873	1099.430	1132.413
AGO steam Temperature (°C)	144.433	141.544	140.129	141.530	144.361	148.692
AGO steam Pressure (kPa)	334.359	327.672	324.395	327.639	334.192	344.218

Variations in the data set 1 are generated by addition of variation of -3%. In data set 2, a change of -2% is introduced. Similarly, for data set 3 until 6, a change of -1%, 1%, 2%, and 3% is applied, respectively. An ANN model is developed using data generated for feed compositions and 10 process conditions for the CDU. From the 198 data sets generated, 148 are used for training the ANN-based model, while 50 data sets are used to validate the model. The ANN model has three hidden layers, where neurons in hidden layers 1, 2, and 3 are 13, 17, and 16, respectively. The correlation coefficient between the target cutpoint and the predicted cutpoint is 0.98.

The reported average value for diesel production in KBPD (kilo barrels per day) in Figure 3 and the individual productions capacity of the sample components are calculated for a set of 50 sample components. The trend for straight runs model shows that production for

33 sample components are reported to be above average, among the sample components for straight runs a maximum of 36.79 KBPD and a minimum of 35.29 KBPD is achieved. For the hybrid model, a maximum of 42.77 KBPD and a minimum of 37.37 KBPD is obtained. For the ANN-based model, the diesel production for 50 sample components reaches an average of 39.38 KBPD for an average energy/volume (E/V) value of 779.67 kW/KBPD from Figure 4 for the ANN-based model. The ANN-based model shows that production for 29 sample components are above average, among the sample components for the ANN-based model a maximum of 41.06 KBPD and a minimum of 36.35 KBPD is achieved. Figure 5 shows the energy to volume ratio of diesel for different sample components. The hybrid model obviously requires a longer time to update the cut points while the ANN-based model is aimed to reduce the time required for updating cutpoints.

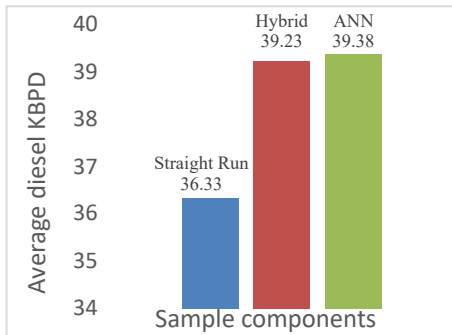


Figure 3: Average diesel production in E/V.

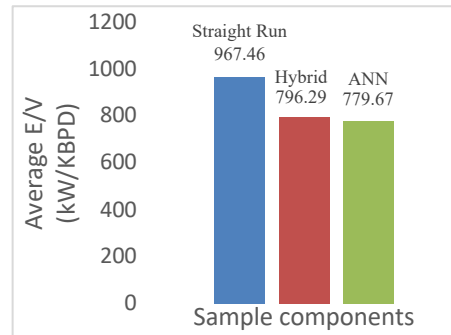


Figure 4: Average E/V values.

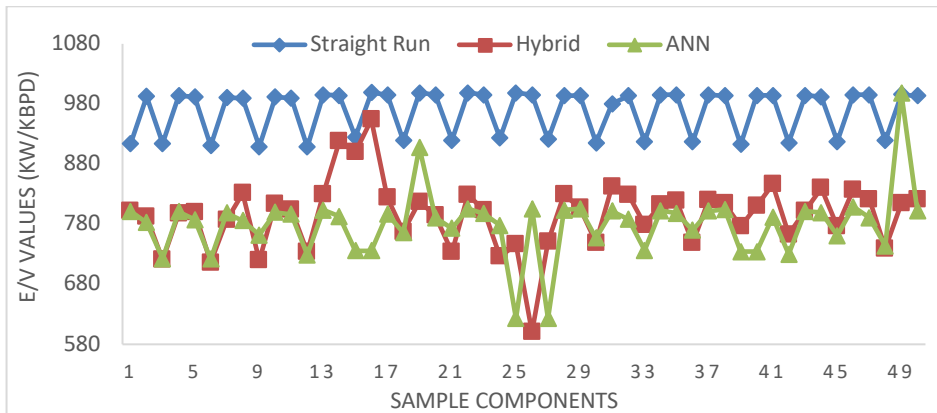


Figure 5: E/V ratio for diesel achieved using the three models.

5. Conclusion

In this study an ANN-based approach for cutpoint optimisation of CDU is proposed. The ANN-based model is trained on data of sets of feed compositions, process conditions, and their corresponding optimised cutpoints derived through hybrid model of Taguchi method and GA. The proposed mechanism is demonstrated via integration of MATLAB and Aspen HYSYS and is applied to a CDU model for Pakistani crude-oils. The ANN model and the hybrid framework are found more reliable and consistent for the diesel production in comparison to the straight-run model, which showed lower production rates of diesel and an overall lower total production. It is found that using the hybrid model, a decrease

of 17.69% in the energy to volume ratio for an increased diesel production of 7.98% is achieved, compared to the straight runs. On the other hand, using the ANN-based model a decrease of 19.44% in the energy to volume ratio for an increased diesel production of 8.39% is achieved compared to the straight-run model.

References

- A.M. Alattas, I.E. Grossmann, I. Palou-Rivera, 2011, Integration of nonlinear crude distillation unit models in refinery planning optimization, *Industrial and Engineering Chemistry Research*, 50(11), 6860-6870.
- S.F. Ali, N., Yusoff, 2012, Determination of optimal cut point temperatures at crude distillation unit using the taguchi method.
- R.W. Brooks, F.D. Van Walsemand, J. Drury, 1999, Choosing cut-points to optimize product yields, *Hydrocarbon Processing*, 78(11), 53-60.
- S. Brueske, C. Kramerand, A. Fisher, 2015, Bandwidth study on energy use and potential energy saving opportunities in us pulp and paper manufacturing (No. DOE/EE-1232). *Energetics*.
- M.A. Durrani, A. Avila, J. Rafael, I. Ahmad, 2018a, An integrated mechanism of genetic algorithm and Taguchi method for cut-point temperatures optimization of crude distillation unit. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 1-6.
- M.A. Durrani, I. Ahmad, M. Kano, S. Hasebe, 2018b, An artificial intelligence method for energy efficient operation of crude distillation units under uncertain feed composition, *Energies*, 11(11), 2993.
- R.E. Franzoi, B.C. Menezes, J.D. Kelly, J.W. Gut, I.E. Grossmann, 2020, Cutpoint temperature surrogate modeling for distillation yields and properties, *Industrial and Engineering Chemistry Research*, 59(41), 18616-18628.
- S.A. Kalogirou, 2001. Artificial neural networks in renewable energy systems applications: a review, *Renewable and Sustainable Energy Reviews*, 5(4), 373-401.
- S.A. Kalogirou, 2004, Optimization of solar systems using artificial neural-networks and genetic algorithms, *Applied Energy*, 77(4), 383-405.
- J.D. Kelly, B.C. Menezes, I.E. Grossmann, 2014, Distillation blending and cutpoint temperature optimization using monotonic interpolation. *Industrial and Engineering Chemistry Research*, 53(39), 15146-15156.
- W. Li, C.W. Hui, I.A. Karimi, R. Srinivasan, 2007, A novel CDU model for refinery planning. *Asia-Pacific Journal of Chemical Engineering*, 2(4), 282-293.
- W.L. Luyben, 2013, *Distillation design and control using Aspen simulation*. John Wiley & Sons.
- B.C. Menezes, J.D. Kelly, I.E. Grossmann, 2013, Improved swing-cut modeling for planning and scheduling of oil-refinery distillation units, *Industrial and Engineering Chemistry Research*, 52(51), 18324-18333.
- D. Panda, M. Ramteke, 2018, Reactive scheduling of crude oil using structure adapted genetic algorithm under multiple uncertainties, *Computers and Chemical Engineering*, 116, 333-351.
- D. Trierwiler, R.L. Tan, 2001, *Advances in crude oil LP modelling*. National petrochemical & refiners association: Singapore; 52-58.
- W. Yaïci, E. Entchev, 2016, Adaptive neuro-fuzzy inference system modelling for performance prediction of solar thermal energy system, *Renewable Energy*, 86, 302-315.
- J. Zhang, X.X. Zhu, G.P. Towler, 2001, A level-by-level debottlenecking approach in refinery operation, *Industrial and Engineering Chemistry Research*, 40(6), 1528-1540.

Application of Constrained EKF based State Estimation to a Coiled Flow Inverter Copolymerization Reactor

Robin Semrau^a, Filippo Tamagnini^a, Alexandru Tatulea-Codrean^a and Sebastian Engell^a

^a*TU Dortmund University, August-Schmidt-Straße 1, 44227 Dortmund, Germany*

Abstract

The performance of the reconstruction of the states of the controlled system is a key factor for the performance of nonlinear model-based control. In this work, the design and experimental evaluation of a Constrained Extended Kalman Filter (CEKF) for a continuous copolymerization process is presented. The experimental set-up and the model are introduced. The performance of the CEKF scheme with a systematic tuning, sampling-based constraint handling approach is tested in simulation studies, and the performance of the CEKF formulation is validated for experimental data.

Keywords: Copolymerization, State Estimation, Constrained EKF, Coiled Flow Inverter

1. Introduction

The production of polymers by free radical solution polymerization is a major branch of the process industry. In most cases, free radical polymerization is carried out in batch or semi-batch reactors. However, these modes of operation have several disadvantages, such as cleaning and transfer times, variations in product quality between different batches, and limited heat removal. Therefore, the transfer to continuous processes has been investigated intensely in recent years as presented by Goerke et al. (2016); Durand and Engell (2016).

The challenge of radical solution polymerization in continuous reactors is to handle the increasing viscosity and to avoid runaway and a resulting blockage of the reactor. This necessitates fast and precise control of the reactor, which is difficult because of the lack of measurements inside the reactor. Therefore, reliable and fast state estimation is a must.

The Extended Kalman Filter (EKF) is the most often used state estimator for nonlinear systems in the field of chemical engineering. The popularity of the EKF results from the simplicity of the implementation and the low computational effort. On the other hand, it is known that the EKF may fail for strongly nonlinear processes, especially in the vicinity validity limits, e.g. negative concentrations. Therefore, a significant amount of research regarding the handling of inequality constraints in EKF based state estimation has been done as presented by Simon (2010). Possible approaches range from simple clipping to sampling-based approaches as presented by Prakash et al. (2014).

2. Experimental Setup

A sketch of the experimental setup is shown in Figure 1. The feed streams are aqueous solutions of the monomers acrylamide (AM) and 2-acrylamido-2-methyl-propane sulfonic acid (AMPS) and the redox initiator system potassium persulfate and sodium formaldehyde sulfoxylate. The different streams are mixed and subsequently fed to a Coiled Flow Inverter reactor (CFI). The CFI provides a homogeneous radial profile due to good radial mixing and low axial mixing even at laminar flow conditions as described by Klutz et al. (2015). The throughput of the reactor ranges from 0.5 kg/h to 1.5 kg/h. The reactor has an inner diameter of $d_i=3$ mm and a total length of 8 m. The reactor is placed inside an oven, which is operated at 343 K. Within the reactor, the cold inlet stream is heated and thereby the polymerization initiated. Additionally, the exothermic polymerization reaction leads to a temperature rise.

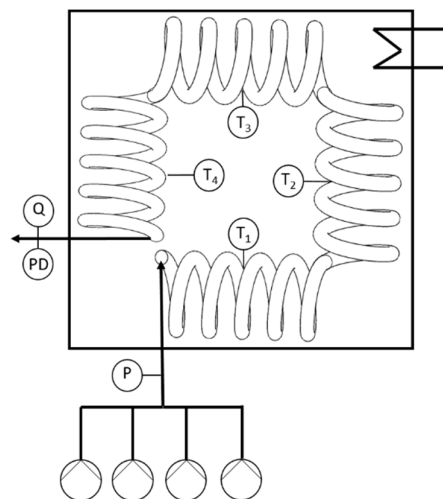


Figure 1: Schematic of the CFI Copolymerization process including measurement set-up

The reaction temperature is monitored by tube temperature measurements. These measurements are located at axial reactor positions of 1 m, 3 m, 5 m and 7 m. The monomer concentration at the outlet is measured using an online Raman spectrometer. With increasing conversion, the viscosity of the reaction mixture increases drastically. The viscosity is inferred from pressure difference measurements. Therefore, the pressure drop over the whole reactor and at the outlet, over a measuring section, are measured.

3. Model

The copolymerization reaction is described by the method of moments and the model accounts for the initiation, propagation, oxygen inhibition, and termination with first and second-order reactions. The energy balances of the reaction fluid and the reactor material are considered including convection, reaction, and heat transfer.

The viscosity dependency on the polymer moments is modeled with a Mark-Houwink approach and the shear dependency with a power law fluid. Back-mixing in the CFI is described with the axial dispersion model. The spatial derivatives of the highly convective dominated mass transport equation and the energy transport equations are approximated using the WENO-3-5 scheme, on 18 grid points. The resulting ode model with 324 states poses a highly nonlinear and stiff problem.

4. Constrained Recursive State Estimation

The Extended Kalman Filter is used as a state estimator. In order to process the available measurement information, the recursive state estimation needs a smaller computation time than the measurement interval, which is 1 s. This cannot be realized with computationally more expensive techniques such as Moving Horizon Estimation or Particle Filtering. Therefore, the EKF is used due to the comparably low computational effort. The performance of the EKF is enhanced by several extensions, such as effective constraint handling and systematic tuning, as described in the following.

4.1. The Extended Kalman Filter

The continuous-discrete EKF employed here consists of a prediction and a correction step. In the prediction, problem (1) is solved to compute the a-priori estimates of the states $x_{k|k-1}$ and the covariance $P_{k|k-1}$.

$$\dot{x} = f(x, u) \quad (1a)$$

$$\dot{P} = \frac{\partial f}{\partial x} P + P \frac{\partial f^T}{\partial x} + Q \quad (1b)$$

In the correction step, the a-priori estimates are corrected with the measurement information y_k using the formula (2) to compute the a-posteriori estimates $x_{k|k}$, $P_{k|k}$.

$$K_k = P_{k|k-1} \frac{\partial h}{\partial x} \Big|_{x_{k|k-1}} \left(\frac{\partial h}{\partial x} \Big|_{x_{k|k-1}} P_{k|k-1} \frac{\partial h}{\partial x} \Big|_{x_{k|k-1}}^T + R_k \right)^{-1} \quad (2a)$$

$$x_{k|k} = x_{k|k-1} + K_k (y_k - h(x_{k|k-1}, u)) \quad (2b)$$

$$P_{k|k} = \left(I - K_k \frac{\partial h}{\partial x} \Big|_{x_{k|k-1}} \right) P_{k|k-1} \left(I - K_k \frac{\partial h}{\partial x} \Big|_{x_{k|k-1}} \right)^T + K_k R_k K_k^T \quad (2c)$$

4.2. Systematic tuning approach

The tuning of an Extended Kalman Filter is not trivial. Especially for large complex models an inappropriate choice of the tuning parameters leads to suboptimal performance and trial-and-error tuning is tedious. A systematic tuning approach that was introduced by Valappil and Georgakis (2000) is used here, as shown in equation (3). The parameter covariance matrix C_p results from a parameter fitting of the model to experimental data. $\frac{\partial f}{\partial p}(x, u)$ is evaluated at every time point in the covariance prediction.

$$Q(x, u) = \frac{\partial f}{\partial p} C_p \frac{\partial f^T}{\partial p} \quad (3)$$

4.3. Inequality constraint handling

In the case considered here, the state estimates are strongly dependent on the pressure drop near the constraints. Therefore, an effective constraint handling stabilizes the EKF and creates the basis a suitable filtering of the pressure drop measurements. In contrast to other approaches that correct only the estimated states, in the sampling-based approach also the state covariance is adapted. In this approach, referred to as MEKF-1 and introduced by Prakash et al. (2014), the constraints are handled in the following way. First, random sample points $\{x_{k|j,i}\}$ from a multivariate normal distribution $\mathcal{N}(x_{k|j}, P_{k|j})$ with the mean $x_{k|j}$ and covariance of $P_{k|j}$ are drawn. Then, the sample points that violate the constraints are clipped to the constraints. Based on this corrected ensemble $\{\tilde{x}_{k|j,i}\}$ the state $\tilde{x}_{k|j}$ and the covariance matrix $\tilde{P}_{k|j}$ are calculated using the equations (4). The

approach is applied after the prediction and after the correction step, accordingly $\tilde{x}_{k|j}$ is the constraint a-priori state estimate $\tilde{x}_{k|k-1}$ or the a-posteriori state estimate $\tilde{x}_{k|k}$.

$$\tilde{x}_{k|j} = \frac{1}{N} \sum_{i=1}^N \hat{x}_{k|j,i} \quad (4a)$$

$$\tilde{P}_{k|j} = \frac{1}{N-1} \sum_{i=1}^N (\hat{x}_{k|j,i} - \tilde{x}_{k|j})(\hat{x}_{k|j,i} - \tilde{x}_{k|j}) \quad (4b)$$

4.4. Computational aspects

The CEKF is implemented in the do-mpc framework, introduced by Lucia et al. (2017), which uses the CasADi 3.5.1 implementation by Andersson et al. (2018). The continuous-time state covariance update is computed using the approach by Mazzoni (2007). To enable the real-time computation, the covariance prediction is performed on a regular grid, with a time step of 0.1 s. The number of sampling points for MEKF-1 was chosen as 3240. The performance was also tested with 32400 sampling points, which did not lead to a different solution. All computations are performed on an AMD Ryzen 9 3950.

5. Results

First, the estimation is tested against a simulation of the plant without model mismatch. The estimator is initialized at the steady state for a throughput of 0.8 kg/h. The plant is initialized at a steady state for a throughput of 0.7 kg/h. At the begin of the simulation, the throughput is increased to 0.8 kg/h. The measurements, which are used in the estimation, are the tube temperatures along the reactor, the mass fraction of the residual monomers at the end of the reactor, the pressure drop over the reactor, and the pressure drop at the reactor outlet. The behavior of the plant and the EKF is shown in Figure 2. The black line displays the mean and the blue line the mean of the a-priori predicted measurements / states over 25 runs with different random measurement noise. The shaded areas indicate the 2σ confidence intervals. In the upper and mid plots, the different measurements are displayed, while in the lower plot the weight average chain length (WACL) and the acrylamide polymer mass fraction at the reactor outlet are shown. In all of the plots, the convergence of the estimator can be seen. However, while the temperature and pressure measurements are predicted accurately within 200 s, the convergence of the residual acrylamide mass fraction and the WACL and the polymer mass fraction takes 400 s. The overall convergence of the estimator takes also about 400 s. The correct prediction of the pressure drop, at the beginning is caused by the fact that the estimator corrects the first and second moment of the polymer chains such that the prediction of the pressure is correct, but the estimates of the states are differing from the true states. The convergence time of the estimator cannot be accelerated by a different tuning, it is an internal limit of the system, which is caused by the convective information transport. Before residence time (316 s) has passed, the system is not completely observable, due to the convective information transport to the pressure and concentration measurements and the limited observability from temperature information. However, while converging, the estimator is able to provide accurate predictions of the pressure drop and consequently the viscosity, which is the main indicator for product quality and safety of the plant. The estimator is fast in comparison to the plant behavior, which does not reach a steady state within 1000 s. The computational time of the estimator is 0.84 ± 0.084 s, therefore, the estimation scheme is real-time feasible.

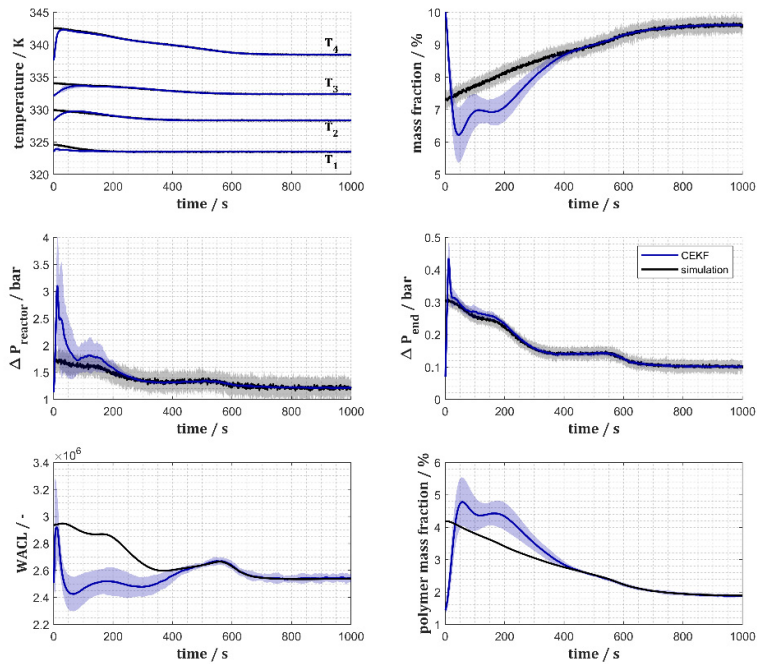


Figure 2: Simulated plant measurements and predicted measurements of the CEKF state estimation

The estimation scheme was also tested with real plant measurements as shown in Figure 3. The results show a fast convergence of the temperatures to the measured values of the tube temperature measurements within 30 s, as can be seen in the upper left plot. The pressure at the end of the reactor also shows similar performance with fast convergence

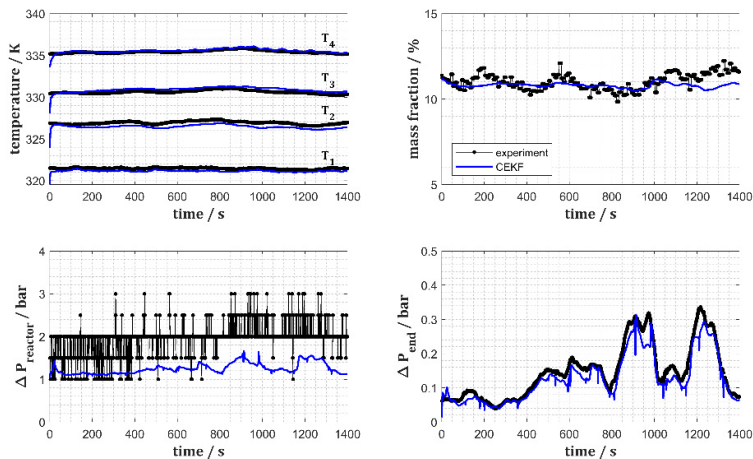


Figure 3: Plant measurements and predicted measurements of the CEKF state estimation

of the estimate. The monomer mass fraction is initialized close to the measured value and the estimates stay close to the measured value. Although there are only minor variations in the monomer mass fractions, the pressure drop shows quite large variations. The reason for this behavior is the high sensitivity of the pressure drop at the outlet to process variations. The total pressure drop over the reactor displays an offset and obviously, the measurement is not reliable. This calls for improvements in the instrumentation. In conclusion, the estimator works reliably with the plant data. In future work, it will be explored whether the performance of the estimator can be improved by a reduced plant model mismatch achievable e.g. through parameter estimation and by using computationally more demanding methods as moving horizon estimation can improve the estimator performance. Nonetheless, the convection dominated behavior of the reactor poses limits to the speed of convergence of the estimation, as also observed in Hashemi et al. (2016).

6. Conclusion

In this work, a CEKF based state estimation scheme for a tubular reactor for a polymerization that leads to a sharp rise of the viscosity was developed and tested in simulation studies and against plant data. The CEKF scheme was applied with a systematic tuning and constraint handling approach and an accurate covariance matrix prediction. Its performance was investigated in simulation studies and with real plant measurements. The speed of convergence is limited by the convection-dominated behavior of the plant. Further extensions of this work include efforts in reducing the plant model mismatch by including parameter estimation and using the estimator in output feedback model-based control of the plant.

References

- Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., Diehl, M., jul 2018. CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation* 11 (1), 1–36.
- Durand, A., Engell, S., jul 2016. Batch to conti transfer of polymer production processes. *Macromolecular Reaction Engineering* 10 (4), 308–310.
- Goerke, T., Kohlmann, D., Engell, S., jun 2016. Transfer of semibatch processes to continuous processes with side injections-opportunities and limitations. *Macromolecular Reaction Engineering* 10 (4), 364–388.
- Hashemi, R., Kohlmann, D., Engell, S., jan 2016. Optimizing control and state estimation of a continuous polymerization process in a tubular reactor with multiple side-streams. *Macromolecular Reaction Engineering* 10 (4), 415–434.
- Klutz, S., Kurt, S. K., Lobedann, M., Kockmann, N., mar 2015. Narrow residence time distribution in tubular reactor concept for reynolds number range of 10–100. *ChERD* 95, 22–33.
- Lucia, S., Tătulea-Codrean, A., Schoppmeyer, C., Engell, S., mar 2017. Rapid development of modular and sustainable nonlinear model predictive control solutions. *Control Engineering Practice* 60, 51–62.
- Mazzoni, T., oct 2007. Computational aspects of continuous–discrete extended kalman-filtering. *Computational Statistics* 23 (4), 519–539.
- Prakash, J., Huang, B., Shah, S. L., jun 2014. Recursive constrained state estimation using modified extended kalman filter. *Computers & Chemical Engineering* 65, 9–17.
- Simon, D., aug 2010. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory & Applications* 4 (8), 1303–1318.
- Valappil, J., Georgakis, C., feb 2000. Systematic estimation of state noise statistics for extended kalman filters. *AIChE Journal* 46 (2), 292–308.

Towards Automated Information Retrieval of Process Data and Knowledge from Academic Databases

Fabian Lechtenberg^a, Javier Farreres^b, Ana Somoza-Tornos^{a,c}, Adrián PachecoLópez^a, Antonio Espuña^a and Moisès Graells^{a,*}

^a*Chemical Engineering Department, Universitat Politècnica de Catalunya, EEBE, C/ Eduard Maristany 16, Barcelona 08019, Spain*

^b*Computer Science Department, Universitat Politècnica de Catalunya, EEBE, C/ Eduard Maristany 16, Barcelona 08019, Spain*

^c*Renewable and Sustainable Energy Institute, University of Colorado Boulder, Boulder, Colorado 80303, United States*

**moises.graells@upc.edu*

Abstract

Process modeling requires both data (chemical reaction yields, kinetic constants, cost estimates, environmental indicators, etc.) and knowledge (operation models and formulations, alternative processes and technologies, etc.). Searching in databases and published research may provide such information, but there is a lack of systematic methods and tools guiding this procedure. The present work describes and assesses an information retrieval methodology that is part of a proposed retrieval and extraction cycle addressing this problem. Two query construction methods for sampling academic databases are proposed assessed and compared. Departing from a seed corpus of a limited number of papers, Scopus® is used as an academic database to retrieve literature containing information associated with pyrolysis processes of waste plastic. It is found that, with minimal human intervention, the methodology is able to return a ranked list of candidate documents that have a considerable (linguistic) relevance.

Keywords: Information Retrieval, Big Data, Text Mining, Academic Databases, Waste-to-Resource

1. Introduction

Due to new communication technologies the world becomes increasingly interconnected.

This enables researchers from all around the globe to publish their work and access publications from their peers. In recent years this has led to a yearly growth of about 9% in academic publications (Landhuis, 2016). The abundance of available information raises the need for strategies to handle this Big Data in order to pinpoint the truly relevant information. In this work we present a methodology to semi-automatically screen academic databases to obtain a set of promising documents related to a given research question.

In Process Systems Engineering (PSE), the modeling of technical systems is a fundamental task. Model parameters and knowledge can be obtained from publications in academic databases but with the ever increasing volume of data this becomes an increasingly difficult task. Text mining and Natural Language Processing are tools that emerged to handle and make use of Big Data. In the field of Systematic Reviews they are being applied to find and classify relevant contributions within a (sub-) field of a discipline in order to include them in a review paper (Usai et al., 2018). A related example where this concept is applied in a natural science field is shown in the work by (Kottmann et al., 2010). They present an IRE system that retrieves, classifies and extracts papers related to metagenomics marine science from the PubMed database.

This work is motivated by the waste-to-resource route assessment work presented by (Pacheco-López et al., 2020). The authors compare the potential of transformation routes of plastic waste to valuable raw materials. By applying an IRE process it is expected to populate a process database for comparison and go beyond what a manual retrieval procedure could achieve, both in terms of retrieved volume and identification of non-intuitive relations. In order to facilitate the search process in a systematic way we propose an Information Retrieval and Extraction (IRE) cycle that departs from an initial set of relevant documents. The main novelty of this contribution is the proposal and validation of a retrieval method to screen access-limited academic databases with the goal of retrieving specific parameters (in this case the ones that characterize chemical transformation processes as part of waste-to-resource routes). Here, we demonstrate our progress in the first step of the cycle, the information retrieval.

2. Problem Statement

To address this objective the problem can be stated as follows:

Given (1) a research question, (2) a set of relevant documents (seed corpus) and (3) an academic database, find an extended corpus (new relevant documents).

The extended corpus must be similarly relevant to the research question and a design objective is to postpone and so to reduce human interaction as much as possible.

3. Methodology

The proposed IRE cycle is presented in Figure 1. First, an academic database is queried using a sampling method that is conditioned to the initial set of documents (seed corpus). Second, the retrieved documents are analyzed in order of decreasing estimated relevance to find the text passages that contain the desired information. So far, all parts of the information retrieval part are implemented in Python 3.8. The database used in this study is Scopus® and is accessed via the Elsevier API. The following subsections describe the hypotheses and individual steps within the information retrieval part of the cycle.

3.1. Hypotheses

1. Linguistically similar texts contain semantically similar information.
2. A seed corpus composed of documents fitting a domain represents that domain.

3. Using relevant keywords extracted from that corpus for constructing queries to structured sets of documents, organized as databases, allows for finding documents having semantically similar information

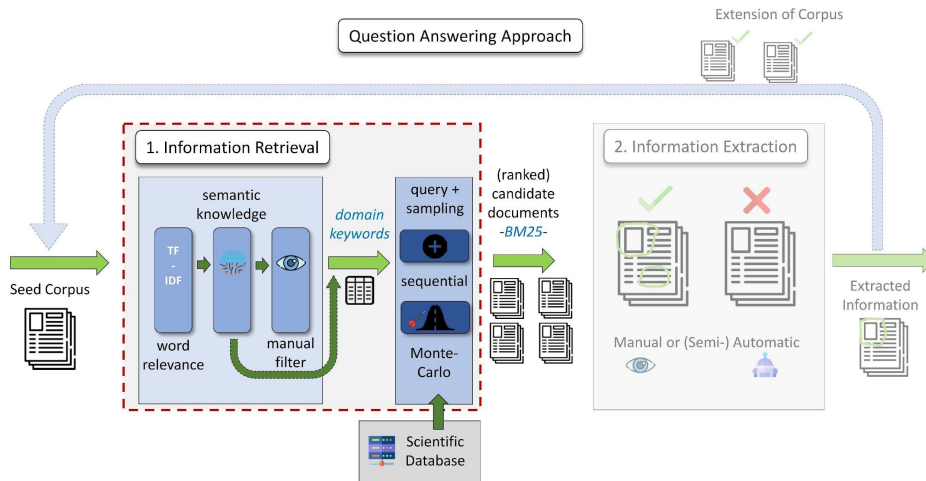


Figure 1. Proposed information retrieval and extraction cycle.

3.2. Extraction of Domain Keywords from Seed Corpus

First, a set of relevant domain keywords must be extracted from the seed corpus to characterize the domain. The Term Frequency-Inverse Document Frequency (TF-IDF) metric (Qaiser and Ali, 2018) determines the relative importance of a word in a set of documents, considering the words with high number of appearances in few documents as relevant while frequent words in all documents and non-frequent words as non-relevant.

3.3. Query Construction Methods for Sampling Academic Databases

In this work we assess and compare two methods:

- Sequential Sampling (SEQ): Add keywords to the query in decreasing relevance order until the number of search results is below a user-defined threshold.
- Monte-Carlo Sampling (MC): Build multiple queries by randomly adding keywords according to their relevance until the number of results is below a user defined threshold. An inherent ranking metric is the appearance frequency (the number of queries a given document appears in, divided by the total amount of queries).

3.4. Ranking of Retrieved Documents

The BM25 ranking function (Robertson and Zaragoza, 2009) is a popular approach in information retrieval to determine the relevance of a document to a query compared to

the other documents in the retrieved corpus. It yields a metric that can be considered as relative (linguistic) relevance.

3.5. *Validation and Limitations*

In order to validate the retrieval method, the normalized BM25 relevance of the retrieved full text documents are compared with each other and with the seed documents. The full text information is chosen because (1) a test on using the abstracts of the documents to determine the relevance of the full documents showed an unreasonable classification and (2) the desired knowledge is contained within the full text.

4. **Case Study**

The methodology is applied to a set of eight seed documents that contain quantitative information about processes for the pyrolysis of plastic waste. The information extracted from these documents was used by (Somoza-Tornos et al., 2021) to assess a process screening framework for the synthesis of process networks from a circular economy perspective. The retrieved documents departing from this seed corpus are expected to help extending the study by identifying additional candidate process conditions and pathways.

4.1. *Targeted Information*

One example of textual information is taken from a seed document (Onwudili et al., 2009) and the keywords are highlighted by general concepts that are searched:

“The compositions of the process (pyrolysis) products of pure low-density waste product (polyethylene), waste product (LDPE) and waste product (polystyrene) and their mixtures have been investigated over a condition (temperature) range from 300 to 500 condition (°C).”

5. **Results and Discussion**

Table 1 shows the top 16 keywords extracted from the seed abstracts. The sequential method was applied with a threshold of 1,000 documents leading to 376 abstracts and full text documents that could be retrieved with the available licensing options. The Monte-Carlo method was performed with 1,000 iterations using 10, 20 and 30 keywords to identify three lists of candidate documents. From the top 2,500 documents of each list a total of 553, 568 and 1,110 abstracts and corresponding full documents could be retrieved respectively. Additionally, from the 30 keywords list (63,186 entries) a randomly selected subset of 2,500 papers was chosen from which a total of 972 could be retrieved and is treated as reference set. The BM25 ranking function was applied to the complete set of retrieved full text documents, as these contain the relevant information. From Figure 2 it can be seen that the documents from the 10 and 20 keywords list have a very similar distribution and a generally higher relevance compared to the 30 keywords list and the sequential method. This trend holds true when using different amounts of keywords for relevance determination. The documents from the randomly selected subset have a considerably lower relevance proving that the methods perform as expected. The number of keywords when sampling the database is a key parameter and the results show that using more “relevant” keywords does not lead to improved

retrieval results. Moreover, it is evident that the Monte-Carlo method retrieved generally more relevant documents than the sequential method.

Table 1. Top 20 keywords and TF-IDF values extracted from the seed corpus abstracts.

Keyword	TF-IDF	Keyword	TF-IDF	Keyword	TF-IDF	Keyword	TF-IDF
pyrolysis	1.23	gas	0.85	bed	0.64	fluidise	0.55
waste	1.12	temperature	0.76	increase	0.64	time	0.50
product	1.01	plastic	0.76	feedstock	0.60	residence	0.50
oil	0.86	yield	0.73	wt	0.60	recycling	0.46
process	0.85	catalyst	0.68	polyethylene	0.58	flash	0.44

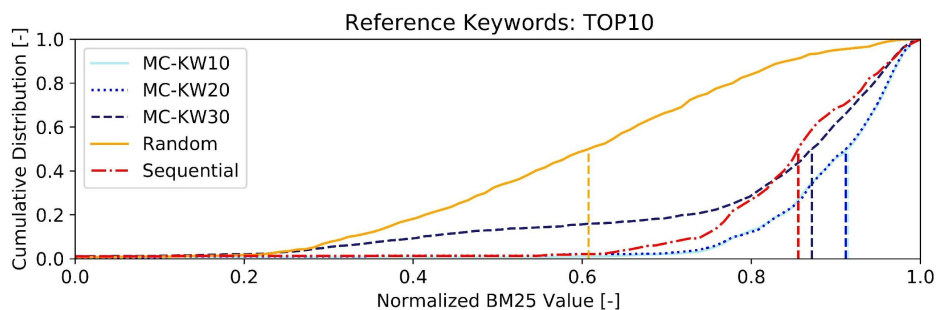


Figure 2. Cumulative distribution of normalized BM25 relevance values.

Figure 3 illustrates that the inherent ranking metric of the Monte-Carlo method correlates to some extent with the document relevance. The point clouds represent the relevance and document frequency of all retrieved documents within the 30 keywords list. It can be identified that, when only choosing documents with a normalized frequency higher than 0.1, a significant amount of irrelevant papers can be discarded (here: 80.3% of set size). Moreover, a large number of newly identified documents has an even higher relevance (22.9%) than the average relevance of the seed corpus, which is a promising starting point for the information extraction step. The sequential method, on the other hand, does not directly have an inherent ranking. Figure 3 (right) shows a derived metric that is the number of keywords contained in the downloaded papers. It appears that there is no clear correlation between the number of keywords included and the BM25 relevance.

The general impression is that the documents are indeed fitting the domain and potentially contain the desired information. A selected truly relevant document appearing in all three MC lists within the upper ranks is titled “Hydrocarbons obtained by waste plastic pyrolysis: Comparative analysis of decomposition described by different kinetic models” (Miskolczi and Nagy, 2012). A relevant passage reads as follows:

“... the parameter (yields) of volatile products were 15.0% [410 condition (°C)], 24.1% [430 condition (°C)] and 55.3% [450 condition (°C)] in case of W-2 marked sample process (pyrolysis), while ...”

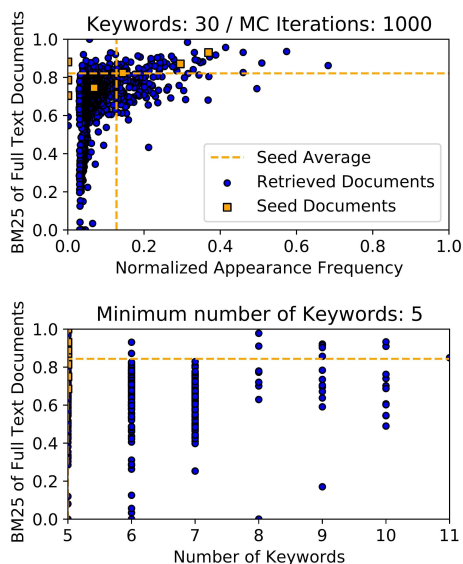


Figure 3. BM25 values plotted against model inherent ranking for Monte-Carlo (left) and sequential (right) method. Shown is the data obtained from the 30 keywords list.

6. Conclusions

Lists of (linguistically) relevant documents were identified with minimal human intervention. These lists have the characteristic of being populated by highly relevant documents in the upper ranks which allows to limit the selection of documents to download for the subsequent information extraction step. The candidate documents proved to have similar or even higher relevance to the domain than the documents in the seed corpus. A first qualitative assessment of the titles and abstracts indicates that these documents are truly relevant to the posed question. Our investigations showed that the proposed information retrieval methodology performs appropriately using the selected database and seed corpus taken from a chemical engineering field. This implies the potential of establishing a systematic machine-assisted search procedure for model parameters and knowledge, effectively reducing the workload of engineers in the PSE community and going beyond what a completely manual procedure could achieve. As of now, the methodology assesses document relevance by means of the BM25 metric. This metric allows for a pre-selection of documents but the next necessary step in the development of the whole information retrieval and extraction cycle is to systematically classify the true relevance of the documents by a machine-assisted information extraction methodology. Moreover, the methodology has been tested using only one database. Further work is in progress to extend the search and improve its efficiency (speed and accuracy).

Acknowledgements

Financial support received from the Spanish Competitiveness, Industry and Economy Ministry and the European Regional Development Fund, both funding the research

Projects AIMS (DPI2017-87435-R) is fully acknowledged. Adrián Pacheco-López thankfully acknowledges financial support received from the Spanish Ministry of Science, Innovation and Universities (grant ref. PRE2018-087135).

References

- R. Kottmann, M. Radom, P. Formanowicz, F. Glöckner, A. Rybarczyk, M. Szachniuk, J. Błazewicz, 2010. Cerberus: A New Information Retrieval Tool for Marine Metagenomics. *Foundations of Computing and Decision Sciences* 35, 107–126.
- E. Landhuis, 2016. Scientific literature: Information overload. *Nature* 535, 457–458.
- N. Miskolczi, R. Nagy, 2012. Hydrocarbons obtained by waste plastic pyrolysis: Comparative analysis of decomposition described by different kinetic models. *Fuel Processing Technology* 104, 96–104.
- J.A. Onwudili, N. Insura, P.T. Williams, 2009. Composition of products from the pyrolysis of polyethylene and polystyrene in a closed batch reactor: Effects of temperature and residence time. *Journal of Analytical and Applied Pyrolysis* 86, 293–303.
- A. Pacheco-López, A. Somoza-Tornos, E. Muñoz, E. Capón-García, M. Graells, A. Espuña, 2020. Synthesis and Assessment of Waste-to-resource Routes for Circular Economy, in: 30 European Symposium on Computer Aided Process Engineering. Elsevier. volume 48, pp. 1933–1938.
- S. Qaiser, R. Ali, 2018. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications* 181.
- S. Robertson, H. Zaragoza, 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends® in Information Retrieval* 3, 333–389.
- A. Somoza-Tornos, C. Pozo, M. Graells, A. Espuña, L. Puigjaner, 2021. Process screening framework for the synthesis of process networks from a circular economy perspective. *Resources, Conservation and Recycling* 164, 105147.
- A. Usai, M. Pironti, M. Mital, C. Aouina Mejri, 2018. Knowledge discovery out of text data: a systematic review via text mining. *Journal of Knowledge Management* 22, 1471–1488.

Prediction of Bioconcentration Factors (BCF) using Graph Neural Networks

E. I. Sanchez Medina^a, S. Linke^a and K. Sundmacher^{a,b,*}

^a*Chair for Process Systems Engineering, Otto-von-Guericke University, Universitätspl. 2, Magdeburg, 39106, Germany*

^b*Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, Magdeburg, 39106, Germany*

* sundmacher@mpi-magdeburg.mpg.de

Abstract

This paper presents a Graph Neural Network (GNN) approach for the prediction of bioconcentration factors (BCF). We show that GNNs are able to exploit structural information of molecules to regress BCF values that are comparable to commonly used quantitative structure-activity relationship (QSAR) models in terms of the coefficient of determination R^2 . However, the main advantage of GNNs is that molecular descriptors do not need to be determined and pre-selected by the user. Instead, they are learned directly by the GNN using backpropagation. A database of 473 molecules was used to train and test the present model. The results obtained suggest that GNNs might be useful for the prediction of other types of sustainable indicators of molecules, which is subject of our further research.

Keywords: Bioconcentration Factor (BCF), Graph Neural Network (GNN), green chemistry, machine learning

1. Introduction

When moving towards a sustainable chemical industry, commonly used toxic solvents have to be replaced with more environmentally responsible options. For this, two main objectives have to be met: the replacements have to be comparable to certain degree in terms of their thermodynamic behavior; and their environmental, health and safety properties (EHS) have to qualify them as safe and sustainable replacements. While great effort has been set to find thermodynamically beneficial solvents, analyzing the EHS properties of the proposed molecules experimentally is quite expensive and time-consuming. To overcome this problem, quantitative structure-activity relationship (QSAR) models are commonly used as prediction methods (Benfenati et al., 2013).

One of the EHS of interest is the accumulation of chemical substances in body tissues (normally measured in fish tissue). This property is usually expressed as a bioconcentration factor (BCF), which measures the difference between the concentration of the substance in the organism's tissue to that in the environment. The QSAR models currently employed to predict BCF are based on several molecular descriptors and regression models such as multiple linear regression (MLR), support vector machines (SVM) and radial basis function neural networks (RBFNN) (Miller et al., 2019). However, all these methods require that the user pre-selects and calculates the appropriate molecular descriptors to serve as inputs to the model, which is a difficult

task given the large number of possible descriptors, e.g. Zhao et al. (2008) analyzed 1022 descriptors. This paper presents a Graph Neural Network (GNN) approach for the prediction of BCF. The advantage of the GNN compared to traditional QSAR methods is that molecular fingerprints are automatically generated. These fingerprints are later used to train a regression model (e.g. a multi-layer perceptron (MLP)). In this way, the complete set-up can be trained using backpropagation and an end-to-end model is obtained that relates the molecular structure (e.g. represented as SMILES strings) to the BCF value. The concept of GNNs is introduced next, before describing the methodology used, and the results obtained in terms of the coefficient of determination R^2 , the mean absolute error MAE and the standard deviation on the error of prediction SDEP. A comparison with traditional QSAR models is then provided.

2. Background

Graphs: A graph G can be defined as a tuple of features corresponding to nodes V and edges E , i.e., $G = (V, E)$. It is also necessary for a graph to specify the connectivity of its nodes, e.g. via an adjacency matrix. In this scheme, some nodes act as receivers and some as senders depending on the direction of connectivity. For undirected graphs, both connected nodes act as sender and receiver for each other.

Graph Neural Networks: Similarly, to artificial neural networks that operate over vectors, Graph Neural Networks (GNNs) operate on graph-structured data. This means that a GNN performs mathematical transformations on a graph as input and returns an updated graph as output (graph-to-graph). The transformations on G can be divided into three operations within a layer l :

- (i) Message passing: Given a node v with features $h_v^{(l-1)}$, for each node w connected to v , there is a differentiable function ϕ_M that sends the information of w to v .

$$m_{v,w}^{(l)} = \phi_M(h_v^{(l-1)}, h_w^{(l-1)}) \quad (1)$$

- (ii) Message aggregation: Given the messages $m_{v,w}^{(l)}$ for all nodes in the neighborhood N of v , there is a differentiable and permutation invariant function ϕ_A that aggregates all messages across all neighbors of v in an aggregated vector $a_v^{(l)}$.

$$a_v^{(l)} = \phi_A(\{m_{v,w}^{(l)} : w \in N(v)\}) \quad (2)$$

- (iii) Features updating: Given the aggregated features of all neighboring nodes $a_v^{(l)}$, there is a differentiable function ϕ_U that combines the features $h_v^{(l-1)}$ with $a_v^{(l)}$ to update the features $h_v^{(l)}$.

$$h_v^{(l)} = \phi_U(h_v^{(l-1)}, a_v^{(l)}) \quad (3)$$

This process of message passing, aggregation and updating can be repeated for many layers, similarly to the convolutional layers used for image processing. Edge features might be also included as part of the inputs of the function ϕ_M . This allows for the edge information to also be part of the message passing scheme. After multiple graph convolutions L , an updated graph is obtained whose nodes contain combined

information of their local neighborhood within an L -node radius. In this way, each node and edge in the final graph possesses information about itself in the context of the whole graph structure. Then, the final graph information can be transformed using a permutation invariant pooling layer to obtain a vectorial representation of the final graph. Finally, this vector might then be used as an input for a regression or classification model depending on the task at hand. Since all the functions involved in this scheme are differentiable, the whole model can be trained using backpropagation. GNNs have been the focus of research since the past decade (Gori et al., 2005), but their popularity has increased rapidly in recent years, especially in the context of molecular structure to property prediction (Coley et al., 2017; Gilmer et al., 2017; Sanchez-Lengeling et al., 2019; Schweidtmann et al., 2020).

3. Methodology

3.1. Database

Several publicly available datasets containing BCF values for different molecules can be found in the literature (Arnot and Gobas, 2006; EURAS, 2007). However, their values depend on many experimental factors, such as the fish species utilized during the experiments. For this reason, a reliability score is sometimes available along with the BCF values. Previous efforts have been made to carefully select high quality data which is suitable for regulatory purposes, such as the REACH legislation. The database reported by Zhao et al. (2008), which was used to construct QSAR models available in the VEGA toolbox (Benfenati et al., 2013), is one of these carefully selected collections of data. This database contains 473 compounds and covers a log-(BCF) range from -1.00 to 4.85, for molecular weights ranging from 68 to 943 g/mol. This same database was used in the present work. The molecule classes proportion in the dataset are shown in Figure 1.

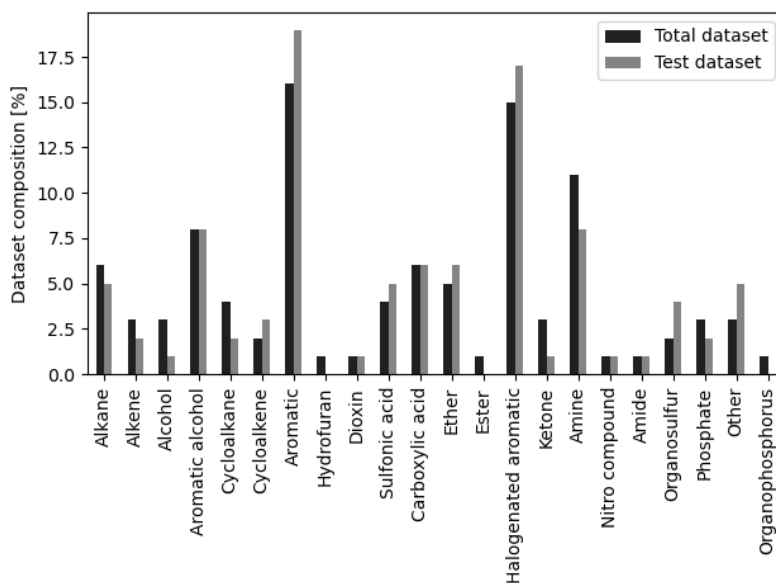


Figure 1: Percentage distribution of molecular classes in the complete dataset and test set.

3.2. Molecule to graph

First, the molecular information has to be integrated into a graph. Starting from Simplified Molecular Input Line Entry System (SMILES) strings, a graph is created for each molecule in the dataset. This is performed using the open-source package for cheminformatics RDKit (<http://www.rdkit.org>) in Python 3.7. In each graph the nodes represent the atoms, and the edges represent the chemical bonds. For simplicity hydrogens were not considered as part of the graphs. Tables 1 and 2 show the type of features used for nodes and edges, respectively. Given the small size of the dataset, a generalized atom type "Other" was included. This accounts for the few instances of rarely appearing atoms in the database while conserving the other features information of the involved molecules.

Table 1: Node features used to define molecular graphs.

Node feature	Description	Dimensions
Type *	Type of atom (C, O, Cl, N, F, Br, S, Other)	8
Ring	Whether the atom is part of a ring	1
Aromaticity	Whether the atom is part of an aromatic ring	1
Hybridization *	Hybridization of the atom (sp, sp ² , sp ³ , sp ³ d, sp ³ d ²)	5
Bonds *	Number of bonds to the atom	6

* Implemented using one-hot-encoding (vector of binary values for each unique integer value).

Table 2: Edge features used to define molecular graphs.

Edge feature	Description	Dimensions
Type *	Type of bond (single, double, triple, aromatic)	4
Conjugated	Whether the bond is in a ring	1
Ring	Whether the bond is in a ring	1

* Implemented using one-hot-encoding (vector of binary values for each unique integer value).

3.3 GNN architecture

The GNN model was implemented using Pytorch Geometric (Fey and Lenssen, 2019) and follows a similar architecture to the one presented by Gilmer et al. (2017). In this structure, the message passing function is denoted by $m_{v,w}^{(l)} = h_w^{(l-1)} \cdot \text{MPL}(e_{v,w})$, where MPL stands for a multi-layer perceptron and $e_{v,w}$ for the edge features vector between the nodes v and w . For this MLP, the rectified linear unit (ReLU) activation function was used with a single hidden layer of 128 neurons. This MLP serves to combine edge information to the node. The aggregation function is just the sum over the

messages in the neighborhood v , $a_v^{(l)} = \sum_{w \in N(v)} m_{v,w}^{(l)}$. And the update function is

$h_v^{(l)} = \text{GRU}(\Theta^{(l)} h_v^{(l-1)} + a_v^{(l)})$ where $\Theta^{(l)}$ is a layer-specific trainable weight matrix.

GRU stands for Gated Recurrent Unit, which allows the sharing of parameters among layers. The use of GRU in GNNs was studied by Li et al. (2016) and was also applied by Schweidtmann et al. (2020) for the prediction of molecular properties. Regarding the method used, it is worth mentioning that, while the node features are being updated from one convolutional layer to the next one, the edge features remain the same.

A *set2set* (Vinyals et al., 2015) pooling layer was used to obtain the vectorial representation of the final graph. Gilmer et al. (2017) suggested that this type of graph embedding should produce a more expressive molecular fingerprint compared to the sum embedding. The number of processing steps was tuned between $0 \leq T \leq 10$. A final value of $T = 3$ was selected. Then, a multi-layer perceptron (MLP) was used to predict the BCF from the molecular fingerprint. This MLP uses the exponential linear unit (ELU) activation function. The model hyperparameters were extensively explored considering the following ranges: initial learning rate $\in \{0.0001, 0.001, 0.01\}$, number of convolutional layers $\in \{1, 2, 3\}$, hidden size of convolutional layers $\in \{21, 32, 59\}$, final MLP layers' size $\in \{64, 32, 16\}$. An initial learning rate of 0.001, 2 convolutional layers with size 21 and a three-layer MLP with 64, 32 and 16 neurons were finally selected based on the models' mean absolute error (MAE). The data splits used during the hyperparameter selection are discussed in the following.

3.4 Training, validation and testing

In order to compare our model with the one reported by Zhao et al. (2008), the same train/validation/test split proportions were used (80% of the whole dataset, 20% of the training set and 20% of the whole dataset respectively). Following the work by Schweidtmann et al. (2020), ensemble learning was used to improve the quality of the predictions given the small dataset in this work. A total of 10 models were trained using randomly selected splits for the validation set. The test predictions of each model were averaged to produce the final prediction. The model was trained for 300 epochs using the optimizer Adam with batches of 30 graphs. A dropout of 50% was used both for the convolutional layers and the final MLP to avoid overfitting. The training data was also normalized to have a zero mean and unit variance. The loss function was selected to be the same one used by Zhao et al. (2008), i.e., mean squared error (MSE). And the reported metrics used were the mean absolute error (MAE) and the R^2 score. A scheduler was specified to reduce the learning rate by 70% whenever the validation error reaches a plateau on 5 consecutive epochs with a threshold of 10^{-4} .

4. Results and discussion

The GNN model presented in this work was compared with the best QSAR model presented by Zhao et al. (2008) (in their original paper referred to as "Model 7" or "Hybrid model"). The results for the test and validation sets can be seen in Table 3. Compared to the QSAR model, our GNN-based model shows similar results. However, the main difference between both modeling approaches is that the GNN is able to map the structural information of the molecule to its property without the need to externally compute any molecular descriptor. As opposed to the 1022 descriptors analyzed by the Zhao et al. (2008), the GNN computes a molecular fingerprint from just 8 relatively intuitive structural parameters for molecules (see Tables 1 and 2). Despite this notable difference in descriptors, QSAR models might provide a physical insight that might be lacking when using GNNs. It is important to mention that both models were constructed using 2D information only. However, including stereochemistry information to a GNN would be relatively straightforward to do (e.g. by adding torsion angles as extra edge features). This is an important advantage given that spatial configuration plays a big role in the EHS properties of molecules. The parity plot for the test set is shown in Figure 2.

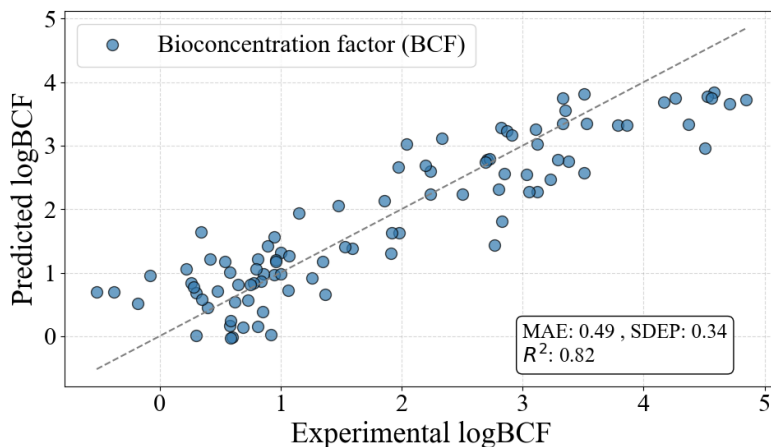


Figure 2: Parity plot for the final GNN model predictions on the test set.

Table 3: Comparison of models for predicting BCF (bold numbers indicate preferred value).

Model	Test set			Validation set			Training set		
	R^2	MAE	SDEP	R^2	MAE	SDEP	R^2	MAE	SDEP
GNN	0.82	0.49	0.34	0.83	0.44	0.32	0.82	0.44	0.35
Zhao	0.79	0.45	0.59	0.79	-	0.66	0.83	-	0.56

5. Conclusion

In this work a GNN-based model to predict BCF is presented. The main advantage of the proposed method is that molecular descriptors do not need to be chosen and calculated externally. Instead, molecular structure patterns are learned by the GNN to map them to the property of interest (e.g. BCF). First, the structure information is exploited by using graph convolutional layers. Then, a pooling method, herein *set2set*, is applied to obtain a molecular fingerprint that would play the role of the commonly used molecular descriptors. Finally, a multi-layer perceptron maps the fingerprint to the property. The investigation of the viability of GNNs with more EHS indicators is subject of future research. Also, the determination of confidence intervals on the GNNs predictions is still missing and an important aspect that needs to be investigated in the future.

6. Acknowledgments

This work is partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - TRR 63 “Integrated Chemical Processes in Liquid Multiphase Systems” (Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - TRR 63 “Integrierte chemische Prozesse in flüssigen Mehrphasensystemen” (Teilprojekt B9) - 56091768). E. I. S. M. and S. L. are also affiliated with the International Max Planck Research School for Advanced Methods in Process and Systems Engineering - IMPRS ProEng at the Max Planck Institute for Dynamics of Complex Technical Systems Magdeburg.

References

- J. A. Arnot, F. A. Gobas, 2006. A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. *Environmental Reviews* 14 (4), 257–297.
- E. Benfenati, A. Manganaro, G. C. Gini, 2013. VEGA-QSAR: AI inside a platform for predictive toxicology.

- In: PAI@ AI* IA. pp. 21–28.
- C. W. Coley, R. Barzilay, W. H. Green, T. S. Jaakkola, K. F. Jensen, 2017. Convolutional embedding of attributed molecular graphs for physical property prediction. *Journal of Chemical Information and Modeling* 57 (8), 1757–1772.
- EURAS, 2007. EURAS bioconcentration factor (BCF) Gold Standard Database, Cefic LRI-ECO7 project. <http://ambit.sourceforge.net/euras/>, Accessed: 2020-10-28.
- M. Fey, J. E. Lenssen, 2019. Fast graph representation learning with PyTorch Geometric. In: *ICLR Workshop on Representation Learning on Graphs and Manifolds*.
- J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl, 2017. Neural message passing for quantum chemistry. arXiv preprint arXiv:1704.01212.
- M. Gori, G. Monfardini, F. Scarselli, 2005. A new model for learning in graph domains. In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005*. Vol. 2. IEEE, pp. 729–734.
- Y. Li, D. Tarlow, M. Brockschmidt, R. Zemel, 2016. Gated graph sequence neural networks. *ICLR*.
- T. H. Miller, M. D. Gallidabino, J. I. MacRae, S. F. Owen, N. R. Bury, L. P. Barron, 2019. Prediction of bioconcentration factors in fish and invertebrates using machine learning. *Science of the Total Environment* 648, 80–89.
- B. Sanchez-Lengeling, J. N. Wei, B. K. Lee, R. C. Gerkin, A. Aspuru-Guzik, A. B. Wiltschko, 2019. Machine learning for scent: Learning generalizable perceptual representations of small molecules. arXiv preprint arXiv:1910.10685.
- A. M. Schweidtmann, J. G. Rittig, A. König, M. Grohe, A. Mitsos, M. Dahmen, 2020. Graph neural networks for prediction of fuel ignition quality. *Energy & Fuels* 34 (9), 11395–11407.
- O. Vinyals, S. Bengio, M. Kudlur, 2015. Order matters: Sequence to sequence for sets. arXiv preprint arXiv:1511.06391.
- C. Zhao, E. Boriani, A. Chana, A. Roncaglioni, E. Benfenati, 2008. A new hybrid system of QSAR models for predicting bioconcentration factors (BCF). *Chemosphere* 73 (11), 1701–1707.

Data and environment based approach for process systems engineering

Nancy Prioux^{a*}, Jean-Pierre Belaud^a, Rachid Ouaret^a and Gilles Hétreux^a

^aLaboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France

nancy.prioux@ensiacet.fr

Abstract

In this paper, a five-step approach coupling data science and life-cycle assessment is detailed. The main purposes of the study are to collect and analyse the process data from scientific literature, to characterize clusters, and to help the decision support in the preliminary choice of biomass or valorisation process according to the environmental indicators. The approach is tested on the comparison of rice straw and corn stover pretreatment processes for glucose production. Following the goal and scope, data architecture, life-cycle inventory, and environmental assessment steps, an impact-process matrix is analyzed in the last step using the multidimensional scaling method (MDS). The visualization of the results shows clusters that allow making a first analysis of the processes.

Keywords: Circular economy, biorefinery, machine learning, life cycle assessment, biomass pretreatments

1. Introduction

In the agriculture field, which is a major source of waste production, the circular economy has become a major study subject over recent years. In France, the waste from the agri-food industries is estimated at 2.6 million tons per year. Globally, this rate will increase with the expected increase in the population. Much of this waste is lignocellulosic byproducts that can be transformed into bioenergy, biomolecules, or biomaterials. A lot of researches are being done to transform these byproducts through sustainable processes, i.e. environmentally responsible, economically viable, and socially accepted (Sammons et al., 2009). Many transformation processes have been studied and published over the last thirty years (Davis et al., 2017). However, there is a lack of criteria to guide the choice between all these processes. The use of environmental, economic, and social assessment in a CE context is a good way to guide this choice. These assessments require a large amount of data, in particular data on processes potentially available in the scientific or technical literature or feeding by off-line simulation (Morales-Mendoza et al., 2012). The huge amount of scientific articles describing the valorization processes of agricultural byproducts constitutes a great and largely under-exploited source of data. This amount can be difficult to process. However, many tools and methods are now available for this purpose thanks to Big Data technics. Environmental evaluation tools such as Life Cycle Assessment (LCA) would benefit from these technologies to analyze process data and results from these evaluations. Some studies utilize massive data to complement LCA. J. Cooper et al. applied massive data to supplement background data (Cooper et al., 2013). The DILCA (Data-Intensive Life Cycle Assessment) method exploits knowledge

engineering (KE) and integrates technological developments that can change the LCA outcome for a given product over time (Bhinge et al., 2015). Divers studies combined Data Envelope Analysis (DEA) and LCA (Kouchaki-Penchah et al., 2017).

The main purposes of our study are (1) to collect and analyze the process data from scientific literature, (2) to characterize trend or clusters, (3) minimize experiments that are costly and energy-consuming, and (4) support the decision in the preliminary choice of biomass or a valorization process according to environmental indicators.

The approach is discussed in section 2 and section 3 deals with a study case: rice straw and corn stover pretreatment processes for glucose production.

2. General approach

The approach is based on multidisciplinary field: data management, sustainability engineering, and life cycle thinking. It supports the analysis of environmental impacts in an automatic way in grouping processes. Following the goal of their study, the researcher or the R&D engineering can compare or choose processes based on our approach's results. It is defined by five steps: (1) goal and scope, (2) data architecture, (3) life cycle inventory (LCI), (4) sustainability assessment, and (5) visualization and analysis of results.

In the first step, the *goal and scope of the study* must be clearly defined. Life cycle thinking is recommended. This thinking encourages a "cradle-to-grave" or "cradle-to-gate" approach if the logistics of a value chain are difficult to get. System boundaries and functional unit significantly influence evaluations. For example, it needs to be clarified whether the upstream biomass supply chain is considered. Once the goal and scope have been properly defined, the supply chain, technologies, and transformation processes should also be described.

The *data architecture* is directly inspired by the construction of big data architecture and consists of five sub-steps: (i) data collection and extraction, (ii) data enrichment and storage, (iii) data processing, (iv) (raw) data analysis, and (v) (raw) data visualization. This step can be automatic, semi-automatic, or manual and it uses data technics e.g. machine-learning methods for the (raw) data analysis. These substeps are detailed in (Belaud et al., 2019).

The *Life Cycle Inventory (LCI)* is completed using the process data (also called foreground data) from the previous step. The background data come from free or commercial LCI databases such as EcoInvent Database.

For the fourth step, one or more impact calculation methods must be determined in accordance with the first step that integrates the nature of the study and the system. Then, the environmental impacts are calculated using these methods. At the end of this stage, the main result is a structure [processes: biomasses: impacts] which is difficult to analyze.

The *visualization and analysis of the results* step include methods derived from artificial intelligence and more precisely from "machine learning" to help in the analysis of environmental impacts. Starting from the statistical literature, traditional dimension reduction (DR) and unsupervised clustering techniques are combined to extract information from environmental impacts. More precisely, this hybrid approach is based on the Multi-Dimensional Scaling (MDS) using the Canberra distance and k-means. The objective is to search for "hidden" structures in multi-dimensional data and to help interpret the matrix. The advantage of this approach is that data-based methods require very little knowledge of processes to perform. Figure 1 summarizes the treatment for a [processes: impacts] matrix. First, DR techniques project the raw process data into a

lower-dimensional space (2 or 3). After a projection of the data by a DR technique, the clustering approach is then applied to consider similar impacts and processes within the lower-dimensional space. Finally, the user (expert) analyzes the points grouped in clusters to link them to significant processes/impacts. The visualization of the data clusters will help the researcher or R&D engineers in the final decision following the goal of their study.

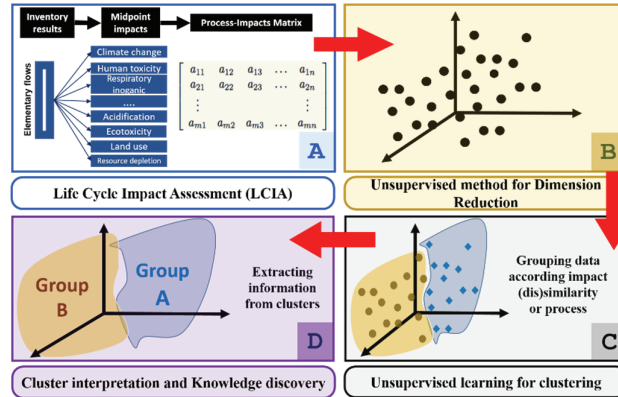


Figure 1: Schematic of data driven processing

3. Case study

In this section, the first results of our approach are presented with an intention to compare different pre-treatment processes for two biomasses: corn stover and rice straw. This study examines twenty processes extracted from twenty scientific articles. The study aims to help the researchers or R&D engineers to analyse processes for glucose production. Life cycle thinking is from “cradle to grave” i.e. the study’s boundaries are from biomass to the enzymatic hydrolysis stage. Biomass is considered as agricultural waste (destined to become a co-product): it has a zero impact - the impacts of the agricultural phase are attributed to the end product of agriculture (corn and rice). The biorefinery is considered to be relatively close to the site. Therefore the transport stage impact is minor. The main function of the system is the production of glucose. The functional unit is “production of 1 kg of glucose” and results are expressed in a functional unit.

Twenty scientific articles have been selected thanks to specific keywords in Science Direct and Web of Science. Process data from articles are extracted semi-automatically using an ontology. The ontology structures the process data and ensures an export in CSV files and supplying internal software. This software developed on Microsoft Excel carries out a first “cleaning” of the data by simulating the processes to calculate and checking the mass balance. The data from three articles are removed from our study because they contain inconsistencies or many missing data that can not be verified by our simulation. The study, therefore, analyses 17 processes of two biomasses. The (attributorial) LCA method applied is ReCiPe 2016, the background database is from EcoInvent v2.2 and the foreground data are “cleaned” process data. The environmental assessment evaluates 17 “midpoint” impacts. The Excel tool has been validated using the commercial simulator ProSim+ for process simulation and SimaPro® for the simulation of the environmental impacts. The matrix of results [impacts: processes] is then analyzed by multi-dimensional scaling using R software. Figure 2 shows the global architecture of our application.

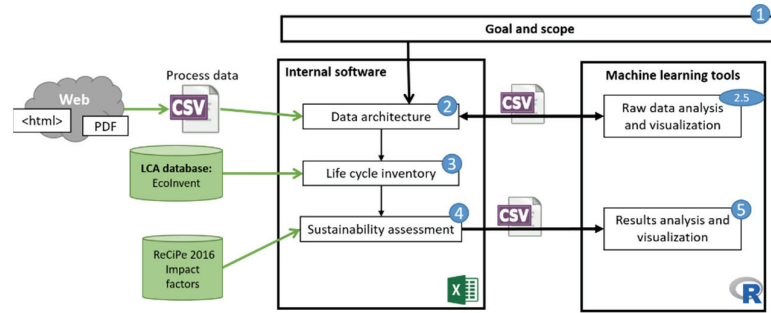


Figure 2: Architecture of the application

Classical Multidimensional Scaling (MDS) is a method for visualizing (dis)similarities between objects in a space of a reduced dimension. It is designed to help understand the structures of proximity and opposition. Starting from information about the mutual similarities of n objects, often with the similarity matrix $\Delta = \delta(i, j)_{0 \leq i \leq n; 0 \leq j \leq n}$, we look for a configuration of n points (in R^2 in general) which would be like that if the object i resembles to j more than the object l to k , we have $\delta(i, j) < \delta(l, k)$. For any distance matrix of size $n \times n$, the MDS allows us to find a set of n points marked by their coordinates whose similarity matrix is equal or very close to the distance matrix according to the data. In our approach, we use the Canberra distance (Lance & Williams, 1966), a weighted version of the Manhattan distance.

Let x_r ($r = 1, \dots, n$) be the coordinates of n points in a p dimensional Euclidean space where $x_r = (x_{r1}, x_{r2}, \dots, x_{rp})^T$ and $[B]_{rs} = b_{rs} = x_r^T x_s$. For an Euclidean distance $\Delta = \delta_{rs}$, from a matrix A of elements $[A]_{rs} = a_{rs} = -\frac{1}{2} \delta_{rs}^2$, deduced from the decomposition $x_r^T x_s$, the matrix B is obtained using the following relation :

$$\mathbf{B} = \mathbf{H}\mathbf{A}\mathbf{H} \quad (1)$$

where H is the centering matrix: $H = I - n^{-1}I \cdot I^T$ with $I = (1, 1, \dots, n)^T$. Elements of A are defined as $a_{r\bullet} = n^{-1} \sum_s a_{rs}$, $a_{\bullet s} = n^{-1} \sum_r a_{rs}$ and $a_{\bullet\bullet} = n^{-1} \sum_r \sum_s a_{rs}$.

The following steps can summarize the algorithmic procedure of multi-dimensional scaling, as illustrated in (Cox & Cox, 2001):

1. Obtain the proximity matrix $\Delta = \delta_{rs}$
2. Find the matrix $A = [-\frac{1}{2} \delta_{rs}^2]$
3. Find the matrix $B = [a_{rs} - a_{r\bullet} - a_{\bullet s} + a_{\bullet\bullet}]$
4. Find the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_{n-1}$ and the eigenvectors v_1, v_2, \dots, v_{n-1}
if B is semi-defined positive (some eigenvalues are negative), either (i) ignore the negative values and continue, or (ii) add an appropriate constant c to the (dis)similarities
5. Choose an appropriate size number l , possibly using $\frac{\sum_1^l \lambda_i}{\sum (\text{positive eigenvalues})}$.
6. The coordinates of the n points in Euclidean dimension space l are given by $x_{ri} = v_{ir}$ ($r = 1, \dots, n; i = 1, \dots, p$).

The interpretation of an MDS result is the same as for any other dimension reduction method: objects that are closer on the scatterplot are more similar than those that are further away. That is, the projected points are arranged in such a way that the grouped points (small distance between them) reflect the original relationships in the data. However, additional information is needed to make the projection more informative. As shown in Figure 1, a clustering algorithm (k-means) was applied to the MDS projection to highlight the most similar objects (Impacts - Processes). Two-dimensional maps of aggregated impacts and processes are shown in Figure 3. For the scatter plot of the impact points (Figure 3 left), we can distinguish three groups using k-means:

- Group 1: This group includes almost all impacts related to the chemical pollution of soils and water bodies.
- Group 2: This group forms three subgroups with superimposed points (in a 2-D perspective). This suggests that these points are very similar based on the Canberra distance. Here, we find a group quite heterogeneous in terms of impact with some not present in groups 1 and 3.
- Group 3: This group mainly includes impacts related to land use and land transformation.

To clarify the discussion, the names of the impacts within groups 1, 2, and 3 are not detailed. The clustering results of projected Process items are similar to those obtained for impacts. Indeed, three distinct clusters have been identified. However, tight and highly separated clusters occur within process data, this may suggest that each cluster is a domain or subdomain which should be analyzed individually. The acronyms captioning the dots represent each type of process which will not be explained here. In group 3, for example, there are two purely mechanical pre-treatments. Going back to the impacts, we find that these two pre-treatments have a very significant impact on the depletion of fossil fuels compared to the others. For Group 1, the three pre-treatments have relatively similar impacts and fossil fuels depletion impact costs around \$10, whereas group 2 pre-treatments have an impact cost of around \$1. However, these results are partial. Indeed, only a representation of dimension 1 is used instead of dimension 2, which gives a low percentage of representativeness and therefore a loss of information. It is planned to extend this study to represent dimensions 3 and 4 and complete the analysis.

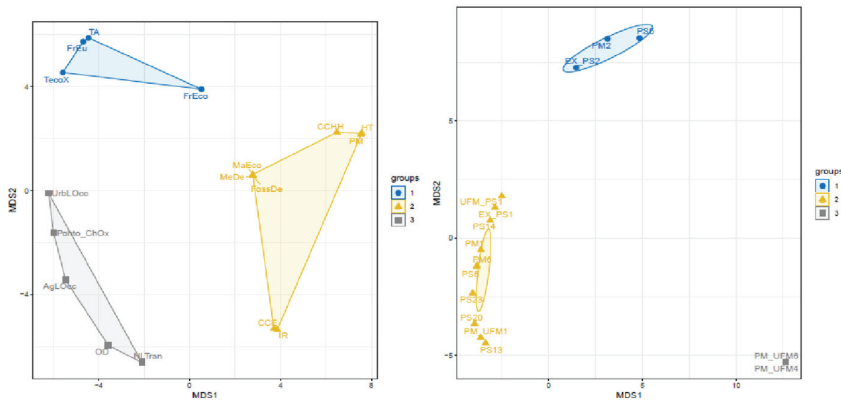


Figure 3: Scatter plot of (MDS) projection (two dimensions) and k-means clustering based on Impacts distance matrix (left) and Process distance matrix (right)

4. Conclusion and perspectives

A five steps approach coupling data science and environmental assessment is proposed and presented as an aid for researchers or R&D engineers during a preliminary study. This approach is being tested by studying the valorization of agricultural waste to promote a circular local economy. It compares different pretreatment of two lignocellulosic biomasses corn stalk and rice straw for glucose production. Following the goal and scope, data architecture, LCI, and environmental assessment steps, an impact-process matrix is analyzed in the last step using the multidimensional scaling method. The approach provides an initial insight to sort by groups and to establish a way to pre-select technologies or biomasses rigorously. The case study application reveals various limitations. Data from the scientific literature are by nature data from experiments in the laboratory and the LCA is therefore carried out for a low level of technological maturity (TRL 1/2). The approach does not incorporate the change in scale required to implement a semi-industrial pilot. Another limitation is the abundance and the quality of the data, which may not be sufficient for new technological pathways.

Different perspectives can be detailed for the approach. We can reconsider the functional unit and the global environmental assessment strategy by integrating the upstream agricultural phase and taking into account the global supply chain according to dynamic analysis, i.e. spatial or temporal. In this study, all the impacts are attributed to the final product (glucose), the policy of impacts relating to effluents can be modified by taking into account the fate of these effluents and their valorization in the circular economy context.

References

- Belaud, J.-P., Prioux, N., Vialle, C., & Sablayrolles, C. (2019). Big data for agri-food 4.0: Application to sustainability management for by-products supply chain. *Computers in Industry*, *111*, 41–50.
- Bhinghe, R., Srinivasan, A., Robinson, S., & Dornfeld, D. (2015). Data-intensive Life Cycle Assessment (DILCA) for Deteriorating Products. *Procedia CIRP*, *29*, 396–401.
- Cooper, J., Noon, M., Jones, C., Kahn, E., & Arbuckle, P. (2013). Big Data in Life Cycle Assessment: Big Data in Life Cycle Assessment. *Journal of Industrial Ecology*, *17*(6), 796–799. <https://doi.org/10.1111/jiec.12069>
- Cox, T. F., & Cox, M. A. A. (2001). *Multidimensional scaling*. Chapman & Hall/CRC. <http://www.crcnetbase.com/isbn/9781584880943>
- Davis, C. B., Aid, G., & Zhu, B. (2017). Secondary Resources in the Bio-Based Economy: A Computer Assisted Survey of Value Pathways in Academic Literature. *Waste and Biomass Valorization*, *8*(7), 2229–2246.
- Kouchaki-Penchah, H., Nabavi-Pelesaraei, A., O'Dwyer, J., & Sharifi, M. (2017). Environmental management of tea production using joint of life cycle assessment and data envelopment analysis approaches. *Environmental Progress & Sustainable Energy*, *36*(4), 1116–1122.
- Lance, G. N., & Williams, W. T. (1966). Computer programs for hierarchical polythetic classification (“similarity analyses”). *The Computer Journal*, *9*(1), 60–64.
- Morales-Mendoza, L. F., Azzaro-Pantel, C., Belaud, J.-Pierre., Pibouleau, L., & Domenech, S. (2012). An integrated approach combining process simulation and life cycle assessment for eco-efficient process design. In I. D. L. Bogle & M. Fairweather (Eds.), *Computer Aided Chemical Engineering* (Vol. 30, pp. 142–146). Elsevier.
- Sammons, N., Yuan, W., Bommareddy, S., Eden, M., Aksoy, B., & Cullinan, H. (2009). A Systematic Approach to Determine Economic Potential and Environmental Impact of Biorefineries. In *Computer Aided Chemical Engineering* (Vol. 26, pp. 1135–1140). Elsevier.

A Data-based Predictive Model for Distillation Column of Bio-based 2,3-Butanediol

Yeongryeol Choi,^{a,b} Hyundo Park,^{a,b} Jiwon Roh,^{a,b} Jongkoo Lim,^c In-su Han,^c
Duk-Ki Kim,^c Sangjun Jeon,^c Hee-Geun Nam,^c Il Moon,^b Hyungtae Cho,^a
Junghwan Kim^{a,*}

^a*Green Materials and Processes R&D Group, Korea Institute of Industrial Technology,
55, Jonga-ro, Ulsan, 44413, Republic of Korea*

^b*Chemical and Biomolecular Engineering, Yonsei University, 50, Yensei-ro, Seoul,
03722, Republic of Korea*

^c*Research and Development Center, GS Caltex Corporation, 359, Expo-ro, Yuseong-gu,
Daejeon, Republic of Korea*

kjh31@kitech.re.kr.82

Abstract

In bio-based chemical processes, stable and efficient process control is challenging because the feedstock composition made by microbial fermentation is uncontrollable. Many researchers have developed data-driven models that predict the operating condition using a correlation between the process data. Since the performance of a data-driven model depends on the quality of the data, data preprocessing techniques such as noise removal and outlier detection are generally considered to enhance the model performance. Thus, research to establish an appropriate data preprocessing method and range is required to increase the performance of the model. In this study, we developed a data-based prediction model for the bio-based 2,3-butanediol distillation column. We prepared sixty train datasets by applying noise removal and outlier detection with different methods and ranges on the raw data and confirmed the correlation between the data using Pearson's correlation coefficient. We considered two kinds of feature selection depending on data preprocessing and compared the performance (R^2 and normalized RMSE (NRMSE)) and reproducibility. The case in which the features were selected by only noise removal showed a similar coefficient value to that of the raw data; therefore, it had a higher performance than the case which applied outlier detection. The best performance provided R^2 and NRMSE values of 0.962 and 0.045, respectively, and it could be enhanced by hyperparameter tuning. Based on this result, we plan to develop a robust predictive model that features adaptable real-time prediction and control throughout the entire process time and period.

Keywords: 2,3-butanediol, 2,3-BDO, data-based prediction model, data preprocessing, feature selection

1. Introduction

Based on the rapid development of computer-related technologies and industries, new technologies integrating artificial intelligence are being established in various fields. The petrochemical industry is based on large equipment and processes, where the connection between unit processes and various sensors is complicated. However, too many unintended internal and external variables hinder efficient operation. Many machine learning-based studies have been conducted to improve prediction and control. Because a machine learning-based model predicts the output based on correlations between input and output data, the performance and stability of the model strongly depend on the data quality. Therefore, it is important to appropriately preprocess the raw data and select correct variables (sometimes called features) to increase the accuracy of the model. In addition, the performance of the predictive model may be inconsistent despite using the same data because the initial weights of the predictive model are randomly determined, and under- and over-fitting can occur during model training. Thus, the reproducibility must be considered to develop a robust predictive model.

2,3-butanediol (BDO) is the only isomer of butanediol that is widely found in nature. It is produced during a bioprocess, i.e. the fermentation of microorganisms, and not by a petrochemical process. GS Caltex is currently developing and operating a demo plant of a microbial-based 2,3-BDO production process (D. Tinôco *et al.*, 2020). However, the raw 2,3-BDO product has a low purity of 10–15 % and includes dozens of minor substances that cannot be detected during the analysis. This causes unstable operating

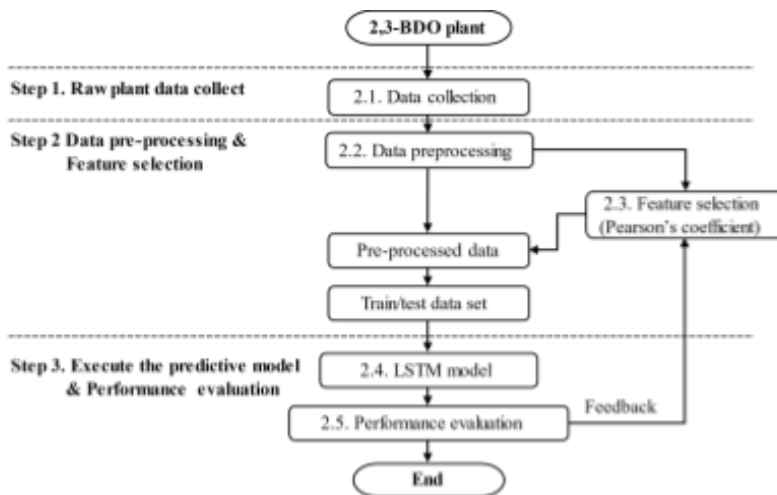


Figure SEQ Figure 1 ARABIC 1. Development strategy for a data-based predictive model

conditions, leading to noise and outliers in the raw data. To solve this problem, a prediction model that delivers a robust performance regardless of the early stopping option or initial weight factor must be established.

In this study, we developed a robust prediction model for the distillation column of bio-based 2,3-BDO by data preprocessing, such as noise removal and outlier detection, and selected features based on the correlation coefficient. The raw data were preprocessed by applying a low-pass filter (LPF) for noise removal and the interquartile range (IQR) and z-score for outlier detection. Subsequently, the features of the

predictive model were selected using the Pearson's correlation coefficient. To evaluate the performance, the coefficient of determination (R^2) and the standard deviation-based normalized root-mean-square error (NRMSE) were used. Each model run was repeated ten times to evaluate the reproducibility of the model. The overall model-development methodology in this study can be divided into three stages, as shown in Figure 1. The methods are detailed in the following section.

2. Methodology

2.1. Data collection

The bio-2,3-BDO plant data used in this study were collected by GS Caltex. The data were collected at one-minute intervals from July 10, 2020, to July 13, 2020 including the process start-up period. As shown in Figure 2, each dataset consists of 20 variables (features) including the control valves and instruments for temperature, pressure, and water-level measurements. The target variable is the temperature of the main product stream (the effluent at the bottom of the distillation column).

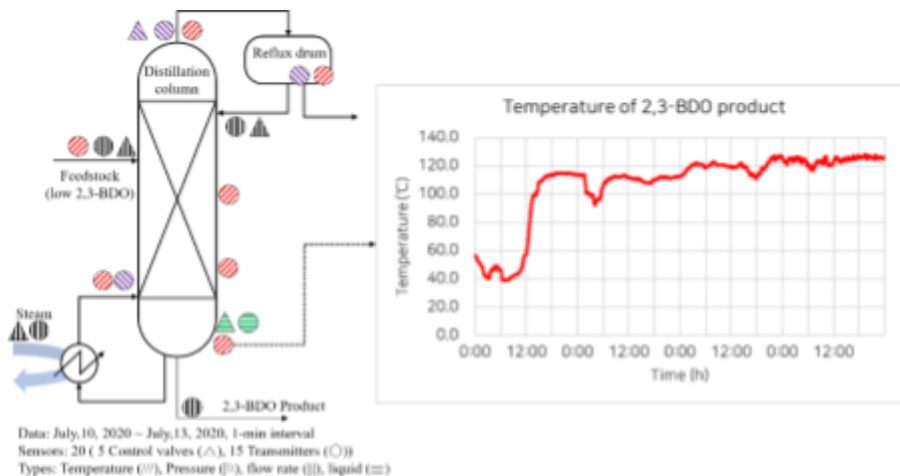


Figure 2. The data type and position in the distillation column

2.2. Data preprocessing

The data used in this study did not contain missing values. Therefore, only noise removal and outlier detection were performed. A first order low-pass filter was used for the noise removal. To compare the changes in correlation coefficient depending on the weight and performance of the prediction model, the weights were divided into the following categories: 0.3, 0.6, and 0.9 (labeled L1, L2, and L3, respectively).

The IQR and z-score, which are the most commonly used statistical methods, were used to remove the outliers (Ivanushkin et al., 2019). Four ranges corresponding to each outlier detection method were used: 2.5, 3.0, 3.5, and 4.0 (labeled I1, I2, I3, and I4, respectively, for IQR, and Z1, Z2, Z3, and Z4, respectively, for z-score). A total of 60 preprocessed datasets, including raw data, were obtained based on the application of individual methods and their combinations. The data were used for the model feature selection.

2.3. Feature selection

The Pearson correlation coefficient was used in this study for feature selection. In general, the coefficients between variables are considered to correlate if the absolute value of the coefficient is greater than 0.3, and this criterion was applied in this study.

2.4. Long short-term memory (LSTM) Model

The LSTM model, a recurrent neural network (RNN)-based algorithm, was used as the machine-learning algorithm (Hochreiter and Schmidhuber, 1997). The LSTM model can be used to solve the vanishing gradient problem; the weight update of the activation function converges to zero when long-term data are learned in an RNN algorithm, which is useful for data with a long learning period.

The predictive model is composed of an input layer containing selected features, a hidden LSTM layer consisting of four hidden units, and an output layer that outputs the temperature of the 2,3-BDO production stage. Seventy percent of the total data were randomly selected and used for training and validation, and 100 % of the data, including the training data, were used to determine the model performance.

The hyperparameters of the model were set based on a previous study (Kwon *et al.*, 2021). The numbers of batches, the number of epochs, the learning rate, and the activation function were set to 10, 1000, 0.01, and Elu, respectively. To prevent under- and/or over-fitting of the predictive model, the “patience” option was used as the early stopping option; it is defined as the number of epochs without an improvement after which the training will be stopped. The number of epochs was set at five for this study.

2.5. Evaluation factors for performance and reproducibility

Because the prediction performance is affected by the initial weight setting of the prediction model and the number of epochs, the reproducibility of the model was confirmed by running the model ten times using the preprocessed data. The R^2 value was used as the accuracy index because the RMSE, which generally represents the precision, significantly differs depending on the scale of the subject. Therefore, the NRMSE—RMSE divided by the standard deviation (σ)—was used in this study.

$$R^2 = \frac{\sum_i^n (y_i - \hat{y}_i)}{\sum_i^n (y_i - \bar{y}_i)} \quad (1)$$

y_i : Actual data

\hat{y}_i : Predicted data

\bar{y}_i : The mean value of actual data

$$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sigma} \quad (2)$$

n : Number of datasets

σ : Standard deviation of actual data

3. Results and discussion

3.1. Feature selection through data preprocessing

Table 1 lists the correlation coefficients between the raw data and those after applying the preprocessing methods. The coefficient values with different ranges corresponding to each preprocessing method showed similar values; the correlation coefficient applied with an LPF of 0.6 and a z-score of 3.5 is listed in Table 1. This result shows that the correlation coefficient obtained after noise removal is similar to that of the raw data. In contrast, the correlation coefficient corresponding to the outlier detection significantly differs from that of the raw data because outlier detection removes anomalous data and retains the data that follow the theoretical law.

We established two feature-selection cases for the predictive model and compared their performances and reproducibility based on different datasets. As described in Section 2.3, features with an absolute value above 0.3 were selected. The 13 features were selected from the raw data and the data obtained after noise removal, and 11 features were selected from the data obtained after outlier detection. The feature-selection cases are called scenario R3 and O3 hereafter.

Table 1. Pearson's correlation coefficients for raw and preprocessed data

Variables (Features)	Raw data		Noise removal (Applied LPF 0.6)		Outlier detection (Applied z-score 3.5)	
	Corre	r	r	difference	r	difference
Input	FCV1	0.19	0.19	0.00	-0.05	-0.24
	FCV2	0.59	0.59	0.00	0.79	0.20
	FCV3	0.24	0.24	0.00	0.76	0.52
	LCV4	0.52	0.52	0.00	-0.23	-0.75
	PCV1	0.03	0.04	0.01	0.06	0.03
	LT1	0.11	0.11	0.00	-0.04	-0.15
	FT1	0.16	0.17	0.01	0.10	-0.06
	FT2	0.77	0.80	0.03	0.69	-0.08
	FT3	0.59	0.60	0.01	0.79	0.20
	FT4	0.54	0.55	0.01	-0.05	-0.59
	PT1	-0.61	-0.61	0.00	-0.07	0.54
	PT2	-0.47	-0.47	0.00	0.96	1.43
	PT3	-0.61	-0.61	0.00	-0.13	0.48
	TT1	0.55	0.56	0.01	-0.43	-0.98
	TT2	0.18	0.18	0.00	0.34	0.16
	TT3	0.54	0.54	0.00	0.62	0.08
TT4	0.90	0.90	0.00	0.68	-0.22	
TT5	1.00	1.00	0.00	0.99	-0.01	
TT6	0.69	0.69	0.00	-0.56	-1.25	
Output	TT1	1.00	1.00	0.00	1.00	0.00

3.2. Prediction performance based on the feature selection

Figure 3 shows the ten datasets with the best performances according to each feature-selection case. The line and bar show the minimal, average, and maximum values for permeance and epoch. As shown in Figure 3, the scenario R3 showed a higher and more stable performance than the scenario O3. Although the best performance exhibited R^2 and NRMSE values of 0.962 and 0.045, respectively, we expect that this result can be enhanced by hyperparameter tuning and algorithm change. This study shows that the process characteristics in the feature selection step affect the performance of data-driven predictive models. It is also possible to develop a prediction model that reflects the characteristics of the actual process through appropriate data preprocessing.

Although the outlier detection method with high performance was applied in both cases, it could not predict the start-up period, which is a limitation of this study. The case using raw data and the L1-L3 dataset could predict the whole process period, including the start-up period, but it showed a low performance, with R^2 and NRMSE as 0.936 and 0.254, respectively, because of the inclusion of the outlier data. Therefore, to improve the overall model performance, it is necessary to tune the hyperparameters or to consider other preprocessing methods that efficiently process outlier data for dynamic systems.

4. Conclusions

In this study, a data-based model was developed for the bio 2,3-BDO distillation process, and an appropriate data preprocessing method and range were established considering the process characteristics. A total of 60 datasets, among which 1 comprised raw data and 59 comprised data preprocessed with noise removal and outlier detection, were prepared. To examine whether the feature selection affects the model performance and reproducibility, Pearson's correlation coefficients of the datasets were compared. Two kinds of feature selection scenarios, scenarios R3 and O3, were established as the case studies. Scenario R3 showed a higher performance than that of scenario O3 because it included the process characteristics. We concluded that feature selection based on correlation of the raw data improves the performance of the data-based predictive model. However, it has a limitation, wherein it shows low performance when using L1–L3 data because the start-up period data is recognized as an outlier. Therefore, future studies should focus on developing a data preprocessing method which can classify outlier data efficiently under dynamic conditions in the whole process.

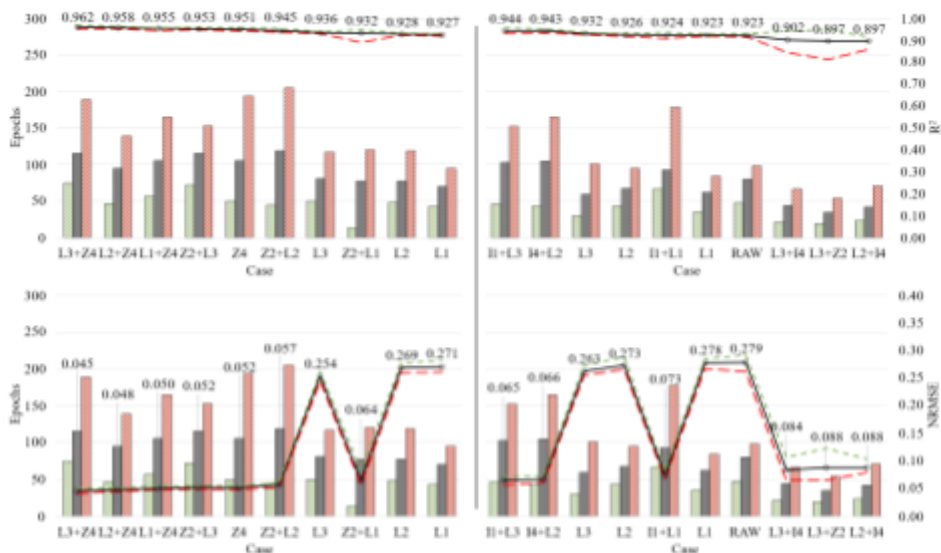


Figure 1. Top 10 results of the R^2 and NRMSE performances based on the feature-selection for scenario R3 (left) and O3 (right).

Acknowledgments

This study was conducted with the support of the Korea Institute of Industrial Technology as “Development of hybrid model and software to optimization of ash removal system in recovery boiler for power generation (EE-20-0014)” and “Development of AI Platform Technology for Smart Chemical Process (JH-21-0005)”.

References

- [1] D. Tinôco *et al.*, 2020, Technological development of the bio-based 2,3-butanediol process, *Biofuels, Bioprod. Biorefining*, 2, 1–20.
- [2] M. A. Ivanushkin *et al.*, 2019, Analysis of statistical methods for outlier detection in telemetry data arrays, obtained from “aIST” small satellites, *J. Phys. Conf. Ser.*, 1326, 1.

- [3] S. Hochreiter and J. Schmidhuber, 1997, Long Short-Term Memory, *Neural Comput.*, 9, 8, 1735–1780.
- [4] H. Kwon *et al.*, 2021, Development and application of machine learning-based prediction model for distillation column, *Int. J. Intell. Syst.*, December 2020.

Data-driven Process Design Exemplified on the Steam Methane Reforming Process

Laurens Lueg^a, Dominik Schack^a, Evrim Örs^a, Robin Schmidt^{a,*},
Patricia Bickert^b, Martin von Kurnatowski^b, Patrick Otto Ludl^b, Michael Bortz^b

^a*AIR LIQUIDE Forschung und Entwicklung GmbH, Innovation Campus Frankfurt, Gwinnerstrasse 27-33, 60388 Frankfurt am Main, Germany*

^b*Fraunhofer Institute for Industrial Mathematics ITWM, Department Optimization – Technical Processes, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany*

robin.schmidt@airliquide.com

Abstract

Process design based on physical models often faces computational problems with respect to convergence, especially if the underlying flowsheets are complex. The use of data-driven surrogate models promises to overcome these challenges. This contribution presents the development of surrogate models and their use for flowsheet simulation. A new sampling strategy consisting of a combination of adaptive and sequential sampling enables the selective placement of new sample points. It is shown, however, that this hybrid strategy does not necessarily lead to higher accuracies than a pure sequential sampling. Surrogates are built for selected key units of the steam methane reforming process, and their individual accuracies are analyzed. When the surrogates are combined to form flowsheets, the prediction errors show a tendency to damp from unit to unit. This proves the suitability of surrogate models for flowsheet simulations. The promising results of this paper pave the way for future work, such as the optimization of flowsheets or superstructure optimization.

Keywords: surrogate modeling, adaptive sampling, artificial neural networks, error propagation, hydrogen production.

1. Introduction

Surrogates replacing flowsheet models have attracted broad attention in process engineering due to their potential to accelerate rigorous simulations for purposes such as exploring the allowed operating window (Heese et al., 2019) and global optimization (Schweidtmann et al., 2019). Their training data is generated with the original simulation according to some sampling strategy. A recent review on the topic can be found in (McBride and Sundmacher, 2019). When dealing with such surrogate models, it is important to distinguish between two different applications: on the one hand, it can be sufficient for the surrogate model to map only subsets of the input space, for example in case of process optimization it is only necessary to be accurate along the trajectory to the optimal solution. Accordingly, the training data does not need to cover the entire input space. On the other hand, the purpose of the surrogate can require to map large parts of the input space. This is for example the case when it is coupled to other sub-models like a unit operation in a flowsheet, as done in this work. Hence, techniques tailored to this

situation are required. A survey on such sampling strategies for global surrogate modeling can be found in (Liu et al., 2018).

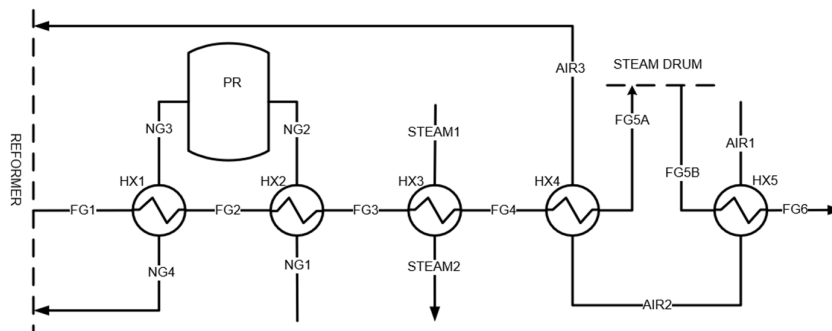


Figure 1: Schematic illustration of the pre-reformer and flue gas waste heat recovery section of an SMR process. The flowsheet is based on Genkin et al. (2010).

This paper continues Air Liquide’s recent activities in the field of surrogate modeling (Soroush et al., 2020; Schmidt et al., 2020). Using the example of the Steam Methane Reforming (SMR) process, the relevant steps for generating a process model for individual surrogate units as well as their combination to a flowsheet are presented. SMR is the state-of-the-art method of producing hydrogen, where a natural gas feedstock (NG) is mixed with steam to produce syngas, and finally hydrogen and CO₂. The focus of this contribution is on the heat recovery from the flue gas (FG) of the SMR process comprising five heat exchangers (HX1-HX5) and one pre-reformer (PR) as illustrated in Figure 1.

2. Methods

2.1. Determination of technically relevant operating conditions of process units

In order to develop a data-driven surrogate model for a specific unit operation, a feasible input region, from which points are sampled for data generation, has to be found. In this work, the definition of these input regions is based on a design simulation of the entire SMR process in Aspen Plus[®] with the following adjustable parameters: (i) The type of natural gas feedstock; (ii) the steam-to-carbon ratio (S/C) in the reformer feed; (iii) the ratio of hydrogen recycle; (iv) the temperature in the pre-reformer feed; (v) the temperature in the reformer feed. Varying these parameters directly influences the composition and/or flow rate of the natural gas and steam streams, and indirectly the properties of the flue gas and air streams in the sub-process in Figure 1. Thus, by varying the adjustable parameters and evaluating the rigorous process simulation accordingly, a set of input points can be recorded at each unit, which is used to define a feasible input region. Here, this is achieved by defining a box domain around the input values and adding a 25 % margin in each dimension to ensure that the boundary regions of the domain are well represented.

For units occurring multiple times within a flowsheet, such as the heat exchanger in the sub-process from Figure 1, there are two modeling options: either to assign an input region and train a surrogate model for each unit (‘one-for-each’), or to define an input region around the combined input domains of all units and use the same surrogate model for each of them (‘one-for-all’). The latter approach will likely result in a more complex surrogate, since a larger input space has to be covered. However, the ‘one-for-all’ model

could also be used more generally, for example in superstructure optimization problems. In this work, the two approaches will be compared with respect to their accuracy in a simulation setting.

2.2. Sampling

In a preceding contribution (Schmidt et al., 2020), the focus was on a Latin Hypercube design (a space-filling strategy) for global metamodeling of the pre-reformer reactor as a sub-model of the overall SMR process. In the current work, this sub-model is used to evaluate an adaptive sampling strategy. Adaptive sampling methods can be based on cross validation, such as the one suggested by Aute et al. (2013). This type of strategy is applicable to any type of model. However, it is prone to rather long computation times because it requires the model to be trained very often. Instead, the variance-based method by Lam and Notz (2008) is adjusted. I.e., the merit function

$$f(\varphi) = (\hat{y}(\varphi) - y^*(\varphi))^2 \frac{\Delta\varphi^2}{\Delta y^2} + d^2(\varphi) \quad (1)$$

is iteratively maximized, the resulting point φ is added to the training data and the model is retrained. Here, $\varphi \in R^{11}$ is the input vector of the pre-reformer simulation. $\hat{y}(\varphi)$ is the model prediction for the response, chosen as the temperature of the product stream, and $y^*(\varphi)$ is the corresponding response value at the training point closest to φ . $\Delta\varphi$ is the Euclidean norm of the ranges spanned by the input variables and Δy is the response range. $d(\varphi)$ is the average Euclidean distance between φ and its k nearest neighbors, where $k = 5$ is chosen. Thus, the first term in $f(\varphi)$ is responsible for local exploitation of strong variations in the response, and the second term ensures a global exploration of the input space. This adaptive sampling method is turned into a hybrid strategy by combining it with a space-filling Sobol sequence. I.e., an adaptive percentage z^* is defined, and in each iteration a random number z is drawn from a uniform distribution. Then f is maximized for $z \leq z^*$, whereas φ is taken to be the next Sobol point for $z > z^*$.

2.3. Development of surrogate unit models

The data used to train the heat exchanger surrogate is generated using the Aspen Plus[®] ‘HEX’ shortcut model, with the added consideration of heat loss. Thus, the following variables are chosen as model input: temperature T , pressure p , molar flow rate F and the composition vector \mathbf{x} for both inlet streams (denoted with c,in and h,in), UA (the exchanger area multiplied by its heat transfer coefficient) and the percentage heat loss $f_{Q,loss}$, as well as the pressure drop dp . The surrogate model predicts the temperatures of both outlet streams (denoted with c,out and h,out). The functional relationships learned from data are given by F_{sur} in Eqs. (2)-(3). Depending on the type of surrogate model used, F_{sur} will have different mathematical forms. In this work, only artificial neural networks (ANNs) are considered:

$$T_{h,out} = F_{sur}^T(\{T, p, F, \mathbf{x}\}_{h,in}, \{T, p, F, \mathbf{x}\}_{c,in}, dp_{c,in}, dp_{h,in}, UA, f_{Q,loss}), \quad (2)$$

$$T_{c,out} = F_{sur}^Z(\{T, p, F, \mathbf{x}\}_{h,in}, \{T, p, F, \mathbf{x}\}_{c,in}, dp_{c,in}, dp_{h,in}, UA, f_{Q,loss}). \quad (3)$$

As mentioned in Sec. 2.2, the model for the pre-reformer reactor is adopted from a previous contribution by Schmidt et al. (2020).

2.4. Connection of surrogate models to form a flowsheet

After the trained ANNs are embedded into the surrogate unit models, it is straightforward to construct a surrogate flowsheet for the entire sub-process in Figure 1. To this end, terminal conditions are introduced that connect the output of one unit with the input of another to represent the process streams. The resulting nonlinear system of equations can be solved either in a sequential-modular or equation-oriented manner. *Pyomo* (Hart et al., 2017) was used to implement the individual units and generate functional representations for the trained ANNs. The *pyomo.network* package allows for the definition of the flowsheet as a connection of units, and it offers sequential decomposition and solution routines. Process optimization is also possible directly in *Pyomo* and will be a future step.

3. Results

3.1. Comparison of sampling strategies

A fully connected feed-forward ANN with a single hidden layer consisting of 100 nodes is used to test the hybrid sampling strategy described in Sec. 2.2. The ANN is implemented in *PyTorch* with a learning rate of 0.002 and with 4000 training epochs. The Python package *sobol_seq* is used to draw the Sobol points, and *scipy.optimize.differential_evolution* is used to maximize the merit function from Eq. (1). The strategy is run with different adaptive percentages z^* , and the root-mean-squared error (RMSE) as well as the maximum deviation of the model predictions with respect to an independent test data set of size $N = 10^4$ (generated with Latin Hypercube) are recorded for each iteration. Since *PyTorch* initializes its net weights randomly, the results are averaged over 10 repetitions of the procedure.

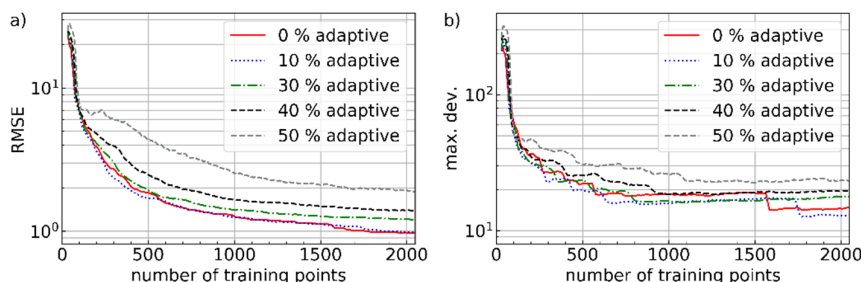


Figure 2: Quality of the ANNs for the pre-reformer reactor trained with the hybrid sampling strategy at different adaptive percentages. a) RMSE, b) maximum deviation with respect to the independent test data set. Curves are averaged over 10 repetitions and then smoothed with a moving average over 20 preceding points.

Figure 2a shows the averaged RMSE of the ANNs for the pre-reformer reactor as a function of the number of training points. The main observation is that a non-zero adaptive percentage does not improve the model accuracy compared to the 100 % Sobol design (solid red curve) for this example. Therefore, purely space-filling sampling methods are used in the subsequent sections. Figure 2b displays the corresponding maximum deviations between prediction and test data at the least accurately predicted test point. Interestingly, this quantity actually slightly increases for most added points. It

only decreases overall due to large steps which occur when a point finally corrects the worst prediction.

3.2. Accuracy of HEX surrogate models

For the ‘one-for-each’ (OFE) approach, a data set of 100,000 points is generated with Aspen Plus[®] for each of the heat exchangers in the sub-process, according to the input-output relationship established in Sec. 2.3. Alternatively, the input data of all heat exchangers are combined to train a single model in the ‘one-for-all’ (OFA) approach. In both cases, the data is then split randomly into training and test sets at a ratio of 85:15. *TensorFlow* is used to train ANNs with one hidden layer on the training set, with their performance being evaluated on the test set. The results are given in Table 1.

It is evident that an increased model complexity (i.e. higher no. of neurons) leads to a better performance for both approaches. However, the mean (‘avg’) and maximum error (‘max’) of the temperature in Kelvin for the OFA method is generally at least three times higher than for the worst performing OFE model. This was also expected, since OFA approximates the HEX model over a larger domain, and will thus be less refined locally, given the same model complexity.

Table 1: Test set accuracy metrics for ANNs of increasing size using one hidden layer. The ‘avg’ and ‘max’ operations are applied across the test set and the two outputs of the surrogate models.

unit	HX1		HX2		HX3		HX4		HX5		OFA	
abs. error (K)	avg	max	avg	max	avg	max	avg	max	avg	max	avg	max
20	0.05	1.01	0.06	0.74	0.03	0.25	0.03	0.38	0.03	0.39	0.28	2.82
50	0.02	0.32	0.02	0.24	0.01	0.08	0.01	0.15	0.01	0.11	0.09	1.22
100	0.01	0.21	0.02	0.24	0.01	0.07	0.01	0.11	0.00	0.06	0.05	1.04

3.3. Error propagation in the flowsheet

To evaluate the performance of the surrogate flowsheet, an equivalent process model in Aspen Plus[®] is used to generate a test data set of 5,000 points by space-filling sampling. The surrogate flowsheet is evaluated at the same input points and the error in the different streams is recorded. The error within the flue gas (‘FG’) stream is of particular importance since it allows the investigation of the error propagation between the surrogate units. An ‘error propagation factor’ ε is introduced, which describes the ratio of the prediction error in the flue gas stream temperature $e_{FG,out}$ to the maximum error in the input temperature for each unit e_{in} :

$$\varepsilon = \frac{|e_{FG,out}|}{|e_{in}|} \quad (4)$$

Table 2: Performance metrics of the surrogate models during evaluation of the connected surrogate flowsheet on a test set of 5,000 points.

unit	HX1		HX2		HX3		HX4		HX5	
metric	ε	max (K)	ε	max (K)	ε	max (K)	ε	max (K)	ε	max (K)

‘One for each’

no. neurons	20	0.54	0.90	1.72	0.46	0.91	0.45	0.78	0.36	-	0.10
	50	0.27	0.44	1.98	0.29	0.98	0.32	1.07	0.25	-	0.08
	100	0.25	0.43	1.51	0.28	0.89	0.31	1.14	0.24	-	0.08
<i>'One for all'</i>											
no. neurons	20	1.73	1.93	1.49	1.87	1.12	1.83	0.86	1.77	-	1.17
	50	1.34	0.62	1.37	0.69	1.06	0.66	0.71	0.57	-	0.27
	100	1.00	0.60	1.42	0.61	0.97	0.53	0.69	0.43	-	0.17

Hence, a unit amplifies the error in its input if $\varepsilon > 1$. In Table 2, the maximum absolute error and the median error propagation factor at each heat exchanger across the test set are listed for both the 'OFE' and 'OFA' approach. Again, a better performance of the OFE models compared to the more general OFA method is observed. Interestingly, the majority of the OFE models achieve propagation factors of less than one, meaning that the prediction error is actually damped along the FG stream.

4. Conclusions

In this work, the development of surrogate unit models for flowsheet simulation was presented. It was shown that an adaptive sampling strategy for the case study under consideration is not advantageous compared to a purely space-filling sampling. Consequently, space-filling designs were used in the following parts of this paper. In general, the "one-for-each" modeling approach leads to a higher accuracy than the "one-for-all" approach. Finally, it was shown that the connection of individual surrogate models to a flowsheet leads to high-quality results in terms of accuracy and error propagation, paving the way for real-time or superstructure optimization. Nevertheless, it has to be verified whether the promising results are applicable to larger processes with more process units. Future work will also focus on the minimum number of training points required to achieve a desired model accuracy as well as on investigating the distribution of the "worst" predicted points of the test set.

Acknowledgement

The financial support of the German Federal Ministry for Economic Affairs and Energy (BMWi) under grant number 01MK20014B (KEEN) is gratefully acknowledged.

References

- V. Aute, K. Saleh, O. Abdelaziz, S. Azarm, R. Radermacher, 2013, Cross-validation based single response adaptive design of experiments for Kriging metamodeling of deterministic computer simulations. *Structural and Multidisciplinary Optimization*, 48, 581–605. doi:<https://doi.org/10.1007/s00158-013-0918-5>
- E. S. Genkin, S. P. DiMartino, M. R. Alvarez, D. Anthony Zagnoli, C. F. Harris, 2010, Steam-hydrocarbon reforming method with limited steam export, U.S. Patent No. 7,850,944 B2.
- W. E. Hart, C. Laird, J.-P. Watson, D. L. Woodruff, G.A. Hackebeil, B. L. Nicholson, J. D. Sirola, 2017, *Pyomo – Optimization Modeling in Python*, Springer, 2017.
- R. Heese, M. Walczak, T. Seidel, N. Aspiron, M. Bortz (2019); Optimized data exploration applied to the simulation of a chemical process; *Comp. Chem. Eng.* 124, 326–342.
- C.Q. Lam, W.I. Notz, 2008, Sequential adaptive designs in computer experiments for response surface model fit. *Statistics and Applications*, 6(1 & 2), 207–233.

- H. Liu, Y.-S. Ong, J. Cai, 2018, A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Struct. Multidisc. Optim.*, 57, 393–416. doi:10.1007/s00158-017-1739-8
- K. McBride, K. Sundmacher, 2019, Overview of Surrogate Modeling in Chemical Process Engineering, *Chem. Ing. Tech.*, 91(3): 228–239.
- R. Schmidt, A. Chattot, A. Bouchrit, M. Mighani, E. Örs, 2020, A Practical Application of Simulation-based Surrogate Modeling for Prereformer Reactor, *Computer Aided Chemical Engineering*, 47, 523–528.
- A. Schweidtmann, A. Mitsos, 2019, Deterministic Global Optimization with Artificial Neural Networks Embedded, *Journal of Optimization Theory and Applications*, 3, 925–948.
- M.Soroush, M. Baldea, T. Eldar, 2020, Smart manufacturing in industrial gas production: A digital transformation, *Smart Manufacturing*, 1, 333–387.

Input transformation for linearization, decoupling and disturbance rejection with application to steam networks

Cristina Zotică^a, Sigurd Skogestad^a

^aDepartment of Chemical Engineering, Norwegian University of Science and Technology, Sem Sælands vei 4, Trondheim 7491, Norway

Abstract

Pressure-flow networks are inherently multivariable, coupled and nonlinear systems that are not straightforward handled by conventional PID controllers. In this work, we analyze the control problem for a steam distribution network in the framework of a recent proposed method, that gives linearization and decoupling, together with a systematic design for nonlinear feedforward control for perfect disturbance rejection.

Keywords: steam networks, linearization, decentralized control, feedforward.

1. Introduction

Steam networks are used to produce and transfer steam as utility for downstream processes such as distillation columns, paper machines, reactors etc. Pressure-flow networks are inherently highly coupled systems where large and fast disturbances such as shut-downs or start-ups become large disturbances both on the generation and the demand side. In addition, the dynamics of steam generators are much slower compared to the dynamics of the steam network. Therefore, to be able to respond fast to load changes, control of the network pressure is commonly implemented in industry (Majanne, 2005).

Both decentralized and centralized control strategies for steam networks are presented in the literature. The work by Bertrand and McAvoy (1986) presents a solution based on PI-controllers that has good performance for disturbance rejection. The work by Kristoffersen et al. (2014) implements model predictive control (MPC) combined with real time optimization to increase the energy efficiency. The work by Majanne (2005) compares the performance of PI-controllers and MPC, and the latter outperforms due to its ability of handling coupled systems.

In this work, we apply the method by Zotică et al. (2020), which transforms a nonlinear system into a first-order linear decoupled system with no effect from disturbances. In addition, we extend the method by explaining how to select the new introduced tuning parameter. Similar methods have been proposed. Feedback linearization linearizes the input-state map of a nonlinear system (Isidori et al., 1981, Khalil, 1992). Input-output linearization linearizes only the input-output map of nonlinear system, while the rest of the system remains nonlinear (Henson and Seborg, 1997). Active disturbance rejection control introduces an observer to estimate unmeasured states and disturbances for

linearization and feedforward disturbance rejection (Huang and Xue, 2014). The work by Lee et al. (2016) combines an extended high-gain observer for unmeasured states and uncertainty together with dynamic inversion that inverts the model using a fast inner loop I-controller. All these methods give a chain of integrators, which brings additional limitations for control, and it is not very robust as in some cases the input might move in the opposite direction initially.

2. New method for input transformation

Figure 1 shows the block diagram for the new proposed method. The key assumptions are that we have the same number of inputs (u) and outputs (y), and that we can measure the disturbances (d). In the example, we will also allow for some measured states ($w = [T_T \ T_L]$) to be treated as measurements. We define a new input (v) from the model equations as a function of the physical input (u), disturbances (d), extra states (w) and output (y) with the objective of transforming the original nonlinear system into a linear first-order system. The resulted system is also decoupled, and therefore, we use SISO-controllers to control y by using v . We find the physical input u by solving as set of nonlinear algebraic equations that give u as a function of v, y, w, d .

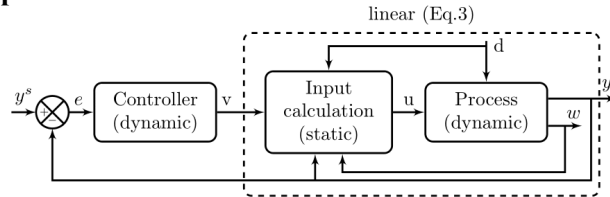


Figure 1. Block diagram for transformed inputs for linearization, decoupling and perfect disturbance rejection.

Consider a general nonlinear system given by Eq. 1.

$$\frac{dy}{dt} = f(y, u, w, d) \quad (1)$$

We define the transformed input v as given in Eq. 2.

$$v = f(y, u, w, d) - Ay \quad (2)$$

where, A is a new tuning parameter, which we discuss in Section 2.1.

By introducing the new input v , the new system becomes first-order, linear, decoupled and with no effect from disturbances, as shown in Eq. 3.

$$\frac{dy}{dt} = v + Ay \quad (3)$$

2.1. New tuning parameter A

One way to select A is such that nominally the positive feedback from y to v is small. Therefore, for each output y_i , we may select A_i as the diagonal elements of the Jacobian of $f_i(y, u, w, d)$ with respect to the output y_i evaluated at the nominal operating conditions,

$$A_i = \text{diag} \left(\left. \frac{\partial f_i(y, u, w, d)}{\partial y_i} \right|_* \right)$$

We may also select a larger A to speed-up the response, or smaller to slow it down. Note that selecting $A = 0$ gives an integrating process similarly to feedback linearization for a model as given in Eq. 1.

2.2. Input calculation block

The input calculation solves Eq. 2 with respect to the inputs u given controller outputs v , extra states w , outputs y and disturbances d . If there is no explicit solution of inverting Eq. 2, we may use a numerical algebraic solver, or an I-controller in a fast inner loop. The second method can be applied to systems with singularities in the transformation.

2.3. Controller tuning

We tune the SISO-controllers based on the SIMC tuning rules in Eq. 4 (Skogestad, 2003).

$$K_c = \frac{1}{k} \frac{\tau}{\tau_c + \theta} = \frac{1}{\tau_c + \theta} \quad (4a)$$

$$\tau_I = \min(\tau, 4(\tau_c + \theta)) \quad (4b)$$

where, K_C is the proportional gain, τ_I is the integral time, k is the process gain, τ is the open-loop time constant, θ is the delay and τ_C is the desired closed-loop time constant. Note that from Eq. (3), $k = \tau = -1 / A$.

3. Case study: steam network

Figure 2 shows the system we analyze within the new proposed method. It is composed of a high-pressure header (i.e. pipelines that physically connect the steam generators and consumers), a turbine and a low-pressure header. High-pressure steam is produced at a pressure p_0 by burning fuel in a boiler. Note that we do not include the boiler in our analysis. The high-pressure steam is supplied as utility to one high pressure consumer with receiving pressure p_{HC} . The remaining steam is expanded to lower pressure steam, either through a fixed-speed back pressure turbine connected to the electric grid to produce electricity, or through a valve that bypasses the turbine (z_{TB}). Note that the fixed-speed turbine is not a degree of freedom available for operation. The low-pressure steam is supplied as utility to two consumers, with receiving pressures p_{LC1} and p_{LC2} respectively.

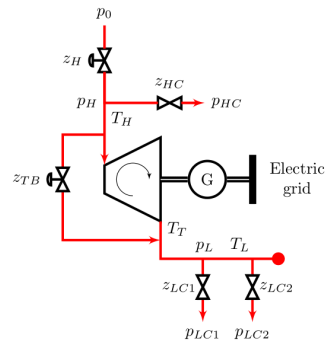


Figure 2. Process flowsheet of the steam network with two pressure headers (high and low) considered in this work.

The manipulated variables are $u = [z_H z_{TB}]$ (the supply of high pressure steam and the turbine bypass). The controlled variables are $y = [p_H p_L]$ (pressure in the high and low pressure headers). The main disturbances are $d = [p_0 z_{HC} z_{LC1} z_{LC2} p_{HC} p_{LC1} p_{LC2}]$ (the high pressure steam supply (p_0), and the consumers demand of high and low pressure steam, given by changes at the receivers pressure or of the valve positions (z_{HC} , z_{LC1} and z_{LC2})). The additional states are $w = [T_T T_L]$ (the temperature at the turbine outlet and in the lower pressure header.)

3.1. Nominal operating conditions

Table 1 shows the nominal operating conditions, typical for a steam network found in a chemical plant. Here, V is the volume of the two headers.

Table 1. Nominal operating conditions

Variable	p_0	p_H	p_{HC}	p_L	p_{LC1}	p_{LC2}	z_i	T_H	T_L	V_H	V_L
Value	42	40	38	7	6	5	0.5	380	200	1	5
Unit	bar	bar	bar	bar	bar	bar	-	°C	°C	m ³	m ³

3.2. Model

We assume ideal gas, constant specific heat capacity, no pressure losses and perfect mixing in both pressure headers. Assuming isothermal conditions in the high pressure header (T_H constant), the dynamic mass balance in pressure form becomes Eq. 5.

$$\frac{dp_H}{dt} = \frac{RT_H}{V_H} (q_H - q_{HC} - q_{TB} - q_T) \stackrel{def}{=} f_H \quad (5)$$

where q_i is the molar flow through a valve.

The low pressure header is not isothermal because work is extracted in the turbine, and therefore the mass and energy balance become coupled. The energy balance in temperature form is given in Eq. 6.

$$\frac{dT_L}{dt} = \frac{RT_L}{V_L p_L} (q_{TB}(T_H - T_L) + q_T(T_T - T_L)) \quad (6)$$

The mass balance in pressure form is given in Eq. 7.

$$\frac{dp_L}{dt} = \frac{R}{V_L} (q_{TB}T_H + q_T T_T - (q_{LC1} + q_{LC2})T_L) \stackrel{def}{=} f_L \quad (7)$$

We assume isentropic expansion in the turbine and that there are no constraints for the power supplied to the electric grid. Therefore, the temperature at the turbine outlet (T_T) is computed from Eq. 8.

$$T_T = T_H \left(\frac{p_L}{p_H} \right)^{\frac{\gamma-1}{\gamma}} \quad (8)$$

where γ is the heat capacity ratio of steam.

To model the molar flows through valves, we use a valve equation with a linear valve characteristic (Eq. 9).

$$q_i = C_{v,i} z_i \sqrt{|p_{in}^2 - p_{out}^2|}, \forall i \in (H, HC, TB, LC1, LC2) \quad (9)$$

where $C_{v,i}$ is the valve coefficient, z_i is the valve opening, p_{in} and p_{out} are the pressures before and after the valve respectively.

To model the molar flow through the turbine, we assume a constant mass flow coefficient (ϕ), equivalent to a choked turbine (Eq. 10).

$$q_T = \phi \frac{P_H}{\sqrt{T_H}} \quad (10)$$

3.3. Input transformation

The new input $v = [v_H \ v_L]$ is defined by applying Eq. 2 resulting in Eq. 11. We assume that the measurements for T_H , T_L and T_T are available.

$$v_i = f_i - A_i p_i, \forall i \in (H, L) \quad (11)$$

$$A_i = \text{diag} \left(\left. \frac{\partial f_i}{\partial p_i} \right|_* \right), \forall i \in (H, L)$$

where, at the nominal conditions from Table 1.

The new system in Eq. 12 is linear, decoupled and has perfect disturbance rejection.

$$\frac{dp_i}{dt} = v_i + A_i p_i, \forall i \in (H, L) \quad (12)$$

3.4. Input calculation

We find the unknown variable $u = [z_H \ z_{TB}]$ by solving the system of linear equations (Eq. 13) resulted from rewriting Eq. 11.

$$\begin{bmatrix} z_H \\ z_{TB} \end{bmatrix} = \begin{bmatrix} \alpha & -\beta \\ 0 & \beta \end{bmatrix}^{-1} \begin{bmatrix} \frac{V_H}{RT_H} (v_H + A_H p_H) + q_{HC} + q_T \\ \frac{V_L}{R} (v_L + A_L p_L) - q_T T_T + (q_{LC1} + q_{LC2}) T_L \end{bmatrix} \quad (13)$$

with $\alpha = C_{v,H} \sqrt{|p_0^2 - p_H^2|}, \beta = C_{v,TB} T_H \sqrt{|p_H^2 - p_L^2|}$

Note that from Eq. 6 and 8, T_L and T_T depend on u , therefore it is not a true disturbance. However, the use of a measured T_L and T_T in the input transformation is not a problem in

this case because the dynamics from the inputs u to the outputs y have a stable inverse (have no RHP-zeros), hence the inverse generated by the input transformation is stable.

3.5. Simulation results

Figure 3 shows the responses for disturbance rejection and setpoint changes for $y = [p_H \ p_L]$ (Figure 3a), $u = [z_H \ z_{TB}]$ (Figure 3b) and $v = [v_H \ v_L]$ (Figure 3c) to $p_0 = 42$ bar at time $t = 10$ s, $p_{HC} = 39$ bar at time $t = 20$ s, $p_{LC1} = 5.5$ bar at time $t = 30$ s, $p_{LC2} = 1$ bar at time $t = 40$ s, $p_H^s = 39$ bar at time $t = 50$ s and $p_L^s = 6$ bar at time $t = 60$ s. We tune the

PI-controllers with $\tau_{C,i} = -1 / 2A_i$, which are only used for setpoint changes. The results in Figure 3 show a decoupled process with perfect disturbance rejection.

4. Discussion

The calculation block is inherently a nonlinear feedforward controller, and therefore we do not need the feedback control in Figure 1 (typically a PI-controller) as long as we have a perfect model and measurements. Setpoint changes can be handled by directly changing the setpoint for v . However, in a real plant we will always have unmeasured disturbances and unmodelled dynamics, and we need the outer PI-controller loop.

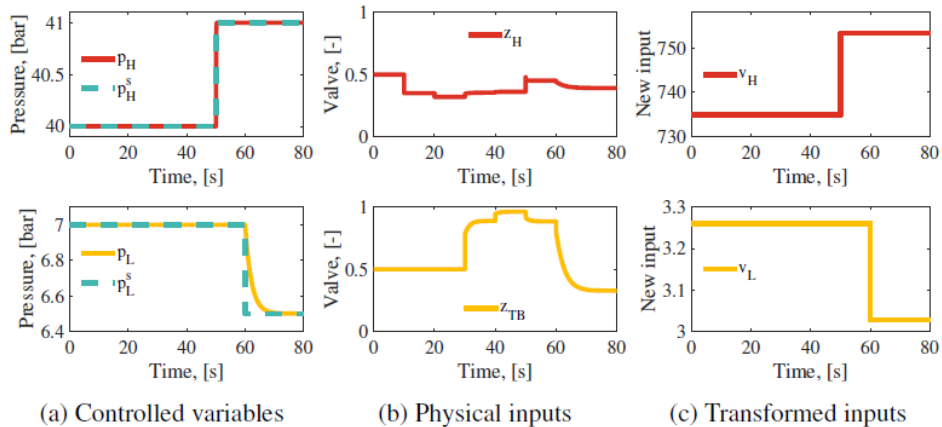


Figure 3. Simulation results for disturbance rejection and setpoint changes.

5. Conclusions

Steam networks are interactive systems, where the main task of the control system is to reject disturbances either on the steam generation or demand side. We design the control structure by using a new method for input transformation that gives linearization, decoupling and perfect disturbance rejection both dynamically and at steady-state (Eq.12), which makes it a good fit for the control structure of a steam network, as seen in Figure 3. The method also transforms a nonlinear system into a first-order linear by introducing a new tuning parameter A .

Acknowledgements

This publication has been partly funded by HighEFF - Centre for an Energy Efficient and Competitive Industry for the Future, an 8-years' Research Centre under the FME-scheme (Centre for Environment-friendly Energy Research, 257632). The authors gratefully acknowledge the financial support from the Research Council of Norway and user partners of HighEFF.}

References

- C. R. Bertrand, T. J. McAvoy, 1986. Short-Cut Analysis of Pressure Control in Steam Headers. In: 1986 American Control Conference. pp. 1757–1763.
- M. A. Henson, D. E. Seborg (Eds.), 1997. Nonlinear Process Control. Prentice-Hall.
- Y. Huang, W. Xue, 2014. Active disturbance rejection control : Methodology and theoretical analysis. *ISA Transactions* 53 (4), 963–976.
- A. Isidori, A. J. Krener, C. Gori-Giorgi, M. Salvatore, 1981. Nonlinear decoupling via feedback: A differential geometric approach. *IEEE Transactions on Automatic Control* 26 (2), 331–345.
- H. K. Khalil, 1992. *Nonlinear systems*. Macmillan, New York.
- T. T. Kristoffersen, D. Snarheim, L. Imsland, M. S. Govatsmark, 2014. Optimal and Robust Production of High Pressure Steam. In: 2014 European Control Conference. IEEE, Strasbourg, pp. 491–497.
- J. Lee, R. Mukherjee, H. K. Khalil, 2016. Output feedback performance recovery in the presence of uncertainties. *Systems & Control Letters* 90, 31–37.
- Y. Majanne, 2005. Model predictive pressure control of steam networks. *Control Engineering Practice* 13, 1499–1505.
- S. Skogestad, 2003. Simple analytic rules for model reduction and PID controller tuning. *Journal of Process Control* 13 (4), 291–309.
- C. Zotică, N. Alsop, S. Skogestad, 2020. Transformed Manipulated Variables for Linearization, Decoupling and Perfect Disturbance Rejection. *IFAC-PapersOnLine*.

SEARCH: A Symptom-based Expert for Advanced Response to Chemical Hazards

Sangwoo Yoo,^a Eunji Shin^a and Dongil Shin^{a,b,*}

^a *Department of Disaster and Safety, Myongji University, Yongin, Gyeonggido 17058, Korea*

^b *Department of Chemical Engineering, Myongji University, Yongin, Gyeonggido 17058, Korea*

dongil@mju.ac.kr

Abstract

Along with the increase in the production, distribution, and consumption of chemicals, the risk of chemical accidents is increasing, and damage by chemicals exposure accidents can be greatly expanded if initial response is inappropriate. Since early determination of chemical substances in the event of a leakage accident is essential for minimizing accidental loss, this study proposes an AI-based system that supports chemical determination by field personnel based on exposure symptoms, in addition to detection using sensors, and provides predictive information on symptoms that may be lacking for new chemicals. In order to analyze and document the symptoms expressed when contacting chemical substances, the knowledge of domestic and foreign symptoms is established as a knowledge base using a knowledge graph. Afterwards, the symptom-based chemical estimation knowledge service is completed by estimating exposed chemicals through reasoning methods such as SPARQL and providing response information that can be immediately taken on site. In order to expand the knowledge of new chemicals, machine learning is used to predict useful symptom information from molecular structure information of chemicals, so that pre-emptive risk management for new materials could be carried out using the proposed system.

Keywords: chemical accident, exposure symptom knowledge, chemical identification, machine learning, intelligent decision support

1. Introduction

Increasing the use of chemicals and improving the quality of human life are inseparable. According to international statistics, 103M chemical substances were developed as of 2015, and about 120,000 chemical substances are actually distributed (Ministry of Environment's Chemical Policy Division, 2016). In addition, about 2,000 new chemical substances are entering the market every year after safety evaluation. In Korea, the distribution volume also increased by 12.4% compared to 2014 as of 2016 (Ministry of Environment press release, 2018), and the risk of various chemical accidents is increasing as the distribution volume increases. The number of domestic chemical accidents is increasing from 233 in 2014 to 277 in 2018.

Chemicals give many benefits to people, but hazardous materials such as hazardous chemicals and chemical accidents have a dual nature that threatens people's lives. In addition, safety awareness and countermeasures for hazardous materials are not keeping up with the growth rate of the chemical industry. In order to solve these problems, various systems and studies are being conducted to respond to chemical accidents. The National

Institutes of Health (NIH) has developed the Wireless Information System for Emergency Responders (WISER) to assist in emergency rescue in case of chemical accidents. In addition, Bhavnani et al. (2007) analyzed the design of a response system to quickly identify chemicals in emergency situations (Bhavnani et al. 2007).

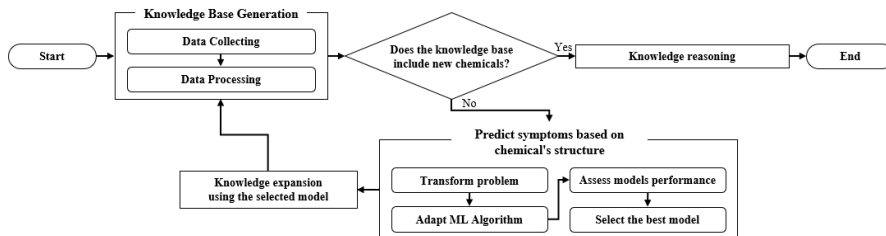


Figure 1. An overall framework of the proposed symptom-based knowledge reasoning.

If the initial response to a chemical exposure accident is inappropriate, there is a risk of spreading to enormous damage including personal injury and property damage. Existing chemical detection technologies are focused on the development of countermeasure technologies for leaks at known sites, analysis of candidates for leaks, and detection studies in dynamic changing environments. If a sensor is not installed at the accident site, an initial estimation can only be made based on qualitative exposure symptoms. However, the knowledge service on hazardous materials considering human bio sensing information (symptoms) is relatively insufficient in development. Therefore, in this study, we propose a real-time intelligent support system for hazardous chemical detection and diagnosis that quickly detects exposed chemicals, responds to dangerous situations early, and protects workers and exposed people from unexpected chemical accidents in the field (see Fig. 1).

2. Example SEARCH Platform: Chemical substance identification knowledge service based on symptom knowledge reasoning

When unexpected symptoms are found in industrial sites, laboratories, etc. or in multi-use facilities, the SEARCH system, which enables the detection and determination of exposed chemicals, is operated as in Fig. 2. The proposed platform supports speech recognition and output of diagnostic results, but in this paper, we will only focus on the

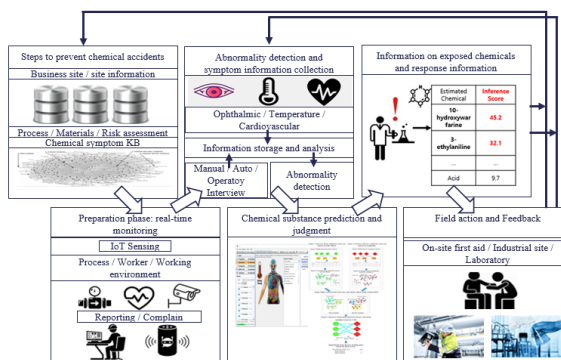


Figure 2. Flowchart of knowledge service on chemical identification and emergency response to exposed chemicals using SEARCH.

collection and storage of chemical substances and symptom information, which are key items in the operation of SEARCH, and reasoning based on the knowledge base.

3. Collection and Preprocessing of Chemical Substance and Symptom Information Data

3.1. Collection of symptom knowledge

The knowledge of symptoms of exposure to chemicals was designed based on the NIH's WISER (Hochstein, C., 2008). The WISER database contains 499 chemical substances and 79 symptom information divided into 10 categories such as temperature, nervous system, oral cavity, etc. We further expanded symptom knowledge by converting 502 data provided by PubChem.

3.2. Chemical data and information preprocessing

In order to transform the collected chemical's structure information into a molecular fingerprint, it was converted into MACCS keys and PubChem keys (see Fig. 3). MACCS keys are the most commonly used structural keys. Among the structure keys of 960 bits and 166 bits, we used a structure key of 166 bits, and a Python package RDKit was used to convert SMILES to MACCS keys. The SMILES structure of 984 chemicals was converted into MACCS keys through RDKit. PubChem keys are a molecular fingerprint used by NIH's PubChem to explore similarities, consisting of 881 bits of structural keys and encompassing a variety of substructure and features. PaDEL-Descriptor was used to convert SMILES into PubChem keys. SMILES of 984 chemical substances were converted into PubChem keys through PaDEL-Descriptor. In addition, each symptom information was transformed into a binary matrix to be applied to the classification model. The binary matrix is expressed as 1 if it has a specific symptom and 0 if it does not and was applied to 1,001 chemical substances and 79 symptoms stored in the knowledge base.

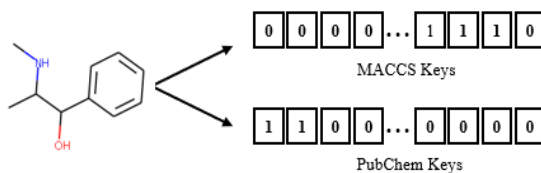


Figure 3. Transform of chemical structure to molecular fingerprints.

4. Knowledge Base Generation and Reasoning

4.1. Knowledge base generation

Since various knowledge services such as determination of exposed chemical substances require effective expression of knowledge, the knowledge base was established using a knowledge graph. Structural and physicochemical information on 449 chemicals of WISER and 502 chemicals of PubChem were extracted through Chempiper, a Python package, and 79 information extracted from WISER were used as symptom information.



Figure 4. An example of triple in knowledge base.

The extracted chemical information was saved in various input triple forms using AllegroGraph API (see Fig. 4). The chemical information stored in the knowledge base includes CAS No., symptoms, synonyms, structural information, and molecular formula etc.

4.2. Knowledge base reasoning

Using SPARQL provided by AllegroGraph, reasoning on exposed chemical substance(s) is performed by inputting a set of symptom information. As a case study, Case 1's inputs are 6 symptom information (headache, runny nose, sweating, arrhythmia, chills), and Case 2's inputs are 7 symptom information by adding throat irritation to Case 1's symptom information. Cases 1 and 2 are like "There is chemical substance B if symptom A is presented", reasoning chemical substances from existing generalized symptom information. If you do not have certain specific symptoms, in the reasoning process, you may increase the speed and accuracy of the reasoning by excluding them. Case 3 is an example: adding information on the no bloody nose to the symptom set in Case 2, in other words, the bloody nose is the excluded symptom.

As a result of the reasoning, it can be showed that 18 chemicals were printed in Case 1, 11 chemicals in Case 2, and 2 chemicals in Case 3: Cases 1 and 2 are the same as the results provided by WISER. It has been shown that if specific symptoms are clearly excluded as in Case 3, which is an unprecedented types of reasoning, less reasoning process may be required, and the speed and accuracy of reasoning for the inferred chemicals can be improved (see Fig. 5).

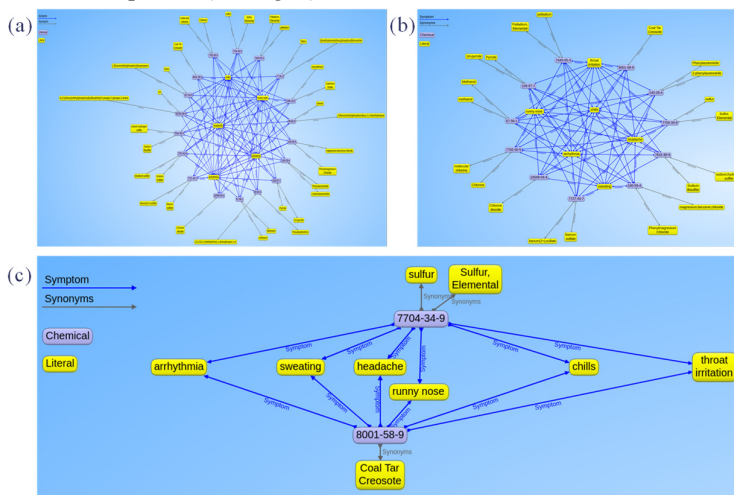


Figure 5. Reasoning results: (a) Case 1; (b) Case 2; (c) Case 3.

5. Symptom Knowledge Expansion Using Machine Learning

5.1. Problem transformation

Since the extracted chemical substances may have more than one symptom, prediction of symptom information based on chemical's structure information is a multi-label classification problem (Tsoumakas, G., 2009). To solve this problem, we apply Label Powerset (LP), Binary Relevance (BR), and Chain Classifier (CC), which are methods of solving the problem of multi-label data with a single label. LP is a method of classifying multiple labels by converting them into one class, and BR is a method of converting into a single label by determining by assigning each classifier to each label. CC is a method

that creates a learner for each classification and sequentially reflects the guess result to the next input when guessing the label (see Fig. 6).

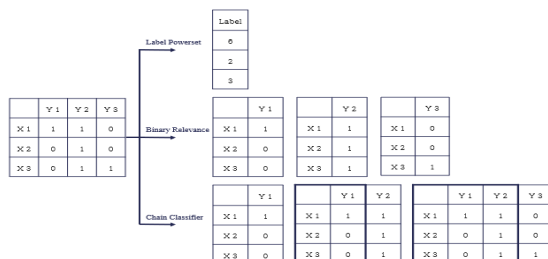


Figure 6. Multi-label classification: Problem transformation method.

5.2. Adaptation of ML algorithm

In this study, k-Nearest Neighbor (k-NN), Decision Tree (DT) and Random Forest (RF) were used as ML methods for knowledge expansion. In order to compare and analyze the effective fingerprinting method, it was divided into MACCS keys and PubChem keys. In addition, LP, BR, and CC were applied to each machine learning method in order to compare and analyze which method is more effective among solving the problem of multi-label data with a single label. In case of Case 1, MACCS keys are input as input values, and PubChem keys are input as input values in Case 2, and symptom information transformed into binary matrix form as output value is displayed.

5.3. Model training method

For accurate training of the model, training data and test data were divided into 8:2 ratio. Since there were little data in the training process, k-fold cross-validation was applied. In this study, k=5 was set.

5.4. Assessment of model performance

The performance evaluation method for a multi-label classification can be divided into example-based and label-based. Example-based is to calculate the average of the difference between the predicted label and the actual label for each case. And the label-based is to calculate the average by calculating the predicted performance individually for each label (Tsoumakias, G., 2009).

Hamming loss was used among the example-based methods because the labels of the data in this study are unbalanced. Hamming loss is the ratio of misclassified labels among all labels (Kim, S., 2017), and the Hamming loss equation is as follows:

$$\frac{1}{N \times L} \sum_{i=1}^N \sum_{j=1}^L XOR(y_{i,j}, \hat{y}_{i,j}) \quad (1)$$

where N is number of data instances, L cardinality of class space, $y_{i,j}$ actual bit of class label j in data instance i , $\hat{y}_{i,j}$ and predicted bit class label j in data instance i .

5.5. Result of model performance assessment

When looking at Cases 1 and 2, it was found that the MACCS keys showed good performance when applying the DT model, and the PubChem keys were effective when the k-NN and DT models were applied. In addition, the method of applying multi-label multiple classification is effective in the order of BR > CC > LP for k-NN and RF models, and LP > BR > CC for DT models regardless of cases. The model that shows the best

performance is the BR(RF) model, which shows the Hamming loss value of 0.2391 in Case 1 and 0.2447 in Case 2 (see Fig. 7).

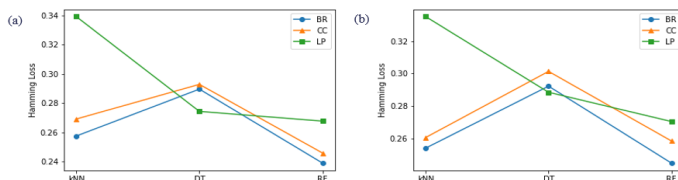


Figure 7. ML model performance for each transformation method: (a) Case 1; (b) Case 2.

6. Conclusion

Based on the input of real-time symptom information, a knowledge reasoning-based chemical estimation system (SEARCH) was designed to predict exposure chemical substance(s) candidates, and its effectiveness was verified through case studies. For chemical substances with high risk at industrial or chemical incident sites, a total of 1,001 expanded symptom information was obtained by adding the existing symptoms by NIH WISER and chemical substances provided by PubChem. This was built into a KB as knowledge graph, and through reasoning knowledge graph using SPARQL, various knowledge services such as discrimination of exposed chemical substances and recommendation of field response are provided. The case study showed the same estimation results as the existing standard system, WISER, and showed that the speed and accuracy of discrimination can be improved through the reasoning method that is not available in WISER.

For the scalability of the system in the future, a function that predicts and provides symptom knowledge that can be used to detect symptoms even for new chemical substances with relatively insufficient information was prosed and implemented. Based on chemical substance information such as structure information, it was confirmed that BR(RF) is the most effective model through the development and performance comparison of a multi-label multi-class ML model that estimates symptoms of new chemical substances. Deep learning models are being developed by further expanding the data, and the developed SEARCH platform is being applied to the Gyeonggi-do Disaster and Safety Headquarters, which is responsible for the safety and accident response of the largest population in Korea.

References

- Bhavnani, S. K., et al., 2007, "Network analysis of toxic chemicals and symptoms: implications for designing first-responder systems," *AMIA Annual Symposium Proceedings*, 51-55.
- Hochstein, C., et al., 2008, "Selected Resources for Emergency and Disaster Preparedness and Response from the United States National Library of Medicine," *Medical Reference Services Quarterly*, 27(1), 1-20.
- Kim, S., Cho, C. and Lee, E., 2017, "Studies on the Chemical Accidents of Korea by the Statistics and Case Review," *Korean Journal of Hazardous Materials*, 5(1), 50-58.
- Ministry of Environment press release, 2018, *2016 chemical distribution volume increased 12.4% compared to 2014, while hazardous chemical increased 8.5%*, <http://www.korea.kr/news/pressReleaseView.do?newsId=156298254>
- Ministry of Environment's Chemical Policy Division, 2016, *Chemical Safety Check Before First Launch*, Ministry of Environment.
- Tsoumakas, G., Katakis, I. and Vlahavas, I., 2009, Mining multi-label data. In: Maimon, O. and Rokach, L. (eds), *Data Mining and Knowledge Discovery Handbook*, 667-685.

A Cloud Computing Application for the Supercritical Carbon Dioxide Extraction Using Coffee Grounds Silverskin

Andrea Galeazzi^a, Rita Nasti^b, Giulia L. Bozzano^a, Luisella Verotta^b, Stefania Marzorati^b and Flavio Manenti^{a,*}

^a*Politecnico di Milano, Piazza Leonardo Da Vinci 32, Milan 20133, Italy*

^b*Università degli Studi di Milano, Via Golgi 19, Milan 20133, Italy*
flavio.manenti@polimi.it

Abstract

This work presents a case study for the application of physical systems modelling techniques on cloud computing infrastructures. The described physical system is the micronized silverskin lipids extraction using supercritical carbon dioxide on laboratory scale equipment. The experimental data of the extractions are modeled using a simplified dual phase kinetic model. The model parameters are further developed until their constituent and constant terms in order to transform the purely regressive kinetic model to a predictive one. The model parameters have been refitted through a non-linear regression using the cloud computing infrastructure purposely created for this case study. The core numerical library used to construct the cloud computing approach is the object-oriented BzzMath library for C++. Cloud computing is a very promising research field and its application to traditional solutions may provide a numerical revolution in the chemical industry as well as academic researches.

Keywords: cloud computing, supercritical fluid extraction, silverskin

1. Introduction

Cloud computing is a very promising technology that moves away from the end-user the burden of performing critical and time-consuming calculations while still retaining its results. This approach is not new and several applications have been developed throughout the years but they are mainly industrial or related to large-scale calculations. More recently, driven by the industrial focus on this topic, an increasing interest in cloud computing applications has spread also in the academy. Several academic fields can benefit from a cloud computing approach to typical problems, one of them is the chemical engineering field, especially in chemical process engineering and modelling. In this work, a cloud computing approach has been applied to the modelling of the supercritical carbon dioxide extraction of lipids from organic matrices, with a focus on coffee grounds silverskin. At first, a custom online platform accessible through a website has been created in order to accommodate the models and calculations required for this case study. This process is a viable candidate to demonstrate an online cloud solution since the mathematical model implemented is useful for the researcher who can analyse the data of different experimental campaigns with different conditions. In fact, with a versatile model for the

supercritical fluid extraction using CO₂, it is possible to interpret different operating conditions, different organic matrices, or different experimental equipment. Thus, the online tool remains the same but the solution can be applied to a plethora of different problems of the supercritical fluid extraction of lipids with carbon dioxide.

2. Cloud Process Engineering Approach

The approach adopted with this implementation makes cloud computing the core that sustains the online interface which, in turn, enables the user to upload the experimental data and perform the desired calculations in order to gain insights and relevant information on the process. The same cloud computing solution could have been deployed with dedicated software that the user should install on their machine. In this scenario, the cloud computing approach would still exist since only the user interface would not be present. However, a website interface bears several advantages with regards to a specifically installed software (Boillat and Legner, 2013; Chen et al., 2017). Some of these benefits are the ease of upgrading the core routines of calculation, the non-necessity of installing any new software other than a web browser, and the cross-compatibility between devices and operating systems.

The first step of the interaction between the user and the online platform is regulated by the upload of the experimental data through preformatted JSON files containing the experimental data of the extraction process and several other data about the system geometry or physical properties of the organic matrix. Now, the online system is able to read the uploaded files and pass all the relevant information to the binary executables which are launched on the calculating server. The executables return the results of the calculations, e.g. non-linear regression of the experimental data, in various additional JSON files structured in such a way that the system will then be able to process them and make them readable by the final user. Additionally, several graphics accompany numerical results. Finally, the resulting materials are shown on the interfacing platform where the user can download them. A schematization of the whole process is shown in Figure 1.

Every single link in this chain may be replaced and relocated elsewhere on the net since a full cloud implementation grants it (Lehrig et al., 2015). If needed, it is possible to directly act on the critical aspects of the infrastructure and fix them. Additionally, if the workload is too heavy it is still possible to serve multiple users at once by instancing several additional workers on the network. This creates an elastic system able to self adapt to computation demand (Al-Dhuraibi et al., 2018). A relevant additional advantage of a cloud implementation, which is useful especially for industrial applications, is the possibility to scale the whole system on-demand thus reducing the overall cost of unused computing power since it will be redirected in real-time to the user calculations who need it the most.

3. Materials and Methods

3.1. Materials

Micronized silverskin has been provided by an Italian roasting company. All HPLC-grade and analytical grade organic solvents were purchased. HPLC-grade water was obtained via a purification system. Carbon dioxide was purchased with a purity of 99.999 %.

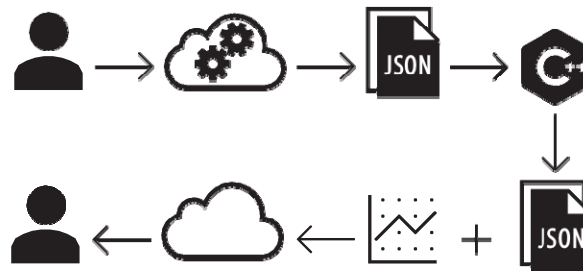


Figure 1: Simplified schematization of the user-cloud interconnection.

3.2. Supercritical CO₂ extraction

The extraction equipment used for the supercritical carbon dioxide extraction consists of a pump connected to the CO₂ cylinder and the extraction vessel mounted vertically inside a thermoregulatory oven. The CO₂ is compressed up to the desired pressure and flows at a constant rate inside the column packed with approximately 40 g of micronized silverskin. Once the CO₂ reaches the extraction vessel it is heated by the oven that contains the column itself. The system is operated continuously for a certain amount of time of maximum 2 h. The coffee silverskin lipid thus extracted has been collected and weighted at several time intervals in order to assess the kinetics of the extraction. As shown in Figure 2, five different extractions have been conducted.

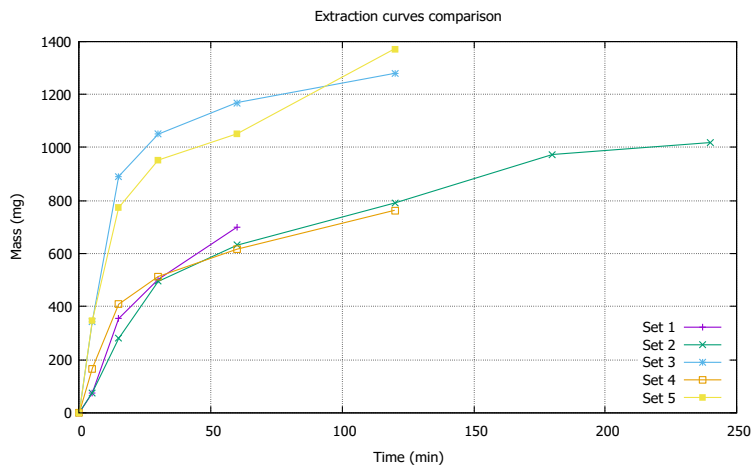


Figure 2: Extraction data points at different operating conditions.

4. Model

The mass transfer model applied to characterize the two-phase system was originally developed by Sovová (1994) with a kinetic simplification introduced by Patel et al. (2011). The choice of such a model is due to the existence of an analytical solution and its versatility in interpreting experimental data (Andrade et al., 2012; Martínez et al., 2003; Mezzomo et al., 2009). These features are needed in order to construct a cloud computing solution that will be able to process different experimental data fast and in a repetitive way without the need for any additional intervention to the modeling aspects. With this model, two different extraction phases are assumed which differ from each other for the mass transfer mechanism. In the first phase, the external mass transfer controls the rate of extraction while in the final phase it is controlled by the internal mass transfer. For this reason, the extraction is faster at the beginning and declines rapidly after the first stage. The two phases' existence is due to the presence of lipids on the surface of the organic matrix that are depleted during the rapid extraction of the first stage (Brunner, 1984; del Valle et al., 2000).

4.1. Simplified dual phase model

The model adopted introduces several assumptions to describe the extraction process, such as: no radial dispersion, pure solvent at extractor entrance, etc. (Patel et al., 2011). The analytical solution to the differential system comprised of the mass transfer equations for the fluid and solid phase results as:

$$e = qy_r [1 - \exp(-A)] \quad \text{for } q < q_m \quad (1)$$

$$e = x_0 - \exp[-BA(q - q_m)][x_0 - q_m y_r (1 - \exp(-A))] \quad \text{for } q \geq q_m \quad (2)$$

where several dimensionless variables are introduced, most importantly the variables e and q which are, respectively, the dimensionless extracted mass of lipid and the dimensionless time related to the amount of solvent flowed in the system. Moreover, the parameter A depends mainly on the fluid dynamics of the system since it is a function of k_f , the fluid mass transfer coefficient, a_0 which is the superficial area coefficient, the superficial velocity U and H , the column height. B is the ratio between the solid mass transfer coefficient (k_s), with the solid density (ρ_s) and the fluid mass transfer (k_f) with the fluid density (ρ_s). x_0 and y_r represent, respectively, the initial solid mass fraction of solute in the solid matrix and the solubility of lipids in the supercritical fluid. Both terms are constituent elements of the variable q_m which represents the dimensionless time of the phase inversion. For a more in-depth analysis of the variables presented, please refer to Patel et al. (2011).

4.2. Fundamental Parameters

In order to allow the model to predict the extraction curve at different operating conditions, it is needed to develop accurate correlations for all the fundamentals parameters that have a significant dependence on the operating conditions. A fine-tuning of parameter A has been performed by using a corrective correlation on the flow rate in a power-law-like form. In fact, a single additional regressive parameter has been added in order to better predict different volumetric flow rates with an empirical approach. For all the parameters that are either constant or with a negligible variation at different operating conditions and whose value is not known experimentally the value is finally evaluated with a non-linear

regression. The fundamental parameters that need an accurate correlation are the solvent density and viscosity and the fluid phase mass transfer coefficient. The density of the supercritical carbon dioxide has been described with the well-known Peng-Robinson's equation of state (Peng and Robinson, 1976) while the viscosity has been calculated with a Jossi-Stiel-Thodos correlation (Jossi et al., 1962). The fluid mass transfer coefficient has been evaluated with the Kunii and Suzuki (1967) mass transfer theory which is valid for finely packed beds characterized by very low Re and Pe numbers. This theory describes the flow of the fluid phase with a channeling effect where the solid phase tends to aggregate to form clumps that are reducing the mass transfer.

5. Results

The cloud computing approach has been successfully applied to solve the non-linear regression problem of the experimental data with the simplified dual-phase extraction model for the supercritical extraction of lipids from micronized coffee grounds silverskin. The non-linear regression has been performed by using a custom implementation of the numerical library BzzMath (Buzzi-Ferraris and Manenti, 2010, 2012) using Visual C++. The global sum of squared errors of the whole experimental set has been minimized with a modified Simplex search algorithm. The modification of the original algorithm implemented in the BzzMath library is aimed at incrementing the robustness of the minima search. The global minima found for the five experimental sets resulted in an average absolute relative deviation of 7.11 % and a standard error of the estimate of 0.00120 (both calculated on the dimensionless extracted mass parameter). These results are reasonably low values as it can be seen from Figure 3. Moreover, the model has a good predictive capability since one of the model parameters is the total solute concentration in the solid phase. This parameter is evaluated experimentally and it acts as a physical limit on the model itself since no more than the total initial concentration can be extracted. Thus, with few experimental points, it is possible to evaluate the extraction dynamic quite accurately (as shown in Figure 3) and with a physical boundary.

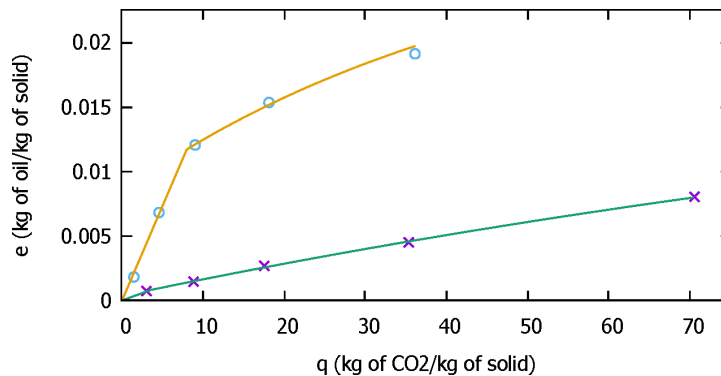


Figure 3: Non-linear regression results of the dual phase simplified model for the first (circles) and fourth (crosses) experimental sets.

6. Conclusions

This case study shows how typical process engineering problems may be solved using a full cloud implementation while retaining the high accuracy results of a local offline solution platform but with all the benefits of a cloud computing approach. Moreover, for the specific case study of the supercritical carbon dioxide extraction process of lipids from micronized silverskin, the implemented model has shown a good versatility in analysing and interpreting experimental data. In addition, the possibility to develop the fundamental parameters transforms the described model from a purely regressive to a predictive one. The ability to predict the behavior of the extraction process at different operating conditions is valuable for the researcher who wants to analyze the system. This numerical tool could be further developed with CAPE-OPEN and OPC protocols to enable it to manage and unify different, already established, process systems engineering tools.

References

- Y. Al-Dhuraihi, F. Paraiso, N. Djarallah, P. Merle, mar 2018. Elasticity in cloud computing: State of the art and research challenges. *IEEE Transactions on Services Computing* 11 (2), 430–447.
- K. S. Andrade, R. T. Gonçalves, M. Maraschin, R. M. R. do Valle, J. Martínez, S. R. Ferreira, jan 2012. Supercritical fluid extraction from spent coffee grounds and coffee husks: Antioxidant activity and effect of operational variables on extract composition. *Talanta* 88, 544–552.
- T. Boillat, C. Legner, 2013. From on-premise software to cloud services: The impact of cloud computing on enterprise software vendors' business models. *Journal of theoretical and applied electronic commerce research* 8 (3), 7–8.
- G. Brunner, sep 1984. Mass transfer from solid material in gas extraction. *Berichte der Bunsengesellschaft für physikalische Chemie* 88 (9), 887–891.
- G. Buzzi-Ferraris, F. Manenti, 2010. *Interpolation and Regression Models for the Chemical Engineer*. Wiley VCH Verlag GmbH.
- G. Buzzi-Ferraris, F. Manenti, 2012. BzzMath. In: *Computer Aided Chemical Engineering*. Elsevier, pp. 1312–1316.
- J. Chen, C. Wang, F. Liu, Y. Wang, aug 2017. Research and implementation of a software online testing platform model based on cloud computing. In: *2017 Fifth International Conference on Advanced Cloud and Big Data (CBD)*. IEEE.
- J. M. del Valle, P. Napolitano, N. Fuentes, dec 2000. Estimation of relevant mass transfer parameters for the extraction of packed substrate beds using supercritical fluids. *Industrial & Engineering Chemistry Research* 39 (12), 4720–4728.
- J. A. Jossi, L. I. Stiel, G. Thodos, mar 1962. The viscosity of pure substances in the dense gaseous and liquid phases. *AIChE Journal* 8 (1), 59–63.
- D. Kunii, M. Suzuki, jul 1967. Particle-to-fluid heat and mass transfer in packed beds of fine particles. *International Journal of Heat and Mass Transfer* 10 (7), 845–852.
- S. Lehrig, H. Eikerling, S. Becker, 2015. Scalability, elasticity, and efficiency in cloud computing. In: *Proceedings of the 11th International ACM SIGSOFT Conference on Quality of Software Architectures - QoSA '15*. ACM Press.
- J. Martínez, A. R. Monteiro, P. T. V. Rosa, M. O. M. Marques, M. A. A. Meireles, mar 2003. Multicomponent model to describe extraction of ginger oleoresin with supercritical carbon dioxide. *Industrial & Engineering Chemistry Research* 42 (5), 1057–1063.
- N. Mezzomo, J. Martínez, S. R. Ferreira, nov 2009. Supercritical fluid extraction of peach (*prunus persica*) almond oil: Kinetics, mathematical modeling and scale-up. *The Journal of Supercritical Fluids* 51 (1), 10–16.
- R. N. Patel, S. Bandyopadhyay, A. Ganesh, jan 2011. A simple model for super critical fluid extraction of bio oils from biomass. *Energy Conversion and Management* 52 (1), 652–657.
- D.-Y. Peng, D. B. Robinson, feb 1976. A new two-constant equation of state. *Industrial & Engineering Chemistry Fundamentals* 15 (1), 59–64.
- H. Sovová, 1994. Rate of the vegetable oil extraction with supercritical CO₂. modelling of extraction curves. *Chemical Engineering Science* 49 (3), 409–414.

Multi-objective optimization of the shape of a fixed-bed reactor

Alexis Courtais^a, François Lesage^a, Yannick Privat^b, Cyril Pelaingre^c and Abderrazak M. Latifi^{a,*}

^a *Laboratoire Réactions et Génie des Procédés, CNRS-ENSIC, Université de Lorraine, Nancy, France*

^b *Institut de Recherche Mathématique Avancée, CNRS, Université de Strasbourg, Strasbourg, France*

^c *Centre Européen de Prototypage et Outillage Rapide, Saint-Dié-des-Vosges, France*
abderrazak.latifi@univ-lorraine.fr

Abstract

In the present paper, a multi-objective shape optimization approach based on the adjoint system method has been developed and tested on a two-dimensional fixed-bed reactor. The optimization problem considered involves two performance indices (i.e. conversion rate and dissipated energy in the reactor) and three constraints consisting of an iso-volume constraint and two manufacturing constraints. The process model, solved using OpenFOAM software, is described by the mass balance and Navier-Stokes equations in laminar flow regime. The multi-objective optimization problem is converted into a single-objective problem using weighting factors and then solved several times to determine the set of optimal solutions, i.e. Pareto front. The best optimal solution is then determined by means of the multi-attribute utility theory (MAUT) method. Finally, the resulting optimal shape which reduces the energy dissipation by 46.7% and improves the conversion rate by 2.7% is manufactured using a 3D printing technique.

Keywords: Multi-objective shape optimization, Multi-attribute utility theory (MAUT), Adjoint system method, Computational fluid dynamics (CFD), Fixed-bed reactor.

1. Introduction

In process optimization, the problems that are very often studied generally aim to determine a solution of a constrained minimization/maximization of a single objective. However, in many cases, multiple conflicting objectives are involved and need to be optimized simultaneously. In this case, the solution is no longer a single solution but a set of optimal solutions, called Pareto front. These solutions are then ranked by means of a decision-aid making method to choose the best one.

In one of our recent works (Courtais et al., 2020), a geometry optimization approach was developed to determine the optimal shape of a two-dimensional fixed-bed reactor. The objective was to maximize the conversion rate while satisfying four constraints: (i) an iso-volume constraint ensuring that the reactor operates at the same residence time, (ii) an inequality constraint which defines a maximum value for the energy dissipated by the fluid, (iii) and two manufacturing inequality constraints that impose minimum values on the pores width and the packing thickness. However, the performance index, i.e. the conversion rate, and the constraint on the dissipated energy are conflicting which means that the solution of the optimization problem is a Pareto set.

In the present paper, a constrained multi-objective optimization of the shape of a fixed-bed reactor where a single-phase liquid flows in laminar regime, is developed. The multi-objective problem is transformed into a single-objective problem using weighting factors, and solved many times to determine the Pareto front of the optimization problem. The multi-attribute utility theory (MAUT) method (Keeney et al., 1979) is then used as a decision-aid making method in order to select the “best” optimal shape of the reactor.

2. Description of the fixed-bed reactor and its modeling

The developed optimization method is used to determine the optimal packing shape of a fixed-bed reactor with a single-phase liquid flow. It is a structured packing made up of obstacles initially elliptical in shape (with half axes of 5mm and 2.5mm) uniformly distributed in the reactor.

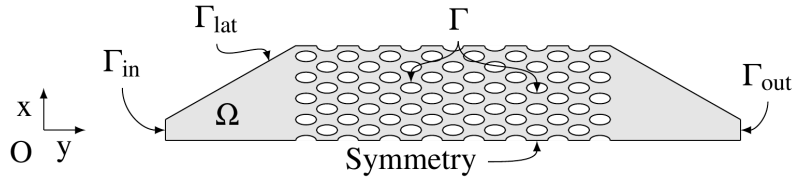


Figure 1 : Initial configuration of the fixed-bed reactor.

The studied domain, whose initial shape is shown in Fig. Figure 1 is denoted by Ω and is delimited by its boundaries $\partial\Omega = \Gamma_{in} \cup \Gamma_{out} \cup \Gamma_{lat} \cup \Gamma$. Γ_{in} and Γ_{out} represent respectively the fluid inlet, the fluid outlet and the side wall of the reactor. Γ is the packing whose shape will be modified during the optimization process and constitutes the decision variable of the optimization problem. The model of the fluid flow in the reactor is described by means of Navier-Stokes and continuity equations (1). It is based on the following main assumptions: laminar flow regime, incompressible fluid, and steady-state conditions. The model equations are expressed as:

$$\begin{cases} -\nu \Delta \mathbf{U} + \mathbf{U} \cdot \nabla \mathbf{U} + \nabla p = 0 & \text{in } \Omega \\ \nabla \cdot \mathbf{U} = 0 & \text{in } \Omega \\ \mathbf{U} = \mathbf{U}_{in} & \text{on } \Gamma_{in} \\ \mathbf{U} = 0 & \text{on } \Gamma_{lat} \cup \Gamma \\ \sigma(\mathbf{U}, p) \mathbf{n} = 0 & \text{on } \Gamma_{out} \end{cases} \quad (1)$$

where $\sigma(\mathbf{U}, p) = 2\nu \varepsilon(\mathbf{U}) - p\mathbf{I}$ with $\varepsilon(\mathbf{U}) = \frac{1}{2}(\nabla \mathbf{U} + (\nabla \mathbf{U})^T)$. ν is the fluid kinematic viscosity, \mathbf{I} the identity matrix, p the kinematic pressure (i.e. the absolute pressure divided by the density), and $\varepsilon(\mathbf{U})$ the strain tensor.

Reagent transport is modeled by the system of PDEs (2). It is assumed that a first order reaction takes place in the bulk of the reactor and no reaction occurs on the walls.

$$\begin{cases} -D\Delta C + \mathbf{U} \cdot \nabla C + kC = 0 & \text{in } \Omega \\ C = C_{in} & \text{on } \Gamma_{in} \\ \frac{\partial C}{\partial n} = 0 & \text{on } \Gamma_{out} \cup \Gamma_{lat} \cup \Gamma \end{cases} \quad (2)$$

D denotes the diffusion coefficient of the reagent in the solvent (i.e. water), C the reagent concentration and k the kinetic constant.

3. Multi-objective optimization problem statement

The multi-objective optimization problem is defined as follows:

- Two **performance indices** are defined by the energy dissipated by the fluid in the reactor and the average reagent concentration at the outlet of the reactor. They are respectively expressed as:

$$J_1(\Omega) = 2\nu \int_{\Omega} |\varepsilon(\mathbf{U})|^2 dx \quad \text{and} \quad J_2(\Omega) = \int_{\Gamma_{\text{out}}} C d\sigma \quad (3)$$

- The **decision variable** is the shape and the position of the packing Γ .
- The **constraints** are defined by the following equality and inequality constraints:
 - The process model described in Section 2 and solved using OpenFOAM software.
 - An iso-volume constraint is considered in order to operate at the same residence time in the initial and optimal shapes of the reactor. It is expressed as follows:

$$C_v(\Omega) = v(\Omega) - v(\Omega_0) = 0 \quad (4)$$

where $v(\Omega)$ denotes the volume of Ω .

- Two inequality constraints related to the reactor manufacturing step. Indeed, the 3D printing technique used to manufacture the reactor imposes minimal values on the local width of the pores and on the local thickness of the packing (Courtais et al., 2021)

To solve the resulting multi-objective optimization problem, it is converted into a single-objective problem by aggregating the two objectives using the linear scalarization method as:

$$J(\Omega) = \tau J_1(\Omega) + (1 - \tau) K_{\text{crit}} J_2(\Omega) \quad (5)$$

K_{crit} is a constant used to ensure the same unit of measure and the same order of magnitude of both objectives J_1 and J_2 .

4. Adjoint system method for the shape gradient computation

The approach used is an iterative method which builds from the initial form a series of shapes that improve the performance of the reactor at each iteration by adapting the position of its boundaries. This approach relies on the concept of shape derivative also called derivative in the sense of Hadamard (Henrot and Pierre, 2005). The method consists in determining at each iteration the sensitivity of the Lagrangian functional (6) with respect to a small perturbation of the domain boundaries.

$$L(\Omega, \lambda_v) = \tau J_1(\Omega) + (1 - \tau) K_{\text{crit}} J_2(\Omega) + \lambda_v C_v(\Omega) \quad (6)$$

This perturbation is given by the following equation:

$$\Omega_{t+1} = (\text{Id} + tV)(\Omega_t) \quad (7)$$

where Id is the identity operator, t is the step of the iterative method, and \mathbf{V} is the vector field representing the displacement of the mesh. Thus, the product $t\mathbf{V}$ stands for the small perturbation applied to the mesh.

Hadamard's derivative is a gradient method which involves adjoint states to compute the shape gradient $G(\Omega)$. This new functional is defined on the free boundary Γ and depends on the solutions $(\mathbf{U}, p$ and $C)$ and $(\mathbf{U}_a, p_a$ and $C_a)$ of the process model equations and their associated adjoint state equations introduced by the method. Finally, the mesh displacement \mathbf{V} leading to a decrease of the Lagrangian (6) depends on the shape gradient and is computed by solving the following system of equations:

$$\begin{cases} -\Delta \mathbf{V} + \mathbf{V} = 0 & \text{in } \Omega \\ \mathbf{V} = 0 & \text{on } \Gamma_{\text{in}} \cup \Gamma_{\text{out}} \cup \Gamma_{\text{lat}} \\ \nabla \mathbf{V} \mathbf{n} = -G(\Omega) \mathbf{n} & \text{on } \Gamma \end{cases} \quad (8)$$

5. Implementation of the optimization algorithm

The algorithm used to solve the multi-objective optimization problem consists of several resolutions of the single-objective problem by modifying the parameter τ . It is implemented within OpenFOAM CFD software which solves the PDEs using the finite volume method. The solution of each single-objective problem is achieved in 6 main steps as:

1. The initial shape of the reactor is meshed using SNAPPYHEXMESH and CFMESH, two mesh generation utilities supplied by OpenFOAM.
2. The systems of PDEs allowing to determine the state variables (\mathbf{U}, p, C) and their associated adjoint states (\mathbf{U}_a, p_a, C_a) are solved. The pressure-velocity couplings in the Navier-Stokes equations and in the equations of its adjoint system are treated using the SIMPLE algorithm.
3. The shape gradient $G(\Omega)$ and the mesh displacement \mathbf{V} are then computed. Since the manufacturing constraints are dealt with during this step, the thickness of the obstacles is defined as the double of the distance between the skeleton and the obstacle. In 2D, the skeleton is defined as the set of points equidistant from the obstacle on each side, i.e. a thin version of the obstacle centered inside it. The local pore distance is computed using an OpenFOAM function named *wallDist* (Courtais et al., 2021).
4. The Lagrange multiplier λ_v is updated as:

$$\lambda_v^{i+1} = \lambda_v^i + \beta C_v(\Omega_i) \quad (9)$$

where i denotes the iteration number and $\beta > 0$ is a small parameter.

5. A test on the mesh quality is carried out by means of 3 criteria very often used in CFD simulation: the mesh aspect ratio, the mesh non-orthogonality and the face skewness (Holzinger, 2015). If the values of these criteria are respectively higher than 10, 65° and 3.8, a remeshing of the shape is performed.
6. At the end of each iteration, a test on the convergence is carried out. If the ratio of the standard deviation to the average of the last 100 values of the Lagrangian is lower than 10^{-4} , the algorithm stops.

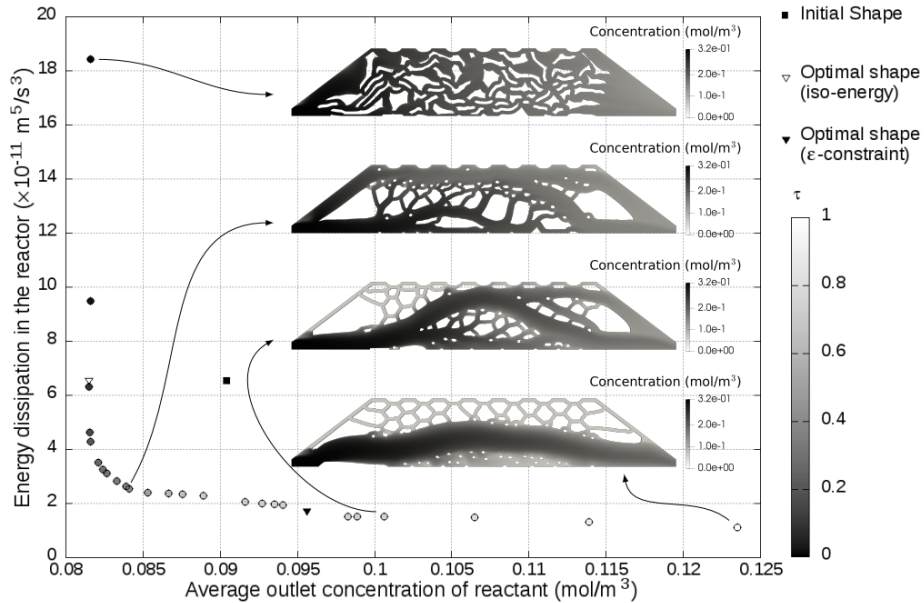


Figure 2: Pareto front of the multi-objective shape optimization problem.

6. Numerical results and discussion

Figure 2 presents the Pareto front where each circle corresponds to the solution of a single-objective problem. As emphasized by Courtais et al. (2020), each single-objective optimization takes between 2 and 3 days to converge on a 3.7GHz Xeon Dell Computer 5810. Consequently, the determination of this estimation of the Pareto front (Fig. 2) made up of 26 single-objective solution is time consuming and has required between 1.5 and 2 months of computations.

Figure 2 shows that the smaller the parameter τ , the greater the conversion rate and the better the homogeneity of the fluid flow in the reactor. Conversely, the closer τ is to one, the more important the energy criterion. In this case, channelings and dead zones appear thus causing the reduction of the energy dissipation in the fluid and the degradation of the conversion rate in the reactor. On Fig. 2, the black triangle of coordinates $(9.56 \times 10^{-2}; 1.7 \times 10^{-11})$ has been determined by the ϵ -constraint method since the Pareto front is slightly concave near this point and consequently the linear scalarization method is unable to determine the concave parts of the front. The performances of the reactor configuration determined by Courtais et al. (2020) are also shown on the Pareto front (white triangle). They highlight the importance of carrying out a multi-objective optimization since it was possible to improve the energy dissipation criterion by 33% with an insignificant degradation of the conversion rate.

Finally, all these optimal shapes of the reactor are ranked using the MAUT method (Fonseca et al., 2020) according to the importance given to the criteria by the decision-maker. The solution at the top of the ranking is then chosen as the best optimal solution; it reduces the energy dissipation by 46.7% and improves the conversion rate by 2.7%. This configuration of the reactor is presented in Fig. 3 and has been manufactured using a 3D printing technique for experimental tests.

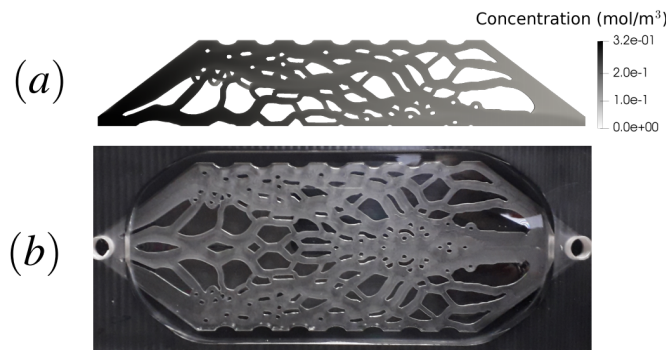


Figure 3: Comparison between the best optimal reactor chosen using MAUT method (a) and the corresponding reactor manufactured by a 3D printing technique (b).

7. Conclusions

A geometry optimization approach based on the adjoint system method has been developed, implemented within OpenFOAM and tested on a 2-dimensional fixed bed reactor where a homogeneous first order reactor occurs. The optimization problem has been formulated with two performance indices (energy dissipation in the fluid and conversion rate in the reactor) and the Pareto front has been determined. Finally, the configuration which offers the best compromise between the criteria has been determined using the MAUT method and manufactured by means of a 3D printing technique.

This paper emphasizes the importance of performing multi-objective optimization of the reactor since the optimal shape determined by Courtais et al. (2020) by a single-objective optimization was not the best optimal configuration of the reactor for manufacturing issues. However, the developed method has shown several limitations. The first one is the method of shape optimization used to solve the optimization problem. Indeed, some dead zones that appear when the energy dissipation criterion is high (i.e. for $\tau > 0.7$) would disappear if topology changes were allowed. Another limitation concerns the width constraints which do not allow their violation during the optimization iterations. These issues are being addressed and significant improvements are expected.

References

- A. Courtais, A. M. Latifi, F. Lesage, Y. Privat, 2021. Shape optimization of fixed-bed reactors in process engineering. Submitted to SIAP Journal on Applied Mathematics.
- A. Courtais, F. Lesage, Y. Privat, C. Pelaingre, A. M. Latifi, 2020. Shape Optimization of a Fixed-bed Reactor using Additive Manufacturing. In: Computer Aided Chemical Engineering. Vol. 48. Elsevier, pp. 1939–1944. <http://dx.doi.org/10.1016/B978-0-12-823377-1.50324-4>
- J. D. Fonseca, A. M. Latifi, A. Orjuela, G. Rodríguez, I. D. Gil, 2020. Modeling, analysis and multi-objective optimization of an industrial batch process for the production of tributyl citrate. Computers and Chemical Engineering 132, 106603. <http://dx.doi.org/10.1016/j.compchemeng.2019.106603>
- A. Henrot, M. Pierre, 2005. Variation et optimisation de forme. Vol. 48 of Mathématiques et Applications. Springer-Verlag, Berlin.
- G. Holzinger, 2015. OpenFOAM: A little User-Manual. CD-Laboratory-Particulate Flow Modelling, Johannes Kepler University: Linz, Austria.
- R. L. Keeney, H. Raiffa, D. W. Rajala, 1979. Decisions with multiple objectives: Preferences and value tradeoffs. IEEE Transactions on Systems, Man, and Cybernetics 9 (7), 403–403. <http://dx.doi.org/10.1109/TSMC.1979.4310245>

Artificial Intelligence Based Prediction of Exergetic Efficiency of a Blast Furnace

Muhammad Salman Arif ^a, Iftikhar Ahmad ^{*a}

^a *Department of Chemical and Materials Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan; msalman.pse1@scme.nust.edu.pk*

^{*} *Correspondence: iftikhar.salarzai@scme.nust.edu.pk*

Abstract

The iron melting furnaces are the most energy-consuming equipment of the iron and steel industry. The energy efficiency of the furnace is affected by process conditions such as the inlet temperature, velocity of the charge, and its composition. Hence, optimum values of these process conditions are vital in the efficient operation of the furnace. Computational methods have been very helpful in the optimum design and operation of process equipment. In this study, a first principle (FP) model was developed for an iron-making furnace to visualize its internal dynamics. To minimize the large computational time required for the FP-based analysis, a data-based model, i.e., Artificial Neural Networks (ANN), is developed using data extracted from the FP model. The ANN model was developed using data sets comprised of the values of temperature of the charge and gasses, velocity, concentration of the oxygen, pressure, airflow directions, energy and exergy profiles, and overall exergy efficiency of the furnace along with its height. The ANN model was highly accurate in prediction and is suitable for real-time implementation in a steel manufacturing plant.

Keywords: exergy analysis, second law of thermodynamics, energy analysis, artificial intelligence, iron, and steel industry.

1. Introduction

Iron ore melting furnaces are a core part of the iron and steel manufacturing sector and are consuming a significant share of the total energy used in a plant. To realize energy-efficient design and operation of the furnaces, exergy-based studies have been getting the attention of the researcher. Compared to the conventional energy analysis methods, exergy-based analysis is more effective in the identification of the location, causes as well as the magnitude of the energy losses in a system. Application of the computational methods in the exergy analysis has been helpful. Mathematical models as well as process simulators, i.e., Aspen, ANSYS, etc., are used for performing exergy analysis of various systems. Several studies based on computer aided exergy analysis of furnaces have also been reported.

For instance, Zhang et al., (2017) performed exergy analysis of recycling oxygen in a blast furnace to determine the optimum composition of charge i.e. feed. In another study, an exergy-based analysis of a blast furnace, using natural gas as a fuel, was performed (Guo et al., 2013). They inferred that operating a blast furnace on natural gas increases productivity, reduces silicon

content in molten metal, and carbon emission but reduces its exergetic efficiency. In another work, high exergy efficiency of the blast furnace was reported through the use of pulverized coal and preheating of the charge material which also helped in saving coke (Ziębik and Stanek, 2006). Similarly, exergy analysis of a silicon furnace has revealed that volatiles and electrodes account for about 8-10 percent of exergy destruction (Børset et al., 2015). Exergy analysis of an annealing furnace helped in identifying the fact that the combustion chamber of the annealing furnace is about 12.9 percent exergy efficient while the overall exergy efficiency of the furnace was found to be 7.3 percent (Hasanuzzaman et al., 2011).

The use of artificial intelligence in exergy analysis is a new trend that is paving way for the digitalization of energy analysis systems of process industries (Akram et al., 2018). The present study is based on the use of Artificial Neural Networks (ANN) in exergy analysis of a Blast Furnace (BF) of a steelmaking plant. Initially, a first principle model of the BF is developed that calculates temperature, velocities, the concentration of oxygen, exergy, and energy efficiencies at 8712 points across the width and height of the furnace. The dataset (8712) is used then used to develop ANN models for prediction of charge temperature, gas temperature, the concentration of oxygen, and total exergy.

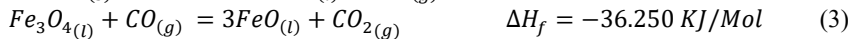
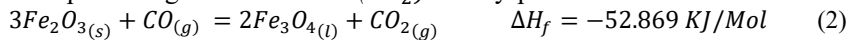
Section 2 describes the process followed by the proposed methodology in section 3. Section 4 elaborate the results while section 5 concludes the present work.

2. Process Description

The process of iron production includes three continuous reduction reactions of iron oxides within the BF. As preheated air enters the furnace from the bottom, it reacts with the coal introduced from the top of the furnace, producing carbon monoxide (CO). These sets of reactions have a specific heat of formations represented below along with the reactions (Sun, 1997).



This carbon mono-oxide reacts with the feed oxide and reduces it into ferrous Oxide (FeO). In the final reduction of iron, this ferrous oxide is further reduced into iron-producing carbon dioxide (CO_2) as a by-product.



3. Methodology

The modelling framework is comprised of three steps. Initially, an FP model of the blast furnace was developed using data from the literature (Dartt, 2011). The FP model was updated to calculate physical, kinetic, and potential exergies

across the furnace. In the third step, data was generated from the FP model to develop the four ANN models. The three steps are briefly discussed below:

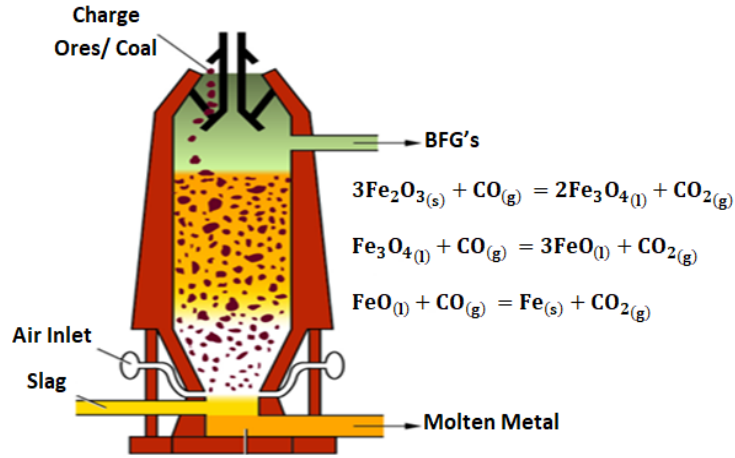


Figure 1: Reaction zones of the furnace (Petrucci et al., 2017)

- FP model development:** For the FP model, the height and diameter of the furnace were kept at 0.605 and 0.36 meters respectively. The inlet temperature of oxides, fluxes, and the air was 298 K, while tuyere and lid diameter were kept 0.05 and 0.11meters respectively. The blast rate of oxygen was 0.224 kg/sec. The mathematical calculations involved Ergun’s pressure drop equation, steady-state equation, reaction rate constant equation, and effective heat transfer coefficient equation.
- Addition of exergy equations in the FP model:** The standard equations used in (Szargut, 2005) were added to the FP model to calculate physical, kinetic, and potential energies. The equations are summarized below:

$$E^{Ph} = RT_o \sum_{i=1}^n Cp_i^{mean} (T_i - T_o - T_o \ln \left(\frac{r_i}{r_o} \right)) \quad (5)$$

$$Cp_i^{mean} = \int_{T_1}^{T_2} Cp_i dT \quad (6)$$

$$Cp_i \left(\frac{j}{mol.K} \right) = a_i + b_i T + c_i T^2 + d_i T^3 \quad (7)$$

where, $a_i, b_i, c_i,$ and d_i are heat capacity coefficients, R is the ideal gas constant at each point of the furnace, P_i is partial pressure, and T_i is temperature.

$$E^K = m_o \frac{c_o^2}{2} \quad (8)$$

$$E^P = m_o g_E Z_o \quad (9)$$

where m_o is the mass flow rate of the bulk stream, C_o is bulk velocity, and Z_o is the altitude above the sea level.

- **ANN modeling:** Data comprised of 8906 samples of 11 process variables/parameters from across the width and height of the furnace was collected for the development of the ANN models. Four ANN models were developed to predict charge temperature, gas temperature, the concentration of oxygen, and total exergy.

4. Results and Discussion

Contours of physical and total exergies of the furnace calculated through the FP model are shown in Fig. 2 while the regression performance of the ANN models is shown in Fig. 3.

Fig. 2 (a) represents the physical exergy of the furnace calculated through the FP model. The physical exergy depends upon the temperature of the furnace. It can be observed that the furnace exergy is high at the points where the temperature profile of the furnace remains higher. In addition to the physical exergy, kinetic and potential exergies were also calculated. The kinetic exergy depends upon the product of the bulk velocity and the bulk flow rate and it was noted that kinetic exergy is much higher at the points where velocity is higher. Potential exergy is the product of height altitude, mass flow rate, and velocity. Hence, velocity is the only factor which is having a higher effect on potential exergy and it is much higher in the high-velocity zones. While at other levels, it has only the effect of height altitude and mass flow rate which is minimized by the lower velocity levels at that position. The sum of all types of exergy termed as total exergy is shown in Fig. 2 (b).

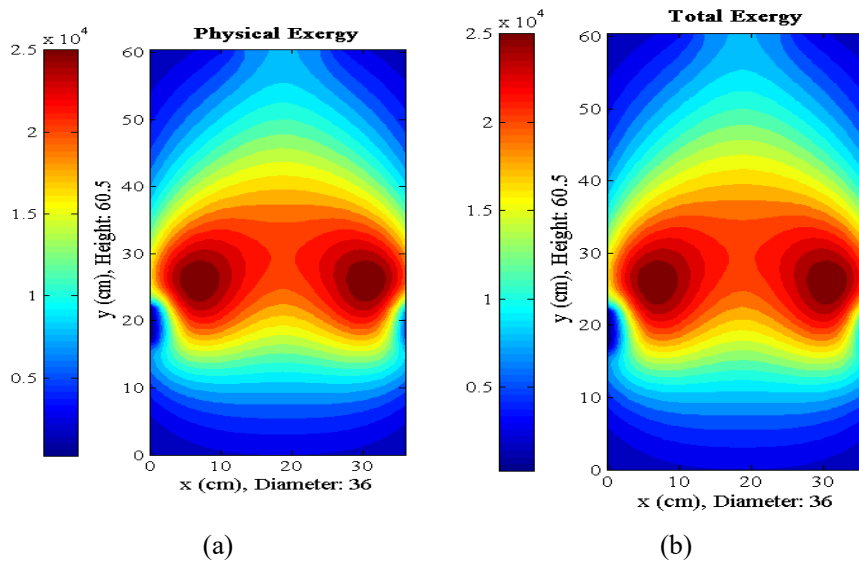


Figure 2: Contour plot (a) physical exergy (b) total exergy along the height of furnace

Fig. 3 (a) represents the prediction of the charge temperature within the furnace. The ANN model had a correlation coefficient of 0.99849 with values obtained through the FP model. Similarly, Fig. 3 (b) represents the correlation between the predicted and the FP model value for the gas temperature. In case of concentration of oxygen, the correlation coefficient was 0.99958 as shown in Fig. 3 (c). Fig. 3 (d) represents the correlation coefficient of the total exergy of the furnace calculated through the FP model and predicted through ANN. The correlation coefficient between the FP values and ANN predicted values were 0.99832.

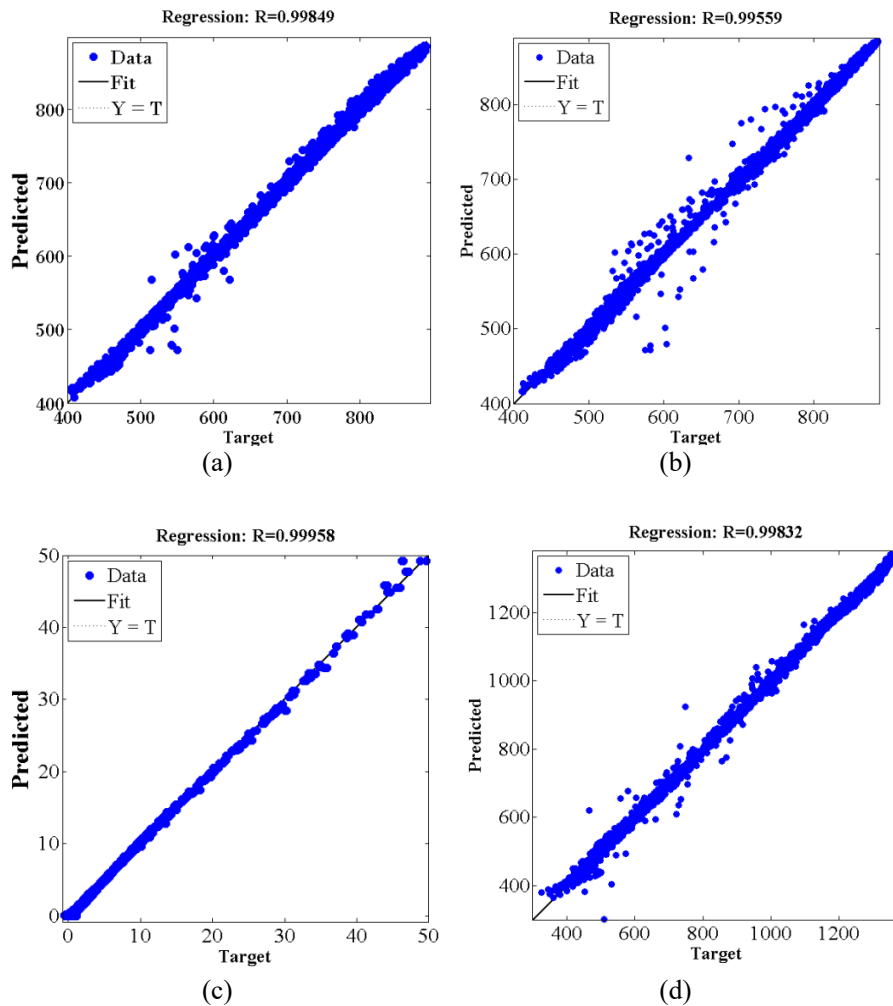


Figure 3: ANN regression plots (a) Charge temperature (b) Gas temperature (c) Concentration of Oxygen (d) Total Exergy

5. Conclusions

A First Principle (FP) was used to calculate kinetic, potential, and physical exergy exergies across the width and height of a blast furnace (BF). The furnace is found to be most exergetic at the points where the temperature of the furnace is higher. To shorten the calculation time for exergy, gas temperature, charge temperature, and concentration of oxygen, ANN models were developed using data extracted from the FP model. ANN models had a correlation coefficient of 0.996, 0.995, 0.995, and 0.993 for prediction of the concentration of oxygen, charge temperature, total exergy, and gas temperature respectively. The ANN-based prediction of process conditions specifically the exergy efficiency provides a platform for realizing a stable and efficient operation of the BF. For a more realistic representation of the internal dynamics of the BF in terms of exergy, the second generation of machine learning methods, i.e., deep learning (DL) can be applied. For training the DL models, further large databases of information may be acquired through infrared-based cameras installed for capturing the internal conditions of the furnaces.

References

- Akram, A. U., Ahmad, I., Chughtai, A. and Kano, M. J. I. J. o. E. (2018) 'Exergy analysis and optimisation of naphtha reforming process with uncertainty', 26(3), pp. 247-262. doi:
- Børset, M., Kolbeinsen, L., Tveit, H. and Kjelstrup, S. J. E. (2015) 'Exergy based efficiency indicators for the silicon furnace', 90, pp. 1916-1921. doi:
- Dartt, K. (2011) *Computational Modeling and Optimization of an Iron Melting Cupollette Furnace*. State University of New York at Binghamton, Thomas J. Watson School of
- Guo, T. L., Chu, M. S., Liu, Z. G., Tang, J. and Yagi, J. I. J. s. r. i. (2013) 'Mathematical modeling and exergy analysis of blast furnace operation with natural gas injection', 84(4), pp. 333-343. doi:
- Hasanuzzaman, M., Saidur, R. and Rahim, N. J. I. J. o. P. S. (2011) 'Energy, exergy and economic analysis of an annealing furnace', 6(6), pp. 1257-1266. doi:
- Petrucci, R. H., Herring, F. G., Bissonnette, C. and Madura, J. D. (2017) *General chemistry: principles and modern applications*. Pearson.
- Sun, S. S. (1997) *A study of kinetics and mechanisms of iron ore reduction in ore/coal composites*.
- Szargut, J. (2005) *Exergy method: technical and ecological applications*. WIT press.
- Zhang, W., Zhang, J. and Xue, Z. J. E. (2017) 'Exergy analyses of the oxygen blast furnace with top gas recycling process', 121, pp. 135-146. doi:
- Ziębik, A. and Stanek, W. J. I. j. o. e. r. (2006) 'Influence of blast-furnace process thermal parameters on energy and exergy characteristics and exergy losses', 30(4), pp. 203-219. doi:

AI System for Supporting Generation of Optimal Synthetic Pathways Based on Chemical Reaction Big Data

Nagyeong Lee, Joonsoo Jeong, Dongil Shin*

Department of Chemical Engineering, Myongji University, Yongin, Gyeonggido 17058, Korea
dongil@mju.ac.kr

Abstract

The selection and design of the appropriate reaction paths has a significant impact on the economics and productivity of the chemical process, enhanced by milder operating conditions, use of cheaper reactants and fewer reaction steps. However, exploration of reaction information is difficult even with reaction databases available, causing path explosion problem due to huge search space. In this study, we propose an AI system (ASICS), which supports synthetic path design at the basic stages of research and process design, based on the hybrid generative exploration and exploitation of reaction knowledge graphs encoding big data of patented reactions and machine learning-based retrosynthetic prediction. ASICS generates an optimal synthetic path that satisfies the given constraints (regulated compounds, etc.), based on A* search using synthetic accessibility and retrosynthetic prediction scores. The preference in searching between confirmed reaction spaces and unexplored reaction spaces through prediction can be selected by the user. The fusion of reaction knowledge base and retrosynthetic prediction model enables to generate optimal synthetic paths beyond the accumulated reaction information.

Keywords: Chemical Reaction Big Data, Synthetic Pathway Design, Knowledge Graph, Inference Reasoning, Decision Support System

1. Introduction

Choosing appropriate reaction pathways is an important part of chemical process development and laboratory research to create high-added value through chemical reactions. Before proceeding with trial-and-error experiments, related compounds and reaction information is usually searched from in-house documents, literature, databases or web. However, it is not easy to narrow down candidate reaction paths and/or to obtain the information required by the target compounds even with commercial databases.

Databases of organic synthesis, e.g., Reaxys and SciFinder, collect and provide information on compounds and reactions from various patents and literature from the past to the present. In addition to that, in order to solve the environmental and energy sustainability and economic issues, it is required to explore a wide range of reaction hyperspaces within a limited time. For example, in the network universe of organic chemistry, 80.2 distinct reactions can be applied to a non-trivial retron on average, translating into 3.5×10^{28} possible 15-step pathways (Szymkuc et al. 2016).

The methods of generating one-step reactions are divided into exploration with database, reaction rule matching and machine learning-based reaction prediction. As retrosynthesis planning methods, there is a heuristic best-first search method that

performs rule matching and exploration based on reaction networks and hand-coded reaction rules (e.g., Synthia, formerly Chematica (Szymkuc et al. 2016)). There is also a template-based, policy-guided tree search method using machine learning-based retrosynthetic prediction model (e.g., AizynthFinder (Genheden et al, 2020)). Depending on the prediction model used, it is divided into template-based or template free.

Exploring and rule matching method based on the reaction network cannot generate result beyond the accumulated information. Retrosynthetic prediction models, however, always has uncertainty and do not supply detailed reaction information such as yield and conditions. Thus, we propose a hybrid system that combines reaction networks search and retrosynthetic prediction to generate synthetic paths by exploring both dimensions. The preference in searching between confirmed reaction spaces and unexplored reaction spaces through prediction can be selected by the user. Reaction information is collected using open data, and the proposed system is intended to be provided as open source.

2. Design of Reaction Knowledge Base: Reaction database sources, essential reaction information, automatic extraction, and representation as knowledge graph

The information required to generate an optimal synthetic pathway includes information on chemical structure, chemical reaction, physical property, and commercially available compounds. We selected PubChem as the main database source for chemical structures, eMolecule for chemical structures for commercially available compounds and the USPTO patents (1976-2016) by Lowe (2014) for chemical reactions. All compounds are stored by expressing them with identifiers, e.g., SMILES (Weininger 1988).

Massive information of chemical structure and physical property was extracted from the source database through automated web crawling. USPTO patents by Lowe contain information such as year, reaction formula, reactant, product, reaction amount, yield, and synthesis procedure. By extracting reaction information and removing duplicates of the same reaction SMLIES, 1,674,945 reaction information out of ~3.3 million reactions was used as the basis for reaction knowledge. Knowledge base was constructed by expressing them as labeled property graphs using Neo4j, an open-source graph database. Knowledge graphs consisting of the knowledge base are shown in Figure 1(a) and (b).

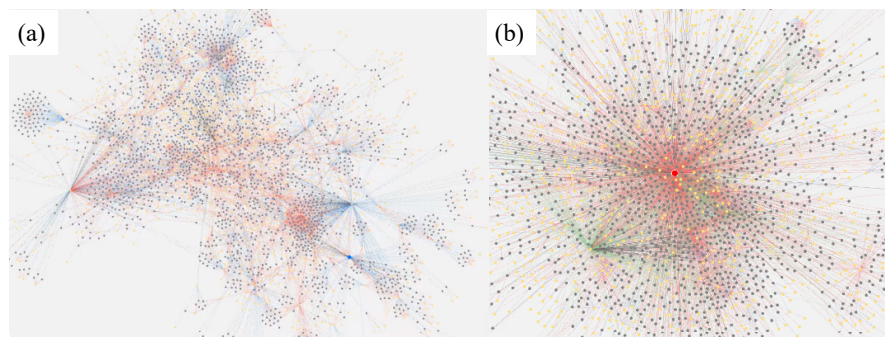


Figure 1. Reaction knowledge graph network: (a) the whole knowledge base, (b) zoomed for dichloromethane. Yellow nodes represent compounds and black nodes reactions. Edges represent relations: blue for products and reactions, red for reactants and reactions, and green catalysts or solvents and reactions.

3. One-step Retrosynthetic Prediction Using an Open-Source Model

With a retrosynthetic prediction model, we have the advantage of considering the reaction of the unreported or unexplored space beyond the definite reactions accumulated in the knowledge base. As the template-based one-step retrosynthetic prediction model, Chen et al. (2020)'s model published at GitHub was imported and used. This model uses a single-layer fully connected neural network of latent dimension 128, with ReLU as activation function. Morgan fingerprint of radius 2 with 2,048 bits is input to this model, and reaction templates are obtained as output. The reaction templates are converted to SMILES using RDKit package to obtain reactants.

4. Generation of Optimal Synthetic Pathway via Simultaneous Exploration of Knowledge Graph and Retrosynthetic Prediction

The most economical synthesis direction among similar reactions is selected based on the Synthetic Accessibility score (SAscore) scaled to be between 1 and 10 (Ertl et al. 2009). A reaction with the lowest SAscore, easier to synthesize, is favored by calculating the average SAscore of the participating reactants. It is assumed that the lower SAscore the lesser cost of reactants.

Table 1. Pseudocode of genRxn() that selects an optimal one-step reaction and returns the results of applying the reaction.

<p>Given target compound t_m and list of commercially or in-house available compounds PCL, generate list of reactants from the selected reaction RL and list containing the synthetic path up to now PL.</p> <p>genRxn(t_m, PL):</p> <p>A. Search one-step reactions in knowledge graph KG for product t_m: Search all candidate reactions and save it as list L_1. # Calculate value function scores. $V_A = [w_1 \times (SAscore(reactants\ set\ of\ x)/10) + w_2 \times (1 - prediction\ score(x)) - availabilityP(reactants\ set\ of\ x)\ for\ x\ in\ L_1]$</p> <p>B. Generate top-50 probable one-step retrosynthetic predictions for product t_m: Run Retro* and select top-50 candidate reactions to save it as list L_2. $V_B = [w_1 \times (SAscore(reactants\ set\ of\ x)/10) + w_2 \times (1 - prediction\ score(x)) - availabilityP(reactants\ set\ of\ x)\ for\ x\ in\ L_2]$</p> <p>Select the optimal reaction as r, out of L_1 and L_2, that has the lowest value of V. $RL = reactants\ set\ of\ r$ $PL.append(r)$ return RL, PL</p> <p>availabilityP(RL): """" Returns 1 only when all the reactants are commercially or in-house available. """" return numpy.prod([int(m in PCL) for m in RL])</p>

A* search proceeds in the same flow as the pseudocodes in Table 1 and 2. It explores optimal synthetic paths having minimum scores of synthetic reaction value function V , which estimates the cost, certainty and readiness of a reaction and the combined value is adjusted by weights put on the exploration of knowledge graph and exploitation by retrosynthetic prediction:

$$V = w_1 \times \frac{SAscore}{10} + w_2 \times (1 - prediction\ score) - availabilityP \quad (1)$$

where weights are decided by user's preference on exploration vs. exploitation. And the *availabilityP* predicate returns 1 only when all reactants are commercially or in-house available. The *prediction score* that represents prediction probability of reaction is 1 when reaction was searched from knowledge graph and likelihood value when predicted. The reaction consisting of commercially available reactants takes precedence in the following

cases: (1) laboratory studies that seeks to complete a target test by rapidly generating the target compound with a small number of reaction steps through active use of commercially available compounds, and (2) process development stage that prioritizes the reaction using by-products by including them in the list of available compounds.

Table 2. Pseudocode of generating the optimal synthetic path.

```

With a target compounds  $t_m$  and list of commercially or in-house available compounds  $PCL$ , generate the optimal synthetic path  $PL$ .

Set weights  $w_1, w_2$  (real numbers between 0 and 1, default = 1).

 $PL = \text{applyGenRxn}([t_m], [])$ 
print( $PL$ )

applyGenRxn( $TL, PL$ ):
    """ Returns the complete synthetic path for the list of target compounds  $TL$ . """
    for  $m$  in  $TL$ :
         $RL, PL = \text{genRxn}(m, PL)$ 
        if not  $RL$ : raise Exception("No possible path exists.")
         $RL = RL - PCL$ 
         $PL = \text{applyGenRxn}(RL, PL)$ 
    return  $PL$ 

```

5. Results and Discussion

5.1. System implementation of ASICS

The proposed system was implemented as open source in Python. Selenium and BeautifulSoup were used for web crawling; Qt Designer and PyQt5 for GUI; Pandas for data pre-processing; Neo4j Python API for knowledge base construction; RDKit for cheminformatics.

5.2. Case study 1: Generation of synthetic path for HENC

The target material 2-hydroxyethyl 5-norbornene-2-carboxylate (HENC) is mainly used as an intermediate in the pharmaceutical and electronic polymer industries. The results of the synthetic paths of SciFinder retrosynthesis, ASICS, Retro* (Chen et al 2020) and AizynthFinder are shown as shown in Figure 2. Comparison of the results of four systems is shown in Table 3.

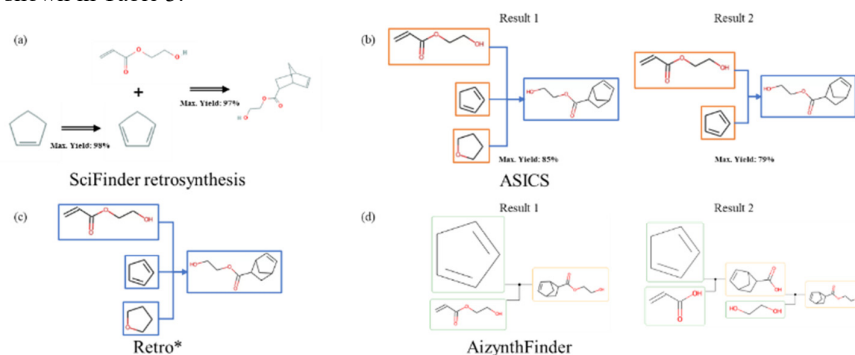


Figure 2. HENC synthetic paths results of (a) SciFinder retrosynthesis, (b) ASICS, (c) Retro* and (d) AizynthFinder.

ASICS intentionally generated two synthetic paths with different value function scores and AizynthFinder generated more number of synthetic paths. SciFinder retrosynthesis generated a two-step path. Retro* and ASICS generated one-step paths. AizynthFinder

generated one-step and two-step paths. The SciFinder retrosynthesis result had the lowest total price of starting materials at \$8.4/g using hydroxyethyl acrylate and cyclopentene, but it is a two-step path. The Retro* result and ASICS's result #1 showed the same one-step path, with hydroxyethyl acrylate, cyclopentadiene and THF being used, and the total prices were same. But the ASICS result contains more detailed additional reaction information and the cost of ASICS's result #2 is low. AizynthFinder's result #1 and ASICS's result #2 are the same, but ASICS with more detailed information. Total price of AizynthFinder's result #2 is slightly lower, but it requires two steps.

Table 3. Comparison of HENC synthetic paths results.

	Number of reaction steps	Number of starting compounds	Total price of starting compounds (\$/g)	Yield (%)	Synthesis procedure	Published year
SciFinder retrosynthesis	2	2	8.4	97	O	-
Retro*	1	3	65.8	X	X	X
AizynthFinder	1, 2	2, 3	46.8, 46	X	X	X
ASICS	1, 1	3, 2	65.8, 46.8	85, 72	O	2002, 2000

5.3. Case study 2: Generation of synthetic path for merimepodib

Results of synthetic path by ASICS and Retro* are the same in that three-step and five starting compounds are used. Both use [(3R)-oxolan-3-yl] carbonochloridate and 3-(aminomethyl)aniline as starting compounds. But in the case of Retro*, it uses a regulated compound (phosgene) and the probability of the first step of retrosynthetic path is low at 0.3336. However, there is no use of regulated compound in ASICS result, and the first step of retrosynthetic path (orange color of Figure 3(b)) is a guaranteed reaction with certainty = 1 since it originated from the knowledge graph. ASICS generated a better synthetic path by using reported reactions instead of an uncertain reaction with low probability.

5.4. Discussion

When using only retrosynthetic prediction results, e.g., Retro* and AizynthFinder, there

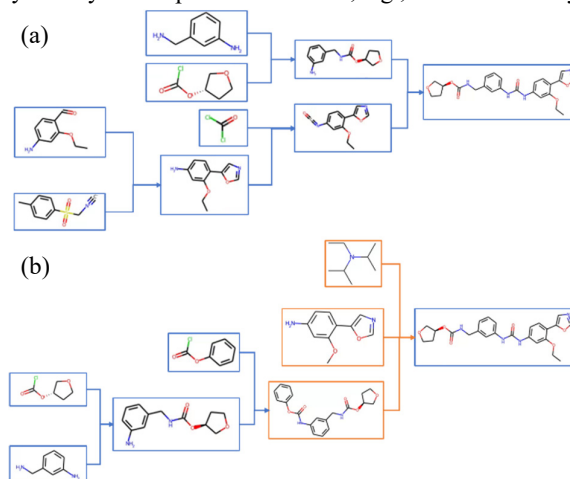


Figure 3. Merimepodib synthetic path results of (a) Retro* and (b) ASICS.

are disadvantages that detailed information (yield, conditions, etc.) for each reaction in synthetic paths cannot be known and the prediction results with low probability may be used. ASICS applied with hybrid generative exploration and exploitation explored

optimal synthetic paths that compensated for these shortcomings. ASICS generated optimal synthetic paths, showing less number of steps or similar to the result of SciFinder retrosynthesis based on ~112 million reaction data, ~100 times of the ASICS data.

6. Conclusion

In this study, we proposed an intelligent support system for reaction path design, generating optimal synthetic paths out of hybrid generative exploration and exploitation of reaction knowledge base and machine learning-based retrosynthetic predictions. The knowledge base, constructed from web crawling of USPTO patents (1976-2016) and related chemical data consists of relationships for 1,674,945 reactions and 923,761 compounds. ASICS generates optimal synthetic paths using lower-cost starting compounds by selecting optimal reactions with the lowest score of synthetic reaction value function. In addition, exploration and exploitation are adjusted according to the user's own preferences to enable optimal search for respective purpose.

Through case studies on HENC and merimepodib, we compared the results against SciFinder retrosynthesis, AizynthFinder and Retro*, the existing computer-aided synthesis planning systems. The ASICS results are similar or use less number of steps to the SciFinder retrosynthesis which works based on the reaction data of bigger size. Compared with the AizynthFinder and Retro* results using only prediction models, ASICS generated optimal synthetic paths consisting of reactions with higher likelihood than prediction-only results with uncertainty in reaction likelihood.

Acknowledgements

This research was partly supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0008475, Development Program for Smart Digital Engineering Specialist) and KIAT grant funded by the Korea Government (MOTIE) (G02N06240000204, Virtual Engineering Platform Program).

References

- Chen, B., Li, C., Dai, H. and Song, L., 2020, Retro*: learning retrosynthetic planning with neural guided A* search, *International Conference on Machine Learning*, PMLR, 1608-1616
- Coley, C. W., Green, W. H. and Jensen, K. F., 2019, RDChiral: An RDKit wrapper for handling stereochemistry in retrosynthetic template extraction and application, *J. of Chemical Information and Modeling*, **59**(6), 2529-2537
- Engel, T. and Gasteiger, J. (Eds.), 2018, *Chemoinformatics: basic concepts and methods*, Wiley
- Ertl, P. and Schuffenhauer, A., 2009, Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions, *J. of Cheminformatics*, **1**(1), 1-11
- Genheden, S., Thakkar, A., Chadimová, V., Reymond, J. L., Engkvist, O. and Bjerrum, E., 2020, AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning, *J. of Cheminformatics*, **12**(1), 1-9
- Lowe, D. M., 2012, *Extraction of chemical structures and reactions from the literature*, Ph.D. Thesis, University of Cambridge
- Szymkuc, S., Gajewska, E. P., Klucznik, T., Molga, K., Dittwald, P., Startek, M., Bajczyk, M. and Grzybowski, B. A., 2016, Computer-Assisted Synthetic Planning: The End of the Beginning, *Angew. Chem. Int. Ed.*, **55**(20), 5904-5937
- Weininger, D., 1988, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. of Chemical Information and Computer Sciences*, **28**(1), 31-36

Application of data science to study fluorine losses in the phosphate industry

Houda Ariba,^{a,b} Paul Vanabelle,^c Salah Benaly,^d Thomas Henry,^b Cédric R. André,^b Grégoire Leonard^{a,*}

^a*Université de Liège, Place du 20-Août 7, Liège 4000, Belgium*

^b*Prayon, Rue Joseph Wauters 144, Engis 4480, Belgium*

^c*Cetic, Avenue Jean Mermoz 28, Charleroi 6041, Belgium*

^d*Université Mohammed VI Polytechnique, Lot 660, Hay Moulay Rachid, Ben Guerir 43150, Maroc*
g.leonard@uliege.be

Abstract

Artificial intelligence has become an attractive science for companies as it allows effective data analysis, which helps to improve the manufacturing processes. The aim of this work is to study fluorine losses in a phosphoric acid unit by applying data science methods to process data. Conductivity was used as an indirect measure of fluorine losses in each recovery cycle. After a pre-processing of the data, a Gaussian Mixture Models (GMM) clustering algorithm was applied. Two clusters were found in the data: one with limited losses, and the other with significant losses. In addition, a ratio (R) was created from measurement data to identify the level of fluorine loss compared to fluorine gain during a time step. This ratio R is used in turn to determine whether the plant generates an acceptable amount of fluorine losses.

Keywords: Clustering, Fluorine losses, Phosphoric acid unit, Data analysis

1. Introduction

In recent years, the interest in data science has risen considerably across industries worldwide (Diez-Olivan et al., 2019; Lee and Shin, 2020). Indeed, numerous sensors record real-time data at all stages of the process. When these data are accumulated over time, they can be used as an input to data science algorithms to model complex, non-linear processes. In turn, these models help process engineers to better understand how different variables affect their plant processes.

Clustering is a data science method for analysing data that groups information into clusters with similar properties or features. Clustering is achieved by many unsupervised machine learning algorithms that can be used in various way in different industries. For example, Liu et al. (2018) applied the clustering method in order to evaluate the energy consumption in different companies. Their study allowed to position companies according to their level (high or low) of energy consumption. Sancho et al. (2020) used a clustering method to classify crude oils according to their physico-chemical properties. Zhang et al. (2017) used clustering to discriminate between operational status, i.e. the distinction between operating in a transitory regime and in a permanent regime.

The present article is an application of data science in the chemical industry on a phosphate production process. More specifically, the study will focus on the fluorine losses in a concentration unit of phosphoric acid. The goal is to qualify these losses and to understand their relationship with the process variables. In parallel, a ratio (R) is defined to quantify the losses and ease the readability and discussion of raw data. This indicator will be discussed and put in relation with the clustering results. Perspectives are a better control of the loss of a product that can be valorised and a reduction of the use of caustic soda necessary to neutralize the acidity generated by the fluorine loss.

This paper is organized as follows. In section 2, the fluorine recovery process in the considered phosphoric acid concentration plant is described. Then, in section 3, the methodology is discussed. Section 4 shows the main results that were achieved, and section 5 concludes and suggests some perspectives.

2. Background

The process of fluorine recovery (Fig.1) is divided into three main steps. First, phosphoric acid is concentrated via a vacuum evaporation, leading to a liquid product that is the concentrated phosphoric acid on one hand, and to the emission of a gaseous effluent that contains fluorine on the other hand. Second, this fluorine is recovered in liquid phase in an absorption tower. Finally, the remaining amount of fluorine from the gas phase is transformed into liquid stream rejects by another recovery system. This final recovery system is itself composed of two main units: the PraySep with a spray at its entrance is used to recover the fluocilic acid droplets contained in the gases, and a condenser is used to condense a maximum of the remaining gas. The fluorine losses are detected by a conductivity measurement in the condenser guard tank. Moreover, the amount of recovered fluorine is estimated from a density measurement in the liquid product of the absorption tower.

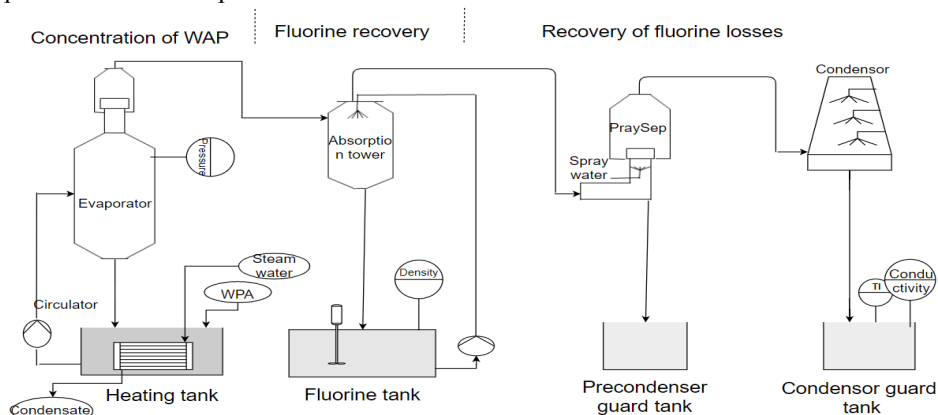


Figure 1: Diagram of the phosphoric acid concentration process.

3. Methods

The data used in this study contain all 2019 data, we focus on the most important variables. Initial data visualization was conducted in JMP® software (SAS Institute) (Yee *et al.*, 2000). The most important signals, which are density and conductivity, are plotted over several cycles in Figure 2. As the fluorine recovery happens in cycles of varying duration (an average of 70 min), the trends of these variables in each cycle was

of interest. During a typical cycle, density in the fluorine tank increases due to fluorine recovery, while at the process exit, conductivity in the condenser guard tank increases due to fluorine losses.

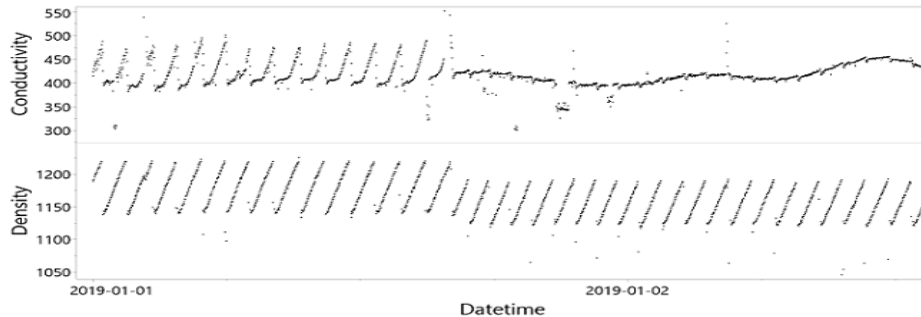


Figure 2: Evolution of the conductivity and the density in the fluorine tank as a function of time.

The figure 2 shows that the absolute values given by the conductivity sensor present a very high variability. Beyond short-term cycles, long-term variations also appear that may be due to incorrect calibration of the conductivity meter, or to other parameters which may affect the conductivity of the mixture in the condensate tank but that are not measured. For instance, the quality of the sprayed water before the PraySep, or even the water used to clean the vapour condensation could cause this long-term variation. To get rid of the long-term trend the conductivity signal has been normalised per cycle. For each conductivity point of a cycle, the value of the conductivity at the start of that cycle was subtracted. Consequently, a new standardised signal was constructed, making losses easier to analyse.

In order to study the behaviour of this conductivity in a more systematic fashion, a clustering method was applied. Gaussian Mixture Models (GMM) were used to automatically classify and label the conductivity cycles according to parameters qualifying their evolution. GMM allow to classify data coming from overlapping normal distribution. In this case, it was assumed that cycles with few losses, with medium losses and with many losses all follow a normal distribution and we want to determine these distributions. Note that, at other times, the conductivity signal is noisier or does not present a regular pattern. From these observations, we decided to fix the number of clusters to 3 in the initialisation parameters, plus a specific cluster for outliers. As the slopes of the conductivity characterise the fluorine losses well, we have used parameters derived from the slopes as inputs to this algorithm. Specifically, conductivity slopes have been calculated over 20-minute ranges at the start, the middle and at the end of each cycle, as well as the slope of the total conductivity cycle.

For the purpose of quantifying fluorine losses, a parameter was needed through which the losses could be described in a more concrete way according to their importance. The variables presenting the fluorine loss and gain were identified. The definition of a ratio R that includes these variables allows to describe the evolution of the losses during a fluorine recovery cycle. Note that this ratio cannot be used to describe losses that may be associated to longer term conductivity variations.

The way in which the density evolves during a cycle gives an idea of the productivity and the efficiency of the fluorine recovery. At the same time, the conductivity behaviour during the cycle reflects the Fluorine losses that are not recovered by the absorption

tower and the PraySep. Therefore, to accurately assess the performance of the fluorine recovery, the density gain and the conductivity loss must be considered in combination. To solve this problem an indicator that measures this recovery has been constructed. It is a ratio between the average of the normalised conductivity and the average of the density gained during a given time. Note that the density signal measures the cumulative density of the fluorine tank mixture which increases during the fluorine recovery cycle, while the conductivity gives values considered as instantaneous since the condenser guard tank is quite small, consequently the residence time of the mixture in this tank is very low. The ratio is defined as

$$R = \frac{\bar{\sigma}}{\bar{\rho}_a - \bar{\rho}_i} \quad (1)$$

Where: $\bar{\sigma}$: Average of normalized conductivity over a period Δt ; $\bar{\rho}_i$: The density of the first point of this Δt ; $\bar{\rho}_a$: Average of density over this Δt

This ratio measures how many conductivity points are generated for density points gained over a given period of time during the fluorine recovery cycle, we tried to calculate the R with different time slots (10 min, 15 min, and 20 min). Theoretically, this ratio increases as a function of time during the recovery cycle. At the beginning of the cycle, the fluorine tank is filled with fresh water. As a result, its recovery capacity is high. Subsequently, the density of the mixture in the fluorine tank increases over time. Consequently, its capacity to recover fluorine decreases, which generates more fluorine losses. As a result, the values of both the numerator and the denominator increase with time. However, the losses (numerator) increase faster than the gains (denominator), so that this ratio can be used as an indicator of when the cycle starts being inefficient in terms of fluorine recovery.

4. Results and discussion

The GMM algorithm was applied with the conductivity slopes as features. The model groups the data into 4 families, the cluster 0 contains the cycles whose probability of belonging to the 3 other groups is negligible, therefore, they can be considered as outliers. As a result of this clustering, the cluster 1 contains 53% of the cycles which corresponds to stable signals with low slopes (low fluorine losses). The cluster 2 includes disturbed signals with very high slopes and large standard deviations between the ratios, this group has over 12% of the conductivity cycles. The cluster 3 includes 33% of the cycles, it gathers signals that do not include too much noise and with high slopes compared to those of the first cluster. In the rest of this study, we will focus on the 1st and the 3rd clusters to quantify the losses of each of them, and to analyse the distributions of the production parameters. The 2nd cluster has been discarded because the signals in this group are too noisy to be interpreted.

Following this, we studied the evolution of the ratio R during the fluorine recovery cycles. To do so, each cycle has been divided into 3 parts, Then, the R value is calculated at the start, the middle, and the end of each cycle. The results by using 20min as time slot appears to be more interesting because several points of the cycle are taken, which allows a better characterization of the R. These values using periods of 20 minute

produced the R_20_Start, R_20_Middle, and R_20_End values, respectively. The figure 3 shows the distribution of these ratios for the two clusters.

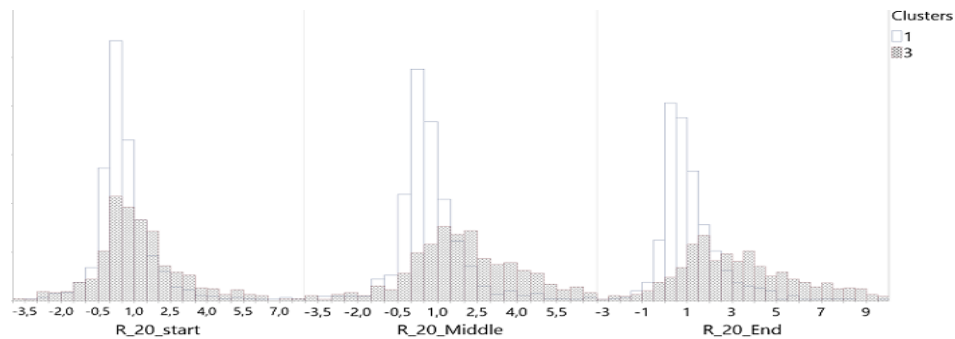


Figure 3: The histogram distribution of the fluorine loss ratios for the two clusters 1 and 3.

The figure 3 illustrates that there is a progressive shift in the histograms of these ratios for the two clusters. In fact, the difference in R ratios between the two clusters starts at the beginning of the cycle, although this difference is relatively small in absolute value (0.5 on average for the first cluster and 0.94 for the third cluster). One can also observe that the standard deviation of cluster 3 is much larger than the one of cluster 1. Over time the difference becomes more and more important until the end of the cycle where the two clusters become clearly separated with a small overlapping section. At the end of the cycles, the ratio for the first cluster remains almost stable (1.08 on average), while the ratio for the third cluster considerably increases, from 0.94 at the start to 3.75 on average at the end. Also, we notice that 90% of the data of the first cluster have an R_20_End less than 3. On the other hand, for cluster 3, 50% of the data are higher than this value. So, the value R_20_End=3 can be used as a limit value which indicates high fluorine losses.

Once these observations have been made, relationships with process variables can be analysed. The important variables that can influence the fluorine recovery are weak phosphoric acid (WPA) feed rate, circulator amperage, steam flow, vacuum, and flow rate of the spray water before the PraySep. The behaviours of these variables have been studied but they do not show a very clear difference depending on whether a data point belongs to one cluster or to the other. This could be explained by the continuous nature of the slopes, by the noise on the conductivity signal and by the overlapping nature of the clusters. Therefore, this prevents a clear distinction between the two clusters in their respective process operating conditions. Nevertheless, it has been remarked that the flow rate of spray water did not change between the two clusters, whereas there is a difference for the steam flow rate. It is possible that the spray water flow rate is not adapted to the productivity of the unit.

To retrieve more information about the influence of process variables, it was decided to work on a sample that encompasses the two extreme types of conductivity cycles. Cycles with intermediate conductivity slope were manually removed from the data set in order to strengthen the differences between the two clusters. The resulting sample is composed of 253 cycles, of which 35% belong to the 1st cluster and 65% to the 3rd cluster. The distribution of R ratios for this sample is similar to the results in Figure 3. We notice that in this case all R_20_End values for the first cluster are lower than 3 and

more than 70% of the cluster 3 R_20_End values are higher than 3. This further confirms the value of R_20_End=3 as a threshold indicator of high fluorine losses.

As a second step, the process operating conditions have been studied for this sample, in the same way as it was done for the whole database. As a result, the difference between the two clusters appears more clearly when doing the study on this sample. The distribution of the WPA flow, the steam flow, the vacuum pressure, and pump amperage show that cycles belonging to the 1st cluster correspond to low values of these process variables. In contrast, cycles belonging to the 3rd cluster have higher parameter values. This confirms the important effect of the process load on the conductivity behavior during a recovery cycle. Indeed, when the flow rate of phosphoric acid to be treated is increased (WPA flow), the operating conditions are being pushed towards the process limits (for instance, more steam is required for the evaporation), but this seems to be insufficient to reach similar Fluor recovery performances. Nevertheless, all these variables present a large zone where the two clusters overlap, which confirms that the process load alone does not explain all the fluorine losses.

5. Conclusion

To conclude, this study was conducted to study fluorine recovery in a phosphoric acid concentration plant. In order to better characterize the losses during the recovery cycles, a ratio R has been defined which considers the gain (density increase in the recovery vessel) and the losses (conductivity increase in the liquid effluent waste). Since conductivity evolves in different ways, a clustering method using GMM algorithm has been used to classify the data according to low or high fluorine recovery performance. Following this methodology, it has been found that from a value of R=3 there are high fluorine losses. So, this ratio can be used as an indicator of fluorine losses.

Finally, it should be mentioned that the same procedure discussed previously was applied to another fluorine recovery unit located in the same phosphate plant. As a result, the same conclusions were reached, which means that this method can be used for any fluorine recovery plant in order to study the fluorine losses. To go beyond this study, a monitoring system that triggers an alert for high fluorine losses when the ratio R is higher than a given value (in this case when $R > 3$) will be studied. In addition, and to clearly identify the causes that influence the behaviour of fluorine losses, a detailed experiment on the flow rate and quality of the spray water injected to the PraySep unit will be carried out.

References

- Diez-Olivan, A., Del Ser, J., Galar, D., Sierra, B., 2019. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Inf. Fusion* 50, 92–111.
- Lee, I., Shin, Y.J., 2020. Machine learning for enterprises: Applications, algorithm selection, and challenges. *Bus. Horiz., artificial intelligence and machine learning* 63, 157–170.
- Liu, G., Yang, J., Hao, Y., Zhang, Y., 2018. Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering. *J. Clean. Prod.* 183, 304–314.
- Sancho, A., Ribeiro, J.C., Reis, M.S., Martins, F.G., 2020. Cluster Analysis of Crude Oils based on Physicochemical Properties, *Computer Aided Chemical Engineering*, 30 European Symposium on Computer Aided Process Engineering. Elsevier, pp. 541–546.

- Zhang, Y., Bingham, C., Martínez-García, M., Cox, D., 2017. Detection of Emerging Faults on Industrial Gas Turbines Using Extended Gaussian Mixture Models. *International Journal of Rotating Machinery*.
- Yee, Christine, et al., 2000. Design of experiment and data analysis by JMP®(SAS institute) in analytical method validation *Journal of pharmaceutical and biomedical analysis*,pp 581-589

Data-driven Branch-and-bound Algorithms for Constrained Simulation-based Optimization

Jianyuan Zhai^a, Sachin Shirpurkar^a, Fani Boukouvala^{a,*}

^a*Georgia Institute of Technology, 311 Ferst Dr, Atlanta, GA, 30332, USA*

fani.boukouvala@chbe.gatech.edu

Abstract

The wide use of detailed simulations for complex systems has led to a growing interest for methods that can optimize simulation-dependent problems using data, without explicit equations or derivatives. Due to the lack of derivatives and the dependence on sampling, simulation-based optimization algorithms lack convergence guarantees and often require a significant number of samples to identify an optimal solution with consistency. Moreover, the presence of black-box constraints is an open challenge because it further complicates sampling and identification of unknown feasible spaces. Previously, we have introduced the Data-Driven Spatial Branch-and-Bound algorithm for box constrained problems, which employs data-driven convex underestimators, finds upper and lower bounds on the optimal objective value and progressively prunes suboptimal subspaces until convergence. In this work, we present recent advances of this framework for handling simulation-based and equation-based constraints. We demonstrate the performance of these features with respect to convergence and sampling requirements through benchmark constrained optimization problems.

Keywords: black-box optimization, simulation-optimization, surrogate modelling.

1. Introduction

An increasing amount of optimization applications in research and industry today require the embedding of various forms of data from simulations of various fidelity and/or scale. Such problems are challenging due to the inability to directly use efficient deterministic optimization solvers that require equation-based formulations. As a result, optimization of such problems is often referred to as “black-box” because it relies on input-output data. Many recent contributions from the engineering literature aim to apply and improve the performance of black-box optimization techniques for a wide variety of applications (Bhosekar et al., 2018). Optimization with embedded simulations can be performed with purely sampling-based algorithms, such as direct-search methods, genetic algorithms, particle swarm optimization, and many more that are reviewed in (Rios and Sahinidis, 2013, Boukouvala et al., 2016). Alternatively, a different class of methods employs approximations of data to perform optimization, and these are called model-based or surrogate-based techniques. Such techniques employ novel mechanisms for data collection, fitting of a surrogate model to represent the data, which then allows one to use equation-based optimization solvers. In order to avoid excessive sampling, most techniques in the literature use a small initial set of samples and subsequently adaptively “fit – optimize – resample” the space until their convergence criteria are met. This

adaptive scheme is always based on techniques that are designed to balance “exploration” (i.e., searching the feasible space just enough to minimize probability of missing the optimal solution), and “exploitation” (i.e., focusing sampling in promising optimal regions where surrogate models need to be accurate). Recent work focuses on the identification and comparison of different types of surrogate models and their performance for optimization (Garud et al., 2019).

Despite recent advances in the simulation-based optimization literature, several open challenges still exist. First, many of these methods suffer from the curse-of-dimensionality, since sampling requirements increase when decision-variables increase. Second, when a surrogate model is used, there are open challenges when it comes to the selection of a surrogate model that is accurate and has a tractable mathematical representation so that it can in turn be optimized (ideally globally). Specifically, some popular surrogate models result to equations with nonconvex terms, and the number of terms and variables rapidly increase with the number of dimensions, model architecture, and even with number of data points. Recent work on global optimization of Neural Networks as surrogate models are tackling this challenge (Schweidtmann and Mitsos, 2019). Third, we have previously shown (Zhai and Boukouvala, 2020) that surrogate model selection and training leads to significant variation in the optimal results, because if a different surrogate model is used, or different training data are used, many algorithms provide a different result. This inconsistency in performance is highly undesirable for any optimization algorithm. Last, earlier work in black-box optimization led to algorithms that treat the problem as a pure “black-box”, often without the ability to consider constraints. However, it is important to treat such problems as “grey-boxes”, which allows one to mix known equations that are typically constraints together with an objective and constraints that are reliant on the simulation.

In our previous work, we have proposed an approach that aims to tackle some of the open challenges mentioned above (Zhai and Boukouvala, 2020). Specifically, instead of relying on a known type of surrogate model, we develop convex quadratic underestimators of data. The data may come from high-fidelity simulations, or multiple low fidelity surrogate models. The reason for this approach is to avoid several of the disadvantages of surrogate modelling, such as: (a) variation in parameters and model type with slight changes in data; (b) computational expense of training and selection of best model out of many; and (c) intractability of complex surrogate models when it comes to their global optimization. Moreover, our approach utilizes a spatial branch-and-bound (b&b) framework to search the space and focus sampling in promising spaces, while pruning spaces that are not optimal. This strategy allows us to incorporate more rigorous convergence criteria, such as absolute or relative ε -optimality gap (as opposed to reaching a maximum sampling or CPU limit). This Data-Driven Spatial Branch-and-Bound (DD-SBB) algorithm, also provides estimates of upper and lower bounds on the optimum at any intermediate stopping point. We have found that this approach is promising and obtains optimal solutions with few sampling points, however, it requires a large amount of samples to close the optimality gap.

In this work, we propose techniques for extending the capabilities of the DD-SBB approach in terms of constraints handling. We consider two different forms of constraints: (1) simulation-based constraints (i.e., constraints that are unknown in equation format but are embedded within the simulation) and, (2) equation-based constraints (i.e., constraints that are explicitly known algebraically as a function of decision variables). The type of formulations we aim to solve in this work is shown in Problem 1 (P1), where the objective

function f and constraints g_s are embedded within a simulation and thus we do not have explicit equations for them. Constraints g_k are available in equation format. We only consider continuous variables that are bounded.

$$\begin{aligned}
 & \text{(P1) } f(x) \\
 \text{s. t. } & g_s(x) \leq 0, \quad s = 1, \dots, S \\
 & g_k(x) \leq 0, \quad k = 1, \dots, K \\
 & x^{lo} \leq x \leq x^{up}, \quad x \in R^N
 \end{aligned}$$

We implement and test different branching and bounds tightening techniques and test this extended capability of the algorithm through a set of benchmark problems. Our approach does not add computational cost to the algorithm because it does not require the fitting of surrogate models for the simulation-based constraints. Concepts from decision-tree techniques are explored for branching the search space cleverly into “feasible” and “infeasible” nodes and the performance of different branching rules are compared with respect to efficiency and convergence.

2. Methods

2.1. Data-Driven Spatial Branch-and-Bound Framework

The DD-SBB framework starts with initial sampling of the entire box-constrained space using Latin Hypercube sampling. Based on these high-fidelity samples from the simulation, a low-fidelity surrogate model can be fit, but this model is only used for ranking the importance of variables and for generation of large amounts of low-fidelity data, to be used to generate the convex underestimators. Convex underestimators are found by solving P2, where M is the total number of samples, and a, b, c are the parameters of the convex quadratic function that underestimates all of the obtained data. In previous work we have reported results on the validity of these bounds as more data is included from high- and low- fidelity samples (Zhai and Boukouvala, 2020).

$$\begin{aligned}
 & \text{(P2) } \sum_i^M (f(x_i) - f_{lb}(x_i)) \\
 \text{s. t. } & f(x_i) - f_{lb}(x_i) \geq 0 \quad \forall i = 1 \text{ to } M \\
 & f_{lb}(x_i) = ax_i^2 + bx_i + c \quad \forall i = 1 \text{ to } M \\
 & a \geq 0
 \end{aligned}$$

For each space or node of the b&b tree, a lower bound (LB) is found through minimization of the convex underestimator and an upper bound (UB) is the best high-fidelity sample collected. The algorithm then proceeds with branching the space with a selected branching rule, resampling within subspaces and updating the LB and UB of the problem. The branching heuristics originally implemented are equal bisection with respect to branch location. When deciding on which variable to branch on first, the algorithm has options prioritizing the longest side or prioritizing the most important variable with respect to the objective. The presence of constraints requires several modifications to this framework, including novel branching and pruning rules, which will be presented in the next section.

2.2. Branching rules in the presence of constraints

Branch rules are essential to a branch-and-bound algorithm. Besides equal bisection on the longest edge and equal bisection with customized variable selection (Zhai and Boukouvala, 2018, 2020), we implemented two other branch rules targeted to handle

constraints. In the presence of constraints, samples are labelled as 0 (infeasible) and 1 (feasible). The proposed branch rules are designed to separate feasible from infeasible regions, with the hypothesis that this will help with faster pruning of infeasible regions. One strategy is to use a weighted Gini impurity score commonly used in classification and regression trees (Kotsiantis, 2013). The Gini impurity score (Eq. (1)) is a measure of the tendency that a randomly chosen sample would be misclassified in a node.

$$G = \sum_{i=0}^1 p(\text{label} = i) * (1 - p(\text{label} = i)) \forall i = 0, 1 \tag{1}$$

where $p(\text{label} = i)$ represents the probability of a randomly chosen sample with label i in the node to be chosen. To select the best location to cut, we minimize the weighted Gini impurity score (Eq. (2)), which minimizes the possibility of misclassification if a cut is generated at x_k on edge d .

$$k, d = (p(x_d \leq x_{k,d}) * G_{x_d \leq x_{k,d}} + p(x_d \geq x_{k,d}) * G_{x_d \geq x_{k,d}}) \tag{2}$$

where k denotes k^{th} equidistant location on edge d and $p(x_d \leq x_{k,d})$ is the possibility of a randomly chosen point lies in the potential subspace.

The second strategy is a customized purity score (Eq. (3)) that measures the difference in the fraction of infeasible samples and fraction of feasible samples in each subspace.

$$\text{Purity}_{k,d} = \left| \frac{N_{\text{infeasible}_{x \leq x_{k,d}}}}{N_{\text{infeasible}}} - \frac{N_{\text{feasible}_{x \leq x_{k,d}}}}{N_{\text{feasible}}} \right| \tag{3}$$

If one subspace contains samples with only one label, the purity score will be 1. Otherwise, the purity score is between 0 and 1. To select the best location, we maximize the purity score (Eq. (4)) to separate purely feasible and infeasible spaces.

$$k, d = \text{argmax}_{k,d} \left(\left| \frac{N_{\text{infeasible}_{x \leq x_{k,d}}}}{N_{\text{infeasible}}} - \frac{N_{\text{feasible}_{x \leq x_{k,d}}}}{N_{\text{feasible}}} \right| \right) \tag{4}$$

where k denotes k^{th} equidistant location on edge d . Note that if one node contains only samples with one label, equal bisection on longest edge will be used.

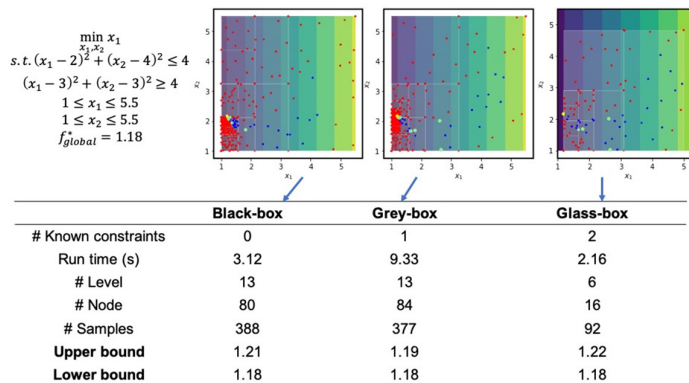


Figure . Motivating example formulation, algorithm performance and results. Red points are infeasible, blue points are feasible.

1.1. Pruning in the presence of constraints

Pruning a search node when the node appears to be less promising to find the global optimum helps the algorithm converge faster. Generally, a node is pruned when the local lower bound is higher than the global upper bound. In the presence of constraints, we implemented additional rules to prune nodes that are infeasible. When the constraints are known, we perform feasibility-based bound tightening at the root node and at each active node. Specifically, we solve a constraint-violation minimization problem to check the feasibility of each node, and if a subspace is found to be infeasible, it is pruned. When the constraints are unknown, node pruning is less straightforward because decisions are made depending on the feasibility labels on the samples instead of explicit mathematical constraints. To reduce the chance of pruning feasible regions, backtracking techniques are incorporated. When the algorithm encounters a node that contains only infeasible samples, it backtracks the parent and grandparent nodes. Only if both the parent and grandparent nodes contain purely infeasible samples, will the node be pruned. By doing so, we avoid pruning nodes prematurely and allow collection of extra samples in that node, to increase the confidence of pruning the node.

2. Results

2.1. Motivating Example

Here we present results on a simple 2-d example with unknown objective function that enables us to highlight the algorithm performance. In Figure 1, we show the formulation of the problem, as well as the performance of the algorithm when the formulation is purely black-box (all constraints are simulation-based), grey-box (first constraint is known, second is unknown), to glass-box (all constraints are known). As expected, as more constraints become known, the algorithm converges to the global optimum with less samples. In addition, we observe that the additional cost of feasibility tightening does not significantly increase the cost of the algorithm. On the contrary, due to faster convergence, the algorithm converges faster with respect to run-time when constraints are present. In this motivating example, we keep the pruning and branching rule fixed to equal bisection with branching on most important variable. A comparison of different heuristics is performed in the next section.

2.2. Algorithm benchmarking

In order to test the performance of different branching rules, we test the algorithm on a larger set of benchmark problems. These are 42 problems from the MINLPlib, with 2-5 dimensions and no limit on the number of constraints. The algorithm terminates if any of the following criteria are met: 1) the absolute gap between the lower and the upper bound is smaller than 0.05, 2) the relative gap is smaller than 0.001, 3) the longest edge in a search area is smaller than 0.05, 4) the number of samples reaches 200,000 and 5) CPU reaches 5 hr. Results are reported for the two extreme cases, namely the black-box case where no constraints are known and the glass-box case where all constraints are known (Figure 2). Results show that the algorithm can locate the global optimum for 80% of the problems in the black-box case, and more than 85% of the problems in the glass-box case. Moreover, the algorithm converges with less samples when constraints are explicitly known. Results show that the novel branching rules, namely purity score and Gini impurity score, proposed in this work expedite convergence (i.e., more problems are solved with less number of samples). However, overall, the equal-bisection branching

rule solves more problems if more samples can be collected. This implies that these decision-tree rules can expedite pruning, however, they are less conservative and thus increase the chance of removing good optimal solutions.

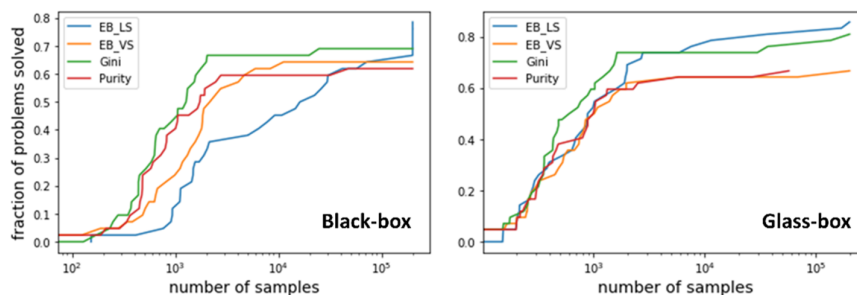


Figure SEQ Figure \^* ARABIC 3. Performance of algorithm and different branching rules for set of 42 benchmarks, (EB_LS: Equal-bisection on longest side, EB_VS: Equal-bisection with variable selection)

3. Conclusions

In this work we presented an extension of our algorithm for data-driven branch-and-bound for constraints handling. We propose and test several branching and pruning rules in the presence of simulation-based constraints that are based on concepts proposed in the literature of decision-trees. In order to treat equation-based constraints, we employ feasibility-based bounds tightening techniques. Results show that it is important to use any known constraints directly, because this expedites algorithm convergence and overall solves more problems without adding computational cost due to reduced sampling requirements. Moreover, branching when using decision-tree heuristics expedites convergence of the algorithm, however, in certain cases prunes valuable solution spaces. As a result, when sampling is not a significant burden, the equal-bisection approach is a more conservative and reliable heuristic.

References

- A. Bhosekar, M.G Ierapetritou, 2018, Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers & Chemical Engineering* 108, 250-267
- S.S. Garud, N. Mariappan, I.A. Karimi, 2019, Surrogate-based black-box optimisation via domain exploration and smart placement. *Computers & Chemical Engineering* 130, 106567
- F. Boukouvala, R. Misener, C.A. Floudas, C.A., 2016, Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO. *European Journal of Operational Research* 252(3), 701-727
- S.B. Kotsiantis, 2013, Decision trees: a recent overview, *Artificial Intelligence Review*, 39: 261-83.
- L. M. Rios and N.V. Sahinidis, 2013, Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization* 56(3), 1247-1293
- A.M. Schweidtmann, A. Mitsos, 2019, Deterministic Global Optimization with Artificial Neural Networks Embedded. *Journal of Optimization Theory and Applications* 180(3), 925-948
- J. Zhai, F. Boukouvala, 2016, Nonlinear Variable Selection Algorithms for Surrogate Modeling. *AIChE Journal* 0(ja), e16601. doi:10.1002/aic.16601
- J. Zhai and F. Boukouvala, 2020, Data-driven Spatial Branch-and-bound Algorithm for Box-constrained Simulation-based Optimization, In.: *Living Archive for Process Systems Engineering*. <http://psecommunity.org/LAPSE:2020.1164>

Reactor scale modelling of gas hydrate-based CO₂ capture (HBCC) process

Hossein Dashti,^a Gerald Fraser,^b Amirpiran Amiri^b

^a*School of Chemical Engineering, The University of Queensland, Brisbane, Australia*

^b*Energy and Bioproducts Research Institute (EBRI), Aston University, Birmingham, UK*

Abstract

During the last decade, the hydrate-based CO₂ capture (HBCC) process has been approved as a promising alternative. The HBCC offers several attractive advantages such as the mild operating pressure and temperature, the ease of regeneration of CO₂, and its low energy consumption. The HBCC is at its infancy stage where the numerical methods play a key role in its research and development. This is to support the experimental research in terms of the model-based experiment design and process analysis/optimisation. In this view, HBCC modelling and simulation have been focused by researchers worldwide. Most of the published models are restricted to the processes at the particle scale that are enormously beneficial for a fundamental understanding of the process science. Nevertheless, a challenging shortfall currently exists for reactor scale models given the process scale research, and development objectives require models at the relevant scale. Accordingly, the reactor level models that may be used in a comprehensive process are required. Ideally, a multi-scale model allows studying both science and technology. This paper presents an HBCC process model at the reactor level for batch and semi-batch operations. The scaled-up model is capable of simulating the CO₂ dissolution, and growth phases. The model evaluates the hydrate layer boundary and estimates the total gas uptake in batch semi-batch reactors. A modified version of the model made it possible to account. The present study establishes a groundwork for the large-scale application of the HBCC.

Keywords: Gas hydrates, CO₂ capture, reactor scale modelling, kinetic model

1. Introduction

During the last decade, there has been growing interest in deploying hydrate-based CO₂ capture (HBCC) due to its moderate operating temperature and pressure, low energy consumption and easy regeneration of the aqueous solution (Dashti and Lou, 2018, Dashti et al., 2015). For full commercialisation, the process computation tools are crucial. To model the HBCC at reactor level, it is essential to achieve fundamental insights at micro and mesoscales. This is vital for computing the kinetics of crystallisation, mass/heat transfer rates. Nucleation and growth are the two interlinked stages involved in the hydrate formation process, making the model development at particle level challenging. Despite a vast amount of research conducted, there is the hydrate kinetics and routes yet to be well understood, revealing the importance of

micro-scale studies. The commercialisation challenges, however, need reactor scale studies and analysis tools.

In a recent study, the modelling of CO₂ hydration growth and nucleation was conducted through the utilisation of the shrinking core method (SCM) (Dashti et al., 2019; Dashti et al., 2019). The models presented in the mentioned works were capable of capturing the conversion trajectory and nucleation phase behaviour over the process time. The boundary condition used in these works was assumed based on the reactor conditions averaged. This may affect the model results due to the non-homogeneity that normally exists in a reactor. This is, in particular, crucial to local temperature and concentration are used as boundary conditions of single particles. Ideally, a CFD model should be employed to capture the spatiotemporal data in a mixed reactor while the particle model is simultaneously deployed. This will be computationally time-consuming due to the giant particle number and the tremendous data exchange between the scales. For a hydrate reactor with no agitator, an applied model does not necessarily need such complexity and computation load. Traditionally, HBCC processes utilised stirred tank reactors as it thought that the mixing provides improved mass transfer kinetics; however, the reactor faces problems of agglomeration of hydrate crystals and the reduced coefficients of mass transfer. It is unfavourable for significant scale-up due to the high stirrer-driving power consumption (Mori, Y. H, 2015). These lead to the application of unstirred reactors with gas hydrate promoters (Dashti et al., 2015, Veluswamy et al., 2018). The packed beds of sand, silica or pulverised coal have been used in some studies (Linga and Clarke, 2017). This paper presents a computationally, fast and fundamentally efficient model for hydrator. This is of interest to couple the particle scale and the reactor scale models. This paper establishes a model to describe the HBCC process during batch and semi-batch operations at the reactor level. Initially, three models, including transient mass transfer, transient heat transfer and CO₂ hydrate formation (Improved model) models for the batch operation, have been developed further, the models adopted for the semi-batch operation. Finally, a scale-up reactor model which is capable of simulating the CO₂ dissolution, nucleation, and growth phases has been developed.

2. Methodology

2.1. Basic and improved model

Gas hydrate formation consists of two main steps: gas dissolution and gas hydrate formation. First, a basic model for CO₂ dissociation has been developed and then this model serves as a reference to the improved model. The basic model for the batch reactor considers transfer rates of CO₂ diffusion in water by first establishing the rate-determining mechanism between the transport phenomena, mass transfer and heat transfer of CO₂ in the stationary water medium. The controlling mechanism will then be utilised as the core of the model to describe the dissolution of CO₂ (A) through the reactor. The concentration and the temperatures are time (t) and location (x) dependent. Only mass diffusion (D) and conductive heat transfer (kVT) are considered in the stationary water. This is a justifiable assumption due to the low temperature and no-agitating system. Table 1 presents the governing equations for basic and improved models. The models' parameters are presented in Table 2.

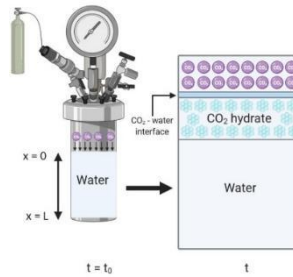

 Figure 1: Reactor schematic for CO₂ hydrate operation

Table 1. Model equations and initial and boundary conditions (BC and IC)

Model	Equations	BC and IC
Basic Model	$\frac{\partial C_A}{\partial t} = D \frac{\partial^2 C_A}{\partial x^2}$	$C_A(0, x) = 0; C_A(t, 0) = C_{A0}; dC_A(t, L) = 0$
	$\frac{\partial C_A}{\partial t} = D \frac{\partial^2 C_A}{\partial x^2}$	
	$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2}, \alpha = \frac{k}{\rho c_p}$	$T(0, x) = T_0; T(t, 0) = T_0; dT(t, L) = 0$
Improved model	$\frac{\partial C_A}{\partial t} = D \frac{\partial^2 C_A}{\partial x^2} - k_0 e^{\left(\frac{-E_a}{RT}\right)} \cdot C_A \cdot \frac{S}{V}$	$C_A(0, x) = 0; C_A(t, 0) = C_{A0}; dC_A(t, L) = 0$ $T(0, x) = T_0; T(t, 0) = T_0; dT(t, L) = 0$
	$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} + k_0 e^{\left(\frac{-E_a}{RT}\right)} \cdot C_A \cdot \frac{S}{V} \cdot \frac{\Delta H}{\rho c_p}$	

 Table 2. Parameters used for the semi-batch process of CO₂ hydrate formation

Parameters	Value
D (m ² /h)	6.912×10 ⁻⁶ (Cussler, E. L., 1997)
α (m ² /h)	5.24×10 ⁻⁴
C _A (J/mol.K)	4181
ρ (kg/m ³)	997
k ₀ (m/h)	2.05×10 ⁶⁹ (Bergeron and Servio, 2008)
E _a (J/mol)	387645.5 (Bergeron and Servio, 2008)
R (J/mol.K)	8.314 (Jensen, 2003)
ΔH (J/mol)	66800 (Claudia, 2013)

2.2. Batch and semi-batch operations

A reactor with a constant volume of water, V, and water surface area, S, that is initially maintained at 25 °C is assumed (Figure 1). In semi-batch operation, a pure CO₂ (1 bar and 5°C) is continuously supplied to apply constant CO₂ pressure at the gas-water interface. Therefore CO₂ concentration estimated based ideal gas law, has been used as a boundary condition. For the batch operation, As CO₂ diffuses through the water to be consumed in hydration over time, its concentration in gas bulk decreases. A time-dependent concentration must be considered at the top boundary. A mathematical expression Eq. (1) was derived from (Raizi, 1997) and (Zarghami, 2017) to calculate the local dynamic concentration, C(t,x). This reactor model uses a stationary non-agitated water volume; therefore, the mathematical expression is suitable for this model. The equation was simplified for the concentration of gas at the top boundary only, i.e., C(0,t).

$$C(0, t) = \frac{P}{H} \exp\left(\frac{t}{\eta^2 D}\right) \operatorname{erfc}\left(\frac{\sqrt{t}}{\eta \sqrt{D}}\right), \quad \eta = \frac{VH}{\sqrt{DART}}, \quad C(0, 0) = \frac{P}{H} \quad (1)$$

in which P is pressure, H is Henry's law constant and R is the molar gas constant.

2.3. Reactor scale-up

With a fully functional batch operation version of the improved model, the reactor scale-up can be attempted to evaluate the profiles and the gas uptake in the hydrate layer of the reactor. In the scaled-up reactor, the reactor length is scaled up from 0.12m to 1m. The theoretical volume of the reactor is 1000m³ and Area is 1000m². The term for Molecular Mass, M of CO₂ is introduced to equations for temperature profile in the CO₂ hydrate formation model and Eq. (2), to indicate the change from molar concentration to mass concentration:

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 C_A}{\partial x^2} + k_0 e^{\left(\frac{-E_a}{RT}\right)} \cdot C_A \cdot \frac{S}{V} \cdot \frac{\Delta H}{\rho C_p M}, \eta = \frac{VHM}{\sqrt{DART}} \quad (2)$$

3. Results and Discussions

The models have been solved using MATLAB ‘pdepe’ solver. The basis model PDEs were numerically solved individually. The improved model PDEs were solved simultaneously to account for the coupled dependence of both concentration and temperature. The reactor length scale and timescale were used; $0 \leq x \leq 0.12\text{m}$ and $0 \leq t \leq 80\text{hr}$. The dynamic concentration and temperature profiles in the reactor are presented in Figure 2a and 2b. Note that independent mass and heat transfers are assumed.

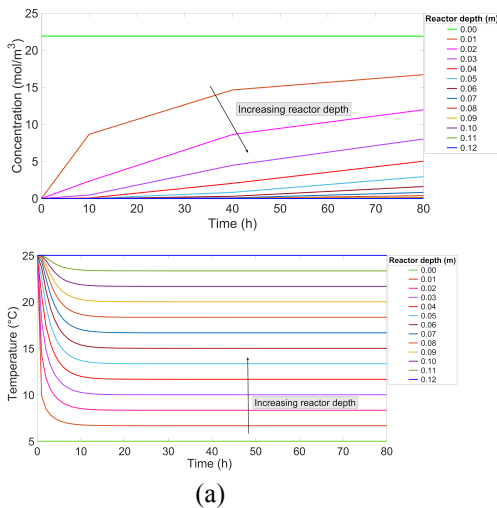


Figure 2. Results of the concentration and temperature changes in the reactor versus time for mass transfer only (2a) and heat transfer only (2b) models

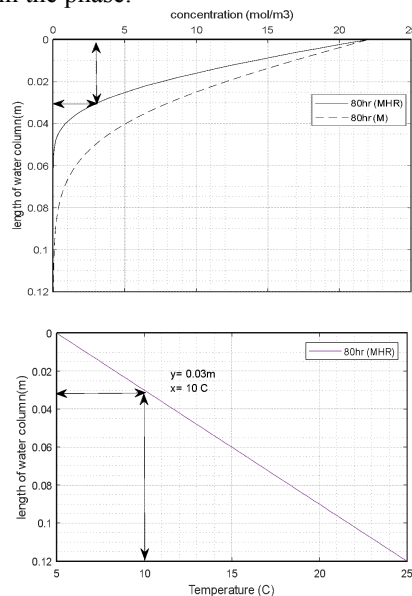
The temperature graph (2b) indicates a faster rate of change over time from 0 to approximately 7 hrs for each length and reaches a steady-state ($dy/dt = 0$) at an earlier time than the concentration in figure 2a. For example, at 0.1m, the concentration has yet to reach a steady-state (indicated by a straight line parallel to the x-axis) but only begins to approach steady-state at around 30-40 hrs. However, at the same point, the temperature has already reached a steady state at a temperature of about 6.7 °C. As a result, this shows that the transient mass transfer process is much slower than the transient heat transfer process and therefore, this mechanism can be considered the rate-controlling mechanism for the CO₂ dissolution process into the water. This is due to higher mass transfer resistances involved with the transport process of CO₂ in the water body over the transport of heat. In a further analysis in this study, the basic model

which uses the only the transient mass transfer compared with the improved model that combines the effects of transient mass transfer, transient heat transfer and the hydrate formation thermodynamics/kinetics (Figure 3a). The basic model is denoted as “M” data line, and the improved model is denoted as the “MHR” data line. The results of this analysis in conjunction with the temperature profile (Figure 3b) in the improved model (MHR) can be used to predict the metastable and stable region for where the hydrate phase can occur within the reactor profile.

It can be seen from Figure 3b that between the top boundary at 0m and 0.03m, the temperature maintains a constant profile of 5-10 °C. Therefore, 0.03m is the maximum length under these parameters for hydrate nucleation and growth to occur. That is 25% of the reactor length. Also, at the maximum length, the concentration is at 3.126 mol/m³, and with the top boundary being 21.917mol/m³, this means that about 86% of the CO₂ diffused into the water remains in the stable region for hydrate growth for the improved “MHR” model. Compared to the basic model “M”, at 0.03m, the concentration is at 8.043 mol/m³. This indicates about 63% of the CO₂ diffused remained in the hydrate stable region.

Further, beyond the maximum of 0.03m for the stable region, the “M” data line continues to maintain a steady linear correlation to approximately 0.04-0.045m. This shows that even after passing through the stable region, there is no change in the concentration gradient between 0 and 0.4m (meaning a constant driving force for incremental changes in x). There is no interference for the transport of CO₂ in the basic model other than the bulk of water itself. In contrast, for the improved “MHR” model after 0.03m, there is a sharp change in the curve. This can be explained as the remaining 14% of CO₂ that diffused beyond the region for hydrate growth has less interference/resistance in the body of the water. This shows that between 0-0.03m, there was an increased constant transfer resistance caused by hydrate growth formation.

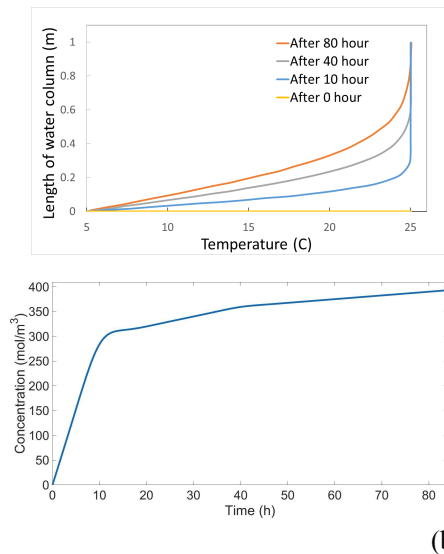
Figure 3 show that the improved model is indeed superior to the basic model for the prediction hydrate even with the appropriate temperature considerations. Furthermore, it is possible to use this type of analysis to predict the stable phase for growth and the proportion of CO₂ within the phase.



(a) (b)
 Figure 3- A comparison between the basic model (M) and improved model (MHR) in terms of concentration (3a) and temperature profile in the improved model (MHR) (3b)

The results of the scale-up reactor model for temperature and the concentration changes are presented in Figure 4. As can be seen in Figure 4a, the temperature profile does not reach steady-state after 80hr. This is due to the change in the length from 0.12 to 1m, so the temperature profile has yet to stabilise. Increasing the time for the process will ultimately push the temperature towards a steady-state profile. Also, because of the transience of the temperature profile in this time range, it means that at 10°C, the stable region for hydrate growth will continue to increase over time until a steady state is reached for that temperature. For example, at 10hr, the stable region barrier for 10°C is at 0.0325m. At 40hr, the barrier is at 0.065m, and at 80hr, the barrier is at 0.095m.

As can be seen in Figure 4b, the reactor model can be used to estimate the amount of CO₂ gas uptake in the hydrate layer. The maximum concentration uptake in the hydrate layer is approximately 380 mol/m³. The volume of the reactor has been stated as 1000m³. Therefore, about 380 kmol (or 16.7 tonnes) of CO₂ are consumed in the hydrate layer after 80hr. 16.7 tonnes per 80hr is much lower than what industry wants for CO₂ capture (typically in the 100's of tons per day). Therefore, in order to increase the rate and capacity of CO₂, considerations about increasing the pressure of the system and the total volume of the reactor to accommodate for higher gas uptake in the hydrate layer. However, researchers seek to find ways to reduce the operating pressure as it entails high costs. This reactor model can be presented as a resource or guideline for anticipating the maximum capacity of gas uptake for set parameters.



(a) (b)
 Figure 4- Temperature profile in the scale-up reactor model (4a) Concentration uptake in the hydrate layer in the scale-up reactor model at depth of 1m (4b)

4. Conclusions

Two models were developed for the HBCC process at the reactors scale. First, a basic model was derived, ignoring the reaction rate to estimate mass and heat transfer roles in the dissolution phase. It was then improved, taking both dissolution and hydrate formation into account. The results indicated that mass transfer was the rate-controlling mechanism as reactor temperature reach a steady value in a shorter time. The comparison between the two models revealed that the consumption/reaction term has a significant influence on the reactor profiles. It was shown that the improved model is more capable in predicting realistic HBCC reactor. The model was applied for batch operation by using a time-dependant boundary condition for the bulk gas. Finally, the model was scaled-up utilising all the improvements and knowledge from the previous case analysis. With the set parameters such as volume, time and length, this model can be utilised to evaluate the position of the hydrate layer boundary and estimate the total gas uptake in the system.

References

- H. Dashti, D. Thomas, A. Amiri, Modeling of hydrate-based CO₂ capture with nucleation stage and induction time prediction capability. *Journal of cleaner production*, 231(2019), pp.805-816.
- H. Dashti, D. Thomas, A. Amiri, X. Lou,. Variations of the shrinking core model for effective kinetics modeling of the gas hydrate-based CO₂ capture process. *Computer Aided Chemical Engineering* 46(2019), pp. 1687-1692. Elsevier.
- H. Dashti, X. Lou, Gas Hydrate-Based CO₂ Separation Process: Quantitative Assessment of the Effectiveness of Various Chemical Additives Involved in the Process. *Energy Technology* (2018), pp. 3-16. Springer.
- H. Dashti, L. Z. Yew, X. Lou, Recent advances in gas hydrate-based CO₂ capture. *Journal of Natural Gas Science and Engineering*, 23(2015), pp.195-207.
- P. Linga, M.A. Clarke, A review of reactor designs and materials employed for increasing the rate of gas hydrate formation. *Energy & Fuels*, 31 (1)(2017), pp.1-13.
- Y. H. Mori, On the scale-up of gas-hydrate-forming reactors: The case of gas-dispersion-type reactors. *Energies*, 8(2)(2015), pp.1317-1335.
- M. R. Riazi, A new method for experimental measurement of diffusion coefficients in reservoir fluids. *Journal of Petroleum Science and Engineering*, 14(3-4)(1996), pp.235-250.
- S. Zarghami, F. Boukadi, Y. Al-Wahaibi, Diffusion of carbon dioxide in formation water as a result of CO₂ enhanced oil recovery and CO₂ sequestration. *Journal of Petroleum Exploration and Production Technology*, 7(1)(2017), pp.161-168.

SMITH: differential homotopy and automatic differentiation for computing thermodynamic diagrams of complex mixtures.

Olivier Cots^a, Joseph Gergaud^a and Nataliya Shcherbakova^{b,*}

^a*Toulouse Univ., INP-ENSEEIH-IRIT, UMR CNRS 5505, 2 rue Camichel, 31071 Toulouse, France*

^b*Laboratoire de Génie Chimique, UMR5503 CNRS, Université de Toulouse, ENSIACET INPT/UPS, BP 84234, Toulouse, F-31432, France*
nshcherb@ensiacet.fr

Abstract

This paper presents SMITH, the working prototype of a new code which significantly improves the accuracy of thermodynamic diagrams thanks to the highly performant numerical technologies including differential homotopy and automatic differentiation.

Keywords: distillation diagrams, univolatility curves, differential homotopy, automatic differentiation

1. Introduction

Extraction of singular compounds from multi-component mixtures through the separation process is the most commonly used technology in chemical, pharmaceutical and food-processing industries (Petlyuk, 2004). Analysis of the thermodynamic diagrams describing the fundamental physical properties of complex mixtures is a crucial step in the conceptual design of industrial process involving separation technologies. In very many cases the topology of these diagrams is non-trivial even for ternary mixtures. In particular, this concerns the computation of certain characteristic curves like univolatility curves, distillation profiles, the phase separation curves, the pinch-point curves, etc. Different types of diagrams are accessible via professional packages, but often they are incomplete or lack in precision, as in the example shown in Fig. 1 about univolatility curves used when designing extractive distillation process.

The insufficiency of the professional softwares creates serious obstacles for the technological transfer from academic research to industrial implementation, so the search of new algorithmic solutions to improve computational efficiency are important for the process engineering community.

The major part of existing algorithms to compute characteristic curves of thermodynamic diagrams use the Newton-Raphson like procedures over the composition space. This implies an important computational effort to access the acceptable precision and requires an a-priori knowledge of the topology of the diagram in order to provide a good initial approximation. On the other hand, algorithms based on the differential homotopy principle provide a very promising alternative to the Newton-Raphson method for path

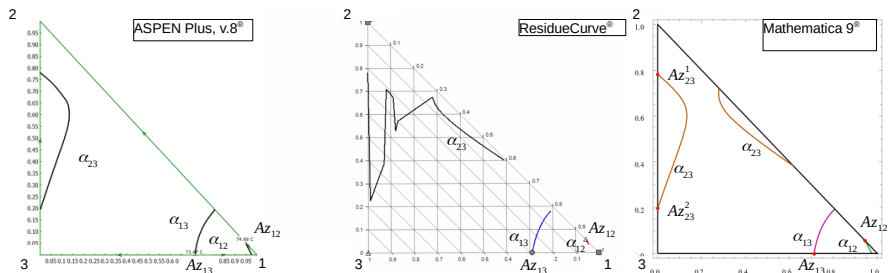


Figure 1: Hexane (1) - Benzene (2) - Hexafluorobenzene (3) at 1.2 atm (generalized NRTL). Univolatility curves computed by ASPEN Plus v.8[®], ResidueCurve[®], and Mathematica 9[®]. Here α_{ij} are the univolatility curves of i and j compounds, Az_{ij} are binary azeotropes. Both professional codes fail to identify the pair of binary azeotropes Az_{23} , ASPEN Plus[®] ignores the azeotrope-free branch of the curve α_{23} . ResidueCurve[®] lacks in precision.

following computations (Allgower and Georg, 2003; Caillau et al., 2012). In process engineering, this approach was proposed by several authors to compute distillation profiles, pinch point curves, univolatility curves (Poellmann and Blass, 1994; Feldbab, 2012; Skiborowski et al., 2016) and phase separation boundaries (Deiters, 2016). Although these codes showed excellent results in comparison with the Newton-Raphson based analogs, still they form a marginal branch in the process engineering computing. The main reason of this poor success is the access to the derivatives of the thermodynamic models required by these methods. Indeed, using analytical expressions of the model and its derivatives significantly increase precision and numerical stability of computations, but their implementation is technically difficult and may form an additional source of error.

In this paper we present SMITH (Separation of Mixtures In Thermodynamics by Homotopy) (SMITH, 2020), a working prototype of a new code allowing highly precise and efficient computation of different types of curves of ternary diagrams. In contrast with codes cited above SMITH realizes the differential homotopy algorithm in coupling with automatic differentiation of the thermodynamic models. Such a coupling allows to easily implement a large class of thermodynamic models, and access the high numerical performance with less computational effort. The current version of SMITH is limited to compute the univolatility curves of ternary mixtures, without an a-priori knowledge of the topological structure of the underlying VLE diagram and independently of the presence of azeotropes. The same computational approach can be developed for many other types of thermodynamic diagrams.

In Section 2 we briefly recall the geometric model of the univolatility curves introduced in Shcherbakova et al. (2017). Section 3 describes the main features of the SMITH code. The paper is concluded with a series of illustration examples.

2. Geometrical model of univolatility curves

Univolatility curves α_{ij} are the sets of points on the ternary residue curve map (RCM) where the pair of compounds i and j have the same relative volatility. They are used as the feasibility indicators in the distillation process design. In extractive distillation, the placement and the terminal points of the univolatility curve on the composition tri-

angle edges determine the order of the withdrawn products as well as the proper column configuration (Gerbaud and Rodriguez-Donis, 2014).

A ternary RCM may contain up to three families of univolatility curves according by their respective index, even in zeotropic case. The topology of univolatility curves is intrinsically related to the topology of the underlying RCM (Kiva et al., 2003; Shcherbakova et al., 2017). Indeed, three univolatility curves of different index intersect at ternary azeotropic points, and each binary azeotrope gives rise to a univolatility curve. Multiple binary and ternary azeotropy can occur. Note that the presence of azeotropes is sufficient but non necessary for the existence of the univolatility curves. A RCM may contain azeotrope-free curves and multiple curves of the same index may coexist. If the topological structure of the RCM is known, the computation of univolatility curves starting at azeotropes is straightforward, while the detection of univolatility curves not associated with azeotropes is a more complicated and time-consuming process, especially in the case of zeotropic mixture.

Consider an open evaporation of a ternary mixture kept at thermodynamic equilibrium at constant pressure. Denote $x_i, y_i, i = 1, 2, 3$ the mole fractions in the liquid and in the vapor phases, and T the temperature of the system. In the absence of chemical reactions a two-phase ternary mixture has three independent state variables. By choosing x_1, x_2 and T , the complete state space of the system is $M = \{(\bar{x}, T) : T \in [T_{min}, T_{max}], \bar{x}_i \in \Omega, i = 1, 2\}$, where T_{min} and T_{max} are the minimum and maximum boiling temperatures of the mixture and $\Omega = \{\bar{x} = (x_1, x_2) : x_i \in [0, 1] \text{ with } x_1 + x_2 \leq 1, i = 1, 2\}$ is the composition space parameterized by the mole fractions of the first two compounds, while $x_3 = 1 - x_1 - x_2$. The distribution coefficients $K_i : M \rightarrow \mathbb{R}$ are the functions over M that describe the vapor-liquid equilibrium (VLE) in terms of molar fractions in the vapor and liquid phases: $y_i = K_i(\bar{x}, T)x_i$ for $i = 1, 2, 3$. Since $\sum_{i=1}^3 y_i = 1$, the following equilibrium condition holds:

$$\Phi(\bar{x}, T) = \sum_{i=1}^3 K_i(\bar{x}, T)x_i - 1 = 0. \quad (1)$$

In 3D state space M , Eq.(1) defines a smooth hyper-surface referred as the boiling temperature surface, namely W -surface. Along with W -surface, M contains three univolatility hypersurfaces defined by equations

$$\Psi_{ij}(\bar{x}, T) = K_i(\bar{x}, T) - K_j(\bar{x}, T) = 0. \quad (2)$$

Possible intersections of the W -surface with univolatility hypersurfaces are smooth curves $\Gamma_{ij} \in M$ called the generalized univolatility curves (Shcherbakova et al., 2017). Univolatility curves $\alpha_{ij} \in \Omega$ are the orthogonal projections of the curves $\Gamma_{ij} \in M$ to the composition space Ω . Since $\nabla\Phi$ and $\nabla\Psi_{ij}$ define the normal vector fields to the 2D surfaces defined by Eqs.(1, 2), the vector field $U^{ij} = \nabla\Phi \times \nabla\Psi_{ij}$ is tangent to the generalized univolatility curve Γ_{ij} . In other words, the curve Γ_{ij} is a solution to the following system of ordinary differential equations in M :

$$\dot{x}_1 = U_1^{ij}(x_1, x_2, T), \quad \dot{x}_2 = U_2^{ij}(x_1, x_2, T), \quad \dot{T} = U_3^{ij}(x_1, x_2, T). \quad (3)$$

In order to compute generalized univolatility curves it is enough to detect their end-points over the border of the composition triangle and use them as the initial points for the numerical integration of Eq.(3). This can be done solving Eqs.(1, 2) on the boundary of

the composition triangle $\partial\Omega$ in the reduced 2D space. For the sake of completeness we remark, that in some cases α_{im} curve starts from the binary edge i, j of the triangle where m compound is missing (see for instance the α_{12} curve in Fig. 1). In this case, as it was proposed in (Kiva et al., 2003), the distribution coefficient K_m must be replaced by the distribution coefficient at infinite dilution, which can be obtained from the ternary distribution coefficient by setting $x_m = 0$. This generalization enables to compute complete univolatility diagrams independently of the presence of azeotropes.

3. SMITH algorithm

3.1. Differential homotopy method

The core of SMITH code uses the differential homotopy method with arc-length parameterization to solve the systems of algebraic equations. This approach is based on the following mathematical result (see in Allgower and Georg (2003) for more details).

Let $F : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^N$, $F(q, \lambda)$, denote the homotopic function where λ denotes the homotopy parameter. Under certain regularity assumptions, the solution to the equation $F(q, \lambda) = 0$ forms a one-dimensional manifold. Indeed, if F is a continuously differentiable function such that

$$F(q_0, \lambda_0) = 0, \quad \text{rank} \frac{\partial F}{\partial \lambda}(q_0, \lambda_0) = N$$

for some q_0 and λ_0 , and if zero is a regular value of F , then a continuously differentiable curve starting from (q_0, λ_0) exists and it is either diffeomorphic to a circle or to the real line. The different branches of $F^{-1}(\{0\})$ form disjoint smooth curves.

As we showed in Section 2, the generalized univolatility curve Γ_{ij} is a "path of zeros" of the function $F_{ij} = (\Phi(q, \lambda), \Psi_{ij}(q, \lambda))$ with $q = (x_1, T)$ and $\lambda = x_2$. SMITH uses a predictor-corrector algorithm via the `nutopy` package with a high order step-size control Runge-Kutta scheme for the prediction, and with a classical simplified Newton method for the correction. The key point of success is the efficient computation of the Jacobian matrix of the homotopic function F_{ij} , which reduces to the computation of the derivatives of the distribution coefficients K_i , $i = 1, 2, 3$. In SMITH code these derivatives are computed via the automatic differentiation tool `tapeenade` (Hasco and Pascual, 2012), which drastically simplifies numerical implementation of the homotopic method.

3.2. Initial points computation

To start the homotopy, the path-following method needs to be initialized at points (q_0, λ_0) verifying $F(q_0, \lambda_0) = 0$. Such points can be chosen on the border of the composition triangle Ω . In fact, according to the index i, j of the couple of compounds, each edge of $\partial\Omega$ may contain up to three types of extremity points verifying univolatility condition, moreover, several points of the same index may co-exist. Such points may be found by applying a standard Newton-Raphson procedure over the border of the composition triangle. SMITH realizes a much more efficient method, based on the following observation. Any edge I of the triangle $\partial\Omega$ can be parameterized by the mole fraction of one of the compounds $x_i = a \in [0, 1]$, $i \in 1, 2$. Denote $\phi(a, T) = \Phi(\bar{x}, T)|_{\bar{x} \in I}$, $\psi_{ij}(a, T) = \Psi_{ij}(\bar{x}, T)|_{\bar{x} \in I}$. Then initial points for the curve α_{ij} are zeros of the function $E_{ij} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by

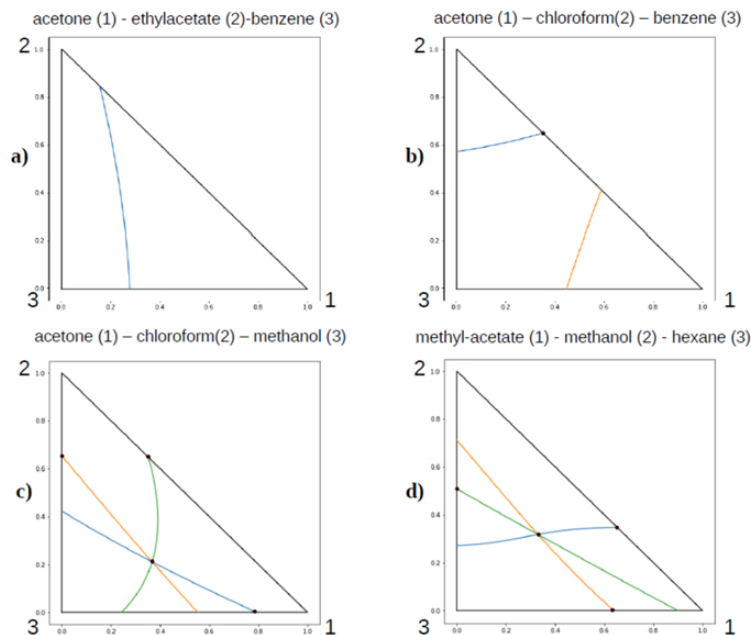


Figure 2: Univolatility diagrams of different Serafimov's VLE topological classes computed by SMITH: a) 0.0-1; b) 1.0-2; c) 3.1-4; d) 3.1-2.

$E_{ij}(a, T) = (\phi(a, T), \psi_{ij}(a, T))$. Due to the uniqueness of the boiling temperature T of a homogeneous mixture, equation $\phi(a, T) = 0$ defines a smooth graph in the 2D plan with coordinates a and T . According to the Implicit Function Theorem, it can be solved in order to express $T = T(a)$. Then the zeros of the function E_{ij} can be detected by finding zeros of the scalar function $\psi_{ij}(a, T(a))$.

SMITH implements the above idea as follows. In order to capture multiple solutions, each edge of $\partial\Omega$ is divided into K sub-intervals $[a_k, a_{k+1}]$, $k = 0, \dots, K - 1$ with $a_0 = 0$ and $a_K = 1$. $T(0)$ is computed by Brent's method in the interval $[T_{min}, T_{max}]$, and going further around the edge, $T(a_k)$ is computed via a standard Newton method using $T(a_{k-1})$ as the initial guess. The sign of $\psi_{ij}(a_k, T(a_k))$ is checked at each k . The change of the sign of ψ_{ij} at k -th step means the existence of a zero of this function in the interval $[a_{k-1}, a_k]$, which can be then found by a standard Newton-Raphson procedure. Realizing such a scheme along every edge of the triangle for all three families of equations (including the equations associated to the compound at infinite dilution) allows to build the complete set of initial points. Observe, that binary azeotropes can be detected a-posteriori by checking the condition $K_i = K_j = 1$ over the computed set of points. The binary bi-azeotropy can be easily detected.

3.3. Examples

The actual version of SMITH code allows computations using NRTL and UNIQUAC models for activity coefficients, and Antoine's and DPPR equations for vapor pressure

computations at various value of process pressure. Fig. 2 shows four examples of ternary univolatility diagrams of different topological classes (see in Kiva et al. (2003) for Serafimov's ternary VLE classes). They were computed at standard atmospheric pressure using NRTL model for activity coefficients and Antoine's equation for vapor pressure calculation. The black points indicate the position of azeotropes. Diagram a) provides an example of a purely zeotropic mixture (Serafimov class 0.0-1) which has one univolatility curve. Diagram b) contains two univolatility curves of different indexes, one of them is azeotrope-free. These two azeotrope-free univolatility curves cannot be detected by ASPEN Plus v.8[®]. Cases c) and d) provide examples of diagrams of three curves of different index, the essential difference is the type of ternary azeotrope: a saddle in c) and an unstable node in d) cases. The intersection points of the uni-volatility curves were used to find the ternary azeotropes.

4. Further perspectives

The current version of SMITH (SMITH, 2020) works on Unix systems. We are no working on the Windows version. Although the current version implements only the standard NRTL model, other thermodynamical models are under development. Soon they will be available, as well as a more efficient method of initial points localization using homotopy. In a long term perspective, we plan to develop new applications to compute phase separation boundaries, pinch point curves, etc. The described mathematical and numerical approach can be easily adapted to these types of computation.

References

- E. Allgower and K. Georg. *Introduction to numerical continuation methods. Classics in Applied Mathematics, vol. 45.* SIAM, Philadelphia, USA, 2003.
- J.-B. Caillaud, O. Cots, and J. Gergaud. Differential pathfollowing for regular optimal control problems, optimization methods and software. *Optimization Methods and Software*, 27(2):177–196, 2012.
- U. Deiters. Differential equations for the calculation of fluid phase equilibria. *Fluid Phase Equilibria*, (428): 164–173, 2016.
- N. Feldbab. An efficient method of constructing pinch point curves and locating azeotropes in nonideal distillation systems. *Ind. Eng. Chem. Res.*, 51(20):7035–7055, 2012.
- V. Gerbaud and I. Rodriguez-Donis. *Extractive distillation chap. 6. "Distillation Book", Vol. II Distillation: equipment and processes.* Elsevier, Amsterdam, 2014.
- L. Hasco and V. Pascual. The tapenade automatic differentiation tool: principles, model, and specification. *Rapport de recherche RR-7957, INRIA*, 2012. URL <http://tapenade.inria.fr:8080/tapenade/>.
- V.N. Kiva, E.K.Hilmen, and S.Skogestad. Azeotropic phase equilibrium diagrams: a survey. *Chem. Eng. Sc.*, 58:1903–1953, 2003.
- Mathematica 9. URL <http://www.wolfram.com/mathematica/>.
- nutopy. URL <https://ct.gitlabpages.inria.fr/nutopy/>.
- F.B. Petlyuk. *Distillation Theory and its Application to Optimal Design of Separation Units.* Cambridge University Press, New York, 2004.
- P. Poellmann and E. Blass. Best products of homogeneous azeotropic distillations. *Gas Sep. Purif.*, 8(4): 194–228, 1994.
- ResidueCurve. URL <http://www.prosim.net>.
- N. Shcherbakova, I. Rodriguez-Donis, J. Abildskov, and V. Gerbaud. A novel method for detecting and computing uni volatility curves in ternary mixtures. *Chem. Eng. Sc.*, 173:21–36, 2017.
- M. Skiborowski, J. Bausa, and W. Marquardt. A unifying approach for the calculation of azeotropes and pinch points in homogeneous and heterogeneous mixtures. *Ind. Eng. Chem. Res.*, 55:6815–6834, 2016.
- SMITH, 2020. URL <https://to-c-ce.gitlab.io/smith/>.

Dynamic modeling of a three-phase gas-solid-liquid fluidized bed absorber for CO₂ capture

Ana-Maria Cormos*, Flavia-Maria Ilea, Simion Dragan

*Babes-Bolyai University, Faculty of Chemistry and Chemical Engineering,
Arany Janos 11, RO-400028, Cluj-Napoca, Romania*

Abstract

Global warming and climate change produced by CO₂ emissions is an important issue today. Carbon capture represents a promising option of reducing CO₂ emissions and allows the continuation of fossil fuels usage for at least a short to medium period of time. This work aims to bring to light an innovative CO₂ capture system that relies on a process intensification technique. The system proposed is a three-phase fluidized absorber, in which the low density inert packing is fluidized by the flow of gas (rich in CO₂) which is flowing in counter current with the liquid phase. A detailed dynamic mathematical model is developed in order to evaluate the post-combustion CO₂ capture process using alkanolamine-based solvents in a gas-solid-liquid capture system. The phenomena which take place in the fluidization column is described by both mass and energy balance equations and includes mass transfer models, effective mass transfer area correlations and hydrodynamic process modeling. The simulation results show the intensification of gas-liquid transfer through the fluidization process, the effective mass transfer area being 8 - 10 times larger than in the case of packed beds. In addition to this the CO₂ molar flow transferred through the gas-liquid interface is up to 8 - 12 times higher than conventional packed bed column.

Keywords: CO₂ capture, three phase fluidized bed, mathematical modelling.

1. Introduction

Carbon capture represents a promising option of reducing CO₂ emissions and allows the continuation of fossil fuels usage for at least a short to medium period of time. The gas-liquid absorption can be considered one of the most attractive, commercially mature and efficient methods for carbon dioxide capture in an attempt to reduce global climate change. Mono-ethanol-amine (MEA) is the most widely used solvent for CO₂ absorption (Rochelle, 2009), and CO₂ capture by MEA in packed bed absorption columns is one of the most common CO₂ capture technologies. Post-combustion CO₂ capture using solvents based on the conventional technology requires very large packed columns. An innovative three phase fluidized design (gas-solid-liquid) would be an interesting and potential promising option to treat high amount of flue gas usually coming from power plants (Dragan 2016). This fluidized system has many advantages, from mass transfer point of view: low pressure drops in the column, very high interfacial contact area per unit volume of the column and capacity to process a large volume of gases. Also, the solid packing is easily handled and can be removed from the column with pneumatic transport, it does not need special expertise and it can easily be made from chemical resistant plastics with low capital cost of equipment.

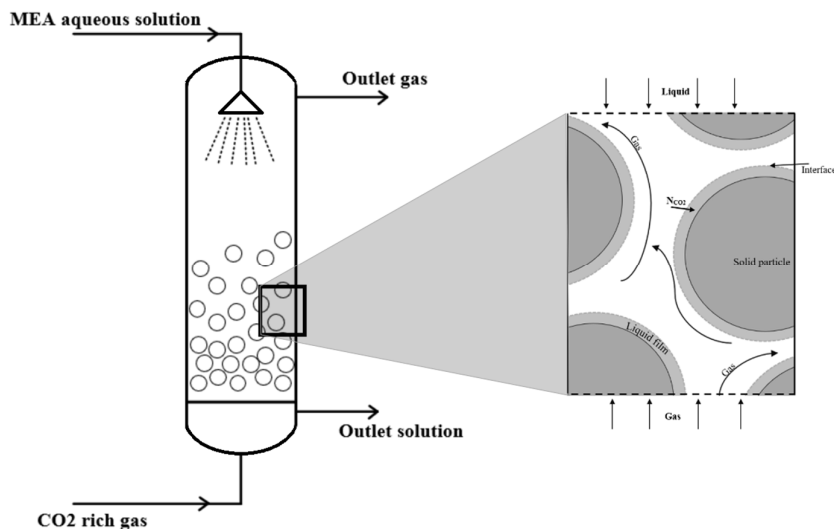


Figure 1. Mass transfer model for the gas-solid-liquid absorption system

The solid phase has no chemical effect on the process, but the intensive mixing of the solid packing in the column determines high turbulence and therefore enhances the mass transfer and increases it significantly comparative to conventional fixed packed beds. The continuous movement of inert solid particles leads to the perpetual renewal of the liquid film that forms on the solid particles and gives an important increase in the effective mass transfer area (Figure 1).

2. Mathematical model

In this work, a dynamic mathematical model is developed in order to evaluate the CO₂ capture process into three phase fluidized bed column. The developed model constitutes in equations that describe the complex nature of the process, referring not only to the mass transfer between the gas and the liquid phase, but also to the hydrodynamics and kinetics of absorption of CO₂.

In terms of the kinetic model for the chemical reaction between MEA and CO₂, the model uses the equations and data revised by Versteed et al (1996). The reaction between carbon dioxide and primary amines, according to the zwitterion mechanism, has an overall order of two, with a value of one with respect to both CO₂ and the amine. Hence, for the process described in this work, the reaction rate can be written as:

$$r = k \cdot C_{MEA} \cdot C_{CO_2} \quad (1)$$

$$k = 4.4 \cdot 10^8 \cdot e^{\frac{-5400}{T}} \quad (2)$$

The mass transfer between the gas phase and the liquid phase can be described using the two-film theory. The resistance to the transfer, concentrated in the thin films of gas and liquid respectively, at the interface, is quantified by the partial mass transfer coefficients described by Billet and Schultes, 1993:

$$k_l = C_l \cdot \left(\frac{g}{\vartheta_l}\right)^{\frac{1}{6}} \cdot \left(\frac{D_{CO_2}^l}{s}\right)^{\frac{1}{2}} \cdot w_{le}^{\frac{1}{2}} \quad (3)$$

$$k_g = C_g * \left(\frac{1}{\varepsilon - h_l}\right)^{\frac{1}{2}} \cdot \left(\frac{a}{s}\right)^{\frac{1}{2}} \cdot D_{CO_2}^g \cdot \left(\frac{w_{ge}}{a \cdot \mu_g}\right)^m \cdot \left(\frac{\vartheta_g}{D_{CO_2}^g}\right)^n \quad (4)$$

An important parameter in the mass transfer model is the effective mass transfer area, which is highly different from both the wetted area and the geometric area of the packing. A literature review shows that there are several correlations available that calculate its value for regular packed bed absorption columns. However, in order to be able to determine the effective mass transfer area obtained in fluidized bed absorption columns, these correlations were adapted based on experimental results (Ilea et al., 2020):

$$a_e = a \cdot e^{-3.8767} \cdot Re_{g,l}^{0.8790} \cdot We_l^{0.2198} \quad (5)$$

Moreover, the mass transfer can be quantified in terms of transferred CO₂ molar flow:

$$N_{CO_2} = a_e \cdot E \cdot k_l \cdot (C_{CO_2}^g - C_{CO_2}^l) \quad (6)$$

Due to the fluidization phenomenon, the mass transfer between the gas and the liquid phase takes place in the fluidized bed height, which is different from the entire height of the column. The model includes also equations that are used to determine the value of the fluidized bed height (Eq 7). Hence, the hydrodynamics of this process are a determining factor when it comes to the efficiency of the absorption.

$$H_{sf} = \frac{H_s^0 \cdot \pi \cdot \frac{d_p^2}{4}}{h_s \cdot s} \quad (7)$$

where $h_s = 1 - h_l - h_g$ (Letzel 1999, Billet&Schultes 1993).

The overall phenomena that take place in the process presented in this work can be described using mass and energy balance equations:

$$F_j^e = F_j^0 \pm v_j \cdot \frac{N_{CO_2} \cdot M_{CO_2}}{\rho_j} \quad (8)$$

$$\frac{dC_{CO_2}^j}{dt} = \frac{F_j^0}{v_j} \cdot C_{CO_2}^{0j} - \frac{F_j^e}{v_j} \cdot C_{CO_2}^j \pm N_{CO_2} - r \quad (9)$$

$$\frac{dC_{MEA}^j}{dt} = \frac{F_j^0}{v_j} \cdot C_{MEA}^{0j} - \frac{F_j^e}{v_j} \cdot C_{MEA}^j \mp N_{MEA} - r \quad (10)$$

$$\frac{dT_j}{dt} = \frac{F_j^0}{v_j} \cdot T_j^0 - \frac{F_j^e}{v_j} \cdot T_j^0 - \frac{\Delta H_r \cdot r}{\rho_j \cdot c_{p_j}} \mp \frac{h \cdot a_e \cdot (T_l - T_g)}{(\rho_j \cdot c_{p_j} \cdot v_j)} \pm \frac{N_{MEA} \cdot \Delta H_{MEA}^v}{\rho_j \cdot c_{p_j}} \quad (11)$$

where j represents the gas/liquid phase, the reaction and vaporization terms are included only in the equations describing the liquid phase.

The operating parameters for fluidized/packed bed system are presented in Table 1.

Table 1. The operating parameters

Model	Liquid flow [m ³ /s]	Gas flow [m ³ /s]	MEA conc. [M]	Column		CO ₂ conc [M]
				Diameter [m]	Height [m]	
Fluidized bed	2.78*10 ⁻⁵	3.25*10 ⁻²	5	0.14	1.1	0.0014
Packed bed					4.36	

3. Results and discussions

In order to be able to analyse the performance of the proposed system, the mathematical model was implemented in Matlab/Simulink and several simulations were conducted to best determine its behaviour. The developed model had been previously validated in terms of hydrodynamics (liquid holdup of fluidized bed absorption column without chemical reaction) and carbon dioxide capture efficiency in fluidized bed system and in packed bed system (for CO₂ capture in MEA aqueous solution), a good correlation ($R^2 > 0.95$) was observed (Gaspar and Cormos, 2011, Ilea et al., 2020).

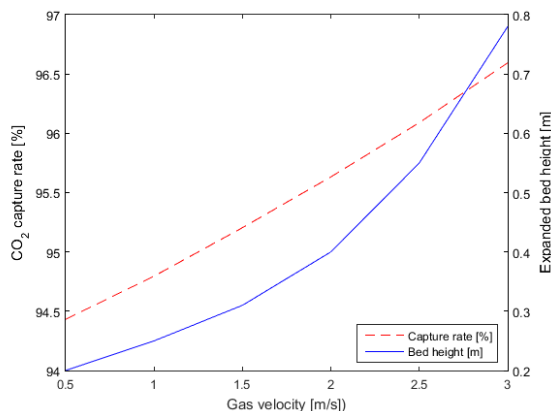


Figure 2. Influence of gas velocity on CO₂ capture rate and expanded bed height

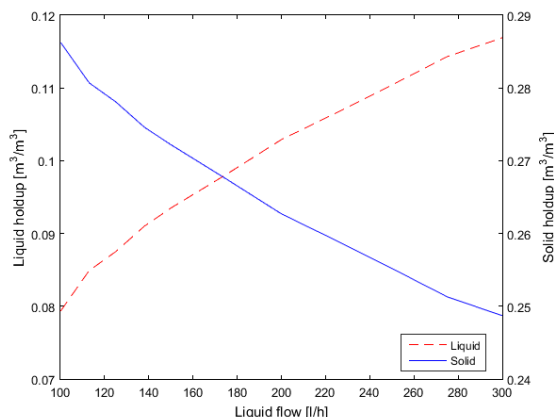


Figure 3. Influence of liquid flow on liquid and solid holdups

The hydrodynamics of the process are highly relevant when trying to achieve better performance. Thus, as seen in Figure 2, the carbon dioxide capture rate increases with the increase of the gas velocity. This dynamic behaviour takes place due to the fact that the increase of the gas velocity favours the fluidization regime and movements of inert solid particles which leads to the renewal of the liquid film and increase in the effective mass transfer area. Moreover, at constant gas velocity, the increase of liquid flow determines an increase in liquid holdup as well as a decrease in solid holdup (Figure 3).

The CO₂ cyclic capacity (quantity of CO₂ absorbed per quantity of solvent) is used to quantify the performance of different capture systems. A comparison between the values of CO₂ cyclic capacity obtained in a pilot scale fluidized bed system and an industrial scale for packed bed absorption column is presented in Figure 4. A 15-20 % increase in the CO₂ absorption cyclic capacity of the system is observed.

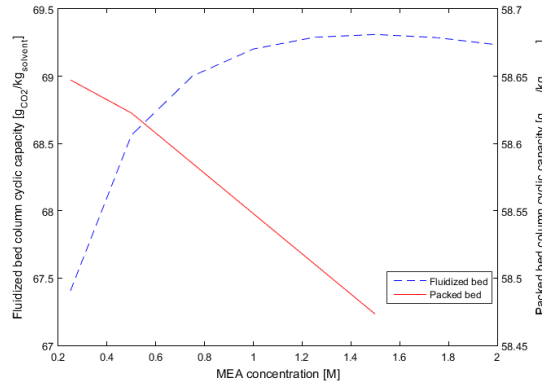


Figure 4. Influence of MEA concentration on CO₂ cyclic capacity

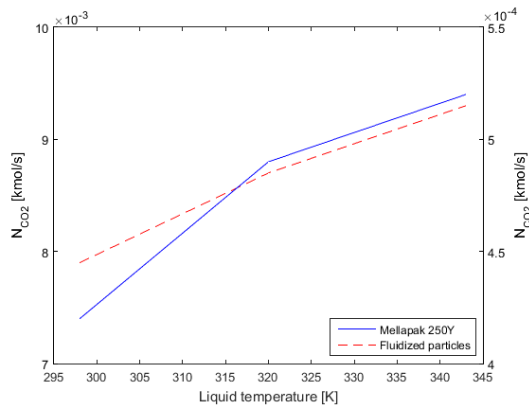


Figure 5. Influence of liquid temperature on CO₂ transferred molar flow

The simulation results show that the three-phase fluidized bed system gives highly improved CO₂ capture capacities compared to regular packed bed columns. The increase of CO₂ absorption efficiency is due to an important increase in the effective mass transfer area, 8 to 12 times higher than packed bed and mass transfer coefficients, k_l and k_g , (3 to 4 times higher). These increased parameters intensify the mass transfer between the gas and the liquid phase, such that the CO₂ molar flow transferred through gas-liquid interface is one order higher in the fluidized bed column's case (Figure 5 and Table 2).

Table 2. Comparison between fluidized bed and packed bed columns

Column	Effective mass transfer area [m ² /m ³]	Transferred flow [kmol/s]
Fluidized bed	1990.70	7.9*10 ⁻³
Packed bed (Mellapak 250Y)	123.26	4.2*10 ⁻⁴

4. Conclusions

A detailed dynamic mathematical model was developed for a three-phase fluidized bed in order to assess the CO₂ capture capacity. The phenomenon which takes place in the fluidization bed was described by mass and energy balance equations and includes mass transfer models, effective mass transfer area correlations and hydrodynamic behaviour modeling. The developed model had been previously validated based on experimental data from a pilot plant.

The simulation results show that the intensification of gas-liquid mass transfer through fluidization process leads to an important increase in the effective mass transfer area (8 - 10 times larger than in the case of packed beds). A three phase fluidized bed absorption not only increases with 8 to 12 times the CO₂ transferred flow between the gas and the liquid, but it also helps achieve a 15-20 % increase in the CO₂ absorption cyclic capacity of the system. This brings essential operational advantages such as larger capacities for CO₂ capture plants.

Acknowledgements

The research leading to these results has received funding from a grant of the Romanian Ministry of Education and Research, CCCDI - UEFISCDI, project number PN-III-P4-ID-PCE-2020-0632, within PNCDI III.

References

- R. Billet, M. Schultes, 1993, Predicting mass transfer in packed columns, *Chem. Eng. and Technol.* 16, 1-9
- S. Dragan, 2016, Calculation of the effective mass transfer area in turbulent contact absorber, *Studia UBB Chemia LXI*, 3, Tom I, 227-238
- J. Gaspar, A.M. Cormos, 2011, Dynamic modeling and validation of absorber and desorber columns for post-combustion CO₂ capture, *Comput. and Chem. Eng.* 35, 2044-52
- F.M. Ilea, S. Dragan, A.M. Cormos, 2020, Assessment of the mass transfer intensification potential for a CO₂ capture process using a three-phase fluidized bed, *Chem. Eng. Process.* 157, 108115.
- M. Kraus, R. Rzehak, 2017, Reactive absorption of CO₂ in NaOH: Detailed study of enhancement factor models, *Chem. Eng. Sci.* 166, 193-209
- H.M. Letzel, J.C. Schouten, R. Krishna, 1999, Gas holdup and mass transfer in bubble column reactors operated at elevated pressure, *Chem. Eng. Sci.* 54, 2237-2246
- N. Ramachandran, R. Ahmed Aboudheir, P. Tontiwachwuthikul, 2006, Kinetics of the absorption of CO₂ into mixed aqueous loaded solutions of monoethanolamine and methyldiethanolamine, *Ind.Eng.Chem.Res.*, 45, 2608-2616
- G.T. Rochelle 2009, Amine scrubbing for CO₂ capture, *Science* 325, p. 1652-1654.
- G.F Versteeg, L.A.J. Van Dijk, W.P.M. Van Swaaij, 1996, On the kinetics between CO₂ and alkanolamines both in aqueous and non-aqueous solutions, *Chem. Eng. J.* 144, 113-158.

Steady and Dynamic State Simulation-based Study of the Coupling of Electrolysis and Methanation Processes within Power to Methane (PtM) Context

Maria Paula Novoa ^a, Camilo Rengifo ^b, Manuel Figueredo ^{c*}

^a *Master in Processes Design and Management. Faculty of Engineering, Universidad de La Sabana, Km. 7 Autopista Norte, Bogotá, Colombia.*

^b *Mathematics, Physics and Statistics Department. Faculty of Engineering, Universidad de La Sabana, Km. 7 Autopista Norte, Bogotá, Colombia.*

^{a, c} *Energy, Materials, and Environment Laboratory (GEMA). Chemical Engineering Department, Universidad de La Sabana, Km. 7 Autopista Norte, Bogotá, Colombia.*

* *Corresponding author: manuel.figueredo@unisabana.edu.co*

Abstract

Through electrolysis and methanation coupling, Power to Methane systems have raised as an attractive alternative for power generation while harnessing renewable sources' potential and reducing carbon dioxide (CO₂) emissions. However, there is a research scarcity in their coupling within a Power to Methane context. This study simulates and analyzes an alkaline electrolysis cell coupled to a methanation process in a Power to Methane context to produce electricity by an open Brayton cycle in both steady and dynamic state. Simulation software is used to analyze the operating conditions and specifications' effects, the effect of the electrolysis power input in a steady-state simulation, and the behavior of a dynamic-simulation with fluctuating renewable energy input for the electrolysis process in two different plant arrangements. Results provide a preliminary glance at the Power to Methane system's behavior finding its contribution to the continuity issues that renewable energy sources have. However, the importance of continuing the research is highlighted to deeply understand the coupling of these processes in order to improve the profitability and performance of a PtM system.

Keywords: *Power to methane, electrolysis, alkaline electrolysis, simulation, Aspen*

1. Introduction

Given the constant changes in a globalized world affected by climate change, energy from fossil sources and polluting gas emissions take on an almost antagonistic role. Hence it is imperative to find solutions to meet the energy demand while contributing to CO₂ emissions reduction through carbon capture technologies and renewable energy sources. A promising alternative that fits these new challenges is the Power to Methane (PtM) technology, in which electrolysis and methanation are coupled for methane production (Boudellal, 2018). This technology aims to harness renewable sources' potential while reducing CO₂ emissions since electrolysis uses renewable energy sources, and methanation contributes as carbon capture sequestration technology. In recent years, electrolysis has shown relevant results in numerous case studies and applications, having considerable investments in projects worldwide (Boudellal, 2018) using technologies such as proton exchange membranes, solid oxide cells, or alkaline solutions (Wang et al., 2020). Thus, it is necessary to study the coupling of these technologies looking to improve the profitability and performance of a PtM system, considering the absence of research in this area. Therefore, this study aims to simulate and analyze an alkaline electrolysis cell (AEC) coupled to a methanation process in a PtM context in both steady and dynamic state to produce electricity by an open Brayton cycle, using two different reactor types for methanation.

2. Model formulation

2.1. Electrolysis model

Alkaline electrolysis (AEC) is the most mature and commercially available technology for this type of process (Boudellal, 2018). It is characterized by commonly using a 25-35% KOH solution and operating at temperatures between 70 and 140°C and pressures up to 30 bar (Sánchez et al., 2018). For its simulation, an algebraic semi-empirical model studied by Sánchez et al. (2018, 2020) was employed. They included the polarization curve, Faraday efficiency, and gas purity in their model, taking current density and operating pressure and temperature as variables. Nevertheless, for this simulation purposes, where higher temperatures will be considered, reversible voltage (V_{rev}) is included as a temperature function shown in Eq.(1), according to Khater et al. (2011).

$$V_{rev} = 1.518 - 1.542 \times 10^{-3}T + 9.523 \times 10^{-5}T \ln(T) + 9.84 \times 10^{-8}T^2 \quad (1)$$

2.2. Methanation reaction and kinetics

A rigorous methanation reactor was modeled considering the Langmuir-Hinshelwood-Hougen-Watson (LHHW) kinetics proposed by Koschany et al. (2016). Simulation parameters were obtained from those implemented by Ortiz et al. (2020) in their analysis within a PtM context.

3. Computational and simulation aspects

AspenTech software was employed to simulate a PtM plant coupling electrolysis and methanation. The electrolysis was modeled in Aspen Custom Modeler® and exported to Aspen Plus®, where additional unit operations (including the methanation reactor) were used. First, a steady-state analysis was performed to find operating conditions and specifications for the unit operations involved in the system. The PtM plant was then exported to Aspen Plus® Dynamics for a dynamic state simulation. However, two different plant arrangements were simulated: ideal and rigorous. The ideal arrangement simulates the methanation process in a Gibbs reactor type, which uses Gibbs free energy minimization to calculate equilibrium and possible reaction products without specifying the reaction stoichiometry or temperature changes (Figure 1). An ideal CO₂ feed is used in this arrangement since other inlet compounds would affect the methanation products. The rigorous arrangement, shown in Figure 2, simulates methanation in a Plug-Flow-Reactor (PFR) type using the catalytic kinetic model mentioned in section 2.2. This arrangement is simulated with a co-current thermal fluid taken from one of the water recirculation streams and a CO₂ feed composition reported by Arenas et al. (2014). Peng-Robinson thermodynamic model was used since its appropriate for methanation (Ortiz et al., 2020) and combustion (Alderetes, 2018).

Figure 1. Ideal PtM arrangement.

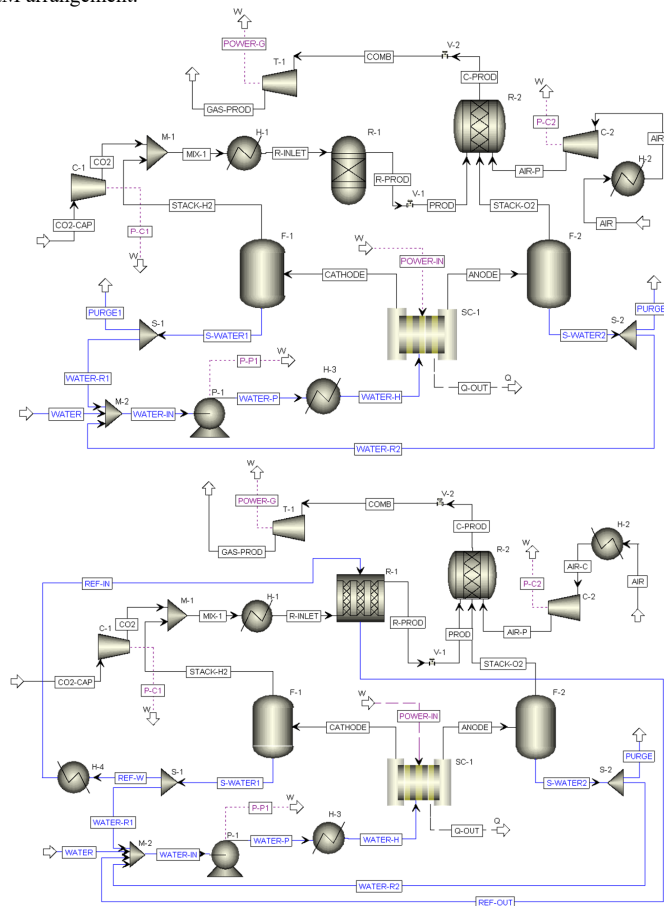


Figure 2. Rigorous PtM arrangement.

4. Results and discussion

4.1. Operating conditions and specifications

4.1.1. Electrolysis

The electrolysis model was analyzed in Aspen Plus to establish the most suitable operating conditions and specifications for the whole PtM plant. Temperature, pressure, power input, number of cells, and cell's area were taken as variables to observe the hydrogen molar flow from cathode output and Faraday's efficiency behavior.

Figure 3 shows the (a) hydrogen molar flow, (b) hydrogen to oxygen fraction (HTO), and (c) Faraday efficiency for different numbers of cells (N) and cell areas at specific operating conditions (100 °C, 7 bar, 10 kW) taken from Sanchez et al. (2020) simulation. Hydrogen molar flow increases as the cell area does, reaching a steady-state value between 0.096 and 0.098 kmol/h for any number of cells between 10 and 20. However, Faraday's efficiency decreases as cell area and cell number increase, the latter having a more significant impact on this variable. Therefore, a considerable hydrogen flow of 0.097 kmol/h, a low HTO fraction (0.018), and an appreciable efficiency of 91% can be obtained using 15 cells with an area of 0.16 m², being this specification appropriate for the cell stack.

Keeping the mentioned cell stack specifications constant, Figure 4 shows the effect of pressure and temperature on the (a) hydrogen molar flow, (b) HTO, and (c) Faraday's efficiency. This analysis shows a lower incidence of pressure (affecting the system when it exceeds 15 bar) and a higher incidence of temperature on these variables: direct for hydrogen flow and proportionally inverse to Faraday's efficiency. Consequently, it is preferable to operate at low pressures to reduce the pump's (P-1) energy requirement and the energy extraction necessary in the phase separators in later stages. Adverse effects and changes in the HTO and efficiency are noted when the temperature exceeds 100 °C, with little increase in the hydrogen flow. Hence, it is preferable to operate at temperatures below 100 °C.

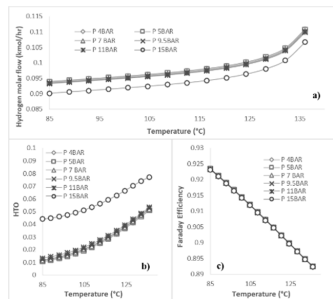


Figure 4. Hydrogen molar flow (a), HTO fraction (b) and Faraday's efficiency (c) at different temperatures and pressure (P).

4.1.2. Methanation and combustion operating conditions

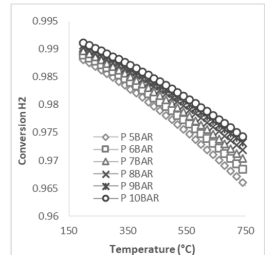


Figure 6. Hydrogen conversion at different temperatures and pressure (P)

For the methanation reactor, hydrogen conversion at different temperatures and pressures is outlined in Figure 6a, showing a decreasing behavior as temperature increases, whereas pressure effects are not sizable at temperatures below 550 °C. Therefore, it is preferred low operating pressures to reduce the compression energy requirement both hydrogen and CO₂ feeds. Hence, reactor operating

temperature and pressure are set to be 226.8 °C and 5 bar, with an H₂/CO₂ ratio of 4, as recommended by Ortiz et al. (2020). A catalytic (*hexaaluminate*) combustion reactor was supposed to reach complete methane conversion since it is suitable for gas turbines and operates at high temperatures (He et al., 2020). Hence, combustion is simulated in a stoichiometric reactor with 100% methane conversion at standard operating conditions of 1230 °C and 15.8 bar (Alderetes, 2018). A 20 % air excess is specified as Eckerlin (2016) suggested for proper methane combustion performance. Table 1 shows the operating conditions for all the units involved in the PtM arrangements simulated.

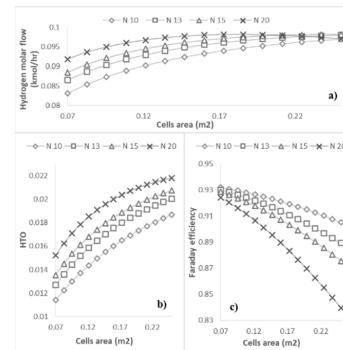


Figure 3. Hydrogen molar flow (a), HTO fraction (b) and Faraday's efficiency (c) at different number of cells (N) and cell area.

Finally, the effect of power input at different temperatures on hydrogen flow (a), HTO (b), and Faraday's efficiency (c) is shown in Figure 5. Hydrogen flow seems to increase significantly as power input does, without strongly affecting HTO fraction nor Faraday's efficiency. However, since power input is intended to be provided by renewable energy sources (e.g., solar panels), necessary arrangements and capacities to maintain more than 10 kW of power in operating times of around 6 hours a day can be costly compared to the hydrogen production that can be achieved in the stack model.

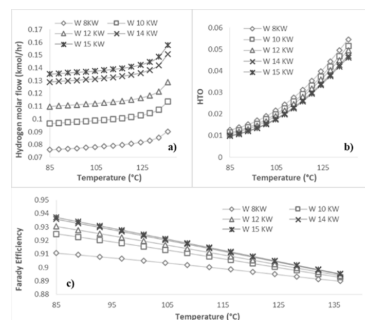


Figure 5. Hydrogen molar flow (a), HTO fraction (b) and Faraday's efficiency (c) at different temperatures and power input (W).

the

for

4.2. Steady-state simulation results

Each stream's temperature, pressure, mass flow, and composition is shown in Table 2 and 3 for ideal and rigorous PtM plants respectively. The results are very similar even for the methane flow produced at the R-1 reactor's outlet, yet the compositions change since oxygen and nitrogen are present from carbon capture. A difference is found in the outlet temperatures since RGibbs keeps it constant while the rigorous reactor considers an increase because of the Sabatier's reaction exothermic nature and a reduction by the heat exchange with the recirculating water stream. For those reasons, a few streams are added to the rigorous arrangement, where the flow of some streams changes (e.g., air, CO₂, and water feed).

Tables 2 and 3 include the power streams as well. It is shown that the turbine's generation is greater than the total power requirement for all compressors and pumps involved, having 0.92 and 1.34 kW of available power for ideal and rigorous plant, respectively. The rigorous plant has more available power because it needs a lower airflow hence a lower power input at C-2, considering that some of the oxygen requirement enters the system from CO₂ capture. The available power is lower compared to the electrolysis power supply; however, PtM systems will serve as an essential contribution to the continuity issues that renewable energy sources (intended to feed the system) have. These fluctuating renewable power sources and their effects on the system cannot be observed in a steady-state simulation (where a value of 10 kW is set), requiring a dynamic-state simulation.

4.3. Dynamic-state simulation results

For renewable energy source supply simulation, POWER-IN is adjusted with sinusoidal ramp tasks to 8 and 10 kW in a time frame of 1 hour, keeping the operating conditions set in section 4.1. The design specifications for the electrolysis water inlet, H₂/CO₂ ratio, excess air, and thermal fluid (the last one only for rigorous simulation) are manipulated using PID controllers.

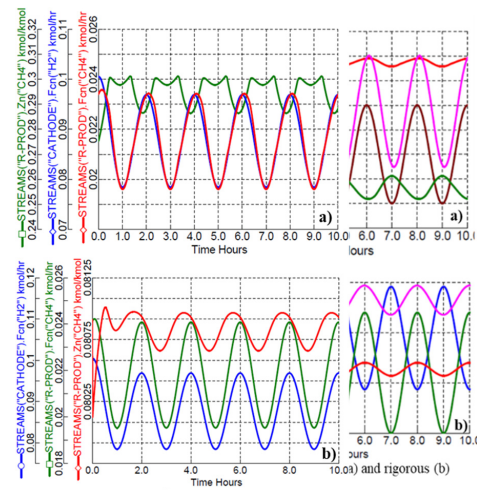


Figure 9. Main products flows and CH₄ molar fraction in ideal (a) and rigorous (b) PtM arrangements

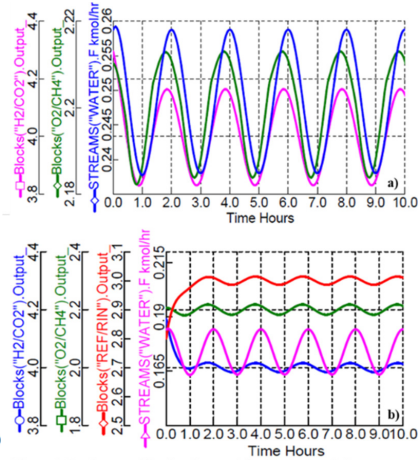


Figure 8. Design specifications' control for ideal (a) and rigorous (b) PtM arrangements

Figure 7 shows the behavior for the renewable power input (POWER-IN), the compressors required power (P-C1 and P-C2), and the power generated by the turbine (POWER-G, which is reported negative since it is leaving the system) for ideal (a) and rigorous (b) PtM arrangement. The incidence of POWER-IN on the rest of the power streams is direct: when the power input increases, the hydrogen flow leaving the cell stack does as well, and CO₂ and air feeds increases because of the ratio controllers. The feeds flow increasing requires more power to maintain the compressors' discharge pressure for the streams to enter the system. For this same reason, these feed ratios are also affected by the input power fluctuations as the system exerts control over them as design specifications. Thus, Figure 8, a) and b) show the design specifications variations with deviations less significant than 6% from the set point for both ideal and rigorous arrangements. Finally, in Figure 9, it is shown that the flows of the main reaction products (hydrogen and methane) obey the direct incidence of power input described previously. However, there is a difference in the methane fraction at the R-1 reactor's outlet from one arrangement to the other since the ideal runs with a pure CO₂ feed, and the rigorous CO₂ feed has the presence of oxygen and nitrogen.

Table 1. Units operating conditions for PtM system simulation

Unit		T (°C)	P(bar)
Heat Exchangers	H-1	226.8	5
	H-2	13.8	1
	H-3	100	5
	H-4	25	5
Separators	F-1	75	5
	F-2	75	5
Methanation Reactor	R-1	226.8	5
Combustion Reactor	R-2	1230	15.8
AEC	SC-1	100	5
Pressure changers		Discharge P (bar)	
Pumps	P-1	5	
	C-1	5	
Compressor	C-2	15.8	
	Turbine	T-1	1
		P drop (bar)	
Valve	V-1		0.3
	V-2		0.3

5. Conclusions

This research compares the simulation results of an AEC coupled to a methanation process (ideal and rigorous) in a PtM context in both steady and dynamic state using two different plant arrangements. This simulation approach contributes to studying and understanding the coupling of electrolysis and methanation technologies for electricity generation by analyzing and setting the operating conditions that favor the PtM system performance. Steady-state simulations' results showed the temperature, pressure, power input, and stack specifications effects in the electrolysis process and how it affects when it is part of a PtM system for power generation. Dynamic-state simulations, considering a fluctuating renewable energy source, showed the direct effects of electrolysis power input on the main reaction products' flows and how it changes the system's power requirement. Steady and dynamic state simulations showed that despite the available power is lower than the electrolysis power supply, PtM systems serve as an essential contribution to the continuity issues that renewable energy sources have and as an attractive carbon capture technology. However, it is essential to enhance the grasp of the coupling of these processes. Different electrolysis technologies might be considered with more rigorous models, improvements in the accuracy of design specifications' control, and Brayton cycle performance.

Table 3. Mass flows, composition, temperature and pressure of steady-state PtM simulation streams – Rigorous arrangement

Stream	T (°C)	P (bar)	Mass Flows (kg/h)	Composition (kg/hr)					
				H2O	H2	O2	CO2	CH4	N2
AIR	25.0	1.0							
AIR-C	13.8		0.94	0	0	0.219	0	0	0.722
AIR-P	474.0	15.8							
ANODE	152.4	5.0	90.77	89.184	0.004	1.583	0	0	0
CATHODE	152.0	5.0	89.38	89.184	0.195	0	0	0	0
C-PROD	1230.0	15.8							
COMB	1230.0	15.5	10.35	1.942	0.004	1.733	1.077	0	5.596
GAS-PROD	747.3	1.0							
MIX-1	201.0	5.0	7.76	0.132	0.195	1.480	1.077	0	4.874
R-INLET	226.8								
CO2	251.6	5.0	7.43	0	0	1.480	1.077	0	4.874
CO2-CAP	25.0	1.0							
R-PROD	535.0	4.7	7.76	1.002	0.001	1.480	0.015	0.387	4.874
PROD		4.4							
PURGE	75.0	5.0	1.34	1.337	1.8E-08	8.1E-05	0	0	0
REF-IN	25.0	5.0							
REF-OUT	37.4	4.7	18.86	18.859	7.3E-06	0	0	0	0
REF-W	75.0	5.0							
S-WATER1	75.0	5.0	89.05	89.052	3.5E-05	0	0	0	0
S-WATER2	75.0	5.0	89.12	89.113	1.2E-06	0.005	0	0	0
STACK-H2	75.0	5.0	0.33	0.132	0.195	0	0	0	0
STACK-O2	75.0	5.0	1.65	0.071	0.004	1.578	0	0	0
WATER	25.0	1.0	3.32	3.316	0	0	0	0	0
WATER-H	100.0	5.0							
WATER-IN	70.2	1.0	180.15	180.145	3.6E-05	0.005	0	0	0
WATER-P	70.5	5.0							
WATER-R1	75.0	5.0	70.19	70.193	2.7E-05	0	0	0	0
WATER-R2	75.0	5.0	87.78	87.776	1.2E-06	0.005	0	0	0
Stream	POWER-G	P-C1	P-C2	P-P1	POWER-IN				
Power(kW)	-2.009	0.474	0.126	0.073	10				
Available Power (kW)		0.044	0.969	0.074	10				
Available Power (kW)									
Available Power (kW)									

References

- C. Alderetes, 2018, *Uso de herramientas de simulación en ciclos combinados*, VI Congreso Argentino de Ingeniería Mecánica.
- C. Arenas, L. Ricaurte, M. Figueredo & M. Cobo, 2014, CO2 capture via barium carbonate formation after its absorption with ammonia in a pilot scale column, *Chemical Engineering Journal*, 254, 220–229.
- M. Boudellal, 2018, *Power-to-Gas: Renewable Hydrogen Economy for the Energy Transition*.
- L. He, Y. Fan, J. Bellettre, J. Yue & L. Luo, 2020, A review on catalytic methane combustion at low temperatures: Catalysts, mechanisms, reaction conditions and reactor designs, *Renewable and Sustainable Energy Reviews*, 119.
- H. Khater, A. Abdelraouf & M. Beshr, 2011, Optimum Alkaline Electrolyzer-Proton Exchange Membrane Fuel Cell Coupling in a Residential Solar Stand-Alone Power System, *ISRN Renewable Energy*, 2011, 1–13.
- F. Koschany, D. Schlereth & O. Hinrichsen, 2016, On the kinetics of the methanation of carbon dioxide on coprecipitated NiAl(O)x, *Applied Catalysis B: Environmental*, 181, 504–516.
- S. Ortiz, C. Rengifo, M. Cobo & M. Figueredo, 2020, Packed-bed and Microchannel Reactors for the Reactive Capture of CO2 within Power-to-Methane (P2M) Context: A Comparison, In *Computer Aided Chemical Engineering* (Vol. 48), Elsevier Masson SAS.

- M. Sánchez, E. Amores, D. Abad, L. Rodríguez & C. Clemente-Jul, 2020, Aspen Plus model of an alkaline electrolysis system for hydrogen production, *International Journal of Hydrogen Energy*, 45(7), 3916–3929
- M. Sánchez, E. Amores, L. Rodríguez & C. Clemente-Jul, 2018, Semi-empirical model and experimental validation for the performance evaluation of a 15 kW alkaline water electrolyzer, *International Journal of Hydrogen Energy*, 43(45), 20332–20345.
- Y. Wang, Y. Du, M. Ni, R. Zhan, Q. Du & K. Jiao, 2020, Three-dimensional modeling of flow field optimization for co-electrolysis solid oxide electrolysis cell, *Applied Thermal Engineering*, 172(January), 114959.
- H. Eckerlin, 2016, The Importance of Excess Air in the Combustion Process", Mechanical and Aerospace Engineering 406 - Energy Conservation in Industry, North Carolina State University.

Upgraded Meta-Learning based Surrogate Selection Paradigm (LEAPS2v2)

Maaz Ahmad,^a Iftekhar A. Karimi^{a,*}

^aDepartment of Chemical & Biomolecular Engineering, 4 Engineering Drive 4, National University of Singapore, Singapore-117585, Singapore

cheiak@nus.edu.sg

Abstract

This work revises and improves our meta-learning based surrogate selection framework, LEAPS2. We perform several modifications to revise LEAPS2 (LEAPS2v2): incorporate noise in response to mimic reality, avoid recommending surrogates that fit noise, add stronger attributes that capture nonlinearities better, propose a novel performance metric that balances surrogate accuracy and complexity, and include larger pool of data sets and surrogate models. Evaluating the recommendation performance of LEAPS2v2 revealed that it successfully recommended at least one of three true best surrogates for more than 94% non-noisy and 89% noisy data sets. It showed large correlation between chances of a surrogate being recommended by LEAPS2v2 and being true best. No specific surrogate accomplished superior performance than others over all data sets. LEAPS2v2 made successful recommendations on a case study involving an industrial compressor.

Keywords: meta-learning, surrogate model, predictive accuracy, surrogate complexity.

1. Introduction

The desire and ability to accurately model complex, nonlinear processes using digital twins has been on the rise. While high-fidelity virtual models offer significant advantages in process analyses and decision-making, they suffer from two main drawbacks. Firstly, it requires rich domain knowledge, expertise, and substantial time to develop robust digital twins. Secondly, it is often computationally prohibitive to use them frequently for real-time analyses. Hence, computationally cheaper surrogate models offer an attractive alternative. They are easy to build by learning the relationships between key inputs and appropriate outputs, using process data. The literature (Queipo et al., 2005) has discussed various surrogate modeling techniques such as Artificial Neural Network (ANN), Support Vector Regression (SVR), Radial Basis Function (RBF), Multivariate Adaptive Regression Spline (MARS), among many others in detail. Identifying an appropriate surrogate is crucial for precise approximation of the system behaviour. One approach for model selection is to employ meta-learning (Rice, 1976), where surrogate selection is guided by accumulating experience of surrogates' performances for diverse data sets. A meta-learning framework is trained by correlating certain features of data or attributes to surrogates' performances, using a learning algorithm. This framework is used for identifying and recommending best surrogates for any data set, based on its learnt knowledge. A recent meta-learning framework, LEAPS2 (Learning based Evolutionary Assistive Paradigm for Surrogate Selection) was developed by Garud et al., (2018). Although LEAPS2 showed good performance in identifying accurate surrogates for a data set, it had some limitations. In this work, we aim to address the limitations of LEAPS2, in addition to upgrading and widening its scope and functionality significantly.

2. LEAPS2 and its Limitations

Garud et al., (2018) developed LEAPS2 (call it LEAPS2v1) for recommending few accurate surrogates, for modeling any data set. It was trained on 264 data sets, generated from 66 test functions, obtained from a virtual library. It extracted 14 attributes and evaluated performances based on prediction error for 25 surrogates. It used regression tree ensembles to learn attributes-performance correlations for each surrogate. Then, it used its stored knowledge for recommending P^* number of surrogates for any given data set. P^* was determined by finding a trade-off between computational burden and success while recommending few surrogates.

LEAPS2v1 had some limitations. First, it was trained completely on non-noisy, synthetic data. Second, attributes such as mean, variance, difference between extreme responses, laid more emphasis in describing and summarizing the response, rather than disclosing the underlying trend and patterns in response which would guide surrogate selection. Third, the performance metric used for evaluating a surrogate's performance was a purely error-based metric. In other words, the complexity of a model was not considered during evaluation. Finally, LEAPS2v1 often recommended as many as $P^* = 5$ or 6 surrogates, which may burden the user and defeat the basic purpose of a recommendation framework.

3. LEAPS2 Version 2 (LEAPS2v2)

Let \mathbf{x} denote N -dimensional inputs x_n ($n = 1, 2, \dots, N$), and $y = f(\mathbf{x})$ denote the response, for any data set. From a pool of M surrogate models, an m^{th} surrogate ($\hat{f}_m(\mathbf{x}), m = 1, 2, \dots, M$) is trained on a set of K sample points $[(\mathbf{x}^{(k)}, y^{(k)}), k = 1, 2, \dots, K]$ to approximate $f(\mathbf{x})$. We revise and upgrade LEAPS2v1, and call our revised framework as LEAPS2v2. LEAPS2v2 has been modified in several ways. First, we consider noise in response and widen the pool of data sets. Second, we remove ineffective attributes and add powerful attributes. Third, we develop and use a novel performance metric that considers both model accuracy and complexity. Fourth, we include many more surrogates as compared to LEAPS2v1. Finally, we revise the recommendation strategy.

3.1. Data Sets

To gather many data sets, we generated synthetic data from different analytical functions having varied characteristic shapes (Multi-modal, Valley, Plate, Bowl and Ridges-shaped). Sobol sampling was used to generate four data sets for each of the following four sample sizes, $K = 50N, 100N, 150N, 200N$. We also added 20 simulation-based data sets pertaining to LNG flash calculations (Coimbatore Meenakshi Sundaram and Karimi, 2021). This gave us 1508 ($93 \times 4 \times 4 + 20$) non-noisy data sets. For noisy data sets, we added gaussian noise with 5% standard deviation, to each non-noisy data set. Moreover, we also included 80 data sets from an industrial compressor and 3 from a machine learning repository (Dua and Graff, 2019), to accumulate 1591 noisy data sets. LEAPS2v2 was trained on data sets with up to 20 input dimensions.

3.2. Attributes

LEAPS2v2 consists of powerful attributes that capture response nonlinearities efficiently. Redundant attributes of LEAPS2v1 that solely summarize the response (such as mean, standard deviations, fluctuations) were removed. LEAPS2v2 has the following attributes:

- Dimensionality (N) of input data.
- Average Absolute Error, Root Mean Squared Error, and Maximum Absolute Error in predictions by 1st order and 2nd order PRSM models, which are simplest models. These attributes indicate the extent of inherent complexity and nonlinearity in the response.

- Degree of Local Nonlinearity (λ): λ gives an idea of local nonlinearity in the neighbourhood of any sample point. To compute λ , we compute the slope angle estimates at sample point k along dimension n , by considering the nearest sample point to k $[NN1(k)]$ and its next nearest neighbour $[NN2(k)]$. Then, we compute degree of nonlinearity ($\lambda_n^{(k)}$) at sample point k along dimension n , as:

$$\lambda_n^{(k)} = \frac{1}{2\pi} \left| \tan^{-1} \left(\frac{y^{(k)} - y^{[NN1(k)]}}{x_n^{(k)} - x_n^{[NN1(k)]}} \right) - \tan^{-1} \left(\frac{y^{(k)} - y^{[NN2(k)]}}{x_n^{(k)} - x_n^{[NN2(k)]}} \right) \right| \quad (1)$$

We calculate the average and standard deviations in $\lambda_n^{(k)}$ over all sample points, in each dimension, to extract $2N$ attributes. We also compute overall local nonlinearity as average of averaged nonlinearities, over all dimensions, as an additional attribute.

- Regional and Inter-Regional Directional Slopes: The directional slope ($m^{(i,j)} = m^{(j,i)}$) between two sample points $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ can be estimated as follows:

$$m^{(i,j)} = m^{(j,i)} = \frac{2}{\pi} \tan^{-1} \left[\frac{|y^{(j)} - y^{(i)}|}{\|\mathbf{x}^{(j)} - \mathbf{x}^{(i)}\|} \right] \quad (2)$$

We create local regions in the input space using k -means clustering technique, where $k = \text{ceil}(K/10)$. For each local region having at least two sample points, regional directional slope is estimated as the average directional slope between all pairs of points within the region. Average and standard deviation of regional directional slopes over all regions constitute two attributes. Inter-regional directional slopes are estimated by computing directional slopes between nearest local regions. Each local region is characterized by its centroid. The response at centroid is approximated by the weighted response of all points within the region. Average and standard deviations in inter-regional directional slopes over the entire input space are two more attributes.

We arranged the aforementioned $(2N + 12)$ attributes in a specific order to ensure proper correlation of attributes with surrogates' performances for any N -dimensional data set. The attributes were ordered as: N , errors after fitting simplest PRSMs, average and standard deviation in regional directional slopes, average and standard deviation in inter-regional directional slopes, overall local nonlinearity, N average local nonlinearities, $(100 - N)$ zeros, N standard deviations in local nonlinearities, and $(100 - N)$ zeros.

3.3. Surrogate Models

LEAPS2v2 has 36 surrogate models, derived from well-known modeling techniques such as Polynomial Response Surface Model (PRSM), Kriging (K), RBF, SVR, ANN, MARS, and Gaussian Kernel Regression (GKR). Each technique provides various mathematical forms such as basis functions, activation functions, kernels, etc. to build a unique surrogate. LEAPS2v2 also consists of additive power law (APL) and multiplicative power law (MPL) models, formed by linear sum and product of power law terms $x_n^{\beta_n}$, $n = 1, \dots, N$, respectively.

3.4. Surrogate Performance Metric

In a quest to evaluate a surrogate's performance based on predictive accuracy as well as complexity, we developed a novel composite performance metric called Surrogate Quality Score (SQS). SQS balances accuracy via a surrogate's coefficient of determination or R^2 -value, and complexity via its degrees of freedom (dof) or the number of independent model parameters. We define and compute SQS_m for $\hat{f}_m(\mathbf{x})$ as:

$$SQS_m = (1 - R_m^2) \times \left\{ \frac{\ln(1 + dof_m)}{\ln(1 + K)} \right\}^{0.5} \quad (3)$$

where, $(1 - R_m^2)$ measures $\hat{f}_m(\mathbf{x})$'s accuracy, while $[\ln(1 + \text{dof}_m)/\ln(1 + K)]^{0.5}$ represents its complexity. Table 1 shows dof_m for the surrogates used in LEAPS2v2.

Table 1: Degrees of freedom for surrogate models used in LEAPS2v2

Surrogate	dof	Surrogate	dof
PRSM	$(O + N)!/(O! \times N!)$ *	GKR, APL	$2N + 1$
Kriging, RBF	K	MPL	$N + 1$
SVR, MARS	$\leq K^{**}$	ANN	$Nn_{HL1} + \dots + n_{HLh}n_0 + n_{HL1} + \dots + n_{HLh}$

* O : Polynomial order; ** dof_m cannot be pre-determined explicitly; *** n_{HLp} : number of nodes in p^{th} hidden layer, h : total number of hidden layers, n_o : number of nodes in output layer

For non-noisy data sets, R_m^2 was evaluated on K sample points of the trained data set, combined with K additional points. However, for noisy data sets, we evaluated R_m^2 on K sample points of training data set alone. This is because real-world data is often limited and does not readily allow additional data sampling. In order to reject surrogates that may fit noise, LEAPS2v2 uses a user-specified limit on maximum allowable R^2 (R_{max}^2) for a surrogate. $R_m^2 > R_{max}^2$ indicates $\hat{f}_m(\mathbf{x})$ overfits, and hence should not be recommended.

3.5. LEAPS2v2 Recommendation Strategy

In order to recommend appropriate surrogates, LEAPS2v2 first requires to learn surrogates' performances (here SQS) over various data sets. Then, for any given data set, it uses its stored knowledge to predict SQS_m for all surrogates, to make recommendations. We used 36 regression tree ensembles (RTEs) to learn the correlations between data attributes and R_m^2 for the 36 surrogates. Since the aim of LEAPS2v2 is to recommend surrogates based on SQS_m , it uses its predicted R_m^2 , and corresponding dof_m (Table 1) to yield predicted SQS_m (Eq.3). However, for surrogates such as MARS and SVR, dof_m cannot be pre-estimated. Hence, predicting SQS_m for such surrogates is not possible unless the respective dof_m is known explicitly. Therefore, we use 5 additional RTEs to learn attributes- dof_m correlations for the 5 SVR and MARS models. This allows LEAPS2v2 to use its predicted R_m^2 and predicted dof_m , to compute its predicted SQS_m , for SVR and MARS as well. Finally, LEAPS2v2 identifies and recommends three surrogates with least predicted SQS_m . For noisy data sets, LEAPS2v2 performs an additional check on its predicted R_m^2 , to reject overfitting surrogates with $R_m^2 > R_{max}^2$.

4. LEAPS2v2 Performance

4.1. Performance Evaluation on Trained and Untrained Data Sets

For any data set, LEAPS2v2's recommendation was flagged as a success, if at least one of its three recommended surrogates were the 1st, 2nd, or 3rd true best surrogates. We define a metric, Degree of Success (DoS) as percentage of data sets for which LEAPS2v2 made successful recommendations. On training LEAPS2v2 on 1130 (75% of 1508) randomly selected non-noisy data sets, it achieved a DoS of 94.7% on trained data sets, and 81.8% on untrained data sets. After training LEAPS2v2 on all 1508 data sets, it showed DoS of 94.2%. For noisy data sets, we used five $R_{max}^2 = \{0.98, 0.96, 0.94, 0.92, 0.90\}$. On training LEAPS2v2 on 1192 (75% of 1591) randomly selected noisy data sets, it achieved $DoS = \{96.6\%, 96.4\%, 96.4\%, 94.6\%, 93.0\%\}$ on trained data sets, and $DoS = \{92.2\%, 90.5\%, 85.0\%, 84.5\%, 84.0\%\}$ on the untrained data sets, for the five R_{max}^2 respectively. Finally, after re-training LEAPS2v2 on all 1591 noisy data sets, it achieved $DoS = \{97.0\%, 94.9\%, 93.8\%, 91.8\%, 89.8\%\}$ for the five R_{max}^2 respectively. DoS deteriorated as R_{max}^2 decreased, as for lower R_{max}^2 , LEAPS2v2 had to reject overfitting models and recommend

true best surrogates. Most data sets for which LEAPS2v2 was unable to make successful recommendations were simple, low dimensional, as such data sets have many surrogates with similar, superior performances. Thus, identifying and recommending one of the top three true best surrogates, while duly rejecting overfitting ones becomes challenging.

We also looked at an alternative way to gauge LEAPS2v2's learning and recommendation capability. We focused on observing, to what extent a surrogate's chance of being recommended correlated with its chance of being one of the 1st, 2nd, or 3rd true best surrogate. In other words, an intelligent framework should recommend true best surrogates more often, and seldom recommend poor ones. Thus, for each surrogate, we observed the number of data sets for which it was one of the top-3 true best surrogates, and number of data sets for which it was recommended. We used Pearson correlation coefficient (*PCC*) to measure the extent of correlation between these two frequencies. LEAPS2v2 showed *PCC* of 91.7% for non-noisy data sets, and *PCC* of {96.8%, 96.3%, 97.0%, 97.3%, 96.2%} for five R_{max}^2 constraints for noisy data sets. High correlations reiterate that LEAPS2v2 can learn and recommend appropriate surrogates efficiently.

4.2. LEAPS2v2 versus LEAPS2v1

A straightforward comparison between LEAPS2v1 and LEAPS2v2 is not fair as the latter has undergone substantial modifications with respect to quality and quantity of data sets, surrogates, attributes, performance metric, and recommendation strategies. However, ignoring these differences, solely looking at respective *DoS* of the two frameworks, LEAPS2v1 achieved a *DoS* of 93% recommending 4 / 25 surrogates, while LEAPS2v2 achieved a *DoS* of 94.2% on more data sets, while recommending 3 / 36 surrogates.

4.3. True Best Surrogates for Non-Noisy and Noisy Data Sets

Observing the true performance of surrogates over various noisy and non-noisy data sets revealed that the true best surrogates for noisy data were very different to those for non-noisy data. This is because in case of noisy data, certain surrogates are prone to overfitting. Fig. 1a depicts the percentage of non-noisy data sets, for which a surrogate is one of the top-3 true best surrogates, while Fig. 1b shows the same for the noisy SDSs. It is evident that ANNs with 2N nodes, K, RBF, and 2nd order PRSMs were preferred for many non-noisy data sets, owing to their combined factors of high accuracies and simplicities. On the other hand, K and RBF were never preferred for noisy data sets as they fit noise at the trained sample points. Rather, MARS, PRSMs, SVRs with polynomial and gaussian kernels, and ANNs were more favoured choices for noisy data sets.

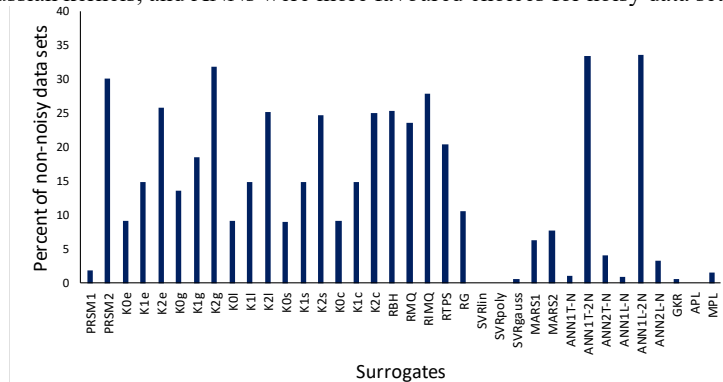


Fig. 1a: Percent of non-noisy data sets for which a surrogate is one of top-3 true best surrogates

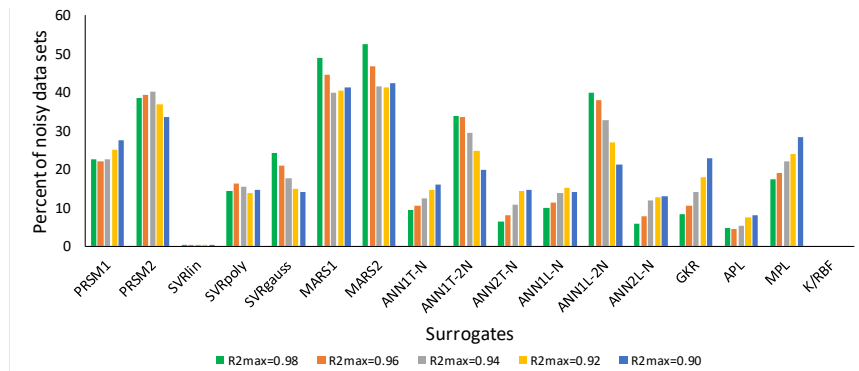


Fig. 1b: Percent of noisy data sets for which a surrogate is one of top-3 true best surrogates

Moreover, no single surrogate stood out as the dominant model for modeling all data sets.

4.4. Performance Evaluation on a Case Study

To further validate the performance of LEAPS2v2, we applied it for modeling volumetric flow across a three stage reciprocating compressor in an LNG (Liquefied Natural Gas) regasification terminal. The flow was modeled for each stage, using suction temperature, suction pressure, discharge pressure, and flow setting capacity as inputs. We considered R_{max}^2 of 0.96 and 0.94. For both R_{max}^2 , LEAPS2v2 recommended one of the top three true best surrogates for stage 1 and stage 3, while two of the top three true best surrogates for stage 2. In other words, LEAPS2v2 achieved 100% *DoS* for all three stages.

5. Conclusions

This work significantly upgrades the scope and utility of our learning-based recommendation paradigm (LEAPS2) through numerous modifications. We incorporate noise in response, update attributes, propose and use a novel composite performance metric that balances surrogate's accuracy and complexity, widen the pool of data sets and surrogates, and revise the recommendation logic. Evaluation of LEAPS2v2 revealed remarkable success in recommending true best surrogates for most noisy and non-noisy data sets. The choice of best surrogates for noisy and non-noisy data sets were very different. We also illustrated LEAPS2v2's successful performance on a case study, involving an industrial reciprocating compressor. LEAPS2v2 clearly proves itself as a smarter, more robust and much refined paradigm for surrogate selection than LEAPS2v1.

References

- Coimbatore Meenakshi Sundaram, A., Karimi, I.A., 2021. State transients in storage systems for energy fluids. *Computers & Chemical Engineering* 144, 107128.
- Dua, D., Graff, C., 2019. UCI Machine Learning Repository [WWW Document].
- Garud, S.S., Karimi, I.A., Kraft, M., 2018. LEAPS2: Learning based Evolutionary Assistive Paradigm for Surrogate Selection. *Computers & Chemical Engineering* 119, 352–370.
- Queipo, N.V., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R., Kevin Tucker, P., 2005. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences* 41, 1–28.
- Rice, J.R., 1976. The Algorithm Selection Problem, in: *Advances in Computers*. Elsevier, pp. 65–118.

Flexibility analysis of chemical processes considering dependencies between uncertain parameters

Christian Langner ^{a,*}, Elin Svensson ^b, Simon Harvey ^a

^a *Division of Energy Technology, Chalmers University of Technology, Gothenburg, Sweden*

^b *CIT Industriell Energi AB, Gothenburg, Sweden*

* *christian.langner@chalmers.se*

Abstract

Chemical processes are usually subject to varying operating conditions. Consequently, the evaluation of the flexibility of a process with respect to variations in inlet conditions is vital to identify bottlenecks in current process flowsheets or new flowsheet design proposals. The flexibility index is a well-established concept to perform flexibility analysis of chemical processes. In this paper, we propose a novel approach to incorporate knowledge regarding dependencies between input parameters when calculating the flexibility index. The aim is to achieve a more accurate indication of a process's flexibility when dependencies between input parameters are present.

Keywords: Flexibility Index, Correlation, Parameter Dependence, Chemical Process Design

1. Introduction

Flexibility analysis of chemical processes has been investigated by the research community since the early 1980s when the Flexibility Index (FI) was introduced by Swaney and Grossmann (1985). The FI indicates the maximum disturbance range within which input parameters may vary while at the same time achieving feasible operation. To be able to solve the FI problem computationally, an active set approach was developed by Grossmann and Floudas (1987) which later was extended to a global solution algorithm by Floudas et al. (2001). Since the active set approach may result in problem formulations which are hard to solve even with state-of-the-art algorithms, alternative strategies to solve the FI problem have been reported in the literature. Li et al. (2015) suggested a framework to calculate the FI by means of an alternating direction matrix embedded in a Simulated Annealing algorithm. Additionally, Kachacha et al. (2018) proposed to conduct Monte Carlo network simulations in the entire expected disturbance range while manipulating degrees of freedom. If sufficiently many operating points are tested, the shape of the feasible uncertainty space is identified. Furthermore, Zhao and Chen (2018) proposed to explicitly calculate the shape of the feasible uncertainty space via cylindrical algebraic decomposition and quantifier elimination. In addition to different strategies to solve the FI problem, Ochoa and Grossmann (2020) extended the traditional FI problem to account also for uncertain input parameters which cannot be measured and therefore recursive actions cannot be taken to respond to variations.

Despite numerous publications on flexibility analysis and on the FI in particular, very little attention has been paid to the influence of dependencies, e.g. correlations, between uncertain input parameters on the flexibility of chemical processes. The reason for this shortcoming is that it is commonly assumed that the correlation between two (or more) input parameters can be expressed via a single algebraic equation, obtained via e.g. linear regression (compare Grossmann and Floudas (1987); Grossmann et al. (2014)). However, single equation (regression) models are only exact if the strongest possible agreement exists between the correlated input parameters (i.e. if the Pearson correlation coefficient is ± 1). Commonly, correlated input parameters in chemical processes agree only to some extent (correlation coefficient $< |\pm 1|$) which means that single equation (regression) models are only able to capture the trend between these input parameters while operating points which deviate from this trend are neglected (i.e. operating points which are caused by “other” sources of variation). However, these outlying operating points (which cannot be captured via single equation regression models) can be vital for actual plant operation. In this paper, we therefore propose a strategy to account for correlations between input parameters when calculating the FI by means of the active set approach (compare Grossmann and Floudas (1987)) which goes beyond single equation regression models.

2. Flexibility Index

As mentioned in the introduction, the FI indicates the maximum disturbance range in which inlet conditions may vary while at the same time achieving feasible operation. The feasible region of a chemical process can be described by the following set of constraints:

$$\begin{aligned} h_i(d, x, z, \theta) &= 0, i \in I \\ g_j(d, x, z, \theta) &\leq 0, j \in J \end{aligned}$$

where d is the vector of design variables, x corresponds to the state variables, z is used for the control variables or degrees of freedom and the varying inlet conditions or uncertain parameters are depicted by θ . In this context, feasibility is achieved when all constraints $i \in I$ and $j \in J$ are satisfied at the point of operation. In the case of independent input parameters, variations in these parameters can be interpreted as a hyperrectangle in the space of the input parameters. The FI is then defined as the ratio between the largest scaled hyperrectangle within the feasible region and the hyperrectangle corresponding to the expected variations. Therefore, feasibility over the expected uncertainty space is guaranteed if the FI is larger than or equal to 1. A visualization of the FI for an example is discussed in Section 3 and is shown in Figure 2.

2.1. Flexibility Index for dependent input parameters

The largest scaled hyperrectangle within the feasible region can be described by the following set of equations:

$$T(\delta) = \{\theta_{i,N} - \delta\Delta\theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta\Delta\theta_i^+\} \forall \theta_i \in \theta \quad (1)$$

For the case where only some input parameters vary independently while other input parameters show dependencies such as correlating trends to the independent input parameters, the total set of input parameters θ must be divided into independent input parameters θ_{ind} and dependent input parameters θ_{dep} . To express the shape in which the input parameters vary (compare hyperrectangle in independent case), it is necessary to formulate mathematical models for the dependent input parameters, i.e. perform the transformation $\theta \rightarrow \theta_{ind}$. In the following, it is assumed that the dependent input parameters can be expressed by a set of functions M of the independent input parameters. Consequently, the space in which the input parameters vary can be described as follows:

$$T(\delta) = \{ \{ \theta_{i,N} - \delta \Delta \theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta \Delta \theta_i^+ \} \forall \theta_i \in \theta_{ind} \{ f_m(\theta_{ind}) \forall m \in M \} \forall \theta_j \in \theta_{dep} \} \quad (2)$$

In general, different mathematical models are available to express the dependency between dependent and independent parameters. In the following section, it is illustrated how the choice of the mathematical model can influence the FI by comparing the results obtained with a single equation and a dual equation model.

3. Example

Let us consider three different cases for a process with two input parameters, θ_1 and θ_2 , operating at the nominal point $(\theta_{1,N}, \theta_{2,N}) = (10, 35)$. In the first case, θ_1 and θ_2 vary completely independently in the range $(\pm 5, \pm 20)$, while in the second and the third case, a significant correlation (positive and negative) between θ_1 and θ_2 is assumed. Figure 1 shows the distributions of θ_1 and θ_2 for each case. It should be noted that for the correlation cases, the variation range is similar to the independent case: $(\Delta\theta_{1,max}, \Delta\theta_{2,max}) = (\pm 5, \pm 20)$. Furthermore, assuming different cases with different distributions is highly theoretical since in reality a dependency between input parameters either occurs or does not occur.

Let us further consider that θ_1 and θ_2 are the input to a process whose feasible region can be described by a set of given constraints (compare Eq. 3a to 3f) where x_1 denotes an additional state variable. For the independent case, the FI can be calculated directly and is equal to 0.36. The graphical interpretation of the FI for the independent case is shown in Figure 2. Since all six constraints are linear, the feasible region is convex and the FI is determined by one (or several) corner points of the largest scaled rectangle inscribed in the feasible region (i.e. one or several corner points of the scaled rectangle intercept with the boundaries of the feasible region). The proportions (length-to-width ratio) of this scaled rectangle are given by the rectangle describing the expected variations. As mentioned in Section 2, the FI can be interpreted as the ratio of the size of the largest scaled rectangle compared to the size of the rectangle describing the expected variations.

As explained in Section 2.1, correlations between input parameters can be expressed using mathematical models. Two approaches have been investigated and are presented in the following two subsections.

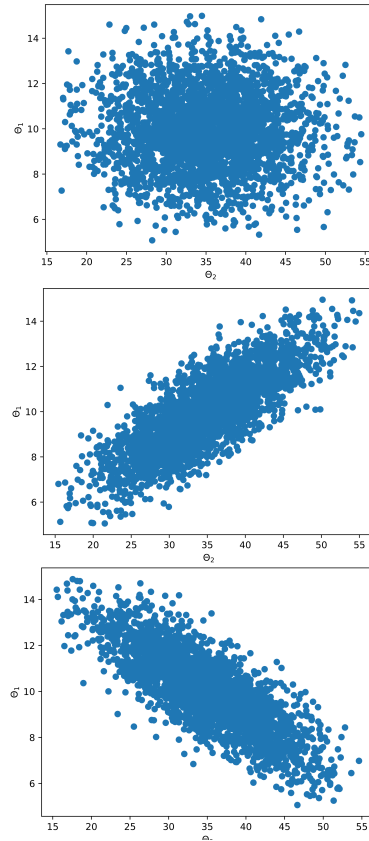


Figure SEQ Figure * ARABIC 1: Distribution of operating points for three different cases (top) independent variation (middle) positive correlated variation (bottom)

$$\begin{aligned} \theta_1 - 2x_1 &= 0 & (3a) \\ 2x_1 - \left(40 - \frac{2}{3} * \theta_2\right) &\leq 0 & (3b) \\ 10 - 0.25\theta_2 - \theta_1 &\leq 0 & (3c) \\ \theta_1 - (0.5\theta_2 + 5) &\leq 0 & (3d) \\ 2.5 + 0.05\theta_2 - 2x_1 &\leq 0 & (3e) \\ \theta_1 - (17.5 - 0.05\theta_2) &\leq 0 & (3f) \end{aligned}$$

3.1. Linear regression model

In the first approach, the positive and negative correlations were expressed using single equation models obtained by linear regression. The linear equation models are shown in Figure 3 together with the feasible region of the problem. Compared to the independent case, the shape which describes the range of the variations (feasible and expected) is no longer a hyperrectangle but a line. The calculated values of the FI are 0.385 (positive correlation) and 0.71 (negative correlation). When using a linear function as model function, the FI is determined by the smallest possible variation for θ_2 at which the model function $f_m(\theta_2)$ intercepts with the boundary of the feasible region. The corresponding points are marked with a star in Figure 3.

From Figure 3 it can be inferred that the used linear models are able to capture the trend between θ_1 and θ_2 while neglecting operating points which deviate from this trend. However, as mentioned in Section 1, these operating points can be vital for actual process operation. To be able to include these operating points, a second approach was studied.

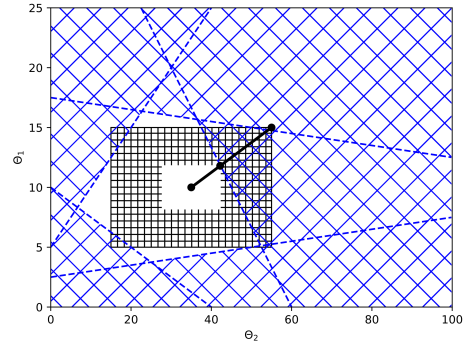


Figure SEQ Figure * ARABIC 2: Visualization of the flexibility index assuming input parameters and to be independent

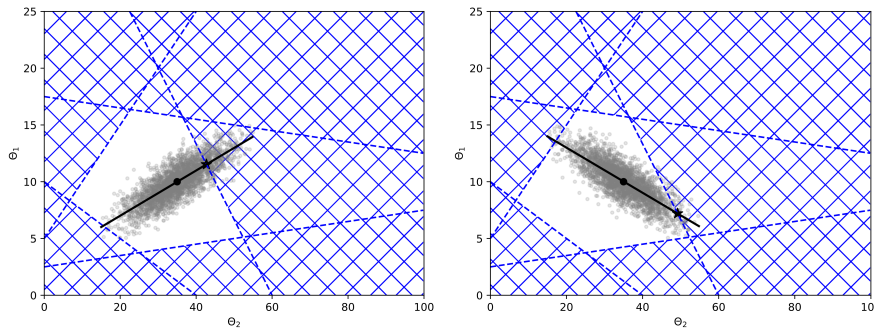


Figure 3: Visualization of the flexibility index when using a linear equation to model (left) positive correlated variations and (right) negative correlated variations.

3.2. Correlation Corridors

Instead of modelling the correlation between the two input parameters θ_1 and θ_2 via a single linear function obtained by regression, an upper boundary function $f_u(\theta_2)$ and a lower boundary function $f_l(\theta_2)$ were defined for each case, yielding a correlation corridor. Practically, this means that θ_1 can take any value between $f_u(\theta_2)$ and $f_l(\theta_2)$ which infers additional uncertainty compared to a single equation model. Different approaches can be followed when defining $f_u(\theta_2)$ and $f_l(\theta_2)$ and a straightforward way is to parallel shift the linear function obtained by regression. This parallel shifting is demonstrated in Figure 4 for the illustrative example and was performed so that 95% of the operating points are in the region between $f_u(\theta_2)$ and $f_l(\theta_2)$. The corresponding values of the FI are 0.27 (positive

correlation) and 0.496 (negative correlation). When using a correlation corridor to model the dependencies between θ_1 and θ_2 , the FI is determined by the smallest possible variation of θ_2 at which either $f_u(\theta_2)$ or $f_l(\theta_2)$ intercepts with the boundary of the feasible region. The corresponding points are marked with a star in Figure 4.

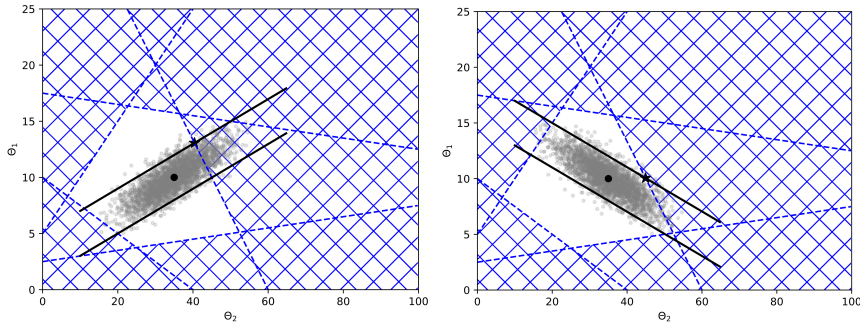


Figure 4: Visualization of the flexibility index when using a correlation corridor to model (left) positive correlated variations and (right) negative correlated variations.

4. Results and Discussion

The results obtained for the FI are summarized in Table 1. It can be seen that the FI obtained with correlation corridors is smaller compared to the FI obtained when using a single equation to model the correlation. This is because the correlation corridor approach leads to a geometric shape describing the range of variations (feasible and expected) that is larger compared to the single equation approach (for the illustrative example: parallelogram vs. line). Figure 3 and Figure 4 show that the geometric shape of expected variations of correlated input parameters (measured operating points) can be emulated more exactly using correlation corridors compared to single equation models. Consequently, it can be implied that the results obtained using correlation corridors give a more accurate indication of a process's flexibility.

Table 1: Flexibility index for different approaches to model correlations between the two input parameters of the illustrative example.

Flexibility index [-]	Linear function	Correlation corridor
Positive correlation	0.385	0.27
Negative correlation	0.71	0.496

Furthermore, the obtained FI can differ substantially depending on the sign of the correlation coefficient. For the presented example, higher values for the FI were achieved when assuming a negative correlation between θ_1 and θ_2 . It should be noted that this observation is highly dependent on the nature of the feasible region. For illustrative purposes, the feasible region was modified by changing the intercept of Eq. 3b from 40 to 50 and the intercept of Eq. 3d from 5 to -2.1. The modified feasible region and the corresponding FI for independent variations of θ_1 and θ_2 are shown in Figure 5. The feasible region was altered in such a way that the FI for the independent case remains 0.36. However, when comparing Figure 5 and Figure 2, it is obvious that in the modified case, a positive correlation would lead to higher FI values. Consequently, no general conclusion can be drawn regarding the influence of the sign of the correlation coefficient on the FI. Generally, it should be noted that in reality a dependency between input parameters of a process either occurs or does not occur. Consequently, assuming different cases and comparing obtained values for the FI is highly theoretical since only one

solution for the FI represents reality (which must include the existence or non-existence of dependencies between input parameters).

In contrast to the presented illustrative example, the dimensionality of the uncertainty space (i.e. number of varying input parameters) of processes in industrial applications is often (much) higher than two. In such situations, it is probable that the dimensionality of dependencies also increases which leads to difficulties when identifying the correlation corridors graphically. A solution approach could be to identify correlation corridors using two-dimensional projections of the multi-dimensional correlation(s). However, this approach may become cumbersome and, thus, further work is necessary to develop automatization strategies for analyzing multi-dimensional dependencies. Furthermore, future work should address the possibility of non-linear dependencies between uncertain parameters and investigate possible impacts on the presented methodology to perform flexibility analysis.

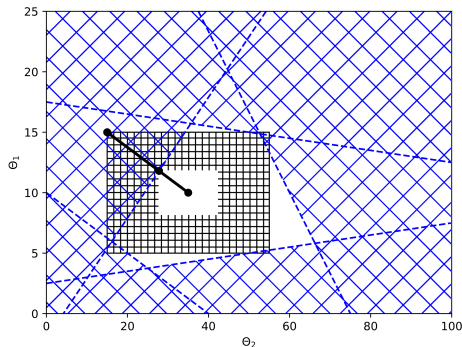


Figure SEQ Figure * ARABIC 5: Visualization of a modified feasible region and the corresponding flexibility (independent case).

However, this approach may become cumbersome and, thus, further work is necessary to develop automatization strategies for analyzing multi-dimensional dependencies. Furthermore, future work should address the possibility of non-linear dependencies between uncertain parameters and investigate possible impacts on the presented methodology to perform flexibility analysis.

5. Conclusion

In this paper, we presented an example to illustrate that if a dependency between (some of) the input parameters is ignored, the obtained FI can be severely wrong, leading to an over- or underestimation of the process's flexibility. In the case of correlations between input parameters, it is necessary to formulate mathematical models in order to include the parameter dependencies in the FI problem formulation. Two approaches were studied, and it was demonstrated that increased accuracy can be achieved by using correlation corridors instead of single equation models if the correlation coefficient is $< |\pm 1|$.

References

- C. A. Floudas, Z. H. Gümüs, M. G. Ierapetritou, 2001. Global Optimization in Design under Uncertainty: Feasibility Test and Flexibility Index Problems. *Industrial & Engineering Chemistry Research* 40 (20), 4267–4282.
- I. E. Grossmann, B. A. Calfa, P. Garcia-Herreros, 2014. Evolution of concepts and models for quantifying resiliency and flexibility of chemical processes. *Computers and Chemical Engineering* 70, 22–34.
- I. E. Grossmann, C. A. Floudas, 1987. Active constraint strategy for flexibility analysis in chemical processes. *Computers and Chemical Engineering* 11 (6), 675–693.
- C. Kachacha, A. Zoughaib, C. T. Tran, 2018. A methodology for the flexibility assessment of site wide heat integration scenarios. *Energy* 154, 231–239.
- J. Li, J. Du, Z. Zhao, P. Yao, 2015. Efficient Method for Flexibility Analysis of Large-Scale Nonconvex Heat Exchanger Networks. *Industrial and Engineering Chemistry Research* 54 (43), 10757–10767.
- M. P. Ochoa, I. E. Grossmann, 2020. Novel MINLP formulations for flexibility analysis for measured and unmeasured uncertain parameters. *Computers and Chemical Engineering* 135.
- R. E. Swaney, I. E. Grossmann, apr 1985. An index for operational flexibility in chemical process design. Part I: Formulation and Theory. *AIChE Journal* 31 (4), 621–630.
- F. Zhao, X. Chen, 2018. Analytical and triangular solutions to operational flexibility analysis using quantifier elimination. *AIChE Journal* 64 (11), 3894–3911.

Computational Characterization of a Pore Network Model by Using a Fast Nitrogen Porosimetry Simulation

*G. Ledezma^{ab}, J.J. Verstraete^a, L. Sorbier^a, D. Leinekugel-Le Cocq^a,
E. Jolimaitre^a, C Jallut^b

^a*IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France.*

^b*Univ Lyon, Université Claude Bernard Lyon 1, CNRS, LAGEPP UMR 5007, 43 boulevard du 11 novembre 1918, Villeurbanne, F-69100, France.*

gabriel-alejandro.ledezma-lopez@ifpen.fr.

Abstract

Nitrogen porosimetry is a characterization technique used to obtain qualitative and quantitative information about the textural properties of porous materials. The physical phenomena exploited in the technique is related to thermodynamic equilibrium and conditioned by pore blocking. The latter phenomenon becomes relevant in highly disordered solids with a wide pore size distribution and a multilevel organization. The aforementioned technique encloses topological information associated with the trend of sorption isotherm. In this work, it is shown how a pore network model can be used to simulate nitrogen sorption process by using an efficient algorithm that combines equilibrium conditions and invasion percolation considerations. Such simulation tool will be the basis for a more comprehensive treatment of experimental characteristic curves and characterization of porous material properties.

Keywords: Digital characterization, Porosimetry, Pore Network Model, Catalyst support, Percolation algorithm, Simulation.

1. Methodology for the Simulation of the Characterization Techniques

The objective of the model is to simulate the evolution of the measured variables during the nitrogen porosimetry characterization experiment. Such an experiment measures, at various nitrogen pressures, the quantity of nitrogen that is present within the porous medium, either adsorbed on the pore walls or filling the pore volume as a liquid phase. An adsorption branch is obtained by increasing the pressure from zero to saturation pressure, and a desorption branch by decreasing the pressure from saturation pressure to zero. If during adsorption, pressure is decreased before reaching saturation pressure, a so-called scanning curve is obtained.

Since the system is equilibrated at every pressure step, a thermodynamic equilibrium model (see 1.1) has to be used to relate the pore size to the volume measured during the sorption process. After generating a geometrical model of the porous medium, it is scanned using an algorithm that determines the desorption sequence of the nitrogen contained in the network pores (see 1.2). The nitrogen sorption simulations can now be performed, both for adsorption and desorption, solving the thermodynamic model equations (see *Algorithm 1*).

```

Read Pore Network Model
Find desorption triggering diameter for each pore
FOR Prelative = 0 to 1
  FOR pore = 1, Npores
    Calculate the quantity of nitrogen inside the pore (Eq.1
    and Eq.2) using the actual diameter (adsorption) or the
    triggering diameter (desorption)
  END FOR
END FOR
Print isotherm branch

```

Algorithm 1. Pseudo-code for the sorption algorithm

The main characteristics of the pore network generation algorithm used in this work were described by Ferreira (Ferreira et al., 2017; Ferreira, 2018). The hypotheses for the sorption simulations are:

- All pores are cylindrical and rigid.
- The adsorption mechanism starts with the adsorption of the nitrogen on the walls followed by sudden condensation.
- The phase change pressure is governed by capillary forces and pore blocking phenomena.
- The meniscus shape is cylindrical for adsorption and hemispherical for desorption.
- The condensed nitrogen needs to be in contact with the vapor phase to evaporate.
- The pore-blocking phenomenon is only dependent on the topology of the system.
- The pore size distribution is within the mesopore range and over 5 nm.
- The system is in equilibrium.
- No other cooperative effects are considered.

1.1. Thermodynamic Model

To model the vapor-liquid equilibrium in a confined medium, the Kelvin-Cohan equation is used, (Equation 1) (Zhang et al., 2006).

$$\frac{p}{p_0} = \exp\left(\frac{-\sigma_{lg} \cdot V_{lm}}{RT} \cdot \left(\frac{1}{r_1(t)} + \frac{1}{r_2(t)}\right)\right)$$

Equation 1. Kelvin-Cohan equation as a function of the two curvature radii of the liquid-gas interface.

In order to calculate the thickness of adsorbed nitrogen, t , we have chosen the Harkins-Jura statistical thickness equation, (Equation 2) (Šolcová et al., 2006).

$$t[\text{Å}] = \left(\frac{13.99}{0.034 - \log_{10}\left(\frac{p}{p_0}\right)}\right)^{\frac{1}{2}}$$

Equation 2. Harkins-Jura statistical thickness equation.

1.2. Pore Blocking Phenomena: Triggering diameter assignment.

Nitrogen adsorption is not supposed to be constrained by the topology of the structure, but nitrogen desorption can be delayed creating a metastable state. Even if the desorption pressure in a given pore is reached, the liquid phase needs to be in contact with the gaseous phase to be able to vaporize. This condition can generate a pore-blocking phenomenon. Indeed, during the desorption step, nitrogen condensed in small

pores can constrain the nitrogen in bigger pores to evaporate at a lower pressure than the one dictated by its vapor-liquid equilibrium condition. Algorithmically, the pore-blocking phenomenon is approached from the external boundary of the pore network as a simultaneous advancing or receding problem. We will define as external pores to those located at this external boundary harboring the liquid-vapor interphase. The pore-blocking is modeled as a sequence of events, each of which corresponds to a given pore. Once the equilibrium constraints are set with Equation 1, the sequence of the events will depend on the pore size and its position within the network. A pore's triggering diameter represents the pressure at which the change of state (vaporization) will happen within the pore. Since the value of the pore's triggering diameter is not necessarily equal to the actual pore diameter, two diameters have to be assigned to each pore: its actual diameter and a triggering diameter for desorption. The actual diameter is used to calculate the quantity of nitrogen subject to the change of state that occurs at the pressure corresponding to the triggering diameter during the desorption process.

To assign a triggering diameter for desorption to each pore, a dedicated assignation algorithm was developed. There are six main rules on which our algorithm is based:

- The triggering diameter (for desorption) is assigned only once to every pore. There is no overwriting of the triggering diameter.
- A pore can be added to the search list only once.
- A node can be visited only once.
- The pore picked up from the search list must always be the pore with the biggest diameter on the list (priority rule).
- The triggering diameters of the external pores correspond to their actual pore diameters.
- The input of the search list must contain all the external pores (boundary conditions).

For the triggering diameter assignation algorithm, we employ four different lists: *visited nodes*, *visited pores*, *searchlist* and *connectivity matrix*. The first two are indexed lists. The third one is a dynamic list handled through binary heaps. The last one is a matrix that contains the information of the pore network.

First, the external pores and the external nodes of the network are identified and listed in *searchlist* and *visited nodes* respectively. *Searchlist* contains the list of pores to be analyzed. The binary heap (max heap) used to handle *searchlist* allows to rapidly find the pore with the biggest diameter, faster than using any kind of sorting algorithm (Masson and Pride, 2014). The pore with the biggest diameter is the first to be taken for analysis and will be called "guide pore". There are two nodes connected to the guide pore: one of them has already been identified and saved in *visited nodes*, so the "new" one (not visited yet) can easily be recognized. The connectivity of the unvisited node is explored to identify the pores connected to the guide pore using the *connectivity matrix*. The newly identified pores will be added to *searchlist*, but first, two actions are needed: a test and the triggering diameter assignation. First, it is verified that the newly identified pores have not been analyzed before by comparing their index to the *visited pores* list. For those that have not been analyzed before, their triggering diameter is obtained using the two following assignation rules:

- If the diameter of the pore is smaller than the triggering diameter of the guide pore, the pore conserves its original diameter as triggering diameter.
- If the diameter of the pore is bigger than the triggering diameter of the guide pore, the pore adopts the guide pore's triggering diameter as its own triggering diameter.

Once the triggering diameter is assigned, the pore is added to *searchlist* and to *visited_pores*. On the other hand, the node is saved in the *visited_nodes* list. Finally, the guide pore is removed from *searchlist*. The loop starts again with a new guide pore selected from *searchlist*. The process is repeated until *searchlist* has been emptied, which means that all the non-isolated pores (pores with a pathway to the surface) have been assigned a triggering diameter. The binary heap created in *searchlist* uses a function for creating the heap, one for *adding* new elements to the heap and one for *eliminating* the guide pore from the heap once it has been treated.

2. Simulation of Nitrogen Porosimetry Curves

The simulation starts from a relative pressure equal to zero. As the adsorption process evolves in two steps, it is necessary to establish a condensation criterion for the cumulated volume of condensed nitrogen. If the diameter of the pore (D_p) is higher than the Kelvin-Cohan equilibrium diameter (D_{eq_n-2t}) for a given relative pressure, then the adsorbed volume in the pore will be equal to:

$$V_{@p_i} = \pi \cdot (D_p - t) \cdot t \cdot l$$

Equation 3. Total adsorbed volume as a function of the adsorbed layer thickness.

Otherwise, it is considered that core volume has already been filled with liquid nitrogen and the condensed volume in the pore is simply equal to:

$$V_{@p_i} = \frac{\pi}{4} \cdot D_p^2 \cdot l$$

Equation 4. Total liquid volume within the pore

For the desorption, as the pore blocking phenomenon exist, the procedure is the same but using the triggering diameter (D_T) of the pore instead of its actual diameter (D_p) for the comparison against the Kelvin-Cohan equilibrium diameter (D_{eq_n-2t}) for the given pressure step.

2.1. Scanning Curves

It is also possible to generate scanning curves accounting for the pore blocking phenomenon by adapting the full Nitrogen Adsorption algorithm. An easy way to do this is by changing the input provided to the invasion percolation algorithm. The initial pressure of the desorption scanning branch is fixed and the corresponding critical pore diameter is calculated. For the pores with a diameter over the calculated critical diameter, they are temporarily assigned a very large diameter.

3. Case Study

3.1. Network Parameters

Networks with zero volume spherical nodes are considered for the case study. A 3D regular cubic lattice of 50x50x50 nodes was employed along with a Pore Existence Probability (PEP) equal to 0.75. The length of the pores is considered constant and equal to 21 nm. A pore size distribution generated by a Gaussian distribution centered at 10 nm and a standard deviation of 2 nm was employed. The consistency between the input and output cumulative distribution functions for the networks with zero volume spherical nodes was verified. An average connectivity equal to 4.2 and a void fraction of 0.38 were obtained.

3.2. Characteristic Curves

The simulated sorption isotherm is shown in Figure 1a. It corresponds qualitatively to typically observed experimental curves for mesoporous alumina. The isotherm

qualitatively corresponds to a type V isotherm (according to IUPAC classification) with a hysteresis loop closing at about 0.42 (Rouquerol et al., 2014). The hysteresis loop is a consequence of the pore blocking phenomenon and the meniscus curvature difference considered between condensation and vaporization in the Kelvin-Cohan equation. The starting and closing volume naturally coincide for both branches.

Figure 1b shows the back-calculated pore size distribution by applying the BJH method (Barrett et al., 1951) to the simulated isotherm. A good agreement was found between the pore size distribution generated from the adsorption branch and the original distribution used for the pore network generation.

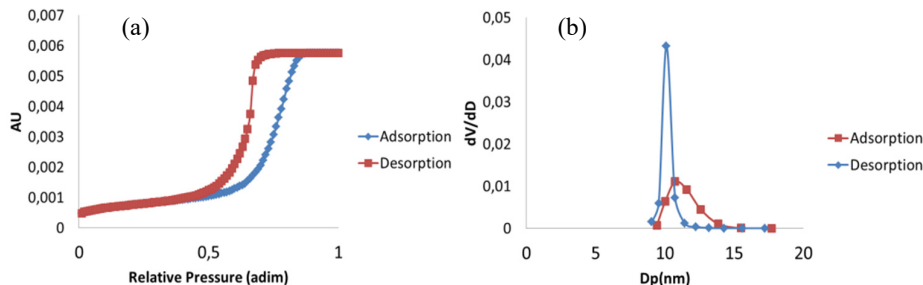


Figure 1: (a) Numerical Nitrogen Isotherm (b) Calculated BJH Pore Size Distribution (Adsorption and desorption branches)

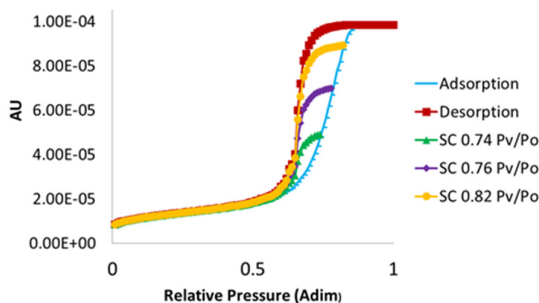


Figure 2: Digital Scanning Curves

Finally, Figure 2 provides an example of the adsorption-desorption scanning curves obtained for this sample. The hysteresis loop is of the H2 type with the scanning curves converging to the closure point of the hysteresis loop.

3.3. Algorithm's Execution Time

In order to test how the execution time of the simulations scales with the size of the network, six cubic lattice networks of different sizes were generated. The pore existence probability (PEP) was set to 1. All the other input parameters have been maintained as described in section 3.1. The results are shown in Figure 3. The execution time is plotted as a function of the total number of pores connected to the surface. As the size of the network increases, the time required for the execution increases as well. The algorithm execution time proceeds with an $O(M\log(N))$ time.

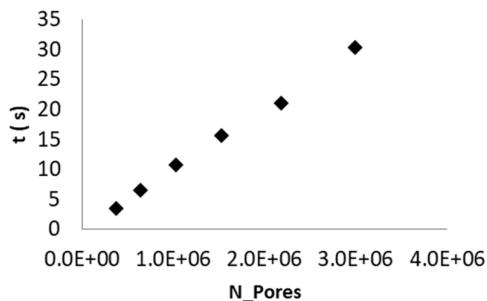


Figure 3: Effect of the number of pores on execution time of the nitrogen porosimetry algorithm.

4. Conclusions

A fast algorithm for the simulation of nitrogen porosimetry within a digital pore network was developed. The thermodynamic equilibrium conditions within the porous solid during nitrogen porosimetry and the associated pore blocking phenomena were considered. The algorithm was sped up by using a Max-Heap. A time execution test for the nitrogen porosimetry algorithm and an octahedral lattice of about 2 million pores only required 6.7 seconds for a single simulation using the binary heap in comparison to 19 hours using conventional sorting algorithms. This shows the advantage of using the binary heap for handling the search list as suggested by Masson (Masson and Pride, 2014). The size of the network is the main parameter that affects the execution time and exhibits an $O(M\log(N))$ behavior.

Valuable information about the topology of the pore network is implicit in the desorption branch of the isotherm and can be exploited by using this type of simulation. Adding an optimizer to this algorithm in order to adjust the digital characteristic curves to those of an actual porous alumina sample will allow obtaining a digital structure that not only exhibits the same macroscopic properties but that could also represent statistically the topology of the material. This algorithm has also been adapted to other porosimetry characterization techniques, such as mercury porosimetry, cryo-porometry, and thermo-porometry.

References

- Barrett, E.P., Joyner, L.G., Halenda, P.P., 1951. The Determination of Pore Volume and Area Distributions in Porous Substances. I. Computations from Nitrogen Isotherms. *J. Am. Chem. Soc.* 73 (1), 373–380. [10.1021/ja01145a126](https://doi.org/10.1021/ja01145a126).
- Ferreira, 2018. Modélisation du transport intragranulaire dans un réacteur catalytique, France.
- Ferreira, S., Verstraete, J.J., Jolimaitre et al., 2017. Random porous network generation and 1D mass transfer simulation for gamma-alumina supports. *Computer Aided Chemical Engineering* 40, 91–96. [10.1016/B978-0-444-63965-3.50017-9](https://doi.org/10.1016/B978-0-444-63965-3.50017-9).
- Masson, Y., Pride, S.R., 2014. A Fast Algorithm for Invasion Percolation. *Transp Porous Med* 102 (2), 301–312. [10.1007/s11242-014-0277-8](https://doi.org/10.1007/s11242-014-0277-8).
- Rouquerol, F., Rouquerol, J., Sing, K.S.W., Llewellyn, P.L., Maurin, G., 2014. Adsorption by powders and porous solids: Principles, methodology and applications. Elsevier/AP, Amsterdam.
- Šolcová, O., Matějová, L., Schneider, P., 2006. Pore-size distributions from nitrogen adsorption revisited: Models comparison with controlled-pore glasses. *Applied Catalysis A: General* 313 (2), 167–176. [10.1016/j.apcata.2006.07.021](https://doi.org/10.1016/j.apcata.2006.07.021).
- Zhang, Y., Lam, F.L.-Y., Yan, Z.-F., Hu, X., 2006. Review of Kelvin's Equation and Its Modification in Characterization of Mesoporous Materials. *Chinese Journal of Chemical Physics* 19 (2), 102–108. [10.1360/cjcp2006.19\(2\).102.7](https://doi.org/10.1360/cjcp2006.19(2).102.7).

Acyclic modular flowsheet optimization using multiple trust regions and Gaussian process regression

E.I. Sanchez Medina^a, D. Rodriguez Vallejo^b, B. Chachuat^b, K. Sundmacher^{a,c},
P. Petsagkourakis^{d*} and E.A. del Rio-Chanona^{b,**}

^a*Chair for Process Systems Engineering, Otto-von-Guericke University, Universitätspl.
2, Magdeburg, 39106, Germany*

^b*Centre for Process Systems Engineering, Imperial College London, Exhibition Rd.,
South Kensington, London, SW7 2AZ, UK*

^c*Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical
Systems, Sandtorstraße 1, Magdeburg, 39106, Germany*

^d*Centre for Process Systems Engineering, University College London, Gower St.
Bloomsbury, London, WC1E 6BT, UK*

* a.del-rio-chanona@imperial.ac.uk

** p.petsagkourakis@ucl.ac.uk

Abstract

This paper presents an algorithm to optimize process flowsheets using Gaussian processes regression and trust regions. We exploit the modular structure of the flowsheet by training separate Gaussian processes (GPs) for each module based on data generated by a process simulator. These GPs are embedded into an optimization model, whose outcome is used to adapt the position and size of the trust region at each iteration. A complication that arises because of the multiple trust regions is that the optimization problem may become infeasible, in which case a feasibility (restoration) problem is invoked. An inherent advantage of this approach is that it removes the need for simulating the complete flowsheet at any point. We demonstrate these ideas on the case-study of an extractive distillation system in order to minimize its total annualized cost (TAC). The performance shows a robust strategy to address flowsheet optimization problems without recycles.

Keywords: Flowsheet optimization, Gaussian processes, trust regions, machine learning

1. Introduction

Process flowsheet optimization carried out using gradient-based methods requires that the gradient information for all the process models is available (i.e., equation oriented modeling). However, the convergence of the equation-oriented method is limited when comparing it to the sequential modular approach. This extra benefit in the convergence comes at the expense of the flowsheet's gradient information, preventing gradient-based optimizers to be applied. In these cases, derivative-free techniques are often paired to the simulation models for optimization. But a flowsheet simulation itself may be computationally demanding, which calls for robust and sampling-efficient methods to

drive the optimization. Existing approaches can be broadly classified into two categories. *Global approaches* proceed by constructing a surrogate model based on a flowsheet simulation before optimizing it, often within an iteration where the surrogate is progressively refined. A number of successful implementations rely on Gaussian processes (Caballero and Grossmann, 2008; Keßler et al., 2019) or artificial neural networks (Schweidtmann et al., 2019). By contrast, *local approaches* maintain an accurate representation of the flowsheet (or separate modules thereof) within a trust region, whose position and size are adapted iteratively. This procedure entails reconstructing the surrogate model as the trust region moves towards the local optimum. Applications of this approach to flowsheet optimization include the work by Eason and Biegler (2018) and Bajaj et al. (2018).

This paper leverages ideas from the real-time optimization field to optimize modular process flowsheets, by training separate Gaussian processes for each module with data generated by the process simulation in a trust region. Background material is reviewed next, before describing the new approach and illustrating it on an extractive distillation system.

2. Background

Gaussian process (GP) regression: Initially developed as an interpolation technique, GP regression can be used to describe an unknown function f using noisy observations. GPs consider a distribution over functions and can be seen as a generalization of multivariate Gaussian distributions, $f(\cdot) \sim GP(m(\bullet), k(\cdot, \cdot))$. The mean function $m(\bullet)$ can express prior knowledge about f . The covariance function $k(\cdot, \cdot)$ accounts for correlations between the function values at different points and has a great impact on the GP's prediction accuracy. The so-called hyperparameters in the covariance functions need to be estimated in order to represent the observations as accurately as possible, e.g., by maximizing a log likelihood function. The predicted distribution of $f(x)$ at an arbitrary input point x , given the input-output data (X, y) , follows a Gaussian distribution, $f(x)|X, y \sim N(\mu_f(x), \sigma_f^2(x))$.

GPs and trust regions in flowsheet optimization: The approach developed by Palmer and Realff (2002) builds GPs around the whole flowsheet, which requires converged flowsheet simulation runs over a large input domain in order to be successful. Caballero and Grossmann (2008) proposed to construct GPs around individual process modules instead, thereby removing the need to simulate the entire flowsheet. But difficulties with this approach arise due to the presence of additional equality constraints (connecting streams) in the optimization problem. del Rio-Chanona et al. (2019, 2020) recently brought trust region concepts into a modifier-adaptation scheme for real-time optimization. The main idea is to use GPs to correct the cost and constraint functions of the optimization problem locally in the trust region. The present work builds upon these ideas to enable separate Gaussian processes in multiple trust regions corresponding to the flowsheet modules.

3. Multiple trust regions and Gaussian processes for optimization

The modular flowsheet optimization problem can be stated as:

$$\sum_{i=1}^m g_{i,0}(x_i, y_i) \tag{1}$$

$$s.t. \quad g_{i,j}(x_i, y_i) \leq 0, \quad j = 1 \dots n_{g_i}, \quad i = 1 \dots m \tag{2}$$

$$c_j(x_1, y_1, \dots, x_m, y_m) = 0, \quad j = 1 \dots n_c \tag{3}$$

$$y_i = \varphi_i(x_i, \theta_i), \quad (x_1 \dots x_m) \in X, \quad (y_1 \dots y_m) \in Y, \quad i = 1 \dots m \tag{4}$$

Each module $i = 1 \dots m$ (m denoting the last connected module) has inputs x_i , process parameters θ_i , outputs y_i , cost contribution $g_{i,0}$, and design/operating constraints $g_{i,j}$. The equality constraints c_j describe the connections between modules. Notice that these connecting constraints are nonlinear in general—think for instance about the input stream to a module being obtained by mixing the output streams from two modules. And the φ_i is the input-output mapping for module i .

The proposed approach entails the use of GP regression to predict the performance of each module separately, $\varphi_i(x_i, \theta_i) \sim N(\mu_{\varphi_i}(x_i, \theta_i), \sigma_{\varphi_i}^2(x_i, \theta_i))$. The trust-region subproblems that are solved at a given iteration k in order to determine the next inputs $x_i^k + d_i^{k+1}$ are given by

$$\sum_{i=1}^m g_{i,0}(x_i^k + d_i^k, \hat{y}_i^k) \tag{5}$$

$$s.t. \quad g_{i,j}(x_i^k + d_i^k, \hat{y}_i^k) \leq 0, \quad j = 1 \dots n_{g_i}, \quad i = 1 \dots m \tag{6}$$

$$c_j(x_1^k + d_1^k, \hat{y}_1^k, \dots, x_m^k + d_m^k, \hat{y}_m^k) = 0, \quad j = 1 \dots n_c \tag{7}$$

$$\hat{y}_i = \mu_{\varphi_i}(x_i^k + d_i^k, \theta_i), \quad (\hat{y}_1 \dots \hat{y}_m) \in Y, \quad i = 1 \dots m \tag{8}$$

$$\|d_i\| \leq \Delta_i^k, \quad (x_1^k + d_1^k, \dots, x_m^k + d_m^k) \in X \tag{9}$$

where $\Delta_i^k \geq 0$ denotes the radius of the input trust region for module i . At iteration k , the GPs for module i are trained with the input-output dataset (X_i^k, Y_i^k) . This dataset is generated using a Sobol sequence sampling within the space delimited by the double of each trust region radius (i.e. $2\Delta_i^k$). Sampling beyond the trust region radius allows for the complete trust region space to be well-approximated (i.e., without large mismatch errors near the trust region boundary).

The decisions about moving the trust region center (current operating point x^k) or changing its radius are based upon:

(i) the ratio of actual cost reduction to predicted cost reduction:

$$\rho^{k+1} := \frac{\sum_{i=1}^m g_{i,0}(x_i^k, \varphi_i(x_i^k, \theta_i)) - g_{i,0}(x_i^k + d_i^{k+1}, \varphi_i(x_i^k + d_i^{k+1}, \theta_i))}{\sum_{i=1}^m g_{i,0}(x_i^k, \mu_{\varphi_i}(x_i^k, \theta_i)) - g_{i,0}(x_i^k + d_i^{k+1}, \mu_{\varphi_i}(x_i^k + d_i^{k+1}, \theta_i))} \tag{10}$$

(ii) the violation of any inequality constraint $j = 1 \dots n_{g_i}$:

$$g_{i,j}(x_i^k + d_i^{k+1}, \varphi_i(x_i^k + d_i^{k+1}, \theta_i)) > 0 \quad (11)$$

A complete statement of the scheme is provided in Algorithm 1. Whenever the optimization model (5)-(9) turns out to be infeasible, the following restoration phase is invoked, that minimizes the violation of the design/operating constraints:

$$\sum_{i=1}^m \sum_{j=1}^{n_{g_i}} g_{i,j}(x_i^k + d_i, \hat{y}_i) \quad (12)$$

$$\text{s. t. } c_j(x_i^k + d_i, \hat{y}_i) = 0, \quad i = 1 \dots m, \quad j = 1 \dots n_c \quad (13)$$

$$\hat{y}_i = \mu_{\varphi_i}(x_i^k + d_i, \theta_i), \quad (\hat{y}_1 \dots \hat{y}_m) \in Y, \quad i = 1 \dots m \quad (14)$$

$$\|d_i\| \leq \Delta_i^k, \quad (x_1^k + d_1, \dots, x_m^k + d_m) \in X \quad (15)$$

Whenever the solution of the optimization subproblem (5)-(9) is feasible, the trust regions are updated according to the value of the search performance factor shown in Eq. 10. The step is accepted whenever this accuracy ratio is large enough. The radii for each trust region are expanded whenever the largest feasible step is taken (i.e., $\|d_i\| = \Delta_i^k$) and a good enough prediction is achieved. If, on the other hand, the accuracy ratio is too small, only the radius of the last connected module (i.e., $i = m$) is reduced. This prevents a possible mismatch among trust regions while still allowing for radius shrinkage as the algorithm converges. For this reason, it is worth mentioning that the proposed algorithm is limited to acyclic process flowsheets only. If a recycle would be present the identification of the last connected module (see line 12 in Algorithm 1) would be problematic. In this case, the optimization subproblem (5)-(9) might become infeasible due to a violation of a connectivity constraint (Eq. 7). For this same reason, the violations considered within the restoration problem (12)-(15) are only coming from the design/operation inequality constraints (Eq. 6) (see line 4 in Algorithm 1).

Algorithm 1: Multiple trust regions and Gaussian process regression for optimization

Input: Input-output mappings φ_i ; process parameters θ_i ; initial trust region centers

x_i^0 ; initial and maximal trust region radii $0 < \Delta_i^k < \bar{\Delta}_i$; trust region parameters $0 < \eta_1 \leq \eta_2 \leq \eta_3 < 1$; shrinkage and expansion parameters $0 < \xi_1 < 1 < \xi_2$; bounds X and Y

Repeat for $k = 0, 1, \dots$

- 1 Retrieve input-output dataset (X_i^k, Y_i^k) using Sobol sequence sampling;
- 2 Train GP for each module i ;
- 3 Solve trust region optimization subproblem (Eqs. 5-9);
- 4 **if** Infeasible due to Eq. 11 **then**
- 5 Solve feasibility (restoration) subproblem (Eqs. 12-15);


```

6    $\Delta_i^{k+1} := \Delta_i^k$ 
7    $x_i^{k+1} := x_i^k + d_i^k$ 
8   else
9     Retrieve output values  $y_i$  for  $\varphi_i(x_i^k + d_i^{k+1}, \theta_i)$ ;
10    Compute  $\rho^{k+1}$  according to Eq. 10;
11    Update trust regions radii:
12      if  $\rho^{k+1} < \eta_2$  and  $i = m$  then  $\Delta_i^{k+1} := \Delta_i^k \xi_1$  ;
13      else if  $\rho^{k+1} < \eta_3$  and  $\|d_i^k\| = \Delta_i^k$  then  $\Delta_i^{k+1} := \min\{\bar{\Delta}_i, \Delta_i^k \xi_2\}$ ;
14      else  $\Delta_i^{k+1} := \Delta_i^k$ ;
15    Update operating point:
16      if  $\rho^{k+1} < \eta_1$  then  $x_i^{k+1} := x_i^k$ 
17      else  $x_i^{k+1} := x_i^k + d_i^{k+1}$ 

```

4. Computational case-study

4.1. Extractive distillation system

The proposed methodology is applied to the minimization of the total annualized cost (TAC) of an extractive distillation system. An isomolar mixture of n-heptane and toluene is separated using phenol as the extractive medium. The input-output mappings are two separate Aspen HYSYS V9 simulations, corresponding to the two distillation columns in the process. These two columns are connected in a serial arrangement. Figure 1 shows a schematic representation of each individual column containing important information for their simulation. Purity constraints of at least 97% mole were imposed to the n-heptane and toluene at the top of the columns.

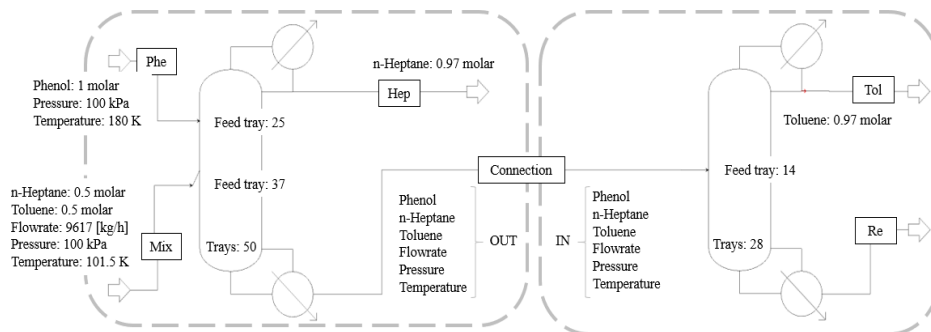


Figure 1: Extractive distillation diagram with some important information. Dashed lines represent the two individual flowsheets used in the proposed methodology.

The decision variables of this problem are the reflux and boilup ratios for both columns and the molar flowrate of phenol at the inlet of the first column. However, the process variables related to the connecting stream (IN and OUT in Fig. 1) are also part of the optimization subproblem that is solved at each iteration. The proposed algorithm requires that the initial trust regions fulfill the equality (connectivity) constraints. Then, at each iteration, a GP is trained around each column using a Sobol sequence sampling on the vicinity of the current trust region. To construct the GPs, the mean function is set to zero and the Matern 3/2 covariance function is used. For this case study, the trust region parameters were set to: $\eta_1 = \eta_2 = 0.1$; $\eta_3 = 0.9$, $\xi_1 = 0.5$ and $\xi_2 = 2$.

We implemented Algorithm 1 in Python 3.7. The optimization subproblems were solved using IPOPT (Wächter and Biegler, 2006) using CasADi (Andersson et al., 2019) for automatic differentiation. At each iteration, we used 10 and 25 samples to construct the GPs for the first and second columns, respectively. We also normalized the input values to be between 0 and 1 and used an initial trust region radius of $\Delta_1^0 = 0.02$ and $\Delta_2^0 = 0.025$. In general, the number of samples and the initial trust regions radii have to be chosen such that a relatively accurate approximation of the original function is obtained, and fast computations can be achieved. Therefore, these two are important problem-dependent hyperparameters. The optimum was reached after 50 iterations. The starting point violates the purity constraints. Therefore, the restoration phase is called

from the beginning as shown by the dashed line in Figure 2. Once the operating point reaches the feasible area, the TAC is continuously minimized at each iteration (solid line). And after the minimum is reached, the trust region radius of the second column starts to shrink to achieve a better approximation of the true problem (dotted-dashed line).

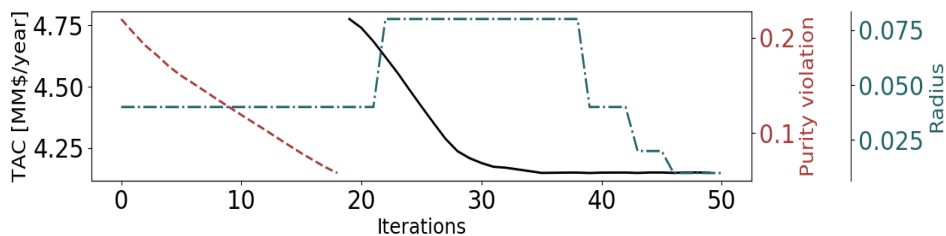


Figure 2: Optimization trajectory for the extractive distillation system. Black solid line: Total Annualized Cost [MM\$/year]. Red dashed line: Purity constraints violation. Blue dotted-dashed line: Trust region radius around second column.

5. Conclusion

In this work an algorithm to optimize process flowsheets was presented. The main advantage of the proposed method is that a complete flowsheet does not need to be simulated at any point. The modular flowsheet structure is exploited so that multiple GPs are trained independently for each individual module. These surrogates are built on separate trust regions and connected within an optimization subproblem using equality constraints corresponding to the connecting streams in the flowsheet. The location and

size of the trust regions are updated according to the search performance at each iteration. A restoration problem is invoked whenever infeasibility is encountered due to an inequality constraint violation. However, the algorithm is unable to account for a mismatch between trust regions. Therefore, its application to process flowsheets with recycles remain an open question and it is subject of future improvement and investigation. The algorithm demonstrated to reach the optimum effectively in the presented case-study.

References

- J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, M. Diehl, 2019. CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation* 11 (1), 1–36.
- I. Bajaj, S. S. Iyer, M. M. F. Hasan, 2018. A trust region-based two phase algorithm for constrained blackbox and grey-box optimization with infeasible initial point. *Computers & Chemical Engineering* 116, 306–321.
- J. A. Caballero, I. E. Grossmann, 2008. An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal* 54 (10), 2633–2650.
- E. A. del Rio-Chanona, J. E. A. Graciano, E. Bradford, B. Chachuat, 2019. Modifier-adaptation schemes employing Gaussian processes and trust regions for real-time optimization. *IFAC-PapersOnLine* 52 (1), 52–57.
- E. A. del Rio-Chanona, P. Petsagkourakis, E. Bradford, J. E. A. Graciano, B. Chachuat, 2020. Modifier adaptation meets bayesian optimization and derivative-free optimization. *arXiv preprint arXiv:2009.08819*.
- J. P. Eason, L. T. Biegler, 2018. Advanced trust region optimization strategies for glass box/black box models. *AIChE Journal* 64 (11), 3934–3943.
- T. Keßler, C. Kunde, K. McBride, N. Mertens, D. Michaels, K. Sundmacher, A. Kienle, 2019. Global optimization of distillation columns using explicit and implicit surrogate models. *Chemical Engineering Science* 197, 235 – 245.
- K. Palmer, M. Realf, 2002. Metamodeling approach to optimization of steady-state flowsheet simulations: Model generation. *Chemical Engineering Research and Design* 80 (7), 760 – 772.
- A. M. Schweidtmann, W. R. Huster, J. T. Lüthje, A. Mitsos, 2019. Deterministic global process optimization: Accurate (single-species) properties via artificial neural networks. *Computers & Chemical Engineering* 121, 67–74.
- A. Wächter, L. T. Biegler, 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming* 106 (1), 25–57.

Systematic generation and targeting of chemical recycling pathways: A mixed plastic waste upcycling case study

Adrián Pacheco-López^{a*}, Ana Somoza-Tornos^{a,b}, Antonio Espuña^{a*}, Moisès Graells^a

^a*Chemical Engineering Department, Universitat Politècnica de Catalunya, Escola d'Enginyeria de Barcelona Est, C/ Eduard Maristany 16, Barcelona 08019, Spain*

^b*Renewable and Sustainable Energy Institute, University of Colorado Boulder, Boulder, Colorado 80303, United States of America*

**adrian.pacheco@upc.edu* **antonio.espuna@upc.edu*

Abstract

The current concerns regarding plastic waste accumulation and resource scarcity are leading to a waste and resource management paradigm shift, and consequently the number of alternatives for treating and upcycling waste keeps increasing. Here, we develop a new approach to determine the most promising processes or process routes to valorize waste by seeking the best environmental and economic performance. Due to the large number of alternatives, we designed this method to select a certain number of choices prior to further assessment and optimization. As a case study, we chose the treatment of mixed plastic waste. Using the methodology, we are able to select a few alternatives among a large number of different alternatives as the most promising towards closing material loops. The use of graph theory and branch and bound algorithms allows building an efficient framework for this purpose.

Keywords: waste-to-resource, chemical recycling, circular economy, short-path algorithms, material upcycling.

1. Introduction

Plastics are widely spread materials due to their versatility, good performance in a wide range of applications, lightweight and low production cost. This fact has led along the last decades to the rapid increase of plastic items production, especially in segments showing low costs and short useful life, leading to increasingly high amounts of plastic waste generation across the globe, mainly in rapidly growing economies. Thus, waste accumulation and resources depletion concerns are set to grow in the following years due to the lack of effective and efficient resource and waste management policies in most countries worldwide. In addition, there are many barriers that prevent the development of suitable long-term solutions to this problem, such as social, technological and financial, among others.

In an effort to address this challenge, there has been an increasing interest in developing waste-to-resources processes to obtain valuable products from plastic wastes. Among them, chemical recycling stands out as one of the most promising alternatives, including a significant number of different implementations, which keeps growing up. This opens up a wide range of options for the upcycling of materials. Within these resource recovery

technologies, the ones drawing more attention are thermolysis processes such as pyrolysis, hydrocracking and gasification, which are being studied under a wide range of conditions (Zhang et al., 2020).

There have been efforts towards model development in order to find optimum solutions for the treatment of available plastic waste, such as mixed plastic waste (MPW), as the one proposed by Somoza-Tornos et al., 2021. Nevertheless, based on the increasing number of choices present, there is a surging need for a methodology to systematically generate and assess new potential pathways as a means to obtain the most promising and profitable alternatives as a previous step to their detailed analysis and optimization.

To address such a problem, the methodology developed in this work extends the framework proposed in previous contributions (Pacheco-López et al., 2020). This framework aims to identify and assess the most suitable routes towards closing materials loops and resources upcycling. For the knowledge modelling of process, waste and resources information in the domain of Process Systems Engineering (PSE), we based our approach on the use of ontologies. The resulting knowledge corpus is queried in order to implicitly generate all possible routes connecting available wastes with tradable raw components. Bounds for partial and complete routes are assessed according to economic and environmental aspects for the identification of the most promising alternatives. However, due to the large number of alternatives available, we propose a pre-selection of the most promising ones, as a previous step to a more thorough assessment.

2. Problem Statement

The problem can be formally posed as follows:

- Given:
 - A large database of processes available for the domain of PSE, including information about processes, such as costs, environmental impacts and yields. Some of them can be potentially used for the treatment of plastic waste and its transformation into eventual intermediate or valuable products.
 - A list of available plastic waste sources, including information about its composition, cost and environmental footprint.
 - A list of potential products demand, including information about their marketable price, quality requirements and technical specifications.
- Find:
 - A set of the most promising routes and alternatives in order to connect waste with tradable products.

3. Methodology

This work focuses on the generation of routes using branch and bound algorithms (B&B). Before applying such algorithms, all the possible processes are obtained from the ontology with the aim of connecting waste sources with tradable products thanks to an input-output matching method, linking each material stream with each one of their tentative processes. Then a tree-like graph with all the possible paths is built using a state-task network approach (Kondili et al., 1993), where operations and processes are considered as “tasks” and feedstock, intermediate and final products as “states”, both included as explicit or implicit network interconnected nodes. Using this kind of algorithm, we are able to find a set of tentative routes avoiding the need of evaluating every single possibility. Consequently, a certain number of paths is generated whose are

fit for further assessment. The methodology can be structured into four sections: knowledge management, database querying, graph building and graph evaluation.

3.1. Knowledge management

As introduced in section 1, we are using ontologies as a knowledge management tool, providing consensual knowledge representation for a specific application domain. For the domain of Chemical Aided Chemical Engineering (CAPE), we are using an open-source, recognized ontology within the domain as it is OntoCAPE (Marquardt et al., 2010). The ontology has been analyzed and adapted to the framework in development; likewise, some concepts have been modified or added to specify aspects that were not previously defined in the ontology, such as the environmental impact indicators among others.

All the processes, plastic wastes, intermediate products and tentative final products have been asserted in the ontology with all the corresponding property relations and values such as costs, prices, yields and environmental indicators.

3.2. Database querying

For the integration of the ontology and the algorithm, an ontology-oriented interface is needed. We are using an object-oriented programming tool for accessing entities from OWL ontologies, namely an Open-Source software called Owlready developed in Python (Lamy, 2017). This tool enables, not only an easy access to all information available in the ontology, such as entities and relations between them, but also adding new information in a systematic way and running a reasoner for the inference of new missing relations or discovery of possible conceptual inconsistencies.

3.3. Graph building

With the information for each entity available in the ontology, an input-output matching method is developed to connect wastes, processes and products, thus building the corresponding graph with the STN approach mentioned above. This representation can be classified as a bipartite graph, which has two kinds of nodes, one for processes (considered in the ontology as “steps”) and another for materials (considered as “states”), which are connected through edges. At this level, the graph stores information about economic, environmental or behavioral aspects in its nodes, but no routing or path finding is performed yet.

3.4. Network evaluation

Once the graph is implicitly built, a branch and bound algorithm is implemented so as to analyze the branches rooting from the initial node, which corresponds with the waste to treat. The algorithm starts building branches and choosing only those with better performance to continue advancing towards the final products or choosing an intermediate point that could be more profitable. The algorithm is adapted to allow the completion of a limited number of paths for an easier further assessment.

With the purpose of calculating the bounds for each branch, economic and environmental aspects are added up, specifically prices with monetized impacts for materials on one side, and costs with monetized impacts for processes on the other. Then, for each process considered, a total profit is obtained and added to the respective branch as shown in equations 1 to 4.

$$EE_price_j = Price_j + Product_EI_j \quad (1)$$

$$EE_cost_i = Cost_i + Process_EI_i \quad (2)$$

$$EE_profit_i = \sum_{j \in outputs} EE_price_j - \sum_{j \in inputs} EE_price_j - EE_cost_i \quad (3)$$

$$Total_profit = \sum_{i \in route} EE_profit_i \quad (4)$$

4. Case Study

In order to test the methodology, we chose a case study for the recycling of Mixed Plastic Waste (MPW). All possible paths for the treatment of MPW are built, including those partial paths that may yield intermediate products and be used for other purposes, such as fuel for energy recovery or the transportation sector.

For this case study, the ontology has been filled with information from processes available in the literature related to this domain, including end-of-pipe processes traditionally used to get rid of this kind of wastes, such as landfilling and incineration for energy recovery, as well as other emerging processes like pyrolysis at a wide range of temperatures with and without previous mechanical sorting, leading to promising circular economy solutions. The information related to those processes which specifications not available in the literature, not reliable or contradictory has been estimated thanks to commercial simulators and standard procedures for the assessment of economic and environmental indicators.

Departing from a total of 140 tentative steps asserted in the ontology which are directly or indirectly related to this specific case study, approximately 300 potential paths are obtained, from which nearly 100 complete paths have been deployed and 200 partial routes have been also analyzed. From this high number of possibilities, we select only 20 paths for illustrative purposes, the partial process tree resulting from these 20 alternatives is schematized in Figure 1. The aggregated amounts for environmental and economic aspects, as well as the yields of each process are shown in the scheme. It is important to note that the values are unitary, i.e. per ton, and they have to be multiplied by the yield to obtain the absolute values.

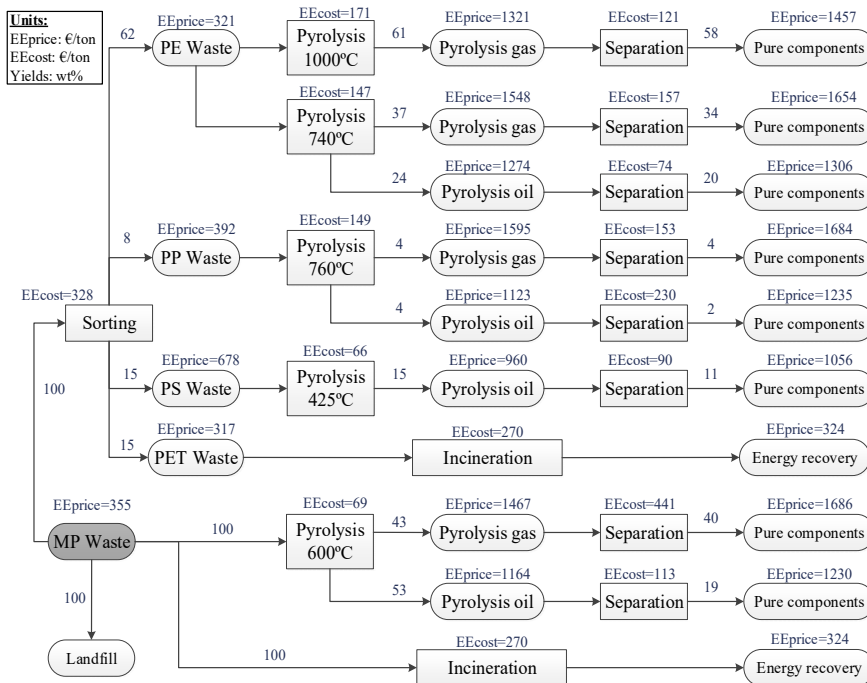


Figure 1. Possible alternatives for MPW treatment

5. Results and Discussion

Among all the possible alternatives, only the selected 20 are shown in Table 1. The algorithm builds the paths with an input-output matching as explained in section 3.3, afterwards it evaluates the branches, node by node, and keeps the alternatives that show better performance, discarding those with lower economic and environmental results.

For this case study, the upper bound corresponds to that where a sorting is applied and a pyrolysis is performed on the polyethylene waste at 1000°C and a mixture of gases is obtained apart from polypropylene, polyethylene terephthalate and polystyrene sorted wastes. As we can appreciate in Table 1 the next upper bound comes when the mixture of gases is separated into its pure components, the third one is the pyrolysis at 740°C of the polyethylene waste, obtaining a mixture of pyrolysis oil and gas, apart from the sorted wastes as mentioned before. The rest of the bounds can be consulted in Table 1.

Table 1. Deployed process paths, outputs and bounds for the illustrative 20 alternatives.

	Processes	Outputs	OF
1	Sorting + Pyro. PE 1000°C	PE 1000°C Pyrolysis gas + PP, PET, PS wastes	1457
2	Sorting + Pyrolysis PE 1000°C + Separation	Butadiene, benzene, ethylene, methane and propene + PP, PET, PS wastes	1084
3	Sorting + Pyro. PE 740°C	PE 740°C Pyro. oil and gas + PP, PET, PS wastes	1048
4	Sorting + Pyro. PS 425°C	PS 425°C Pyrolysis oil + PP, PET, PE wastes	985
5	Sorting + Pyrolysis PE 740°C + Separation(Gas)	PE 740°C Pyrolysis oil, ethylene, ethane, methane and propene + PP, PET, PS wastes	769
6	Pyrolysis MPW 600°C + Separation(Gas)	MPW Pyrolysis oil, butane, ethylene, 1-butene, propene, propane, hydrogen, methane and ethane	619
7	Sorting + Pyro. PE 740°C + Separation(Oil)	PE 740°C Pyrolysis gas, pyrene, indane, benzene and toluene + PP, PET, PS wastes	423
8	Pyrolysis MPW 600°C	MPW Pyrolysis oil and gas	356
9	Pyrolysis MPW 600°C + Separation(Both)	Butane, ethylene, propane, hydrogen, methane, ethane, ethylbenzene, styrene, benzene, toluene...	288
10	Sorting	Sorted plastic waste (PE, PP, PET, PS)	236
11	Sorting + Pyrolysis PS 425°C + Separation	Ethylbenzene, cumene and toluene + PP, PET, PE wastes	185
12	Sorting + Pyrolysis PE 740°C + Separation(Both)	Ethylene, ethane, methane, propene, pyrene, indane, benzene and toluene + PP, PET, PS wastes	144
13	Sorting + Pyrolysis PP 760°C + Separation(Gas)	PP 760°C Pyrolysis oil, ethane, methane, propene, and ethylene + PE, PET, PS wastes	123
14	Sorting + Pyro. PP 760°C	PP 760°C Pyro. gas and oil + PE, PET, PS wastes	115
15	Sorting + Pyrolysis PP 760°C + Separation(Both)	Ethane, methane, propene, ethylene, benzene, toluene and naphthalene + PE, PET, PS wastes	94
16	Pyrolysis MPW 600°C + Separation(Oil)	MPW Pyrolysis gas, ethylbenzene, styrene, benzene and toluene	25
17	Sorting + Pyrolysis PP 760°C + Separation(Oil)	PP 760°C Pyrolysis gas, benzene, toluene and naphthalene + PE, PET, PS wastes	86
18	Direct Downcycling MPW	MPW	0
19	Incineration	Energy Recovery	-17
20	Landfill	None	-238

Among the most promising possibilities for the upcycling of MPW, we find pyrolysis at different temperatures, with or without previous sorting. The number of alternatives grows exponentially with the number of available processes. Therefore, it is not efficient to assess all these alternatives individually, in addition to being more time consuming and computing resources demanding. Hence, this previous selection guarantees that the subsequent assessment and in-depth analysis is performed only on the most promising alternatives, and not all possible alternatives.

Results show that this methodology is useful to identify all the possible routes for closing the materials loops and select only those that are most promising, prior to more detailed route assessment. This contribution represents an important previous step and extension of the previously developed scheme (Pacheco-López et al., 2020), with the purpose of narrowing down the number of available alternatives.

6. Conclusions

This approach enables the systematic identification of the available opportunities for the treatment of waste and therefore closing the material loops, reducing resources exploitation and depletion, as well as the disposal of waste and reducing the environmental impact this entails. The methodology allows the pre-selection of paths according to their performance, reduces the amount of choices to those that are more promising and facilitates next steps. Moreover, the method allows the pre-selection of the most relevant paths and removes those that are economically and environmentally less profitable.

Acknowledgements

Financial support received from the Spanish Competitiveness, Industry and Economy Ministry and the European Regional Development Fund, both funding the research Projects AIMS (DPI2017-87435-R) is fully acknowledged. Adrian Pacheco-Lopez thankfully acknowledges financial support received from the Spanish Ministry of Science, Innovation and Universities (grant ref. PRE2018-087135).

References

- Kondili, E., Pantelides, C.C., Sargent, R.W.H., 1993. A general algorithm for short-term scheduling of batch operations-I. MILP formulation. *Comput. Chem. Eng.* 17, 211–227. [https://doi.org/10.1016/0098-1354\(93\)80015-F](https://doi.org/10.1016/0098-1354(93)80015-F)
- Lamy, J.-B., 2017. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. *Artif. Intell. Med.* 80, 11–28. <https://doi.org/10.1016/j.artmed.2017.07.002>
- Marquardt, W., Morbach, J., Wiesner, A., Yang, A., 2010. OntoCAPE: A Re-Usable Ontology for Chemical Process Engineering. <https://doi.org/10.1007/978-3-642-04655-1>
- Pacheco-López, A., Somoza-Tornos, A., Muñoz, E., Capón-García, E., Graells, M., España, A., 2020. Synthesis and Assessment of Waste-to-resource Routes for Circular Economy, Computer Aided Chemical Engineering. Elsevier Masson SAS. <https://doi.org/10.1016/B978-0-12-823377-1.50323-2>
- Somoza-Tornos, A., Pozo, C., Graells, M., España, A., Puigjaner, L., 2021. Process screening framework for the synthesis of process networks from a circular economy perspective. *Resour. Conserv. Recycl.* 164, 105147. <https://doi.org/10.1016/j.resconrec.2020.105147>
- Zhang, F., Zhao, Y., Wang, D., Yan, M., Zhang, J., Zhang, P., Ding, T., Chen, L., Chen, C., 2020. Current technologies for plastic waste treatment: A review. *J. Clean. Prod.* 124523. <https://doi.org/10.1016/j.jclepro.2020.124523>

Calibration of an EPANET model by using monitoring points in the bulk distribution of Sedibeng Water, for predicting disinfectant decay

Charl van der Walt,^a Prof Evans Chirwa,^a Danie Traut,^b

^a*University of Pretoria, 2 Lynnwood Road, Pretoria, 0002, South Africa*

^b*Sedibeng Water, Protea Street Balkfontein Bothaville 9660, South Africa*

Abstract

By using the EPANET software, the distribution system of Sedibeng Water in the Republic of South Africa was modelled and populated with actual free chlorine data. It was found from the model that the maximum age of the water from leaving the plant up to a point 70 kilometres away from the plant was 160 hours. With a calculated summer chlorine decay rate of 0.052 h^{-1} for summer conditions and 0.027 h^{-1} for winter conditions, it was demonstrated that interim booster chlorination can be discontinued during the winter months, while these booster stations should be operational during the warmer periods to ensure adequate disinfection in the overall 300 kilometre pipe supply system.

In addition to optimising the current disinfectant conditions, it was also demonstrated that a switch to ClO_2 for final disinfection in the network is possible, with the benefit that interim chlorination can be discontinued. This option will require an additional USD 27,000 to be spend per annum, with provision made for a once-off capital expense of USD 65,000.

Keywords: Chlorine, Chlorine dioxide, EPANET, decay, disinfectant.

1. Introduction and Context

Sedibeng Water is a Water Board in South Africa and purifies approximately 210 MI/d raw water from the Vaal River to potable standards. This raw water is characterized by seasonal elevated levels of algae which leads to unpleasant taste and odours in the water. Sedibeng Water employs standard clarification and filtration technologies at their two water treatment plants (WTP's) namely the Balkfontein WTP and the Virginia WTP. Use is made of chlorine as disinfectant while they also make use of interim chlorination at three areas in their network due to the vast distances being covered by their supply system which uses over 300 kilometres of piping. Chlorine depletion occurs in the network and a need was established to predict this occurrence using suitable software.

EPANET is a public domain software application which can be used to track amongst others the flow of water, the pressure at nodes, chemical concentrations and the decay thereof, and the age of the water in the network.

Another aspect of this research included the evaluation of chlorine dioxide (ClO_2) as an alternative to using chlorine as final disinfectant in the network. Due to the superior ability of ClO_2 in controlling iron, manganese, and taste and odour causing compounds, the option to replace chlorine with ClO_2 becomes an attractive possibility, and once the

EPANET model was calibrated, a “what-if” scenario was modelled where the efficacy of ClO₂ to provide for sufficient residuals was investigated.

2. Disinfectant Decay

Disinfection is regarded to be the most important process in water treatment, and with known dangers associated with this action around the formation of disinfectant by-products, the formation of these compounds need to be minimised (Monteiro et al, 2014). At the same time, it is recognised that water without disinfection, will result in far greater dangers around health, and cost implications of cleaning out overgrown reticulation systems. It is also difficult to predict chlorine decay in extensive and complex distribution systems, due to all the variables such as pH, organic content, hydraulic conditions and temperature to name but a few.

Since these large number of variables limit extensive applications of universal chlorine decay models, each distribution system should be considered individually, depending on its different environmental conditions (Pasha & Lansey, 2010). The decay of disinfectants have been reported on by numerous authors and it is generally accepted that the decay of any disinfectant is affected by two main items namely the bulk reaction in the water mass and the surface reactions (wall decay) in the pipe network (Fisher, et al., 2011). The bulk reaction is influenced by chemical properties of the water such as temperature, organic content and pH, while the surface reactions are more complex and influenced by biofilm growth, sediments in the pipe, type of pipe as well as its condition and age.

Laboratory studies can be conducted on-site to determine the bulk decay rate of chlorine in treated water, while the determination of the wall decay rates can best be obtained by analysing actual data in the network. By obtaining these values, and applying a suitable mathematical model to it, it is possible to populate the EPANET model with these decay rates and calculate theoretical free chlorine values at selected points in the network. EPANET is the most well-known process based model to be used for chlorine decay as it makes provision for separate measurement and programming for bulk decay and wall reactions (Rossman, 2000)

Comparing theoretical values to actual values subsequently remains to be done in delivering a calibrated model of EPANET whereby the decay of disinfectants can be modelled and used in “what-if” scenarios for planning purposes

3. 3. Model Calibration

3.1. On-site Experimental Decay Work

The on-site decay tests were done by taking a sample of the treated chlorinated water from the sampling point at the pumping station, placing it in amber, dark glass bottles in a controlled temperature environment (13 °C and 23 °C) and testing for free chlorine readings on hourly intervals for a period of 14 hours. Free chlorine analysis was done by the laboratory using the same method employed for the weekly monitoring points in the network. These tests were done for the final chlorinated water at the Balkfontein WTP and at the Virginia WTP. [Figure 1](#) indicates the results for these tests.

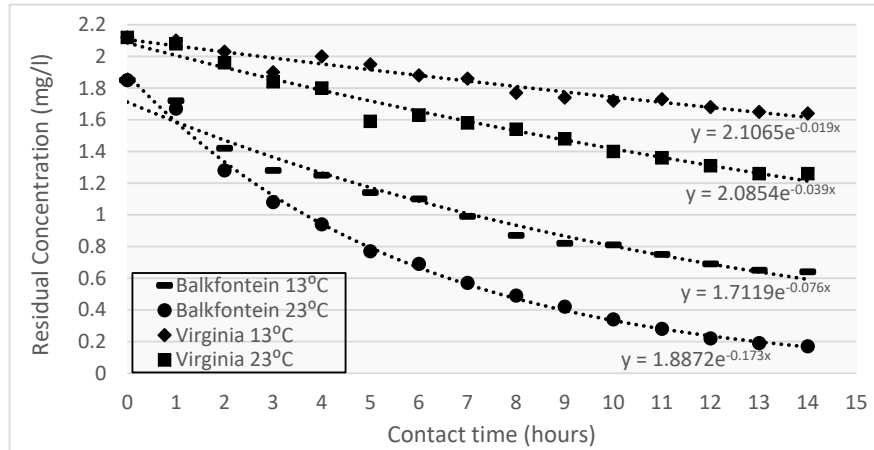


Figure 1: Experimental decay rates for Balkfontein and Virginia final water

Decay rates for chlorine were calculated using the following equation:

$$C_t = C_o e^{-kt}$$

Where :

- C(t) = Calculated disinfectant concentration after time t (mg/l)
- C_o = Initial disinfectant concentration (mg/l)
- k = Decay rate (1/h)
- t = time (h)

In general it is seen that the 10 °C difference in temperature leads to an approximate doubling in decay rates, which is not nearly as drastic as the suggested 5 °C increase as per Fisher, et al., 2011, and also lower than the 16–17 °C temperature rise as per the (Sathasivan, et al., 2009) findings for doubling of decay rates.

3.2. Calibration with Network Data

Weekly monitoring of water quality is done at selected sites in the Sedibeng Water network and measured amongst others for free chlorine, pH and turbidity. This data were analysed to obtain site specific bulk- and wall decay coefficients. The method followed for every EPANET pipework section is set out below:

- Obvious errors regarding the measured values were discarded. These were mostly where an increase in free chlorine value was found following a period of decay. In discussions with Sedibeng Water it was indicated that this is acceptable and dealt with instances where interim emergency adding of chlorine was made;
- A first order equation was fit on the data to determine bulk decay coefficients and wall decay coefficients. Where data points only constitutes a pipe section, a

combined bulk and wall decay coefficient was established and where only a reservoir was monitored for inlet and outlet sections, a bulk decay coefficient was established;

- The wall decay coefficients and bulk decay coefficients were plotted on a graph as a function of time of year to determine any obvious correlations with temperature;
- Statistical evaluations were done on the data to determine a median for the data, a standard deviation of the data as well as a coefficient of variation (CV). A CV larger than one (1) was considered to be statistically less usable in the calculation of a decay coefficient, and was hence discarded.
- Finally, the coefficients for the various sections were introduced into the EPANET model and an average residual chlorine concentration was calculated for the various monitored nodes in the system and compared to an average value of chlorine residual actually monitored at the node over a period of a year (See [Table 1](#))

Two clear results stood out from the data in that:

- the decay rate during the winter months dropped to 50-60% of the decay rate experienced during the summer months. This is in line with the on-site test results and it was therefore necessary to make provision for a winter model and a summer model; and
- the bulk decay rates in the system for both winter and summer conditions approximated the value of the combined decay coefficient. This is in line with findings from (Fisher, et al., 2017) where it was found that bulk decay is predominant when pipe diameters exceeds 500 mm, which is the case in most of this network.

[Figure 2](#) below shows a typical distribution of calculated decay coefficients as a function of the time of year experienced at two of the monitoring points in the network. With South Africa being in the Southern hemisphere, the colder months are typically from April to July, and it is seen that the decay coefficient decreases in the colder months

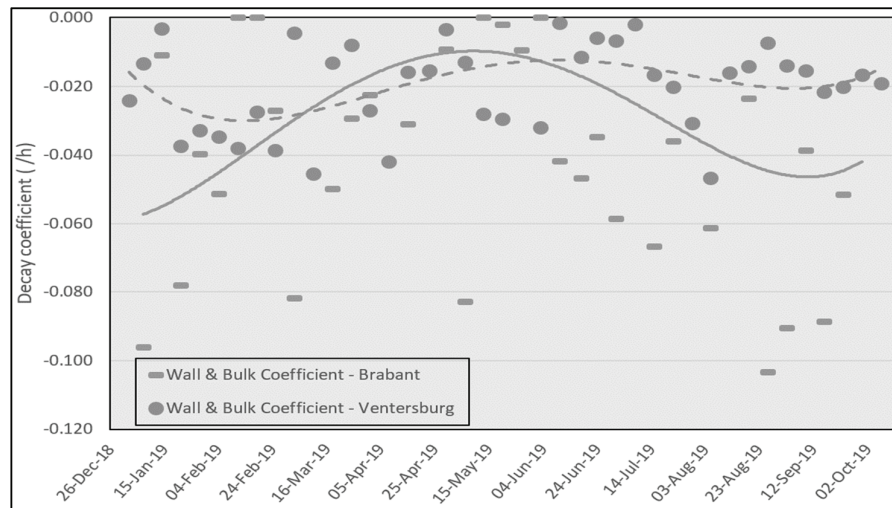


Figure 2: Calculated decay coefficients for two monitoring points in the network

Table 1 below shows the average actual measured free chlorine values per node in the supply system, compared with the calculated EPANET free chlorine values.

Table 1: Calculated EPANET model values compared with actual measured values

Item	Summer data			Winter data		
	Actual	EPA-NET	Absolute variance (%)	Actual	EPA-NET	Absolute variance (%)
Balkfontein Old Reservoir	2.38			2.68		
Balkfontein New Reservoir	2.11			2.70		
Bothaville Reservoir in	1.17	0.71	39.3 %	1.49	1.24	16.8 %
Leeudoringstad Reservoir in	0.73	0.72	1.46 %	1.44	1.43	0.7 %
Re-chlorine Leeudoringstad	1.19			1.69		
Buisfontein Reservoir out	0.51	0.69	35.3 %	1.19	1.40	17.6 %
De Erf reservoir out	1.64	1.33	8.9%	2	2.02	1.0 %
Allanridge Reservoir in	1.32	0.95	28.0 %	1.56	1.63	4.5 %
Allanridge Reservoir out	0.29	0.31	6.9 %	1.1	0.86	21.8 %
Re-chlorine Koppie Alleen	1.79			1.91		
Brabant Reservoir out	1.36	1.36	0.0 %	1.62	1.66	2.5 %
Ventersburg Reservoir out	0.74	0.58	21.6 %	1.06	1.06	0.0 %
Wesselsbron draw-off	1.27	1.28	0.8 %	1.48	1.62	9.5 %
Wesselsbron Reservoir in	0.91	0.74	18.7 %	1.32	1.22	7.6 %
Wesselsbron Reservoir out	0.34	0.30	11.8 %	0.93	0.69	25.8 %
Virginia Reservoir	2.04			1.34		
Dirksburg Reservoir 1 out	0.32	0.33	3.1 %	0.49	0.55	12.2 %
Leeubult reservoirs out	0.18	0.13	27.8 %	0.25	0.22	12.2 %
AVERAGE VARIANCE			16.4 %			10.0 %

4. Discussion

The above sections demonstrated that EPANET can be used successfully to simulate actual conditions in the bulk distribution network regarding hydraulic behaviour and chlorine decay. After calibration of the model, an 84-90 % correlation with actual measured values could be reported. To use this model in practice, a number of “what-if” scenarios have been identified to assist Sedibeng Water in planning purposes for:

- the possible change to an alternative disinfectant like chlorine dioxide and the financial implications thereof;
- the determination of minimum levels of chlorine measured at the two WTP’s to not warrant any additional chlorination in the network; and
- the determination of minimum levels of chlorination at the WTP’s and the booster chlorination stations to ensure adequate free disinfectant levels in the system.

5. Conclusions

The aim of this research was to develop a mathematical model for chlorine decay based on actual conditions prevalent in the bulk distribution system of Sedibeng Water. This model can then be used for operational and planning purposes. The following conclusions were reached following the investigation:

- a) Although the effect of wall decay rates was expected to have a similar impact of bulk decay rates, it was found that the bulk decay coefficient approached the total decay rate coefficient. This can be attributed to the large number of reservoirs in the system, mostly laminar flow conditions being experienced in the pipes and the fact that the majority of bulk pipelines has a diameter larger than 500 mm.
- b) A definite effect of temperature could be seen on the calculated decay rates with the so-called winter decay rate being 50-60% of the so-called summer decay rate.
- c) Using the monitoring values in the reservoirs, a reasonably accurate simulation of hydraulic and disinfectant decay conditions of the Sedibeng Water bulk distribution network could be done by means of the EPANET software. This model has an 84-90% correlation with actual conditions in the network;
- d) It was found that Sedibeng Water already to a large extent has optimised their booster dosing conditions. There is a possibility to further reduce average chlorine consumption at the Balkfontein WTP by 25% and save about USD 65,000 per annum in the process;
- e) It will not be possible to do away with interim chlorination during summer periods due to the impact which high levels of chlorine will have on nearby communities and due to the relatively fast decay rate of chlorine in the network; and
- f) The implementation of chlorine dioxide as secondary disinfectant at the Balkfontein WTP is an option, while the use of interim chlorine booster stations can be discontinued. This option will require an additional USD 27,000 to be spend per annum, with provision made for a once-off capital expense of USD 65,000

References

- Monteiro, L. et al., 2014. Modeling of chlorine decay in drinking water supply systems using EPANET MSX, *Procedia Engineering*, Issue 70, pp. 1192-1200.
- Pasha, M. & Lansley, K., 2010. Effect of parameter uncertainty on water quality predictions in distribution systems-case study. *Journal of hydroinformatics*, Volume 12, pp. 1-21
- Fisher, I. et al., 2011. Suitability of Chlorine Bulk Decay Models for Planning and Management of Water Distribution Systems. *Critical Reviews in Environmental Science and Technology*, Volume 41, pp. 1843-1882.
- Rossman, L., 2000. EPANET 2 user's manual. U.S. Environmental Protection Agency.
- Sathasivan, A., Chiang, J. & Nolan, P., 2009. Temperature dependence of chemical and microbiological chloramine decay in bulk waters of distribution systems. *Water Science & Technology: Water Supply-WSTW*
- Fisher, I., Kastl, G. & Sathasivan, A., 2017. New Model of Chlorine-Wall Reaction for Simulating Chlorine Concentration in Drinking Water Distribution Systems. *Water Research*, Volume 15, pp. 427-437

Compartmental Modelling of Shell Side Fouling in a Shell and Tube Heat Exchanger

Renat Targalinov^a, Federico Lozano Santamaria^a and Sandro Macchietto^{a*}

^a*Imperial College London, South Kensington Campus, London SW7 2AZ, UK*

s.macchietto@imperial.ac.uk

Abstract

Accurate and fast models are required to optimise the design, control and operations of shell and tube heat exchangers (STHE) subject to fouling. To date, research effort has mainly focused on tube side fouling. Techniques proposed for the shell side fouling are limited to i) mechanistic thermo-hydraulic models with simple Flow Stream Analysis (FSA) and ii) Computational Fluid Dynamics (CFD) models. The former ignore flow dynamics on the shell side. The latter cannot quantitatively predict fouling, are computationally very heavy and cannot be utilised for optimisation and control.

In the paper, we combine the benefits of FSA and CFD methods by creating a hybrid Compartmental Model (CM). For an actual exchanger subject to crude oil fouling we analyse its outcomes and compare them to CFD and FSA methods. A preliminary CFD study yields detailed shell side velocity field information, based on which an appropriate compartments network is created. A simulation is performed using a two dimension (2D) distributed dynamic STHE model which utilises a FSA for the shell side. A dynamic model for the compartments with a shell side threshold fouling model is developed, utilising selected results from the CFD (velocity data) and 2D-FSA (heat duty) studies. Results from the CM model and comparison with the CFD and 2D-FSA models, for the same operation, provide valuable insights regarding the role of shell side velocities in predicting overall exchanger performance.

Computationally, for this case study, the CM model is solved 5.5 times faster than the distributed dynamic model and 36 times faster than the CFD model, indicating the approach has good potential for use in design and operations optimisation.

Keywords: Compartmental multi-zonal hybrid model, CFD model, Heat exchanger shell side, crude oil fouling, Flow Stream Analysis.

1. INTRODUCTION

Accurate and robust models of STHEs under fouling are required for design, control and operations support purposes, with large economic and safety implications (Diaz-Bejarano et al., 2018). A choice must be made between model detail and usability. Mechanistic thermo-hydraulic models can simulate mass and heat transfer together with complex reaction kinetics and separations in relatively simple geometries and over long periods in reasonable computer times and are therefore useful for control and operational studies. Computational Fluid Dynamics (CFD) models, on the other hand, can describe complex fluid dynamics in far greater detail in complex geometries, however only over short periods and at the expense of long computation times. Compartmental Models (CM) aim to achieve a compromise between the two. There is a

rich literature on their use (Jourdan et al., 2019), however, not for shell and tube heat exchangers (STHE) subject to fouling. Here, a CM model, specifically capable of dealing with fouling on the shell side, is presented. A direct comparison of three approaches (CFD, a 2D distributed thermo-hydraulic and CM) is also presented.

1.1. Shell and tube heat exchanger models

To date, research effort has mainly focused on fouling on the tube side. Shell side fouling is very common but has not been adequately studied. Various mechanistic thermo-hydraulic heat exchanger models can deal with fouling (see Diaz-Bejarano et al. 2017 for a review). Lumped parameter type models do not consider spatial variations in the heat exchanger. Distributed parameter models account for spatial changes axially along the tubes and radially inside a tube (e.g. Coletti and Macchietto, 2011, Diaz-Bejarano et al., 2016) resulting in much better predictions.

Estimating the thermal and hydraulic performance on the shell side is challenging due to the complex flow distribution induced by the equipment geometry (e.g. segmental baffles, bundle supports, flow bypasses) (Figure 1). The Bell-Delaware and Flow Stream Analysis (FSA) (Hewitt, 2008) approaches, extensively used in industry for the shell side in a clean state, do not account for fouling. All the thermo-hydraulic lumped models and most of the distributed ones discussed above neglect shell side fouling. A notable exception is Diaz-Bejarano et al. (2018), who couple a shell side Flow Stream Analysis with a typical reaction fouling model, to calculate the progressive occlusion of flow bypasses. This model, implemented in Hexxcell Studio (2018) is used later. By its nature, this approach does not reflect the spatial distribution of velocities and all related quantities (e.g. Reynolds no., deposition, heat transfer) on the shell side.

CFD models calculate detailed fluid dynamics and heat transfer taking into account irregular flow distribution patterns (e.g. vorticity, recirculation joints, etc.), wall effect and boundary layer conditions. For particulate fouling in crude oil, Emani et al. (2017), calculated the particles transportation, adhesion and removal trajectories. This gives only a qualitative assessment of mass, thickness and locations of fouling. For chemical reaction fouling CFD still lacks the ability to predict accurately reaction kinetics and fouling rates. Chambon et al. (2019) calculated a-posteriori steady state fouling rates based on velocity and temperature profiles. For dynamics, CFD fouling simulations is normally restricted to seconds or minutes. Many simplifications are used, such as decoupling shell and tube sides (Wang et al., 2012, Yang et al, 2015, Emani et al., 2019; Gounder & Emani, 2018). Altogether, this presently prevents the use of CFD models for control and operations optimisation.

Compartmental Models (CM) achieve a compromise between the above approaches. CFD views an object as a very large number of small cells, CM as a small number of large, approximately homogeneous, compartments (macro cells), mapped to a finer CFD grid (Figure 2 left). Equations in a compartment detail the processes within it (e.g. heat exchange, fouling), while convective and turbulent fluxes between them define their interaction (Figure 2 right) (Bezzo *et al.*, 2000; Haag *et al.*, 2018).

1.2. Case study, CFD model, Compartment model and 2D distributed model

A case study is drawn from a well-documented CFD paper (Chambon et al., 2019), from which exchanger data (single pass, 4 baffles), properties of both fluids (residue in the shell side and crude oil in the tube side) and inlet operating conditions (Table 1), were obtained (additional details in Taurgalinov, 2020). A CFD model of the exchanger

was developed in COMSOL Multiphysics 5.4 (COMSOL Inc., 2019). A simulation without fouling gives the velocity fields on tube and shell sides (Figure 3). As expected, the shell side velocity field is greatly affected by the baffles, with low velocity wakes and recirculation joints. All central and outlet baffle spacing have a low and a high velocity zones. This CFD simulation took 72 minutes. Additional cylinders (around the tube bundle and avoiding the bypass clearance) of various lengths allowed calculating average shell side velocities in the corresponding domains (Figure 4, top left).

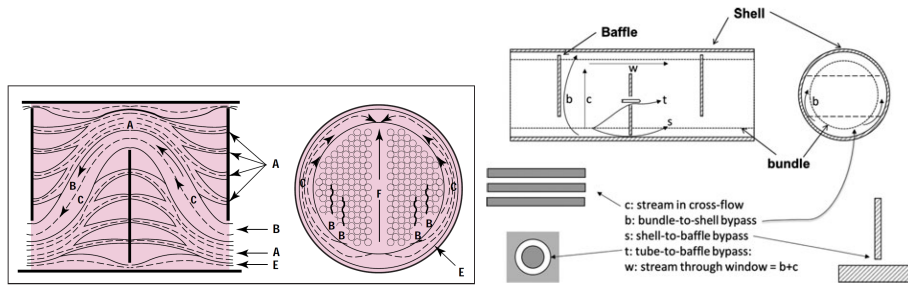


Figure 1 Shell-Side Flow Distribution (left, Mukherjee, 1998) and Flow Streams (right, Diaz-Bejarano et al., 2018)

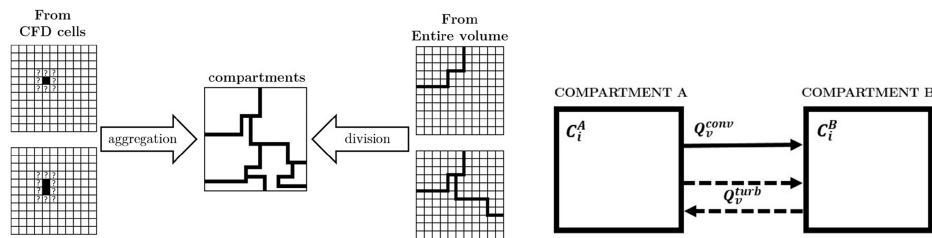


Figure 2 Compartment Formation through Aggregation or Division (left) and Exchange Fluxes between Compartments (right) (Haag et al., 2018)

Based on these flow distributions, the shell side was divided into 9 cylindrical compartments (1 for the residue inlet section, 2 for each baffle section and 2 for the exit section), mapped onto the exchanger and linked between them as shown in Figure 4 (bottom). Each compartment model is structured as detailed in Figure 4 (top right). A single tube is considered to represent all tubes in the bundle, with a fouling layer deposited on the outside (shell side) according to the prevailing conditions in that compartment (for ease of assessment, no fouling is assumed to occur on the tube side). Due to space limitations selected equations only are given in Table 2. Deposition on the outside of the tube is assumed to follow the Ebert and Panchal model (Panchal et al., 1997), with conductive heat transfer through the tube wall and the fouling layer in series in the radial direction. Velocities of shell-side flows in each compartment are based initially (clean bundle) on averages from the CFD simulation, and then adjusted for fouling according to the gap reduction between the tubes (Eq. 5 in Table 2). Shell side heat transfer is modelled as forced convection across cylinders (Hilpert, 1933). The CM model was implemented in gPROMS (2018). The same exchanger was simulated using the 2D-FSA dynamic model in Hexxcell Studio (labelled HS). The CFD model does not include fouling, HS does, using FSA on the shell side. Comparison runs assumed no fouling on either side, and fouling only on the shell side with the same fouling parameters. Full details are reported in Taurgalinov (2020).

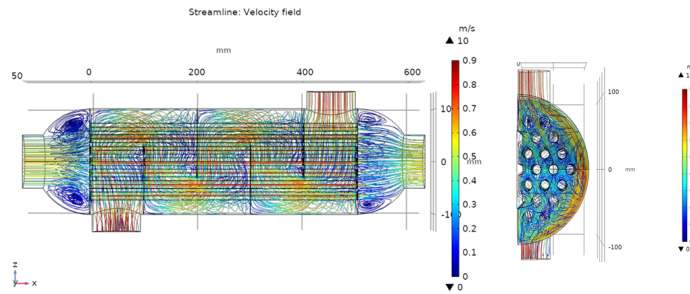


Figure 3 CFD Streamline Plots of Velocity Field: z-x plane and z-y plane views

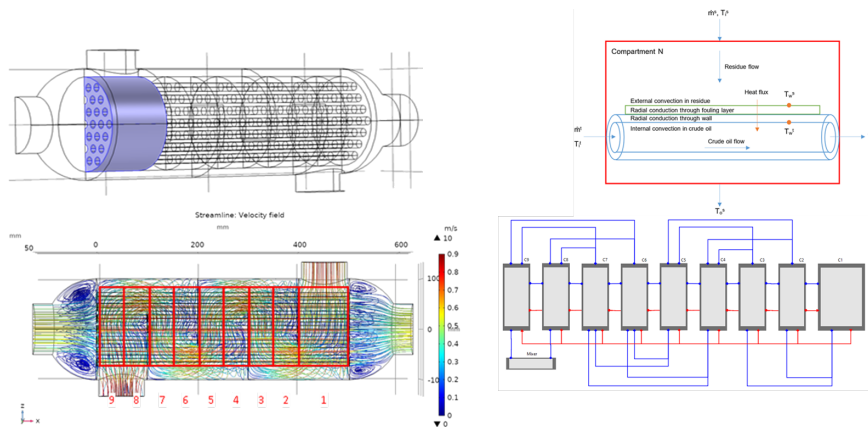


Figure 4 A sectional fluid domain for calculation of average velocities (top left); Mapping between CM and CFD grids (bottom left); Structure of CM tube and shell side model (top right); Network structure linking all CM modules (bottom right)

Table 1 Properties of Materials and Inlet Conditions

	Units	Residue	Crude	Carbon Steel
Density	kg/m ³	730	675	7800
Dynamic viscosity	Pa s	$7.23 \cdot 10^{-4}$	$3.75 \cdot 10^{-4}$	
Thermal conductivity	W/m/K	$8.94 \cdot 10^{-2}$	$9.18 \cdot 10^{-2}$	45
Specific heat capacity	J/kg/K	2925	2980	470
Inlet velocity	m/s	0.7	0.5	
Inlet temperature	°C	314.3	293.9	

2. Results and conclusions

The dynamic CM model runs well and calculates all variables. Without fouling, CM predicts an overall heat duty (5.949 KW) within 17.9% of the CFD result (5.045 KW), while the HS duty (7.451 KW) is 25.2% and 47.7% higher than with CM and CFD. Notably, the FSA model in HS gives a shell side velocity of 0.52 m/s, much larger than the CFD velocities for the flow through the tube bundle (0.24-0.34 m/s), also used in CM. The difference in velocities leads to differences in velocity-dependent quantities (e.g. Reynolds No., fouling deposition, deposit depth, exit temperatures). As the current CM model does not include the faster bypass flows in the gap between bundle and shell,

a by-pass flow split factor (f_b in eq. 5, Table 2) was adjusted so as to match the clean HS shell velocity, achieved with $f_b = 1.74$. This factor, assumed constant, was used in the dynamic simulation with shell side fouling. Figure 5, left shows the bulk temperature of the shell fluid along the exchanger at Days 0 and 40. The HS profile is linear, while the CM profile is “discretised” in the 9 different compartments. There are visible differences in the two compartments corresponding to parallel flows within two baffles. Differences are reduced as fouling progresses. Figure 5, right shows the predicted shell-side Fouling Thickness and Velocity for CM and HS. As before, differences in predicted velocities result in different deposit thickness.

Table 2 Selected equations in Compartment model (superscripts are omitted for clarity)

Nº	Equations	Notes and assumptions
Tube wall +Shell-side Deposit Domain		
1	$\frac{dR_f}{dt} = \alpha Re^{\beta} Pr^{\gamma} e^{-\frac{E}{RT_f}} - \gamma \tau$	Ebert and Panchal threshold fouling model (Panchal et al., 1997), with α , β , γ , and E given. R is the gas constant.
2	$\delta = R_f k_f$	k_f is assumed uniform, constant and known.
3	$Q = \frac{2\pi L(T_w^s - T_w^t)}{\ln\left(\frac{D_o}{D_i}\right)/k_w + f_c \ln\left(\frac{D_o + 2\delta}{D_o}\right)/k_f}$	Conductive radial heat transfer through the tube wall and fouling deposit layer (i.e. conduction in series). Adjustment factor (f_c) is applied to control the effect of fouling on heat conduction. k_f , k_w and all geometric parameters are assumed known.
Shell Side Domain		
5	$u = \frac{u_{in} f_u f_b}{\left(1 - \frac{2\delta}{x}\right)}$	Shell side local flow velocity u depends on: i. initial velocity relative to inlet velocity u_{in} ii. clearance between tubes due to fouling deposit of depth δ iii. compartment split factor (f_u), from CFD iv. bypass flow split factor (f_b).
13	$\frac{h_o(D_o + 2\delta)}{k} = C Re^m Pr^{1/3}$	Forced convection across cylinders (Hilpert, 1933). Constants C and m based on shell side Reynolds no. (range 4000 – 40000).

Compared to models based on FSAs, CM enables calculating the shell side fouling rates along the exchanger as a function of the local (average) conditions in each compartment. The initial results in this work suggest that some refinements may be useful in respect of selection of compartments (e.g. adding compartments for the parallel flow portions or the radial gap between bundle and shell) and velocity split factors. While the FSA methods mainly aim at getting the overall thermal behaviour right, the method proposed aims mainly at incorporating the shell side hydraulics behaviour, with an input from CFD calculations. A way of adjusting the compartment split and bypass flow split factors to improve model predictions matching (if desired), beyond their initial setting may be useful. Independent experimental verification of shell side velocities would be useful to improve both CM and fully distributed models.

Computationally, the case study presented was solved with the CM model 5.5 times and 36 times faster than with the 2D distributed and CFD models, respectively, indicating the approach has good potential for use in design and operations optimisation.

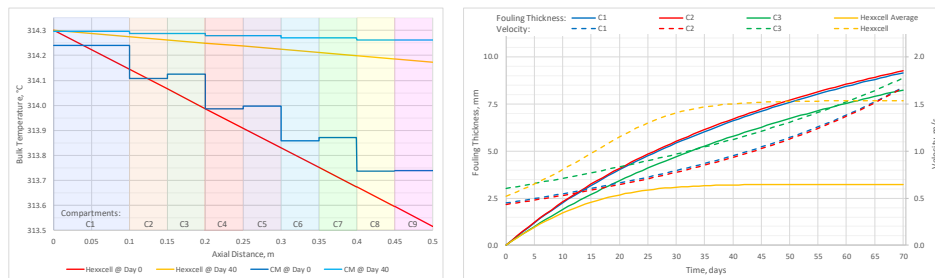


Figure 5 Simulation results with CM (bypass split factor $f_b = 1.74$) and HS: Bulk Temperature along STHE at Days 0 and 40 (left); Fouling Thickness and Shell Velocity (right)

References

- Bezzo F, S. Macchietto, C. C. Pantelides, 2000. A general framework for the integration of computational fluid dynamics and process simulation, *Comput Chem Eng*, 24, 653-658.
- Chambon, A., Anxionnaz-Minvielle, Z., Fourmigué, J.F., Guintrand, N.A. Davaille' and F. Ducros, 2019. Use of Distributed Threshold Fouling Model at Local Scale in Numerical Simulations. *Ind. Eng. Chem. Res.*, 58 (32), 15003–15013.
- Coletti, F. and S. Macchietto, 2011. A dynamic, distributed model of shell-and-tube heat exchangers undergoing crude oil fouling. *Ind. Eng. Chem. Res.*, 50 (8), 4515–4533.
- COMSOL Inc., 2019. *COMSOL Multiphysics 5.4 (Build 246)*. <https://www.comsol.com>.
- Diaz-Bejarano, E., F. Coletti and S. Macchietto, 2016. A New Dynamic Model of Crude Oil Fouling Deposits and Its Application to the Simulation of Fouling-Cleaning Cycles. *AIChE J.* 62 (1), 90–107.
- Diaz-Bejarano, E., F. Coletti and S. Macchietto, 2017. Thermo-hydraulic analysis of refinery heat exchangers undergoing fouling. *AIChE J.* 63, 3, 984-1001.
- Diaz-Bejarano, E., F. Coletti and S. Macchietto (2018) Modelling and Prediction of Shell-Side Fouling in Shell-and-Tube Heat Exchangers. *Heat Transf. Eng.*, 40 (11), 845-861.
- Emani, S., M. Ramasamy and K.Z.K Shaari, 2017. CFD modelling of shell-side asphaltene deposition in a shell and tube heat exchanger. *AIP Conference Proceedings*, 1859(1):020118, doi:10.1063/1.4990271.
- Emani, S., M. Ramasamy and K.Z.K Shaari, 2019. Discrete phase-CFD simulations of asphaltene particles deposition from crude oil in shell and tube heat exchangers. *Appl. Therm. Eng.*, 149, 105–118.
- Gounder, R.M. and S. Emani, 2018. CFD Simulations of Crude Oil Fouling on Heat Transfer Surfaces. In *Computational Fluid Dynamics - Basic Instruments and Applications in Science*, A Ionescu Ed., doi:10.5772/intechopen.71886.
- Haag, J., Gentric, C., Lemaitre, C. & Leclerc, J.P. (2018) Modelling of Chemical Reactors: From Systemic Approach to Compartmental Modelling, *Int J Chem React Eng*, 16 (8), 1–22.
- Hewitt, G.F. (2008) Flow stream analysis method for segmentally baffled shell and tube heat exchangers. In: G F Hewitt (ed.). *Heat Exchanger Design Handbook*, Begell House, Redding, NC, USA.
- Hexxcell Limited (2018) *Hexxcell Studio*. <https://www.hexxcell.com>.
- Hilpert, R. (1933) Wärmeabgabe von Geheizten Drähten und Röhren. *Forsch. Gebiete. Ingenieurw.* 4, 220.
- Jourdan, N., T. Neveux, O. Potier, M. Kanniche, J. Wicks, I. Nopens, U. Rehman, Y. Le Moulec, 2019. Compartmental Modelling in chemical engineering: A critical review. *Chem. Eng. Sci.*, 210, 115196.
- Mukherjee, R., 1998. Effectively design hell-and-tube heat exchangers. *Chem. Eng. Prog.* 94 (2), 21–37.
- Panchal, C.B., W.C. Kuru, C.F. Liao, W.A. Ebert and J Palen, 1997. Threshold conditions for crude oil fouling. In: T R Bott (ed.). *Understanding Heat Exchanger Fouling and its Mitigation*. Lucca, Italy, Begell House. Redding, NC, USA, pp. 273–281.
- Process Systems Enterprise Ltd. (2019) *gPROMS (Build 6.0.2.55126)*, <https://www.psenterprise.com>.
- Taurgalinov, R., 2020. *Modelling of Shell Side Fouling in a Shell and Tube Heat Exchanger*. Imperial College London.
- Wang, Y., Q. Lu, Y. Zheng, B. Yuan and T. Hanzhong, 2012. A CFD-based analysis on trends of heat exchanger fouling. Power and Energy Engineering Conference (APPEEC), Asia-Pacific. 1-4. 10.1109/APPEEC.2012.6306963.
- Yang, J., Serratos, M.G.J., Fari-Arole, D.S., Müller, E.A., et al. (2015) Crude Oil Fouling: Fluid Dynamics, Reactions and Phase Change. *Procedia IUTAM*. 15, 186–193.

Optimization-based digital design of a commercial pharmaceutical crystallization process for size and shape control

Ayşe Eren^a & Botond Szilágyi^a, Justin L. Quon^b, Charles D. Papageorgiou^b, Zoltán K. Nagy^{a*}

^a*Davidson School of Chemical Engineering, Purdue University, 610 Purdue Mall, West Lafayette 47907, United States*

^b*Takeda Pharmaceuticals International Co., 40 Landsdowne St, Cambridge, MA 02139, United*

znagy@purdue.edu

Abstract

This work demonstrates a systematic development of a digital twin of a commercial active pharmaceutical ingredient (API) - Compound A - crystallization processes to be used for in-silico design of experiments (DoE) and process optimization. Eleven highly correlated kinetic parameters were estimated by exploiting an optimization approach with an evolutionary algorithm, to fit to concentration profile, mean size and the shape distribution of the product. Two more experiments were then performed to validate the model showing that the developed model could predict the product size and shape properties. The digital twin of the process was used to perform in silico design of experiments to understand the attainable operating space of the system and provide a framework for rapid quality-by-design (QbD). Subsequently the model was used in a dynamic optimization framework for the digital design of the API crystallization process to achieve the critical quality attributes (CQAs) without extensive experimentation.

Keywords: Digital twin, crystallization, process optimization, digital design

1. Introduction

Prediction and control of the product properties in crystallization processes represent significant challenges in the pharmaceutical industry. To meet the critical quality attributes (CQAs) with the least variation among batches, effective crystallization process design and operation techniques are needed. Hence, monitoring, modelling, and control of CSD has become the focal point of the QbD driven crystallization process design (Nagy and Braatz, 2012). The paradigm shifts in pharmaceutical crystallization from Quality-by-Testing (QbT) to QbD promoted the extensive usage of process analytical technology (PAT) tools for the direct measurement of CQAs. PAT tools not only enabled more profound process understanding but also opened the way for the application of PAT-based process control strategies and model development and have become standard instruments in crystallization process design (Saleemi, Onyemelukwe and Nagy, 2013).

From an experimental perspective, DoEs are simple, legacy tools for design space identification. The number of necessary experiments depends on the number of factors, and the levels associated to each factor. As the number of design variables increase, for

example, temperature cycles are included or combined anti-solvent and cooling crystallization process is considered, the number of experiments will increase substantially. Being labour and material intensive, this is undesired in the pharmaceutical industry. Moreover, material availability can become a bottleneck in the case of APIs under early phase clinical trials for designing such complex crystallization processes, in which case the application of mathematical modelling tools for *in silico* experimental design can be advantageous (Rosenbaum *et al.*, 2019).

The objective of this work is a process model development for the batch cooling crystallization of Compound A that is suitable for quick process simulations (for *in-silico* DoE) and process optimizations. The Compound A has high nucleation tendency and forms slow-growing high aspect ratio crystals, which requires the implementation of temperature cycles for internal fines removal and for aspect ratio improvement. Secondary nucleation, growth and dissolution of crystals is considered in the one-dimensional (1D) PBM, which is solved numerically with a high-resolution finite volume method.

2. Experimental Methods, Model Development, Parameter Estimation, and Process Optimization

Six batch cooling experiments were performed with varying operating parameters such as temperature cycle number, seed loading, seeding temperature and cooling rate to be used in model training. Two more experiments with same parameters varied were conducted after parameter estimation to validate the model in the same batch system. The experimental space is given in table 1 summarizing these changing operation conditions for these experiments.

Table 1. Experimental space for the parameter estimation calculations.

Name	C.R. (°C/min)	# Cycles	Seed load (%)*	T_{sat} (°C)	T_{seed} (°C)	T_{min} (°C)	T_{max} (°C)
Cal #1	0.1	1	1	70	65	40	60
Cal #2	0.1	2	1	75	68	45	65
Cal #3	0.1	4	1	70	65	50	70
Cal #4	0.1 ... 0.3**	3	1	75	68	55	69
Cal #5	0.2	1	3	75	68	20	68
Cal #6	0.2	2	1	70	65	40	65
Val #1	0.2	1	3	70	65	20	65
Val #2	0.2	2	2	75	68	20	70

*% of expected product mass, **parabolic cooling was applied between the given rates to understand better the crystallization behaviour.

In each experiment, PAT tools were utilized for data acquisition and monitoring. UV/Vis spectrophotometer, focused beam reflectance measurement (FBRM), and Particle Vision Measurement (PVM) were used to measure concentration, particle count, morphology and size. In addition, Malvern Mastersizer 3000 and Nikon SMZ1500 microscope were used for offline characterization of size distribution and crystal morphology. A supervisory control software, CryMOCO, was used to implement temperature cycles with

different cooling/heating rates (Simone, Zhang and Nagy, 2015). Population balance model (PBM) was used to describe the crystallization of Compound A considering nucleation, size dependent growth (SDG) and dissolution (SDD) mechanisms based on the experimental data as shown below. The monovariate population density function is introduced and noted by $n(L, t)n(L)$, which gives the number of particles in time t within the size domain $L, L + dL$.

$$\frac{\partial n(L)}{\partial t} + \frac{\partial [Gn(L)]}{\partial L} = B_s \delta(L - L_n) \tag{1}$$

$$\frac{\partial n(L)}{\partial t} - \frac{\partial [Dn(L)]}{\partial L} = 0$$

$$\frac{dc}{dt} = \begin{cases} \{-k_v \rho_c \left(3 \int_0^\infty GL^2 n(L) dL + B_s L_n^3 \right), & \text{if } \sigma \\ > 0 \ 3k_v \rho_c \int_0^\infty DL^2 n(L) dL, & \text{else} \end{cases} \tag{2}$$

$$B_s = k_s \sigma^s V_c, G = k_g \sigma^g \exp\left(-\frac{E_A}{RT}\right) G = k_{g,0} \frac{1}{2} (\tanh \tanh(k(L_c - L)) + 1) (1 + \alpha L^Y) \sigma^g \exp\left(-\frac{E_a}{RT_K}\right), \beta(\lambda, L) = D = k_d (-\sigma)^d D = k_d(L) \sigma \tag{3}$$

$$D = (1 + \alpha_d L)^{-Y_d} \sigma$$

Finding the values of parameters was carried out by solving an optimization problem as shown:

$$O(P) = \frac{1}{I_c J_c} \sum_{i=1}^{I_c} \sum_{j=1}^{J_c} (c_{e,ij} - c_{s,ij})^2$$

$$+ \frac{w_L}{I_L J_L} \sum_{i=1}^{I_L} \sum_{j=1}^{J_L} \int_0^{L_{max}} [v_{e,ij}(L) - v_{s,ij}(L)]^2 dL \tag{4}$$

$$+ \frac{w_N}{I_N J_N} \sum_{i=1}^{I_N} \frac{1}{\mu_{0,i}^2} \sum_{j=1}^{J_N} R_{ij}^2$$

where the first term minimizes the differences between the measured ($c_{e,ij}$) and simulated ($c_{s,ij}$) concentrations in all experiments having concentration data (I_c), in all concentration sampling instance (J_c). The second term minimizes the difference between the measured ($v_{e,ij}(L)$) and simulated ($v_{s,ij}(L)$) cumulative CSDs in all experiments (I_L) in all CSD sampling instance (J_L). The last term is another novelty introduced in this work. It introduces the maximization of correlation between the FBRM count and the simulated number density function, in all experiments (I_N) and in all FBRM sampling instance (J_N). The parameters w_L and w_N are weight factors. In the last term of the objective function, $\mu_{0,i}$ denotes the mean simulated crystal number density in the i^{th} experiment, and R_{ij} is defined as:

$$R_{ij} = \mu_{0,ij} - F_{ij} \tag{5}$$

where F_{ij} is the number density calculated by the best fitting first order polynomial on the points in the FBRM count-simulated number density diagram. Therefore, R_{ij} is the

j^{th} residual of the corresponding first order polynomial fit in the i^{th} experiment. The parameters P were determined solving the multivariable minimization problem:

$$O(P) \quad (6)$$

where R_+^0 denotes the set of non-negative real numbers.

In this work, a novel SDG formulation was used combining the different behaviours of commonly used SDG rate equations as shown with G in equation 3 (Mydlarz and Jones, 1990). The reason to develop a new SDG formula was that the experimentally observed broadening in the CSD and obtaining product size distributions was not reachable with the seed CSD used in the experiments. Given the experimental observations that the small particles can get confined in the eddies and deplete the local supersaturation, one might hypothesize that strong size dependent character can be expected for the crystal size range comparable to the Kolmogorov length scale of turbulence, and there is a critical crystal size under which the particles are virtually not growing. Therefore, we looked for an equation that provides significant size dependency in the small crystal size range. The different behaviours of all the SDG formulas were examined and these were tested in the model. The behaviours of these models are shown in Figure 1 and these models are named as: Quasi power-law (QPL), exponential, hyperbolic tangent dependency (HT), and extended hyperbolic tangent formula (E-HT) (Equation 3 in this paper).

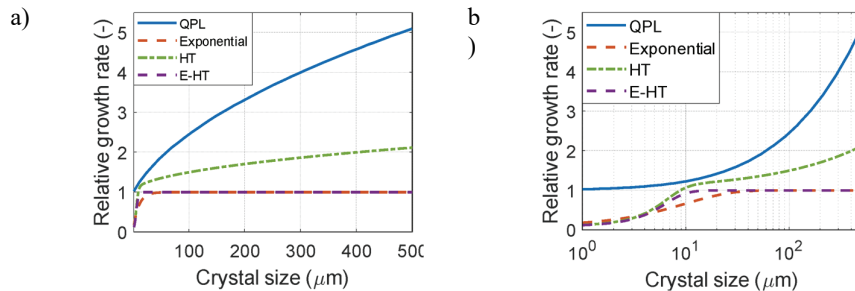


Figure 1. Comparison of the SDG functions considered in this work: the hyperbolic tangent decouples the size dependency in the small and large size domains. a) linear and b) logarithmic size scale.

3. Results and Discussion

Four SDG models are applied: QPL and exponential, as well as HT and E-HT. HT was also considered as it has the same number of parameters as model QPL and exponential. Two parameter estimations were carried out for all model structures: with and without using the FBRM count data. Although the objective function values differ depending on the usage of FBRM data, the general conclusions remain unchanged. Namely, amongst the three parameter SDG models the QPL one gave the worst calibration performance. Exponential model gave an order of magnitude lower objective function value, whereas another order of magnitude decrease was achieved by the hyperbolic tangent model. For these models, the difference in achievable performance is explained with the shape of relative growth rate curves (see Figure 1). Extending the tangent hyperbolic model HT to E-HT, improved the CSD shape fit, but less prominent objective function value decrease

is achieved. The same trends exist in the case of validation experiments, which is a good indication that there is no overfitting. These results underline the applicability of hyperbolic tangent function family as SDG model. Table 2 lists the kinetic parameters estimated using the FBRM count data in the proposed way. The first observation is that the width of confidence intervals (CI) became better i.e. less narrow for the QPL and exponential models and significantly narrower for the proposed SDG formula (E-HT). It is hypothesized that in the case of using E-HT formula, the FBRM count term contributed to de-coupling of nucleation and growth kinetics, which reduced greatly the inter-correlation of parameters between these mechanisms.

Table 2. Parameter estimation results for the considered model structures with the 95 % CI limits for the parameter estimation using the FBRM data.

Parameter	(units)	Nominal parameters (\pm confidence interval limits)			
		QPL	Exponential	HT	E-HT
$\log_{10}(k_s)$	(#/m ³ s)	$-2.8 \pm 7.6 \cdot 10^{-4}$	-2.55 ± 0.36	$-3.04 \pm 6.1 \cdot 10^{-2}$	-3.06 ± 0.21
s	(-)	$6.56 \pm 9.5 \cdot 10^{-4}$	1.85 ± 0.58	$2.74 \pm 5.0 \cdot 10^{-2}$	2.25 ± 0.37
$\log_{10}(k_g)$	($\mu\text{m/s}$)	$9.46 \pm 5.2 \cdot 10^{-4}$	10.08 ± 0.24	$10.86 \pm 1.9 \cdot 10^{-2}$	10.79 ± 0.79
g	(-)	$2 \pm 8.0 \cdot 10^{-4}$	$1.05 \pm 8.6 \cdot 10^{-2}$	$1.95 \pm 8.5 \cdot 10^{-3}$	1.04 ± 0.17
$\log_{10}(E_A)$	(J/mol K)	$4.90 \pm 5 \cdot 10^{-8}$	$4.83 \pm 8.7 \cdot 10^{-3}$	$4.88 \pm 1.4 \cdot 10^{-3}$	$4.92 \pm 2.4 \cdot 10^{-2}$
$\log_{10}(\alpha)$	(-)	$1.7 \cdot 10^{-3} \pm 2 \cdot 10^{-6}$	$-2.93 \pm 8.9 \cdot 10^{-2}$	-	-0.90 ± 0.39
γ	(-)	$0.5 \pm 3 \cdot 10^{-6}$	$-0.73 \pm 7.1 \cdot 10^{-2}$ *	-	0.84 ± 0.14
$\log_{10}(k)$	(-)	-	-	$-0.84 \pm 5.7 \cdot 10^{-3}$	$-0.81 \pm 2.9 \cdot 10^{-2}$
$\log_{10}(L_{cr})$	(μm)	-	-	$1.19 \pm 5.6 \cdot 10^{-3}$	$1.05 \pm 1.4 \cdot 10^{-2}$
$\log_{10}(k_d)$	($\mu\text{m/s}$)	$-0.52 \pm 2 \cdot 10^{-7}$	-0.41 ± 0.24	$-1.35 \pm 7.2 \cdot 10^{-2}$	$-1.90 \pm 3.0 \cdot 10^{-2}$
$\log_{10}(\alpha_d)$	(-)	$0.3 \pm 3 \cdot 10^{-6}$	0.057 ± 0.138	-0.54 ± 0.35	-3.19 ± 105.6
γ_d	(-)	$0.5 \pm 2.5 \cdot 10^{-8}$	$1 \pm 5.9 \cdot 10^{-3}$	$0.34 \pm 8.8 \cdot 10^{-2}$	1 ± 237.6

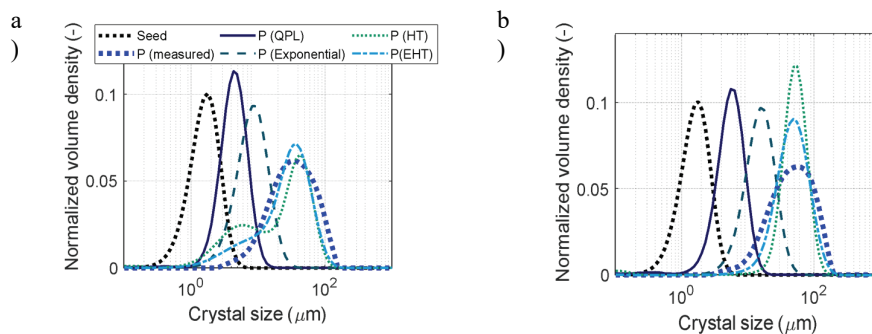


Figure 2. Model performance to predict the product CSD denoted as P (model) in the validation simulations (not used in the kinetic parameter estimations). a) Validation expt. #1 b) Validation expt. #1 (see table 1 for details).

Figure 2 presents the model performance for the two validation experiments for product CSD prediction. According to both validation simulations the QPL model showed minimal size increase throughout the process, whereas the exponential model predicted slightly larger product sizes, but still much below the measured values. Both hyperbolic tangent models captured the mean product size reasonably. The width of the distribution

given by the E-HT is broader than the HT, but still narrower than the measured product CSD, while the shape could be captured better. In addition to the model simplifications, the Mastersizer's tendency to overestimate the width of the CSD should also be mentioned as a contributing factor to this mismatch (Rønneest *et al.*, 2012).

4. Conclusions

The paper demonstrated a workflow for the crystallization process model development and calibration of a commercial API, Compound A. This model was needed for executing numerical experimentations and simulation-based DoE that permits to analyse and design the Compound A crystallization process with less experimental burden. The model developed in this work introduced two novelties: 1) new SDG formulation and 2) integration of crystal count in the objective function for parameter estimation shrinking the confidence hyper ellipsoid during parameter estimation and enhancing the search process, while making FBRM a mainstream PAT tool for modelling purposes.

The traditionally applied growth rate models are compared with a newly proposed hyperbolic tangent based equation, that is well-known in the field of crystal breakage and shown superior ability to capture the experimentally measured CSD dynamics in this case. Additionally, a graphical user interface (GUI) is developed that enables quick and coding-free (standalone application) process simulation for the batch cooling crystallization processes. This is not only suitable to simulate the crystallization of Compound A, but by changing the appropriate kinetic and thermodynamic parameters it can be used as a generic, simple to use cooling batch crystallization simulator.

Acknowledgements

Financial support from Takeda Pharmaceuticals Int. Co. is gratefully acknowledged.

References

- Mydlarz, J. and Jones, A. G. (1990) 'On modelling the size-dependent growth rate of potassium sulphate in an msmpr crystallizer', *Chemical Engineering Communications*, 90(1), pp. 47–56.
- Nagy, Z. K. and Braatz, R. D. (2012) 'Advances and New Directions in Crystallization Control', *Annual Review of Chemical and Biomolecular Engineering*, 3(1), pp. 55–75.
- Rønneest, N. P. *et al.* (2012) 'Comparison of laser diffraction and image analysis for measurement of *Streptomyces coelicolor* cell clumps and pellets', *Biotechnology Letters*, 34(8), pp. 1465–1473.
- Rosenbaum, T. *et al.* (2019) 'Population Balance Modeling to Predict Particle Size Distribution upon Scale-Up of a Combined Antisolvent and Cooling Crystallization of an Active Pharmaceutical Ingredient', *Organic Process Research and Development*, 23(12), pp. 2666–2677.
- Saleemi, A., Onyemelukwe, I. I. and Nagy, Z. (2013) 'Effects of a structurally related substance on the crystallization of paracetamol', *Frontiers of Chem. Sci. & Eng.*, 7(1), pp. 79–87.
- Simone, E., Zhang, W. and Nagy, Z. K. (2015) 'Application of Process Analytical Technology-Based Feedback Control Strategies To Improve Purity and Size Distribution in Biopharmaceutical Crystallization', *Crystal Growth & Design*, 15, pp. 2908–2919.

CONDITION NUMBER AS A QUANTITATIVE MEASURE OF FLEXIBILITY IN A PROCESS

César Ramírez-Márquez,^{a*} Luz Gabriela Arreola-Nájera,^a Juan Gabriel Segovia-Hernández^a

^aUniversity of Guanajuato, Campus Guanajuato, Division of Natural and Exact Sciences, Department of Chemical Engineering, Noria Alta S/N, 36050, Guanajuato Gto., Mexico.

cesar.ramirez@ugto.mx

Abstract

Sustainable and green industrial chemical processes enhance the profitability, production of merchandises, inherent safety, and minimize the generation of waste. Among them, profitability, and especially inherent safety, can be characterized indirectly through the controllability and flexibility of the processes. Although several indices have been proposed to measure these parameters, the flexibility index and the controllability measured by the condition number stand out. However, these two indicators are usually quantified individually, losing information inherent to the synergies by the link between profitability and process safety. The present work aims to find a quantitative parameter, similar to the condition number, to simultaneously measure the flexibility and controllability of the processes with the purpose of generating a simple measurement of the inherent safety of the processes. This indicator can be applied for designing processes enhancing their profitable, green and sustainable dimensions. To show the advantages of the use of the metric proposed, the results binary and multicomponent reactive distillation are evaluated, showing that the condition number measuring controllability at low frequencies can be related to flexibility. This results in a double indicator which can measure both controllability and flexibility simultaneously.

Keywords: Controllability, Condition Number, Flexibility, Green Chemistry.

1. Introduction

Nowadays sustainable and efficient processes are needed in the chemical industry to address the great challenges of resource depletion, energy consumption and environmental impact. Therefore, the identification, design, and development of appropriate processes are important. For the industry to remain competitive (Wang, et al., 2019), the development of sustainable and efficient processes, along with inherent safety is a primary objective. In order to achieve this, safe and reliable systems are required (Cheng & Zhao, 2019). Indirect measures of safety are the controllability and the flexibility. Therefore, the evaluation of the controllability of the system, together with the evaluation of the process flexibility can indicate the safety level of the process. An integrated indicator including these aspects, would facilitate the optimization of the

process at the design stage. Achieving that the indirect indicator provides values of inherent safety would favor the objective of having a more effective risk management, since actions can be taken to avoid accidents instead of mitigating their consequences (Medina-Herrera et al., 2019).

Inherent safety has gained popularity thanks to its ability to assess the level of safety of the plant at an early stage of design by presenting four main objectives: intensification, substitution, attenuation and limitation. Considerations of these principles allow the designer to give safety aspects the same importance than economic and environmental parameters, thus safety is considered as a key variable design, rather than a late occurrence (Vázquez et al., 2019). Having an indicator that can assess both the controllability of the system and its flexibility at the same time also contributes to the fact that the indicator can indirectly measure objectives of the metrics of green chemistry and the circular economy, such as inherent safety.

Processes designed from a control perspective to respond fast to disturbances, being less disturbed from approach of flexibility, influence the process safety. Control and flexibility are parameters with a great impact on the design of processes. Although they seem similar aspects, control and flexibility correspond to different technical concepts. On one hand, the condition number, is a method commonly used to measure control, it has to do with the stability and quality of the process in dynamic state. On the other hand, flexibility is related to the fact that the operation of the process is viable in different operating conditions in state stationary. The condition number and the minimum singular value provide important information in the theoretical control for a given frequency in a transfer function matrix in a multivariate system, it is the relationship between the maximum and minimum singular value, and represents the sensitivity of the system to the uncertainty that is presented in the input. Therefore, it is possible to have the condition number as a reference to measure the flexibility of a process (Cabrera et al., 2017).

This work proposes an indicator that relates flexibility with controllability. Having a single indicator, it would be simpler to implement in optimization processes, and obtain cheaper processes as a result, and at the same time ensure compliance with environmental and safety indicators.

2. Methodology

The study of the condition number as a quantitative measure of the flexibility of the processes, was carried out in five different processes. The three of these processes are the production of solar grade silicon (0.99999 mole fraction) following the (a) Siemens, (b) Union Carbide, and (c) Hybrid processes, as reported by Ramírez-Márquez et al. (2019). To meet the performance of 2000 t/y, the processes are fed with the components and quantities with 532.32 kg/h of SiO₂ and 369.84 kg/h of Carbon. These processes, shown Figure 1, were studied due to their complexity. In addition, two separation processes by means of distillation columns are studied, a binary distillation column that separates pentane and hexane (0.98 mole fraction), Figure 1 (d), and the intensified Petlyuk scheme to separate propane-butane-pentane (0.99 mole fraction in dome and bottom, 0.98 mole fraction in side stream), Figure 1 (e).

In Figure 2 (a) we can see the methodology for calculating the flexibility index. Equations [1] and [2] correspond to inscribing the largest proposed polytope within the feasible region, then the flexibility index is equal to the distance between the vertex (V) and the nominal operating point [0] (Di Pretoro et al., 2019). Furthermore, in Figure 2 (b) we can see the methodology for calculating the condition number (Cabrera et al.,

2017). Once both the flexibility index and the condition number have been calculated, it will be possible to see the trend of both and solve if there is a guideline that both in complex processes and in simple processes the condition number can be used as a qualitative measure of the process flexibility.

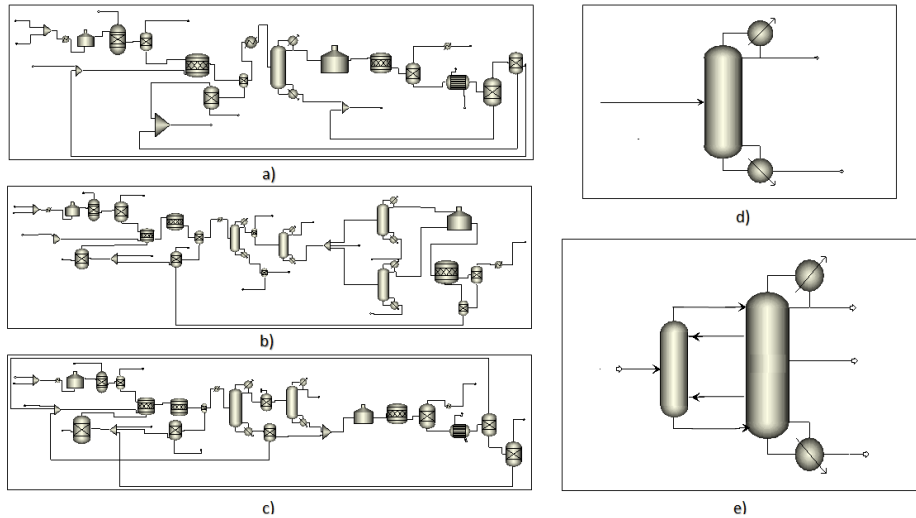


Fig. 1 a) Flowsheet of the Siemens Facility, b) Flowsheet of the Union Carbide Facility, c) Flowsheet of the Hybrid Facility, d) Flowsheet of the binary distillation column, e) Flowsheet of the Petlyuk distillation column.

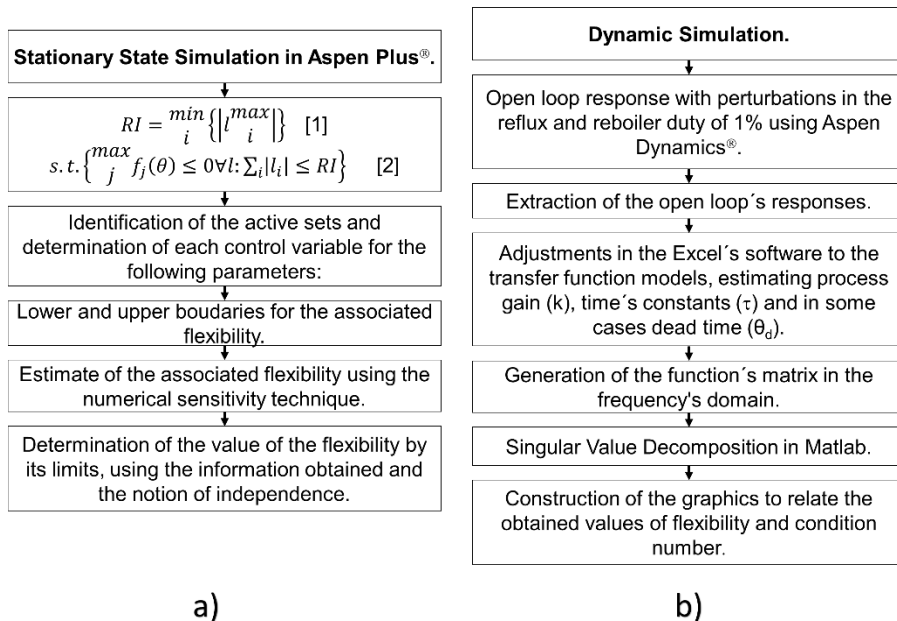


Fig. 2. a) Methodology to calculate the flexibility index. b) Methodology to calculate condition number.

3. Results

The results are shown in Figure 3. It can be observed that, in each one of the processes studied at low frequencies ($w \leq 1$), there is a very similar behavior trend between the condition number and the flexibility index. It is evident that with increasing frequency this trend changes, to the extent that at very large frequencies ($w \geq 1000$), the condition number is not useful as a quantitative measure of the flexibility of a process.

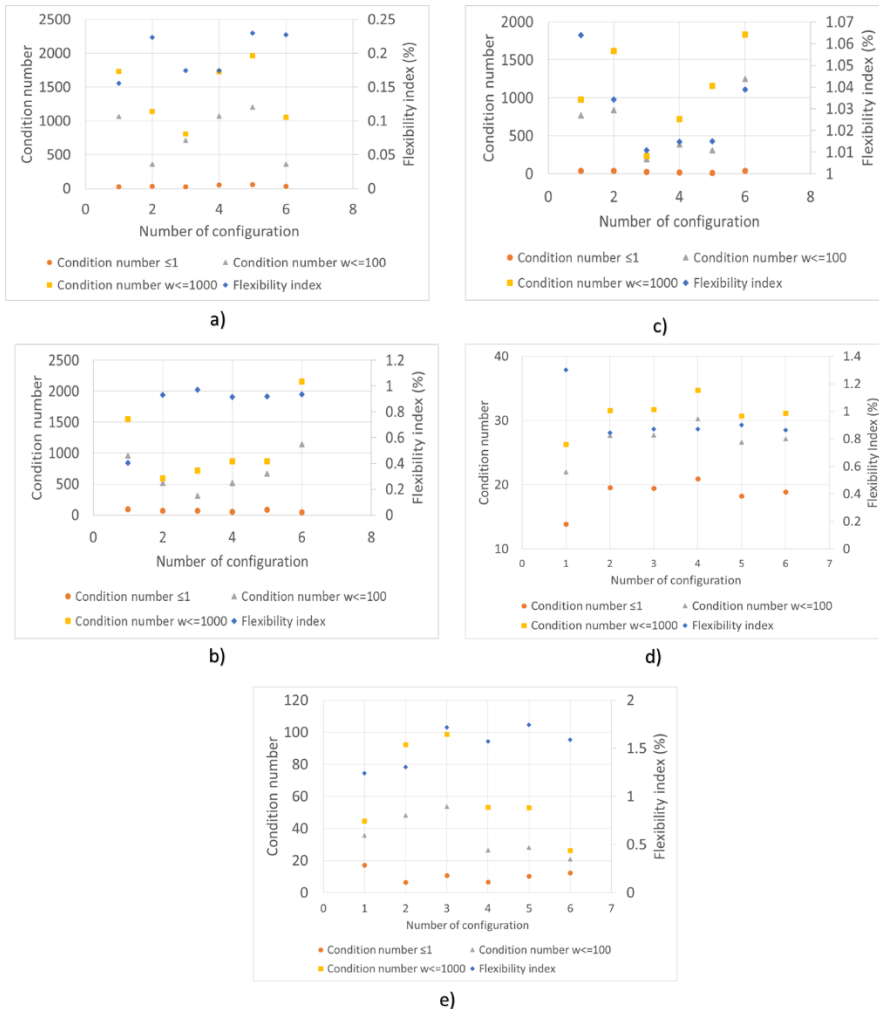


Fig. 3 a) Results of the Siemens Facility, b) Results of the Union Carbide Facility, c) Results of the Hybrid Facility, d) Results of the binary distillation column, e) Results of the Petlyuk distillation column.

The results of the processes for obtaining silicon are shown in Figure 3 (a-c). We can observe that at low and medium frequencies ($w \leq 1000$) the condition numbers have a behavior antagonistic to the flexibility index, and it is precisely the trend that we wanted to observe. That is, at low condition numbers (which represent systems with better control properties), the value of flexibility is greater (that is, processes that support a

variety of changes in steady state). The similar trend is observed in Figure 3 (d) and (e), where for the binary distillation columns and Petlyuk distillation column, at low frequencies ($w \leq 1$), the condition number can be considered as a quantitative measure of the flexibility of each of the processes. Therefore, we can observe that the flexibility has to do with the issue of ensuring a viable steady state operation under a variety of operating conditions, while controllability has to do with the quality and stability of the dynamic response of the process. This provides a suitable index for the processes control and the flexibility could be the condition number.

Table 1 shows the design parameters for the separation within the Siemens Process, such as the values of the condition number and the flexibility index. A strong relationship in the design parameters of the most relevant equipment in the processes it can be observed. In other words, the larger the height of a column or its diameter, the better control properties the process has, and subsequently, also better performance in the flexibility of the process. With parameters such as reboiler duty and reflux ratio, a similar behavior is displayed if the value is higher, a disturbance could represent such a small percentage that it would not alter the process. The performance of each configuration studied was maintained.

Table 1. Results of flexibility index, condition number (different frequencies), and other parameters for the Siemens Process.

Point	Flexibility index	Condition number	Condition number $w \leq 1000$	Stages	Reflux	Reboiler duty [kW]	Diameter [m]
1	0.155	26.97	1729.96	40	29.9	1219.94	0.89
2	0.223	31.42	1141.48	99	29.9	1220.89	0.89
3	0.174	24.58	805.37	22	71.2	2951.68	4.75
4	0.174	55.22	1727.97	39	30.0	1223.33	3.59
5	0.229	62.09	1967.96	50	39.4	1615.49	2.55
6	0.227	34.24	1051.27	54	93.5	3884.55	2.45

The results show that in each scenario it is necessary to ensure that the process is controllable in an adequate minimum time, even when the conditions of the process may change. Since the one that takes a long time to stabilize represents more operational costs, as well as not reaching the desired product in the oscillation period, stoichiometry problems would arise, more waste would be generated, resulting in a low safety profile. It would differentiate it from an unstable and unsafe process by not incorporating green metrics is an economical process from a green perspective. Therefore, not well controlled processes produce a lot of waste, consume more materials and energy per unit of finished product, and in some cases lead to a reduction in performance and life cycle. The process must also operate in the most flexible way possible in order to adapt to consumer demands, in this context the use of appropriate operational strategies for the process and control will increase production and increase the performance of high-end products value, in addition to minimizing energy consumption, pollution and, as a direct consequence, environmental impact.

According to the study, the condition number, for these processes, near the steady state has an effect directly proportional to the flexibility behavior. Therefore, it indicates that there is a direct correlation between the condition number and flexibility, and revealing the option that the condition number that can be used as a control objective function. The limitations to highlight of these results, is that for the correlation to be used, it is necessary to work with low frequencies ($w \leq 1$), because with high frequencies ($w \geq 1000$) the behavior changes, however, working with low frequencies does not represent a drawback, since Cabrera, et al. (2017) mention that a real process behaves at such frequencies ($w \leq 1$). Having relatively large diameters in the design, and a high number of stages, becomes a necessary condition, and this can associate controllability with flexibility simultaneously.

4. Conclusions

In this work, a quantitative measure is presented to evaluate simultaneously both the controllability of a process and its flexibility through the condition number. The results show that the condition number serves as a quantitative index of the controllability of the processes, and that in turn, this correlates with flexibility. However, this correlation presents problems when working with relatively high frequencies ($w \geq 1000$), but at low frequencies ($w \leq 1$) it could be a reliable index to measure both items. With the above, this index could be provided that would ensure in future optimizations to have economic, affordable, green and controllable processes, respecting and also enhancing an objective that is currently as important within sustainable processes as safety is. Considering the results obtained, the fusion of the flexibility and controllability indicators in a single indicator for the first time, such as the condition number, seems to be an important advance in the area of green and sustainable process design.

5. References

- A. Di Pretoro, L. Montastruc, F. Manenti, & X. Joulia, 2019. Flexibility analysis of a distillation column: Indexes comparison and economic assessment. *Computers & Chemical Engineering*, 124, 93-108.
- C. Ramírez-Márquez, G. Contreras-Zarazúa, M. Martín, & J. G. Segovia-Hernández, 2019. Safety, Economic, and Environmental Optimization Applied to Three Processes for the Production of Solar-Grade Silicon. *ACS Sustainable Chemistry & Engineering*, 7(5), 5355-5366.
- D. Vázquez, R. Ruiz-Femenia, & J. A. Caballero, 2019. A Novel Optimizable Inherent Safety Index Based on Fuzzy Logic. In *Computer Aided Chemical Engineering*, Elsevier, 46, 559-564.
- F. Cheng, & J. Zhao, 2019. A novel process monitoring approach based on Feature Points Distance Dynamic Autoencoder. In *Computer Aided Chemical Engineering*, Elsevier, 46, 757-762.
- J. A. Vázquez-Castillo, J. G. Segovia-Hernández, & J. M. Ponce-Ortega, 2015. Multiobjective optimization approach for integrating design and control in multicomponent distillation sequences. *Industrial & Engineering Chemistry Research*, 54(49), 12320-12330.
- J. Cabrera-Ruiz, M. A. Santaella, J. R. Alcántara-Ávila, J. G. Segovia-Hernández, & S. Hernández, 2017. Open-loop based controllability criterion applied to stochastic global optimization for intensified distillation sequences. *Chemical Engineering Research and Design*, 123, 165-179.
- J. Wang, P. Ji, X. Chen, A. Tula, & R. Gani, R., 2019. Integrated Process and Controller Design Software Tool-ProCADC. In *Computer Aided Chemical Engineering*, Elsevier, 46, 745-750.
- N. Medina-Herrera, S. Tututi-Avila, & A. Jiménez-Gutierrez, 2019. A new index for chemical process design considering risk analysis and controllability. In *Computer Aided Chemical Engineering*, Elsevier, 46, 373-378.

Nonlinear Model Predictive Control of Food Sterilization Processes

Chrysovalantou Ziogou^a, Margaritis Kostoglou^b, Michael C. Georgiadis^{a,c,*}

^a*Chemical Process and Energy Resources Institute (CPERI), Centre for Research and Technology Hellas (CERTH), 57001, Thessaloniki, Greece*

^b*Department of Chemistry, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece*

^c*Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece*
mgeorg@auth.gr

Abstract

This work presents a framework for the optimal control of several sterilization units operating in parallel to satisfy the thermal processing needs of a process motivated by a food industrial facility producing canned fish (Georgiadis et al., 2020). The objective is to consider all relevant control challenges that arise during the operation of the sterilizers such as the required microbial lethality set-point while minimizing the overall processing time. Furthermore, energy consumption minimization is also considered. To this end, a Nonlinear Model Predictive Control (NMPC) framework is developed and applied in multiple sterilization units operating in parallel and sharing the same steam network. A novel reduced order process model is derived and validated. Simulation experiments reveal significant benefits by using the NMPC framework with respect to the availability of shared resources without any loss of performance.

Keywords: optimum control, shared resource efficiency, time-varying operation, nonlinear dynamic optimization

1. Introduction

Today food industries are facing critical challenges in response to consumers' needs, which, in addition to health and safety awareness, demand products with high quality standards. Food process industries that deliver canned products are trying to satisfy the needs of new markets and population sectors, relying on more efficient processes in order to gain an increased market share (Simpson et al., 2019). Thermal processing of canned food is one of the most critical steps in the overall production chain. The duration of the sterilization process is usually dictated by the microbial lethality evolution, which is a product dependent parameter (Alonso et al., 2013). The sterilization process in the food industries must guarantee a uniform processing temperature throughout the cans and achieve the respective lethality levels. In order to systematically improve and optimize the efficiency of the sterilization processes, advanced process control can significantly enhance the efficient operation of the sterilization units by optimizing the utilization of shared resources such as the steam network.

2. Process description and modeling

The process under consideration has been inspired by a large-scale, real-life food industrial facility producing canned fish (Vilas and Alonso, 2018). Sterilization units use steam from a common plant network and their operation is constrained by the availability of this resource (Fig. 1). The cans, which are thermally sterilized, are heated at a specified temperature for a prescribed duration in a retort. The sterilizing duration is defined by a safety index, the microbial lethality evolution.

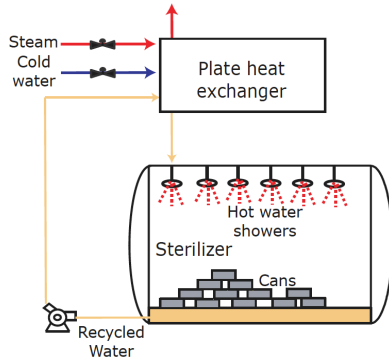


Figure 1 Overview of the sterilizer unit

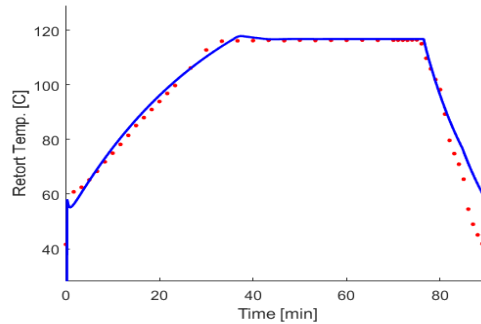


Figure 2 Model response vs. experimental data

The operation cycle begins by the insertion of preheated water (~ 60°C) into the sterilizer and recycled to the Plate Heat Exchanger (PHE). The water temperature in the retort increases until it reaches the operation point ($T_{retort} = 110 - 130^{\circ}\text{C}$). Subsequently the temperature is kept constant for a given amount of time and the steam valve in the PHE is closed. After the lethality specifications are met, cold water is introduced to cool down the cans. The duration of the 2-stage cooling is usually 13mins. Initially the inlet water temperature is set at 65°C while the retort temperature is above 75°C and then the inlet water temperature is set at 20°C for the remaining cooling time.

2.1. Modeling of the sterilization process

A reduced order process model is derived and validated against a detailed model using measurements from an industrial plant. The model consists of three parts and was used by the control framework. The aim is to capture the dynamic behavior of the involved subsystems and to derive a model, suitable for control purposes. The governing equations for the can temperature (T_{can}) evolution profiles are:

$$\frac{dT_{can}}{dt} = Bi_2(T_{w,R} - (c_0 + c_1 + c_2)(1 + q/2)) - Bi_1(c_0(1 + q/2) - T_{w,R}) + 2MBi_R(T_{w,R} - (1 + q)(c_0 + c_1/2 + c_2/3)) \tag{1}$$

$$q = \frac{Bi_R(T_{w,R} / (c_0 + c_1/2 + c_2/3) - 1)}{2 + Bi_R} \tag{2}$$

$$Bi_R = \frac{h_R R}{k}, Bi_1 = \frac{h_{Z1} L}{k}, Bi_2 = \frac{h_{Z2} L}{k}, M = \frac{L^2}{R^2} \tag{3}$$

where $T_{w,R}$ is the retort water temperature, k is the thermal conductivity, L and R are the length and the radius, $h_{(R,Z1,Z2)}$ is the convective heat transfer coefficient right, top and

bottom of the can. In each time step c_o , c_1 and c_2 have to be found from the solution of the following 3x3 system of nonlinear algebraic equations:

$$T_{can} = (c_o + c_1 / 2 + c_2 / 3)(1 + q / 2) \quad (4a)$$

$$c_1(1 + q / 2) = Bi_1(c_o(1 + q / 2) - T_A) \quad (4b)$$

$$(c_1 + 2c_2)(1 + q / 2) = Bi_2(T_A - (c_o + c_1 + c_2)(1 + q / 2)) \quad (4c)$$

The integration of a single ordinary differential equation with a system of three algebraic equations describes the evolution of a two-dimensional temperature profile in the can.

2.1.1. Parallel Plate Heat Exchanger Model (PPHE)

Let us assume a parallel plate heat exchanger (PPHE) with mass flow rate of water M_w and mass flow rate of steam M_s . A steady state problem for the countercurrent parallel heat exchanger is considered. The equations can be integrated in closed form and scaled back to the whole heat exchanger to give:

$$Q = M_s c_{ps} (T_{s,in} - T_{s,out}) \quad (5)$$

$$Q = M_w c_{pw} (T_{w,out} - T_{w,in}) \quad (6)$$

$$Q = UA(T_{w,out} - T_{s,in} - (T_{w,in} - T_{s,out})) / \ln((T_{w,out} - T_{s,in}) / (T_{w,in} - T_{s,out})) \quad (7)$$

where Q is the rate of heat transferred in the PPHE. The heat transfer coefficient and the (one side) surface area of the plate is U and A respectively. Q can be eliminated between the above equations to leave with a system of two equations with two unknowns (the outlet temperatures). The system of Eq.(4) - (6) is highly non-linear and can be replaced by the linear system:

$$T_{w,out} = \frac{T_{w,in} + p2 \left((1 - p1) T_{s,in} - T_{w,in} \right)}{1 - p1p2} \quad (8)$$

$$T_{s,out} = T_{w,in} + p1 \left(T_{s,in} - T_{w,out} \right) \quad (9)$$

$$p1 = \exp \left(UA \left(\frac{1}{M_w c_{pw}} - \frac{1}{M_s c_{ps}} \right) \right) \quad (10)$$

$$p2 = \frac{M_s c_{ps}}{M_w c_{pw}} \quad (11)$$

For given inlet temperatures to PHE the outlet temperatures are computed by Eq. (8) and (9) and then Q by Eq. (5).

2.1.2. Sterilizer Model

The heat balance in the sterilizer is the leading equation of the model. The water temperature $T_{w,R}$ is dominated by the equation:

$$M_{wR} c_{pw} \frac{dT_{w,R}}{dt} = Q_{PHE} - Q_{can} - Q_{env} \quad (12)$$

$$Q_{PHE} = M_s c_{ps} (T_{s,in} - T_{s,out}) \quad (12a)$$

$$Q_{can} = n_{can} M_{can} c_{p,can} \frac{dT_{can}}{dt} \quad (12b)$$

$$Q_{env} = h_c A_c (T_{env} - T_{wR}) \quad (11c)$$

where M_{wR} is the water mass in sterilizer, Q_{can} is the heat adsorbed by the cans and Q_{env} is a simple loss term where h_c , A_c are the corresponding heat transfer coefficient and the metal cover area of the retort respectively and T_{env} is the environment temperature. The total model consists of the main Eq. (12) which must be solved simultaneously to the can heating submodel (one ODE and three algebraic equations) and to the PPHE model (two algebraic equations). Everything else is post processing (e.g. lethality, color) or control actions (e.g. steam valve) of the core model. Model predictions are in excellent agreement with experimental results (Vilas and Alonso, 2018) as illustrated in Fig. 2.

3. Control framework

To optimize the operation of the sterilization processes, advanced control can provide significant benefits by satisfying all operating constraints and managing the availability of shared resources such as the steam network. Thus, a number of time-varying parameters must be taken into account along with dynamic changes, which occur as multiple sterilizers operate in parallel. The control objective is to address, in a centralized manner, several challenging control issues that arise during the parallel operation of the sterilizers and to monitor and evaluate their performance at real time. More specifically, the operational objectives are to achieve the required microbial lethality set-point for each batch of cans while minimizing the overall processing time to avoid quality deterioration of the canned food. Furthermore, it is crucial to minimize energy consumption during the heat-up stage of the sterilization process. To achieve these objectives a Nonlinear Model Predictive Control (NMPC) framework is proposed and demonstrated in this problem involving 16 units operating in parallel.

NMPC computes online a finite-time constrained optimization problem over a prediction horizon (T_p), using the current state of the process as the initial state. The optimization yields an optimal control sequence ($u_k \dots u_{k+N_c}$) over a control horizon (T_c), which is partitioned into N_c intervals and only the first control action (u_k) for the current time is applied to the system. We consider the following formulation of the NMPC problem:

$$\min J = \sum_{j=1}^{N_p} (\hat{y}_{k+j} - y_{sp,k+j})^T QR (\hat{y}_{k+j} - y_{sp,k+j}) + \sum_{l=0}^{N_c-1} \Delta u_{k+l}^T RI \Delta u_{k+l} \quad (13)$$

$$\text{s.t.: } \dot{x} = f_d(x, u), \quad y = g(x, u) \quad (14a)$$

$$e_k = (y_{pred} - y_{meas})_k \quad (14b)$$

$$\hat{y}_{k+j} = y_{pred,k+j} + e_k \quad (14c)$$

$$u^L \leq u(t)_{k+j} \leq u^U, y^L \leq y(t)_{k+j} \leq y^U \quad (14d)$$

where u, y, x are the manipulated, the controlled and the state variables, $y_{pred}, y_{meas}, y_{sp}$ are the predicted, the measured variables and the desired set-points and QR, RI are the output tracking and the input move weighing matrices. The minimization of functional J (eq. 13) is also subject to constraints of u, x and y . The controlled variables are the retort temperature ($T_{w,R}$) and the lethality (F_0) while the manipulated variable is the steam

flow. A centralized approach is applied for the control of the sterilization process that uses the nonlinear model described by Eq. (1)-(12).

4. Operation analysis and discussion of results

A set of scenarios where multiple sterilizers operate concurrently is explored. The sole purpose of this study is to evaluate the flexibility of the NMPC framework under various conditions and different operation requirements. We assume that each sterilizer processes the same amount and type of cans. A nonlinear programming problem (NLP) is solved at each iteration while the NMPC parameters are $T_c=1$, $T_p=2$, $t_s=60$ sec. The framework is developed at the Mathworks MATLAB environment. The optimization problem is solved using a Sequential Quadratic Programming (SQP) method with constraint tolerance $1e-6$ and step tolerance $1e-8$. In order to explore the response of the NMPC architecture a scenario where three sterilizers start to operate concurrently is presented.

Fig. 3 presents the retort temperature profile while Fig. 4 illustrates the duration of batches at each sterilizer. The steam flow rate is adjusted according to the number of sterilizers that are concurrently employed during the heat-up stage. It is clearly illustrated that the heat-up time is affected accordingly when multiple sterilizers are activated. For example, the slope of the 2nd (yellow) sterilizer changes when the 1st sterilizer reaches the heat maintenance stage and when the 3rd sterilizer enters the heat-up stage.

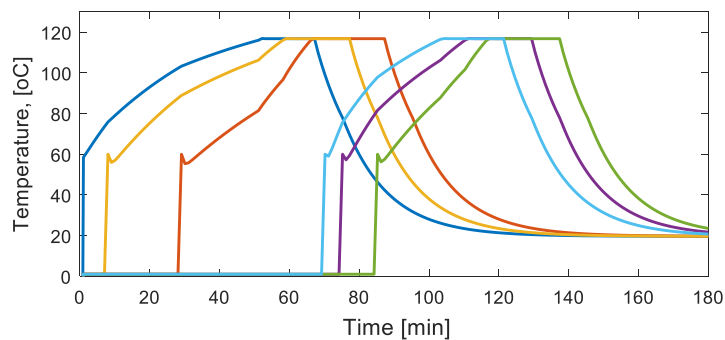


Figure 3 Operation of three sterilizers that start to heat-up concurrently

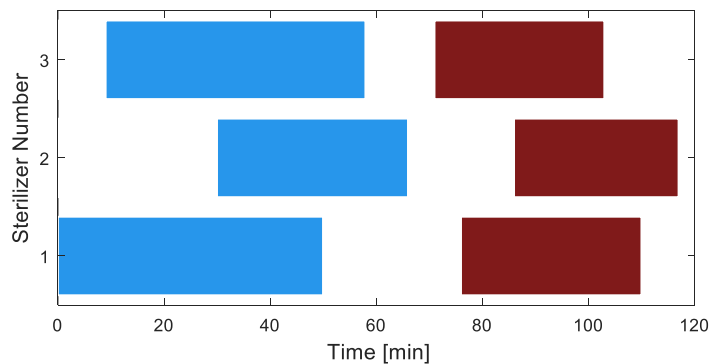


Figure 4 Batch timing for the three sterilizers

Finally, Fig 5 shows the distribution of steam between the 3 sterilizers. Initially the 1st sterilizer operates. The 2nd sterilizer starts to heat-up at $t=8\text{min}$ and the steam is shared equally between the 2 sterilizers. Finally, the 3rd sterilizer is activated at $t=30\text{min}$ and the shared steam is further reduced between the 3 sterilizers. At $t=50\text{min}$ the 1st sterilizer reaches the temperature target and the steam requirement is reduced. After an hour only the 3rd sterilizer is still operating at the heat-up stage and is allowed to use the maximum steam flow rate from the network. Overall, it can be seen that the steam is adjusted according to the operating stage of the sterilizers.

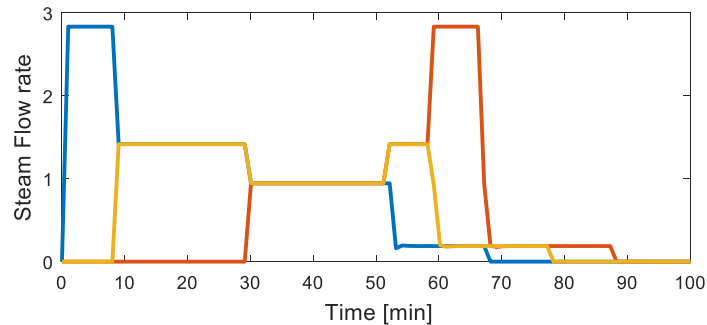


Figure 5 Steam flow rate distribution among the three sterilizers

The sterilization time of the processing units is dynamically adjusted over time according to the availability of steam. The operation follows the optimal temperature profiles, thus ensuring that the required degree of sterilization is met, and the cost is minimized.

5. Conclusions

An NMPC framework that relies on a reduced order model is developed and validated against realistic conditions for the sterilization units of a canned food process. The predictive power of the model is clearly illustrated to capture the dynamic behaviour of the parallel plate heat exchangers, the retort temperature, the canned food product temperature evolution and the effect of food temperature on safety and quality. A case study used to illustrate the applicability and efficiency of the proposed NMPC approach. The Results illustrate that the optimal operation of the sterilizers is achieved under tight operating constraints and taking into account the availability of steam.

References

- A. A. Alonso, A. Arias-Mendez, E. Balsa-Canto, M. R. Garcia, J. I. Molina, C. Vilas, M. Villafin. 2013, Real time optimization for quality control of batch thermal sterilization of prepackaged foods. *Food Control*, 32(2), 392-403
- G. P. Georgiadis, B. Pampín, D. Cabo, M. C. Georgiadis, 2020, Optimal production scheduling of food process industries, *Computers & Chemical Engineering*, 134, 106682
- C. Vilas and A.A. Alonso, 2018. Real time optimization of the sterilization process in a canning industry. *Actas de las XXXIX Jornadas de Automática, Badajoz, 5-7 de Septiembre de 2018*
- R. Simpson, C. Ramirez, D. Jiménez, S. Almonacid, H., Nuñez, A. Angulo, 2019. Simultaneous multi-product sterilization: Revisited, explored, and optimized. *Journal of food Engineering*, 241, 149-158

Design and control of triple-column pressure swing distillation process for acetone-methanol-hexane separation

Tan Akinciturk , Devrim B. Kaymak*

Department of Chemical Engineering, Istanbul Technical University, 34469 Maslak, Istanbul

Abstract

The mixture of acetone-methanol-hexane is used in the industry as a solvent, an extraction agent and for cleaning purposes. This mixture belongs to the topological class of 3.1-2 in Serafimov's classification with three binary and one ternary minimum boiling homogeneous azeotropes (Shen et al., 2016). To our knowledge, steady-state design studies for this class of azeotropes are very rare, and there is no dynamic control study in the literature. In this study, a triple-column pressure swing distillation (PSD) process is designed for separation of three components with a 99.9 mole % purity. The process is optimized based on total annual cost (TAC), and a plant-wide control structure is designed for the optimal steady-state case. The control structure including inferential temperature controllers shows an effective base-level regulatory control with acceptable offsets in compositions. This work hereby demonstrates that a ternary mixture which forms three binary and one ternary azeotropes can be separated using a triple-column PSD process, and this process system can be controlled robustly using a decentralized control structure.

Keywords: Ternary azeotropes, Pressure-swing distillation, Process design, Dynamic control.

1. Introduction

Distillation is the most attractive method of separation that is based on vapor-liquid equilibrium (VLE), and studied widely in the literature. In recent years, the interest has shifted to more complex, multicomponent distillation systems with non-ideal vapor-liquid equilibrium. Especially, distillation boundaries that are imposed by azeotropic mixtures are challenging, and require complex distillation strategies.

Pressure swing distillation (PSD) is a method that makes use of the pressure sensitivity of azeotropic mixtures for separation. Thus, by designing more than one column operating at significantly different pressures, the distillation boundaries are shifted so that the separation can be achieved. There is an extensive library on the separation of binary azeotropic mixtures using PSD processes in the literature (Luyben and Chien, 2010). Nowadays, the interest of academia is on the separation of ternary mixtures including more than one binary azeotropes and/or ternary azeotropes. Although the non-ideal thermodynamics and distillation concepts for multicomponent azeotropic mixtures have been discussed since the nineties (Doherty and Malone, 2001; Stichlmair and Fair, 1998), the papers on rigorous design and control of triple-column PSD processes are started to be published after the second half of 2010's, and there are still a very small number of papers in the literature.

Separation of acetonitrile-methanol-benzene mixture which has three binary azeotropes has been investigated by three papers (Zhu et al., 2017; Luyben, 2017; Zhang et al., 2019). Among these studies, Luyben (2017) focused on the process control aspects, while other studies dealt only with steady-state design. On the other hand, Wang et al. (2019a, 2019b) studied the separation of tetrahydrofuran-ethanol-water mixture. They compared alternative PSD sequences for the ternary mixture having three binary azeotropes.

In none of the aforementioned papers, the studied mixtures form ternary azeotropes. Nevertheless, to our knowledge, there are only two papers studied the separation of mixtures with ternary azeotropes (Cui et al., 2019; Guang et al., 2019). In the first study, benzene-isopropanol-water mixture has been investigated, while the second study focused on diisopropyl ether-isopropanol-water mixture. Both mixtures form a heterogeneous ternary azeotrope besides two homogeneous and one heterogeneous binary azeotropes. Although these papers discussed the steady-state design of triple-column PSD processes, they did not investigate the controllability of the proposed configurations.

On the other hand, Gmehling and Böltz (1996) presented in their experimental work that acetone, methanol and hexane form a ternary azeotrope besides three binary azeotropes. This mixture is used in laboratories for several purposes such as a solvent for chromatography, an extraction agent and for cleaning purposes. Although there is a patent by Forman (1952) where four columns are used for lab scale separation of this mixture, there is no integrated design and control study of this mixture in the open literature. In this study, as an alternative to the above-mentioned patent, a novel triple-column PSD process is proposed based on ternary diagram analysis. Furthermore, a robust plant-wide control structure is developed and tested for the optimum design of the proposed flowsheet.

2. Process Studied and Methods

The feed flowrate of the process is 100 kmol/h with mole fractions of 47 % methanol, 43 % acetone and 10 % hexane. The desired product specifications are 99.9 mole % for each component. First, ternary diagrams at different operating pressures are used to develop a triple-column PSD process flowsheet. Several thermodynamic model results are investigated and compared with the experimental data in the literature. Since the results of UNIFAC model are in coherence with the experimental work of Gmehling and Böltz (1996), this thermodynamic model is used in this study. Then, the steady-state simulations are conducted by Aspen Plus. The product purities are achieved using the “Design Spec/Vary” feature of the simulator. The flowsheet is optimized based on total annual cost (TAC). A sequential iterative procedure is used for the cost evaluation of the triple-column PSD process. Minimal TAC is determined by sequentially changing the values of variables such as recycle flowrate, recycle feed location for the first distillation column, and number of trays and feed tray locations for each column. The values of these variables when all the local minima are combined are presented as the best case. Equations used for capital cost calculations are taken from Douglas (1988) and Dimian (2003). Data for operating costs are found from Luyben (2013) and Patrascu et al. (2016). Next, plant-wide control structures are designed for the optimized steady-state design. Level and pressure controllers are used for inventory control purposes. Feed flowrate acting as the production rate handle is manipulated by a flow controller. PI controllers are used for temperature and/or composition control. A dead time of one minute for temperature, and three minutes for composition loops is included

into these loops. Steepest slope method is used to decide the tray location in case of temperature control. These loops are tuned by Auto Tune Variation (ATV) test following Skogestad's sequential loop method (Hovd and Skogestad, 1994). Controller parameters are calculated using Tyreus-Luyben parameter settings. All control valves are designed half-open at the nominal conditions. Finally, the robustness of the proposed control structure is tested against disturbances in feed flowrate and composition.

3. Results and Discussion

3.1. Steady-State Design

In Figure 3.1, ternary diagrams of acetone-methanol-hexane mixture are examined at two different pressures. The dashed curve represents the distillation boundary at 1 bar, while the solid curve represents the distillation boundary at 5 bar. It is seen that the distillation boundaries differ considerably as pressure changes. The F point is the composition of fresh feed stream. It is combined with the distillate of the third distillation column, which is recycled back to the first column, using a straight line on the figure. It means that the composition of the total feed into the first distillation column is anywhere on this line, shown with F^1 .

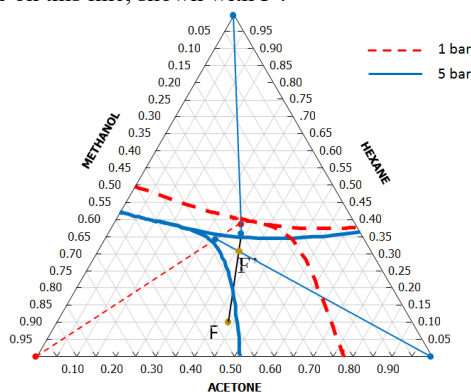


Figure 3.1: Ternary diagram for acetone-methanol-hexane

As dictated by Lever Rule, the feed, bottoms and distillate compositions of a distillation column should lie on a straight line, and this line should stay within the distillation boundary. Based on this idea, for each distillation column, a straight line starting from 99.9% pure component is connected to the point representing the column feed, and then elongated up to the composition of distillate staying within the corresponding distillation boundary.

As the result, a triple-column PSD flowsheet is developed with operating pressure of 5 bar, 1 bar and 5 bar, respectively. In this sequence; acetone, methanol and hexane are drawn from the bottoms of distillation columns, respectively. Table 3.1 shows the economic optimization results for this process configuration.

Table 3.1 : Summary of variables and costs of optimal case			
	C1	C2	C3
Number of Trays	52	52	8
Recycle Location	30	-	-
Feed Location	4	4	4
Reflux Ratio	3.02	0.03	0.23
Diameter (m)	2.84	1.42	2.00
Recycle (kmol/h)	450	-	-
Fixed Column Cost (10^3 \$)		1,778.5	
Fixed Heat Exchanger Cost (10^3 \$)		1,776.1	
Operating Cost (10^3 \$/yr)		5,765.8	
Total Annual Cost (10^3 \$/yr)		6,947.3	

3.2. Plant-wide Control Design

Several control structures are evaluated for this process, and the results of the one with the best performance are given in this study. Figure 3.2 shows the proposed control structure including inferential temperature controllers with reflux to feed (R/F) ratio. The location of the trays to control the temperature is decided via steepest slope analysis as controlling the temperature on these stages would end up with a more robust composition control. As the result of the analysis, the temperature on stage 44 is selected to be controlled for first and second distillation columns, while the temperature on stage 5 is chosen for the third distillation column.

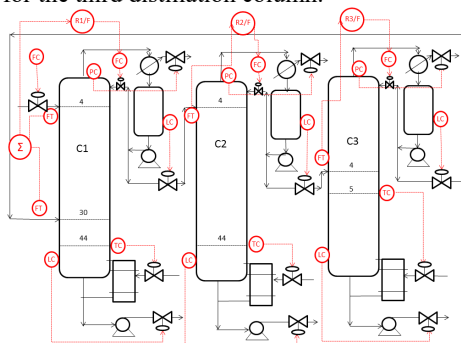


Figure 3.2: Proposed Control Structure

Dynamic simulations are run in Aspen Dynamics, and the performance of control structures are tested by applying two types of disturbances. Response of the control structure for $\pm 20\%$ throughput change is illustrated in Figure 3.3. The solid line represents positive change and the dashed line represents the negative change. It is seen that temperatures settle down into their set-points easily. Additionally, the overshoots of temperatures are very small for all columns. Since inferential temperature controllers are used, an offset in compositions is expected. However, as seen from the figure, the composition of hexane recovers back to the steady-state, while there are only very small offset in compositions of acetone and methanol.

An improvement is done by adding composition controllers in the control structure. As the result, the temperature-composition cascade controllers eliminate the offset in compositions. However, it is observed that the settle time is approximately twice of the time spent with temperature only controllers. Additionally composition control requires analyzers, which are expensive equipment with hefty maintenance processes. Since the offset obtained with the inferential temperature controllers is tolerable, it is selected as the proposed control structure in this study.

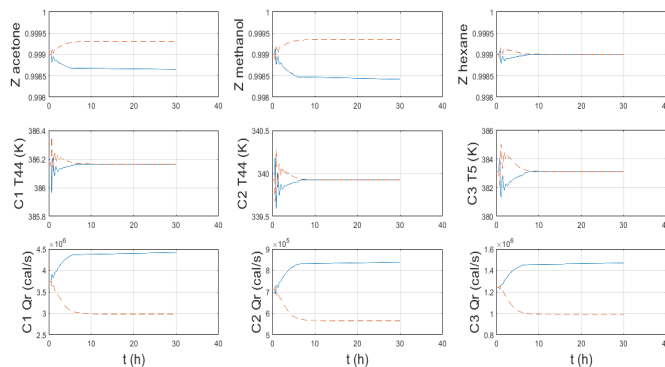


Figure 3.3: Response of control structure for feed flowrate changes

Second disturbance given to the system is in feed composition. The magnitude of composition changes in the feed stream is selected based on the distillation boundaries of ternary diagrams. The fractions are first changed to 0.42, 0.43 and 0.15 for acetone, methanol and hexane, respectively. Then, the feed composition is changed to 0.45, 0.5 and 0.05 for acetone, methanol and hexane, respectively.

The dynamic response for the feed composition changes is illustrated in Figure 3.4. The solid line represents the first composition set, while the dashed line represents the second disturbance. Similar to the throughput change results, the temperatures settle down into their set-points easily. In addition, hexane composition does not show any offset, while there are small offsets in the compositions of acetone and methanol. Compared to the responses of throughput disturbance, the response here is more sluggish. However, offsets of the product purities are smaller. These results illustrate that the proposed control structure is capable of handling both disturbances, and provides an effective base-level regulatory control.

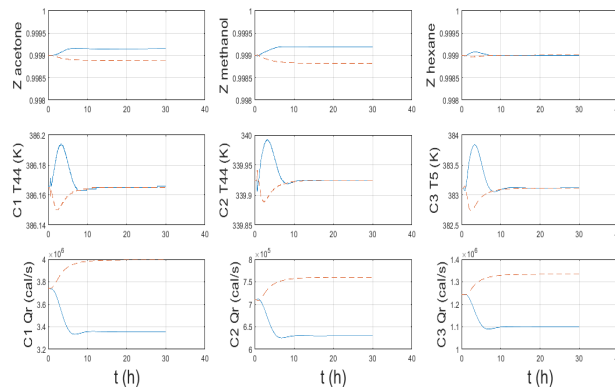


Figure 3.4: Response of control structure for feed composition changes

4. Conclusions

This work covers the steady-state design, cost optimization and process control of a triple-column PSD process for the separation of acetone-methanol-hexane mixture. Ternary diagrams reveal that there is significant change in distillation boundaries with the change in operating pressure for the studied process. A triple-column PSD process is designed based on ternary diagrams, where the columns operate at 5-1-5 bar,

respectively. Total annual cost of the steady-state process with 99.9% product purity is optimized using Aspen Plus. The optimal TAC of this process configuration is calculated to be 6,947,319 \$. Different control structures are designed for the optimal process design. As the result of their responses against disturbances in feed flowrate and composition, it is concluded that the control structure including the inferential temperature controllers with reflux to feed ratio provides the best performance. In the future, additional studies will be carried out in which optimal TAC of this process configuration is further reduced by applying heat integration. Furthermore, the process control of the heat integrated PSD configuration will be studied. Moreover, alternative triple-column PSD process configurations are going to be obtained based on the further analysis of ternary diagrams at different pressures, and their results will be compared in terms of total annual cost and process control aspects.

References

- Y. Cui, X. Shi, C. Guang, Z. Zhang, C. Wang, and C. Wang, 2019, Comparison of pressure-swing distillation and heterogeneous azeotropic distillation for recovering benzene and isopropanol from wastewater, *Process Saf. Environ. Prot.*, vol. 122, pp. 1–12.
- A. C. Dimian, C. S. Bildea, A. A. Kiss, 2003, *Integrated Design and Simulation of Chemical Processes*, Second Edition, Elsevier,
- M. F. Doherty, M. F. Malone, 2001, *Conceptual Design of Distillation Systems*, McGraw-Hill, New York.
- J. M. Douglas, 1988, *Conceptual Design of Chemical Processes*, McGraw-Hill, New York.
- S. E. Forman, 1952, Azeotropic Distillation of Methanol and Acetone with Hexane, 2,581,789.
- J. Gmehling and R. Böltz, 1996, Azeotropic data for binary and ternary systems at moderate pressures, *J. Chem. Eng. Data*, vol. 41, pp. 202–209.
- C. Guang, X. Shi, Z. Zhang, C. Wang, C. Wang, and J. Gao, 2019, Comparison of heterogeneous azeotropic and pressure-swing distillations for separating the diisopropylether/ isopropanol/ water mixtures, *Chem. Eng. Res. Des.*, vol. 143, pp. 249–260.
- M. Hovd and S. Skogestad, 1994, Sequential design of decentralized controllers, *Automatica*, vol. 30, pp. 1601–1607.
- W. L. Luyben, 2013, *Distillation Design and Control Using Aspen™ Simulation*, Second Edition, Wiley, Hoboken, New Jersey.
- W. L. Luyben, 2017, Control of a triple-column pressure-swing distillation process, *Sep. Purif. Technol.*, vol. 174, pp. 232–244.
- W. L. Luyben, I. L. Chien, 2010, *Distillation Systems for Separating Azeotropes*, Wiley, New Jersey.
- I. Patrascu, C.S. Bildea, A. A. Kiss, 2017, Eco-efficient butanol separation in the ABE fermentation process Iulian Patras, vol. 177, pp. 49–61.
- W. F. Shen, H. Benyounes, and J. Song, 2016, A review of ternary azeotropic mixtures advanced separation strategies, *Theor. Found. Chem. Eng.*, vol. 50, pp. 28–40.
- J. G. Stichmair, J. R. Fair, 1998, *Distillation Principles and Practices*, Wiley-VCH, Weinheim.
- C. Wang, C. Guang, Y. Cui, Z. Zhang, and X. Zhang, 2019, Separation of a ternary mixture with multiple azeotropes via pressure-swing distillation, *J. Chem. Technol. Biotechnol.*, vol. 94, pp. 2023–2033.
- C. Wang, Z. Zhang, X. Zhang, C. Guang, and J. Gao, 2019, Comparison of pressure-swing distillation with or without crossing curved-boundary for separating a multiazeotropic ternary mixture, *Sep. Purif. Technol.*, vol. 220, pp. 114–125.
- Q. Zhang, M. Liu, W. Li, C. Li, and A. Zeng, 2019, Heat-integrated triple-column pressure-swing distillation process with multi-recycle streams for the separation of ternary azeotropic mixture of acetonitrile/methanol/benzene, *Sep. Purif. Technol.*, vol. 211, pp. 40–53.
- Z. Zhu, D. Xu, Y. Wang, X. Geng, and Y. Wang, 2017, Effect of multi-recycle streams on triple-column pressure-swing distillation optimization, *Chem. Eng. Res. Des.*, vol. 127, pp. 215–222.

Design and control of a Biobutanol Purification Process through IBE Fermentation: Basic Design Configuration

Ilayda N. Oksal,^a Devrim B. Kaymak^a

^a*Department of Chemical Engineering, Istanbul Technical University, Istanbul 34469, Turkey*

Abstract

The decrease in energy sources, the increase in energy demand and the environmental concerns have caused to be in search of renewable and cleaner sources compared to fossil fuels. Biobutanol stands out as an alternative which has superior fuel properties compared to other alcohol-based biofuels. In this study, a novel flowsheet configuration providing high-purity alcohol products is developed for isopropanol-butanol-ethanol (IBE) purification process. First, the flowsheet is optimized by minimizing the total annual cost (TAC). Then, a plantwide control structure is designed for the optimized process configuration. The proposed control structure is evaluated by exposing it to disturbances. The results show that the developed control structure provides a robust control for this process configuration.

Keywords: biobutanol purification, plantwide control, IBE fermentation, azeotropic mixture.

1. Introduction

In recent years, renewable energy sources have started to draw attention of industry and academia because of the increasing energy demands and environmental concerns. To meet these demands and concerns, several studies have been conducted on alternative energy sources including biofuels. Among several biofuels, biobutanol comes forward with its advantages such as high energy content, less solubility in water and the ability of blending with gasoline in any percentage without any engine modification. The production of biobutanol by Clostridia microorganisms is called either acetone-butanol-ethanol (ABE) or isopropanol-butanol-ethanol (IBE) fermentation depending on the formation of acetone or isopropanol, respectively. There are several studies in the literature on biobutanol production based on ABE fermentation (Kreamer, et al. 2010, Bildea, et al. 2016, Patraşcu, et al. 2017, Kaymak, 2019). However, biobutanol production by IBE fermentation has some other advantages compared to ABE fermentation. First of all, isopropanol produced by IBE is a more requested by-product compared to acetone, because its energy density is higher than acetone. In addition, acetone has some unwanted corrosive effects on engine parts. On the other hand, downstream processing of IBE is more complex because it includes two homogeneous azeotropes between isopropanol-water and ethanol-water in addition to a heterogeneous azeotrope between butanol-water. Besides, isopropanol and ethanol form a closed-boiling mixture.

As far as we know, there are only a few studies on the downstream processing of IBE fermentation in the literature. In their pioneering study, Pyrgakis et al. developed a process flowsheet including a seven column distillation sequence to purify isopropanol,

butanol and ethanol from IBE fermentation broth (2016). Results of this study show that this process flowsheet has a very large energy consumption. In another study, a separation system including two distillation columns has been proposed for IBE dehydration, but this process flowsheet does not include any distillation sequence to purify isopropanol, butanol and ethanol (Grisales Diaz and Olivar Tost, 2017). Recently, Zhang et al. proposed a six column distillation sequence to obtain high purity biobutanol with lower energy consumption (2020). However, this process does not purify other components of IBE, and obtain them as an isopropanol-ethanol mixture. Furthermore, none of the above mentioned papers studied the controllability of the proposed process flowsheets.

The aim of this study is to develop a novel process configuration to purify all three components of IBE fermentation broth with a lower energy consumption compared to the pioneering study mentioned above. The study proposes also a plantwide control structure for this process configuration, which provides a regulatory level control.

2. Process Studied

As mentioned above, the process flowsheet reported by Pyrgakis and co-workers includes seven distillation columns in total, where they proposed a distillation column operating at 0.05 atm to separate ethanol from isopropanol-water mixture, while the azeotropic mixture of isopropanol-water is separated using a pressure swing distillation system (2016). Since the azeotropic composition of isopropanol-water mixture shifts slightly with the change of operating pressure, the use of pressure swing distillation significantly increases the energy consumption of the process. In addition, the ethanol purification column operating at 0.05 atm requires frozen brine to condense the top stream, which significantly increases the operating cost. On the other hand, the six-column configuration suggested by Zhang and co-workers focuses on obtaining high purity biobutanol, but it does not deal with the purification of isopropanol-ethanol mixture (2020). Since isopropanol and ethanol form a closed-boiling mixture, if they were aiming to separate this mixture, they had to use two more distillation column.

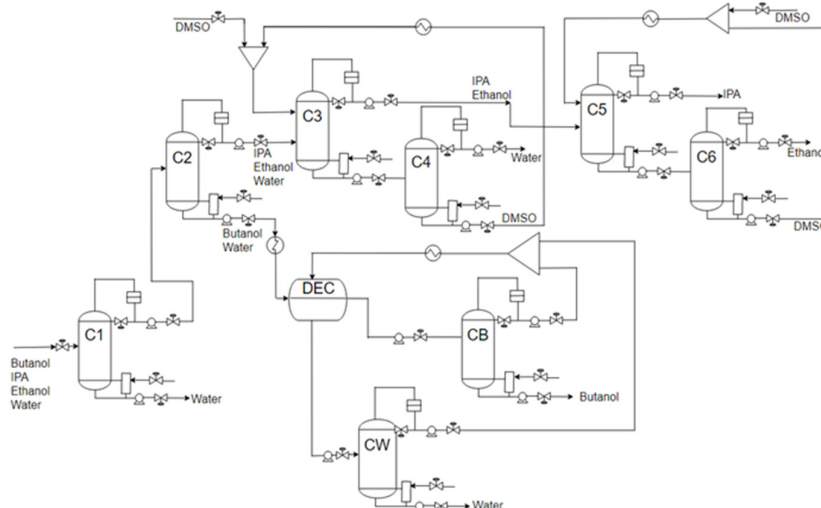


Figure 1. Steady-state design flowsheet

Thus, the configuration proposed here includes eight distillation columns as given in Figure 1 with an aim of purifying all three components of IBE mixture. The flowsheet

consists of a preconcentration column, C1, to take away the excess amount of water, a conventional distillation column, C2, to separate homogeneous and heterogeneous azeotropic mixtures from each other, a decanter-distillation hybrid system, DEC+CB+CW, to separate butanol-water heterogeneous azeotrope, and two extractive distillation systems in series, C3+C4 and C5+C6 to purify water, isopropanol and ethanol, respectively. DMSO is used as the entrainer in extractive distillation part of the process. Version 10.0 of Aspen Plus is used for the simulation of steady-state configuration. The feed properties of the downstream process are obtained from the study of Pyrgakis et al. (2016). The feed flowrate is 31282 kg/h with mass fractions of 0.024, 0.045, 0.003 and 0.928 for isopropanol, butanol, ethanol and water, respectively. NRTL is selected as the suitable thermodynamic model after comparing its results with experimental data. The basic configuration of the IBE purification process discussed in this study is economically optimized using total annual cost (TAC) as the objective function. TAC consists of operating cost and capital cost with a payback period of 3 years. Capital cost includes the cost of column shells, trays, heat exchangers and decanter, while operating cost includes the cost of steam and cooling water. The minimum value of TAC is searched using a sequential iterative optimization method. In this method, design variables such as number of trays and feed tray locations are changed until to find their optimum values minimizing TAC. For each run, “Design Spec/Vary” function of Aspen Plus is used to keep the product purities at their desired values by varying manipulated variables such as reflux ratio and product flowrates. Figure 2 gives the sequential iterative optimization procedure for the extractive distillation part of process.

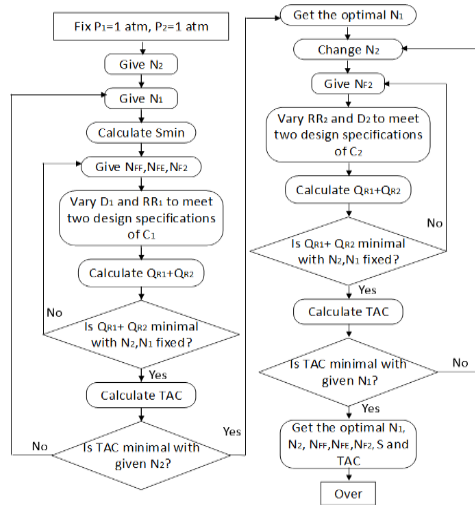


Figure 2. Sequential iterative optimization procedure for extractive distillation

In addition to economic evaluation, CO₂, SO₂ and NO_x emissions are investigated as an environmental sustainable metric, where standard coal is used as the fuel to generate steam. Gas emissions are calculated based on the following formula, where A is the standard coal conversion factor for different gases (2.493, 0.075 and 0.0375), T is the operating time (8000 hr/year), Q_{std} is the calorific value of the standard coal (29307.6 kJ/kg) and Q_R is the reboiler heat duty (Zhang et al. 2020):

$$G_E = A * M_{coal} \tag{1}$$

$$M_{coal} = Q_R * T / Q_{std} \quad (2)$$

3. Results and Discussion

3.1. Steady-State Simulation Results

The design parameters of the optimal case are given in Table 1. N_T , is the total number of trays, while N_F , and $N_{F,S}$ are the feed trays for mixture and entrainer, respectively. P, RR and D_C represent operating pressure, reflux ratio and column diameter. Heat duty of condensers and reboilers are given by Q_C and Q_R , respectively. Results indicate that the total reboiler heat duty of this configuration is 6309 kW. On the other hand, the total reboiler heat duty of seven column configuration of Pyrgakis et al. is 17370 kW, which is reduced to 10990 kW by applying heat integration (2016). Based on these results, it is seen that the basic process configuration proposed here reduces the energy consumption by 42.5%, even compared with the heat-integrated flowsheet of the benchmark study. The total annual cost of this configuration is 1881.7×10^3 \$/yr. In addition, the annual CO_2 , SO_2 , and NO_x emissions of the configuration are calculated as 15455.93, 464.98 and 232.49 t/yr, respectively.

Table 1. Design parameters for basic configuration

	C1	C2	C3	C4	C5	C6	CB	CW	DEC
N_T	27	34	55	24	50	20	19	16	-
N_F	4	6	49	15	35	8	2	2	-
$N_{F,S}$	-	-	5	-	4	-	-	-	-
P (atm)	1	2	1	1	1	1	1	1	1
RR	0.49	3.00	1.11	2.36	2.00	5.00	0.043	0.15	-
D_C (m)	1.01	0.59	0.36	0.47	0.56	0.31	0.61	0.27	1.06
Q_C (kW)	2189	1048	341	368	416	-129	577	153	-
Q_R (kW)	2617	1260	348	398	752	134	606	194	-
TAC (10^3 \$/yr)	706.1	364.6	136.7	96.9	242.8	63.7	191.1	68.5	11.3

3.2. Design of Plantwide Control Structure

Aspen Dynamics is used to run the pressure-driven dynamic simulations. Figure 3 illustrates the plantwide control structure designed for the basic process configuration. A decentralized multi-loop control system employing several single-input and single-output (SISO) feedback controllers are used in this control structure. Feed flowrate acting as the production rate handle is controlled by a flow controller. All reflux drum levels are controlled by manipulating distillate flowrates, while the column base levels are controlled by manipulating bottoms flowrates except C4 and C6. Since these are entrainer recovery columns, their base levels are controlled by arranging make-up streams. Bottoms flowrates of C4 and C6 are manipulated to control the entrainer flowrates to C3 and C5, respectively. For all columns, a feed/reflux ratio controller is used. The operating pressures are controlled by manipulating the heat removal of condensers. The organic and aqueous phase levels in the decanter are controlled by manipulating the corresponding effluent stream flowrates, respectively. The temperatures of the entrainer feed streams are held by manipulating heat duties of HX1 and HX4, respectively. Similarly, the temperatures of decanter feed streams are controlled by arranging the heat duties of HX2 and HX3. Inferential temperature controllers are used to keep the product purities at their

desired values. Tray locations for temperature control are selected based on steepest slope criteria.

All level controllers are P-only with a gain of 2. PI controllers are used for flow, pressure and temperature controls. Temperature controllers with 60 s lag implemented are tuned using ATV test, and controller parameters are calculated using Tyreus-Luyben tuning

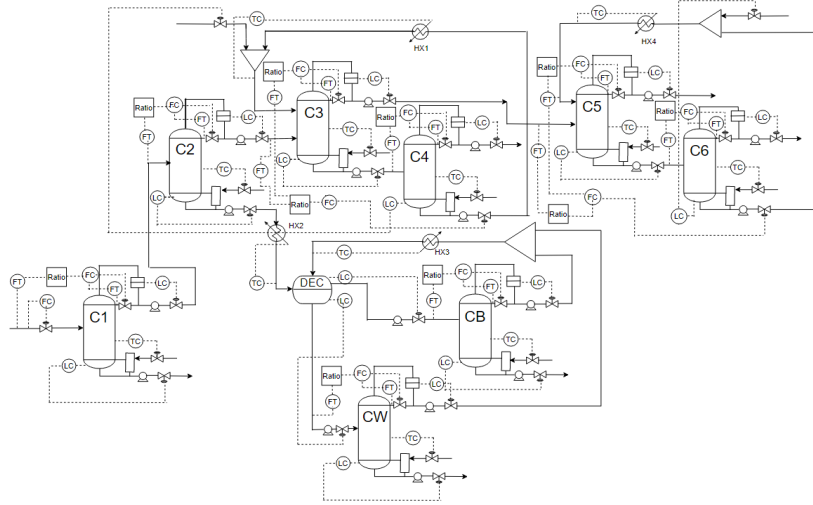


Figure 3. Plantwide control structure

relations. All valves are half opened at the nominal conditions, and the robustness of the control structure is checked against disturbances in feed flowrate and composition.

3.3. Dynamic Simulation Results

Figure 4a illustrates the dynamic response of product purities, controlled tray temperatures and manipulated reboiler heat duties against $\pm 10\%$ changes in production

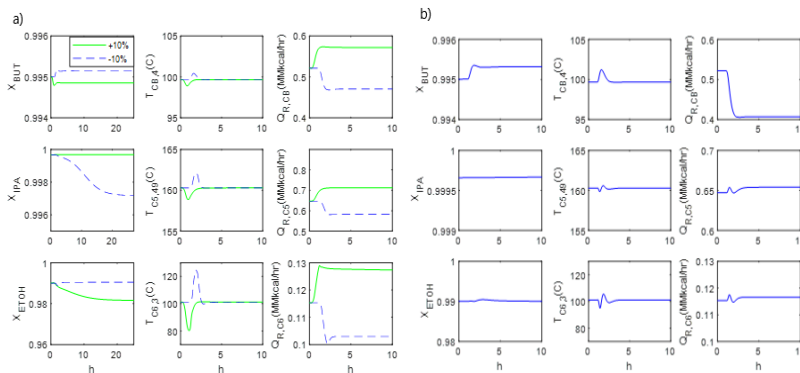


Figure 4: Dynamic response of control structure against disturbances in a) feed flowrate, b) feed composition

rate handle. It is seen that it takes approximately 4 hours for temperatures to recovery back to their set points with proper changes in their manipulated variables. However, it is not the same for product purities. Although the compositions of butanol and isopropanol settle down into close vicinity of their set point, a larger deviation in the ethanol

composition is observed in case of +10% disturbances. It is seen that there are small transient deviations in temperature controllers of butanol and isopropanol purification columns, while transient deviation of temperature controller related to ethanol purification is large. A direct composition control at the top of C6 instead of an inferential temperature control could improve dynamic behaviour ethanol composition.

The response of control structure against a change in feed composition is given in Figure 4b. It is seen that the controlled temperatures recovery back to their set points in less than 5 hours without any big transient deviation. Similarly, the product compositions settle down into their steady-state values with very small deviations in a short transient time.

4. Conclusions

In this study, a new downstream process configuration for isopropanol, butanol and ethanol purification is investigated in terms of design and control. The complex system includes two homogeneous and one heterogeneous binary azeotropes is purified using extractive distillation and decanter-distillation hybrid parts. Optimized process flowsheet reduces the energy consumption significantly compared the benchmark configuration given in the literature. A decentralized plantwide control structure designed for the optimal flowsheet configuration provides an effective base-level regulatory control against disturbance in feed flowrate and composition.

Acknowledgment

This research is financially supported by Research Fund of the Istanbul Technical University (Project Number: MGA-2019-42280).

References

- C.S. Bildea, I. Patraşcu, J.G. S. Hernandez, and A. A. Kiss, 2016, Enhanced Down-Stream Processing of Biobutanol in the ABE Fermentation Process, *Computer Aided Chemical Engineering*, 38, 979-984.
- V. H. Grisales Diaz, and G. Olivar Tost, 2017, Energy efficiency of a new distillation process for isopropanol, butanol, and ethanol (IBE) dehydration, *Chemical Engineering and Processing*, 112, 56-61.
- D.B. Kaymak, 2019, Design and control of an alternative process for biobutanol purification from ABE fermentation, *Industrial & Engineering Chemistry Research*, 58, 1957-1965.
- K. Kraemer, A. Harwardt, R. Bronneberg, and W. Marquardt, 2010, Separation of butanol from acetone-butanol-ethanol fermentation by a hybrid extraction-distillation process, *Computer Aided Chemical Engineering*, 28, 7-12.
- I. Patraşcu, C. S. Bildea, and A. A. Kiss, 2017, Eco-efficient butanol separation in the ABE fermentation process, *Separation and Purification Technology*, 177, 49-61.
- K. A. Pyrgakis, T. Vrije, M. A.W. Budde, K. Kyriakou, A. M. Lopez-Contreras, and A. C. Kokossis, 2016, A process integration approach for the production of biological iso-propanol, butanol and ethanol using gas stripping and adsorption as recovery methods, *Biochemical Engineering Journal*, 116, 176-194.
- H. Zhang, Q. Ye, L. Chen, N. Wang, Y. Xu, and Y. Li, 2020, Purification of isopropanol-butanol-ethanol (IBE) from fermentation broth: process intensification and evaluation, *Chemical Engineering & Processing*, 158, 108182.

Fault Diagnosis of Electrical Power Systems with Hybrid Quantum-Classical Deep Learning

Akshay Ajagekar, Fengqi You

Cornell University, Ithaca, NY 14853, USA

asa273@cornell.edu

Abstract

In this work we develop hybrid QC-based deep learning models for fault diagnosis of electrical power systems that effectively extract suitable features from time-series data using Conditional restricted Boltzmann machine (CRBM) based network. Generatively training the CRBM network using classical learning techniques can be computationally intensive. Therefore, we train the CRBM network with a learning algorithm assisted by a quantum computer that yields better quality of optimal model parameters of the CRBM network. To demonstrate the applicability of the proposed hybrid QC-based deep learning model, we use a standard case study based on IEEE 14-bus that is commonly used to test concepts and methodologies in power systems. The obtained fault diagnosis results show that the proposed hybrid QC-CRBM fault diagnosis model clearly outperforms state-of-the-art classical fault diagnosis models with low missed detection rates and false alarm rates.

Keywords: Quantum computing, deep learning, power systems, hybrid approach

1. Introduction

Critical problems like blackouts and unwanted variations in voltages and currents can be avoided by taking timely protective actions, which require fast and accurate fault diagnosis approaches (Gao et al., 2015). This need motivates the research and development of new fault detection and diagnosis methods capable of detecting and identifying potential abnormalities in electrical power systems to minimize performance degradation (Shang and You, 2019). Pattern recognition techniques have been widely used in fault detection and diagnosis of power systems owing to their ability of extracting useful features from process signals. The ability of an artificial neural network (ANN) to learn new nonlinear relationships and generalize previously unseen data can be applied to solve complex problems in power systems (Haque and Kashtiban, 2000). Support vector machines (SVMs) due to their ability to use multiple classes and mapping to higher dimensions have also been used in fault diagnosis of power systems (Salat and Osowski, 2004).

Quantum computing (QC) which has attracted widespread attention from the scientific community in recent years, is a field that applies quantum mechanical theories to solve complex problems in a multitude of areas. There are various QC applications in process systems optimization (Ajagekar et al., 2020) and energy systems (Ajagekar and You,

2019). Quantum machine learning has also been gaining traction with technological advancements in quantum hardware (Biamonte et al., 2017) and quantum deep learning algorithms (Wiebe et al., 2014). Fast and accurate fault diagnosis can be achieved by quantum advantages offered by QC in terms of both speed and methods of operations (Ajagekar and You, 2020). Harnessing the complementary strengths of both quantum and classical computers to develop hybrid pattern recognition algorithms and overcome QC's limitations is a promising strategy for fault analysis and diagnosis. The objective of this paper is to develop hybrid QC-based deep learning models for fault diagnosis of electrical power systems that effectively extract suitable features from time-series data using conditional restricted Boltzmann machines (CRBM) based network. The superior feature extraction capabilities of the CRBM are combined with discriminative learning to identify the unknown state of measured data samples with high fidelity. To demonstrate the applicability of the proposed hybrid QC-based deep learning model, we use a standard case study based on IEEE 14-bus system.

2. Background

Restricted Boltzmann machines (RBMs) are generative neural networks used to model the unknown data distributions. CRBM is a nonlinear generative model typically used for time-series data. It uses an undirected RBM model with visible neurons and hidden neurons that represent the observable data and dependencies between the latent variables, respectively (Taylor and Hinton, 2009). In a CRBM, the visible variables \mathbf{v} and binary latent variables \mathbf{h} receive directed connections from neurons in the conditioning layer \mathbf{u} that account for the visible data at the last N timesteps. Model parameters of the CRBM network are the weight and bias vectors associated with its visible, conditioning, and hidden layers. Like the RBM, an undirected weight W_{ij} is assigned to the edge between the i^{th} visible unit and the j^{th} hidden unit of the CRBM network. Similarly, weights A_{ki} and B_{kj} are directed from the k^{th} conditioning unit to the i^{th} visible unit and the j^{th} hidden unit, respectively. The static bias terms b_i and c_j are associated with the visible and hidden units, respectively. A CRBM models the

distribution $p(\mathbf{v}|\mathbf{u})$ by using an RBM network to model \mathbf{v} and using the conditioning vector \mathbf{u} to dynamically calculate the biases of that RBM. The energy function of the CRBM for a joint configuration of visible, hidden, and conditioning units is denoted by $E(\mathbf{v}, \mathbf{h}, \mathbf{u})$ as given in Eq. (1). A joint probability distribution over \mathbf{v} and \mathbf{h} and conditional on \mathbf{u} is assigned by the CRBM model through its energy function and is analogous to the Gibbs or Boltzmann distribution.

$$E(\mathbf{v}, \mathbf{h}, \mathbf{u}) = -\sum_{i \in \text{vis}} b_i v_i - \sum_{j \in \text{hid}} c_j h_j - \sum_{i \in \text{vis}, j \in \text{hid}} w_{ij} v_i h_j - \sum_{k \in \text{cond}, i \in \text{vis}} A_{ki} u_k v_i - \sum_{k \in \text{cond}, j \in \text{hid}} B_{kj} u_k h_j \quad (1)$$

$$\frac{\partial \mathcal{L}}{\partial w_{ij}} = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}} \quad \left| \quad \frac{\partial \mathcal{L}}{\partial A_{ki}} = \langle u_k v_i \rangle_{\text{data}} - \langle u_k v_i \rangle_{\text{model}} \quad (2)$$

Learning in CRBMs generally involves performing gradient ascent to maximize the log conditional likelihood of the observed or visible data. Computing the exact model expectations could be very computationally expensive. These model expectations can be approximately estimated through the contrastive divergence (CD) learning algorithm performed on a classical computer that is extensively used to train RBM and other energy-based models (Hinton, 2002).

3. Hybrid Quantum-Classical Fault Diagnosis Model

Fault diagnosis tasks largely depend on feature extraction from the measured signals. Energy based models like RBM and CRBM are capable of learning hierarchical representations of measured data and could be used for such feature extraction tasks. CRBMs can effectively model short-term temporal structures and are suitable choices to model multivariate time-series data (Taylor and Hinton, 2009). Implementation and use of the proposed fault diagnosis model involve two sequential phases, offline training, and online fault classification. Primary components of the offline training phase in building the hybrid QC-based fault diagnosis model are quantum generative training followed by discriminative training. Quantum generative training is an automatic unsupervised learning technique that uses unlabeled measured data. The trained fault diagnosis model can then be used to classify the states of the measured data. The amount of training data required to achieve maximum performance depends on the model complexity and can be determined using statistical heuristics.

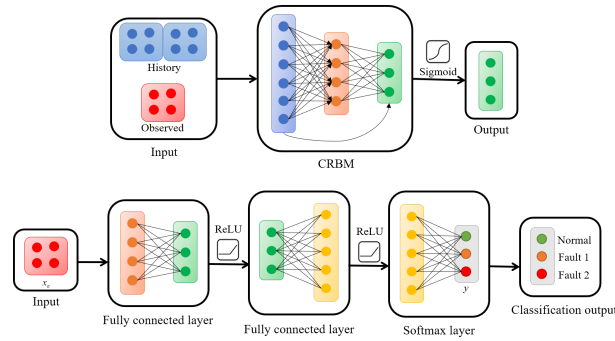


Figure 1. Model architecture of the hybrid quantum-classical fault diagnosis model in generative training and discriminative training phases

As seen in Figure 1, the model architecture consists of several components that include the input, the CRBM network, and the classification network. Input data to the CRBM network consists of two sets of inputs, the observed data samples x_o and the historical data samples x_h . History data sample x_h captures N observed data samples immediately prior to the current observed data sample. CRBMs can effectively capture higher order regularities in the input feature space. For d dimensional observed data samples and their corresponding Nd dimensional conditional data samples, a CRBM network can be constructed with associated weights and dynamic biases as the model parameters.

Weights $A_g \in \mathbb{R}^{Nd \times d}$ and $B_g \in \mathbb{R}^{Nd \times m}$ denote the directed weights between units in the conditional layer and the visible units, and between conditional units and hidden units, respectively, where m latent units are present in the hidden layer. $W_g \in \mathbb{R}^{d \times m}$ represent the weight matrix associated with the connections between the visible and the hidden layer. Additionally, the visible and hidden bias vectors $b_g \in \mathbb{R}^d$ and $c_g \in \mathbb{R}^m$ together with the weights form the CRBM model parameters. The local conditional output probabilities can be easily computed using standard sigmoid function as shown in Eqs. (3) and (4), where $\sigma(x) = 1/(1 + e^{-x})$. With all defined sets of inputs,

outputs, and model parameters, the CRBM network in the hybrid QC-based fault diagnosis model is trained in an unsupervised generative manner. Discriminative training is performed on the classification network and is achieved by attaching classification layers on top of a subgraph of the CRBM network to form the classification network.

$$P(\mathbf{h}=1|\mathbf{v},\mathbf{u})=\sigma(\mathbf{v}\cdot\mathbf{W}_g+\mathbf{u}\cdot\mathbf{B}_g+c_g) \quad (3)$$

$$P(\mathbf{v}=1|\mathbf{h},\mathbf{u})=\sigma(\mathbf{h}\cdot\mathbf{W}_g^T+\mathbf{u}\cdot\mathbf{A}_g+b_g) \quad (4)$$

Energy based models like the CRBM are often trained with the CD algorithm and its variants. However, CD is a biased algorithm that could result in poor estimates of the log conditional likelihood gradients (Carreira-Perpinan and Hinton, 2005). In some cases, CD may not even converge and fail to yield maximum likelihood estimates of the model parameters. To overcome such limitations of the CD algorithm, we train the CRBM network in the hybrid QC-based fault diagnosis model with a new quantum generative training technique proposed in this work. Quantum sampling can be carried out on an adiabatic quantum computing (AQC) device. The use of AQC devices as a sampling engine has been inspired by some of the previous works in pattern recognition (Adachi and Henderson, 2015). Problem formulation step for programming the QPU involves translating the nodes and edges of the objective or energy function to qubits and coupler strengths in the Chimera lattice of the QPU. The energy function is required to take the form of a quadratic unconstrained binary optimization problem. Model expectations required to estimate the log conditional likelihood gradients are calculated through the quantum sampling technique described above. Evidence from experiments conducted with AQC devices has confirmed that AQC devices approximately sample from a Boltzmann distribution under certain conditions (Perdomo-Ortiz et al., 2016). A subgraph of the CRBM network is an RBM network with an energy function given in Eq. (5). Sampling the RBM energy function from an AQC device requires embedding of E_{RBM} on the Chimera lattice of the QPU. As the RBM energy function takes the form of a quadratic unconstrained binary optimization problem, the visible and hidden variables can be mapped as physical qubits on the QPU. Final states of the qubits can be effectively described by a Boltzmann distribution (Benedetti et al., 2016), which allows us to approximately model the joint probability distribution of visible and hidden units. Assuming that N_q samples are drawn from the QPU for the corresponding energy function, the model parameters of the CRBM network are updated based on learning rule in Eq. (10). At every step of the quantum generative training process, the model parameters are updated for every mini-batch of training data to introduce stochasticity in gradient ascent.

$$E_{RBM}(\mathbf{v},\mathbf{h})=-\sum_{i\in vis}b_g^i v_i-\sum_{j\in hid}c_g^j h_j-\sum_{i\in vis,j\in hid}W_g^{ij}v_i h_j \quad (5)$$

$$W_g^{ij}\leftarrow\alpha W_g^{ij}+\varepsilon\left(\left\langle v_i h_j \right\rangle_{data}-\frac{1}{N_q} \sum_{QC} v_i h_j\right) \quad (6)$$

The architecture of the classification network of the hybrid QC-based fault diagnosis model used to detect the state of the measured data is shown in Figure 1. Only observed measurement data and their labels are considered for discriminatively training the

classification network. The bipartite subgraph of the CRBM network with model parameters $W_g \in \mathbb{R}^{d \times m}$ and $c_g \in \mathbb{R}^m$ forms the first fully connected layer of the classification network. An additional fully connected layer is connected to the first layer, where each input neuron is connected to every hidden unit, in order to learn nonlinear combinations of the extracted features. Model parameters of the classification network are fine-tuned and optimized by updating them sequentially, utilizing the gradients of the loss function computed at each step with the backpropagation technique. In supervised learning, categorical cross entropy between the predicted and true labels is minimized. Discriminatively training the classifier network in the hybrid QC-based network yields the model parameters that minimize the total categorical cross entropy loss. The trained model can then be used to predict the unknown state of measured data samples by identifying one of the previously defined faults.

Table 1. Computational results of the IEEE 14-bus case study

Fault Type	Support vector machine		Hybrid QC-based fault diagnosis model	
	MDR (%)	FAR (%)	MDR (%)	FAR (%)
1) ABC / ABCG	0	0	5.3	0
2) AG	5.1	13.78	3.6	1.02
3) BG	0.7	16.4	3.1	3.86
4) CG	12.7	8.89	2.2	6.02
5) AB	0	18.6	8.4	0.91
6) AC	0	22.7	3.5	2.95
7) BC	9.6	7.11	3.3	3.58
8) ABG	0	0	7.3	0
9) ACG	0	0	4.1	0
10) BCG	0	0	10.4	0

4. Computational Results : IEEE 14-Bus System

The performance of the proposed hybrid QC-based fault diagnosis model is evaluated with the IEEE 14-bus system. A MATLAB simulation of this case study is used to collect training and testing data for building and evaluating the hybrid QC-based fault diagnosis model. Three phase voltage data collected at each bus account for the 42 process variables in the IEEE 14-bus simulation. Ten different types of faults are simulated at bus 14 in the case study. We carry out computational experiments with the simulated IEEE 14-bus test system and the proposed hybrid QC-based fault diagnosis model, in order to illustrate the applicability. Quantum generative training of the CRBM network in the hybrid QC-based fault diagnosis model, which is initialized with the above configuration, is performed on an AQC device with a D-Wave 2000Q quantum processor. This QPU provides access to 2,048 qubits and 5,600 couplers that can be used to sample the energy function. We compare the training performance of the CRBM network with CD algorithm (Hinton, 2002). Another instance of the same CRBM network is trained by CD implemented on a classical computer. From the free-energy curves it can be seen that quantum generative training converges much faster than CD learning. This also implies that less computational effort is required by the quantum technique to achieve similar performance. After training the hybrid QC-based fault diagnosis model for the IEEE 14-bus system, we test its fault diagnosis performance by

identifying the state of measured data samples in the testing dataset. Computational results obtained with both methods are reported in Table 1. Symmetrical faults ABC and ABCG that rarely occur in practice are detected by both techniques easily with comparable performance. Severe faults like the line-to-line faults AB, AC, and BC are identified with slightly higher MDRs by using the hybrid QC-based fault diagnosis model for the cases of AB and AC. However, as evident from the computational results, their FARs are significantly low as compared to those of SVM. The MDRs achieved with the hybrid QC-based fault diagnosis model are less than 11%. With low MDRs for both symmetrical and unsymmetrical system faults paired with some of the lowest FARs, the proposed hybrid QC-based fault diagnosis model demonstrates a viable and superior alternative to SVM based fault diagnosis model for the IEEE 14-bus system.

5. Conclusions

In this paper, we proposed a hybrid QC-based fault diagnosis model for analysis and diagnosis of faults in electrical power systems. We performed unsupervised learning of the CRBM network using quantum generative training with quantum sampling facilitated by an AQC device. Discriminative training of the classifier network in the hybrid QC-based fault diagnosis model further allowed us to identify several types of faults from normal operating states. The obtained fault diagnosis results showed that the proposed hybrid QC-based fault diagnosis model clearly outperformed state-of-the-art classical fault diagnosis models.

References

- S. H. Adachi and M. P. Henderson, 2015. Application of quantum annealing to training of deep neural networks. arXiv preprint arXiv:1510.06356.
- A. Ajagekar, F. You, 2019, Quantum computing for energy systems optimization: Challenges and opportunities. *Energy*, 179, 76-89.
- A. Ajagekar, T. Humble, F. You, 2020, Quantum computing based hybrid solution strategies for large-scale discrete-continuous optimization problems. *Computers & Chemical Engineering*, 132, 106630.
- A. Ajagekar, F. You, 2020, Quantum computing assisted deep learning for fault detection and diagnosis in industrial process systems. *Computers & Chemical Engineering*, 143, 107119.
- M. Benedetti, J. Realpe-Gómez, R. Biswas and A. Perdomo-Ortiz, 2016. Estimation of effective temperatures in quantum annealers for sampling applications: A case study with possible applications in deep learning. *Physical Review A*, 94, 022308.
- Y. Chu, F. You, 2015, Model-based integration of control and operations: Overview, challenges, advances, and opportunities. *Computers & Chemical Engineering*, 83, 2-20.
- M. A. Carreira-Perpinan, G. E. Hinton, 2005. On contrastive divergence learning. *Aistats*, Citeseer.
- Z. Gao, C. Cecati, S. X. Ding, 2015. A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches. *IEEE Transactions on Industrial Electronics*, 62, 3757-3767.
- M. T. Haque, A. Kashtiban, 2000. Application of neural networks in power systems; A review. *Power*, 2005.
- G. E. Hinton, 2002. Training products of experts by minimizing contrastive divergence. *Neural Computation* 14(8): 1771-1800.
- C. Ning, F. You, 2019, Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*, 125, 434-448.

- A. Perdomo-Ortiz, B. O'Gorman, J. Fluegemann, R. Biswas and V. N. Smelyanskiy, 2016. "Determination and correction of persistent biases in quantum annealers. *Sci Rep*, 6, 18628.
- R. Salat, S. Osowski, 2004. Accurate fault location in the power transmission line using support vector machine approach. *IEEE Transactions on Power Systems*, 19, 979-986.
- C. Shang, F. You, 2019, *Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era*. *Engineering*, 5, 1010-1016.
- G. W. Taylor, G. E. Hinton, 2009. Factored conditional restricted Boltzmann Machines for modeling motion style. *Proceedings of the 26th Annual International Conference on Machine Learning*. Montreal, Quebec, Canada, Association for Computing Machinery: 1025–1032.
- N. Wiebe, A. Kapoor, K. M. Svore, 2014. Quantum deep learning. *arXiv preprint arXiv:1412.3489*.
- J., P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe and S. Lloyd (2017). "Quantum machine learning." *Nature* 549: 195-202.

Data-Driven Robust MPC for Controlled Environment Agriculture

Wei-Han Chen and Fengqi You

Cornell University, Ithaca, New York 14853, USA

wc593@cornell.edu

Abstract

In this work, a data-driven robust model predictive control (DDRMPC) framework for greenhouse climate control is proposed to minimize the total control cost, while minimizing the constraint violation on system states. First, the state-space model of the greenhouse is generated. Next, historical data on weather forecast errors which represent uncertain disturbances are collected. Support vector clustering with weighted generalized intersection kernel is adopted for constructing data-driven uncertainty sets of ambient temperature, solar radiation, and humidity. The historical data information can then be incorporated into robust model predictive control (RMPC). Uncertainty sets constructed in the proposed DDRMPC framework balance between reducing control cost and conservatism. A case study of controlling indoor temperature, CO₂ concentration, and humidity of a semi-closed greenhouse in New York City is presented to demonstrate the advantages of the proposed DDRMPC framework for greenhouse climate control over rule-based control, certainty-equivalent model predictive control, and RMPC strategies.

Keywords: Model predictive control, greenhouse climate control, uncertainty.

1. Introduction

The purpose of greenhouse is to protect plants from outdoor weather (Bakker et al., 1995). Temperature, CO₂ concentration, and humidity are the most important factors of the greenhouse's indoor climate that should be carefully regulated in controlled environment agriculture. Model predictive control (MPC) is an ideal framework for greenhouse control (Chen, 2018), because building dynamics are slow and the system model incorporates disturbances and constraints that can be derived from first principles models (Serale et al., 2018). However, most existing studies on MPC for greenhouse climate control do not consider robustness (Blasco et al., 2007), given that uncertain disturbances could deviate the greenhouse climate from the optimal condition (Shamshiri et al., 2018). For those studies using robust MPC (RMPC), the possibility of over-conservatism due to the oversized uncertainty set is not considered (Shang et al., 2019). Thus, to fill this knowledge gap, the objective of this work is to develop a novel RMPC framework for greenhouse climate control that can effectively hedge against uncertain disturbances, leverage the value of historical weather forecast data to reduce over-conservatism, and simultaneously control multiple system states of greenhouse climate to minimize the total cost and the constraint violation probability.

2. Greenhouse Dynamic Model Formulation

In this work, the states we consider are greenhouse air temperature, floor temperature, ceiling temperature, wall temperature, absolute humidity, and greenhouse CO₂ concentration. The control inputs are the heating and cooling power, dehumidifier, and CO₂ enrichment. The disturbances are solar radiation, ambient temperature, ground temperature, ambient humidity, and ambient CO₂ concentration. Forecast errors are only from solar radiation, ambient temperature, and absolute humidity. The structure of the proposed dynamic model for greenhouse climate control is shown in Figure 1.

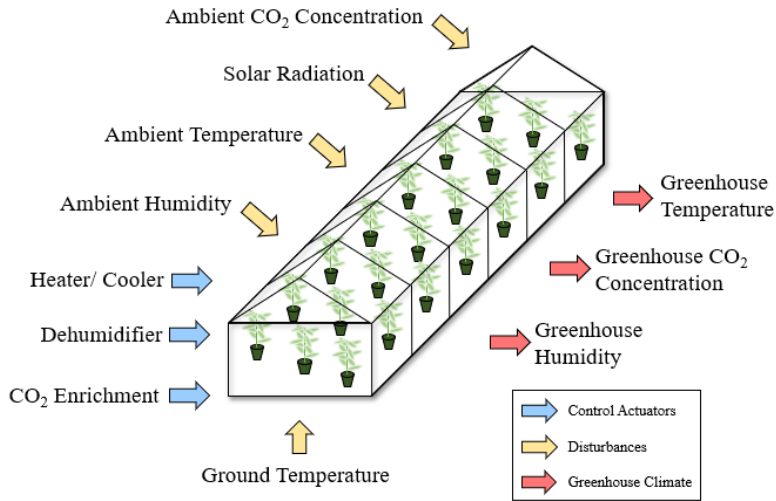


Figure 1. Semi-closed greenhouse structure model that shows control actuators, disturbances, and greenhouse climate system states.

To control greenhouse CO₂ concentration, a model to calculate net uptake rate of CO₂ by crops per unit greenhouse area is required, and it can be estimated by an empirical model of net photosynthesis as follows (Klärning et al., 2007).

$$P_{net} = a_0 (f_I(I) \cdot f_C(C) \cdot f_{TP}(T) + f_{TR}(T)) f_L(LAI)$$

where P_{net} is the net photosynthesis, I is the photosynthetically active radiation (PAR), C is the CO₂ concentration in the air, T is the air temperature, LAI is the leaf area index of the crop, and a_0 are coefficients that can be obtained from the literature (Klärning et al., 2007). The functions f_I , f_C , and f_{TP} show the effects of solar radiation, CO₂ concentration and temperature on photosynthesis, respectively. f_{TR} is for the temperature effect on respiration, and f_L describes the effect of the leaf index on both photosynthesis and respiration.

For greenhouse humidity control, humidity is strongly affected by temperature. Since the temperature would be controlled within a small range in the greenhouse that is most beneficial for the crop's growth, absolute humidity could be taken into account in the MPC to avoid the complexity of dependency between temperature and relative humidity.

$$\frac{dQ_{gr}}{dt} = \frac{1}{l}(Q_{tran} - Q_{cond} + Q_{amb})$$

where Q_{gr} is the greenhouse absolute humidity, Q_{tran} is the plant transpiration, Q_{cond} is the condensation, Q_{amb} is the flux of ambient absolute humidity, and l is the greenhouse height.

The dynamic greenhouse climate model in this work combines the greenhouse temperature model generated by Building Resistance-Capacitance Modeling (BRCM) toolbox, CO₂ concentration model, and absolute humidity model, given in Eq. (1)-(2). When the length of prediction horizon H is given, a compact form of the dynamic greenhouse climate model is given by

$$\mathbf{x} = \mathbf{A}\mathbf{x}_0 + \mathbf{B}_u\mathbf{u} + \mathbf{B}_v\mathbf{v} + \mathbf{B}_w\mathbf{w}$$

where \mathbf{x} includes greenhouse air temperature, ceiling temperature, floor temperature, wall temperature, greenhouse CO₂ concentration, and absolute humidity. \mathbf{u} consists of heating/cooling power, CO₂ enrichment, and dehumidifying rate. \mathbf{v} contains forecast of ambient temperature, ground temperature, solar radiation, and ambient humidity. \mathbf{w} has forecast error for ambient temperature, solar radiation, and ambient humidity.

The constraints on control inputs and system states in the prediction horizon H can be stacked together. The compact forms are shown as

$$\mathbf{F}_x\mathbf{x} \leq \mathbf{f}_x, \quad \mathbf{F}_u\mathbf{u} \leq \mathbf{f}_u$$

where \mathbf{F}_x and \mathbf{f}_x are vectors that define the state constraints. \mathbf{F}_u and \mathbf{f}_u are vectors that define the control input constraints.

3. Control Strategies

3.1. Data-Driven Uncertainty Set for RMPC

In this work, we adopt support vector clustering (SVC) with weighted generalized intersection kernel (WGIK) approach to form uncertainty sets of temperature, solar radiation, and humidity due to its capability of handling asymmetry and correlation in the uncertainty distribution, as well as tackling outliers (Shang and You, 2019). The SVC approach tries to find the minimal sphere radius that can capture data (Shang et al., 2017); WGIK is implemented when solving the dual form of SVC optimization problem, and it is well suited for robust optimization due to its piecewise linearity (Shang et al., 2020). To construct the uncertainty set for ambient temperature, training data have to be first obtained. By adopting SVC with WGIK approach, the uncertainty set for temperature is formed as follows,

$$w_{temp} \in D_{temp} = \left\{ w_{temp} \mid \sum_{i \in SV} \alpha_i \left\| \mathbf{Q}_{temp} (w_{temp} - w_{temp}^{(i)}) \right\|_1 \leq \theta \right\}$$

where D_{temp} is the data-driven uncertainty set for temperature forecast error, \mathbf{Q}_{temp} is a weighting matrix that can be obtained from the covariance matrix of temperature forecast error w_{temp} , model parameters α_i and uncertainty set parameters θ are determined after solving the dual form of SVC using WGIK. Since Eq. (5) is a polytope, solving the

resulting robust optimization problem could be accomplished without difficulties. The uncertainty sets for solar radiation and absolute humidity are constructed in a similar way. Other types of data-driven uncertainty sets (Ning and You, 2019), such as those based on principle component analysis (Ning and You, 2018), could be used as well.

3.2. *Soft-Constrained Data-Driven Robust Model Predictive Control*

DDRMPC strategy in this work adopts SVC with WGIC to construct uncertainty sets that could tackle outliers of weather forecast errors. Furthermore, the performance guarantee is ensured after tuning uncertainty sets by the calibration data set. The approach to solving the optimization problem in DDRMPC also uses affine disturbance feedback (ADF) policy to obtain the approximate solution given that the original problem is intractable (Goulart et al., 2006).

Although ADF policy could transform the original optimization problem into a tractable problem, it is impossible to guarantee that it is always feasible at each time step due to the presence of system state constraints, control input constraints, and disturbances. A common approach to guarantee the recursive feasibility of MPC is to soften constraints that might be violated. The corresponding soft-constrained MPC is formulated as,

$$\begin{aligned} & \min_{\mathbf{M}, \mathbf{h}, \boldsymbol{\varepsilon}} \mathbf{c}^T \mathbf{h} + \boldsymbol{\varepsilon}^T \mathbf{S} \\ \text{s.t. } & \mathbf{F}_x [\mathbf{A}x_0 + \mathbf{B}_u \mathbf{h} + \mathbf{B}_v \mathbf{v} + (\mathbf{B}_u \mathbf{M} + \mathbf{B}_w) \mathbf{w}] \leq \mathbf{f}_x + \boldsymbol{\varepsilon}, \forall \mathbf{w} \in D \\ & \mathbf{F}_u [\mathbf{M} \mathbf{w} + \mathbf{h}] \leq \mathbf{f}_u, \forall \mathbf{w} \in D \\ & \boldsymbol{\varepsilon} \geq \mathbf{0} \end{aligned}$$

where \mathbf{S} is constraint violation penalty weight matrix, and $\boldsymbol{\varepsilon}$ is slack vector for system state constraints. Since this is a minimization problem and the penalty weights in \mathbf{S} are all positive, slack variables would be zero when hard-constrained DDRMPC is feasible.

4. **Case Study**

4.1. *Problem Description*

In this work, a semi-closed greenhouse located in Brooklyn, New York, USA for year-round tomato production is simulated for closed-loop temperature, humidity, and CO₂ concentration level control under four different control strategies: rule-based control, certainty equivalence MPC (CEMPC), RMPC, and DDRMPC. The dimension of this greenhouse is 40 m × 13 m × 4 m. The main material for roof and walls is 10 mm twin-wall polycarbonate which provides good insulation against heat. The floor is made of concrete. Historical temperature forecast data, historical temperature measurement data, and historical solar radiation data from May 2018 to June 2018 are collected from (Meteogram Generator, 2018). Historical solar radiation forecast data is hard to access, so we adopt the following model from literature (Kasten and Czeplak, 1980) to estimate solar radiation from cloud coverage, since historical cloud coverage forecast data is easier to obtain, compared to historical solar radiation forecast data.

4.2. *Results and Discussion*

Figure 2 presents the temperature control profile of May under all four control strategies. The trajectory of DDRMPC leaves some margin from the temperature constraints, so as to avoid constraint violation when the weather forecast errors are large. Therefore, due to the margin saved for prediction error, temperature constraint

would not be violated, or there is at most only very minor violation when extreme cases strike. Similar to DDRMPC, RMPC also leaves some margin for the temperature constraints. However, the margin left by RMPC is usually larger than DDRMPC's, especially from 10pm to 5am. The comparison clearly demonstrates the ability of DDRMPC to reduce conservatism. On the other hand, both CEMPC and rule-based control violate the constraints more severely. The reason for the violation is that CEMPC does not consider prediction error when the MPC optimization problem is being solved at each step, and rule-based control does not take the weather forecast into account. When the actual ambient temperature turns out to be lower than the predicted ambient temperature, greenhouse temperature could not be maintained to be above the minimum acceptable temperature, and this is when constraint violation happens. The results demonstrate the robustness of RMPC.

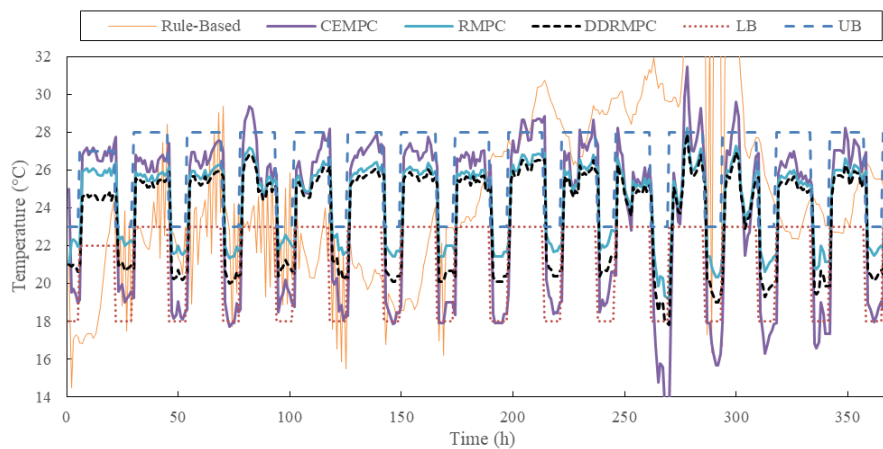


Figure 2. Greenhouse temperature profile in first half of May 2018 under different control strategies with upper bound constraint and lower bound constraint that change according to time of the day. Higher constraint during daytime and lower constraint during nighttime.

Table 1. Percentage of time violating temperature constraint in each month

	May	June	Average.
Rule-Based (%)	69.1	68.1	68.6
CEMPC (%)	76.5	79.14	77.82
RMPC (%)	0.13	0.28	0.21
DDRMPC (%)	0.27	0.41	0.39

Table 2. Cost of controlling temperature, CO₂ concentration, and absolute humidity under different control strategies in each month

	May	June	Total
Rule-Based (\$)	6,237	5,144	11,381
CEMPC (\$)	4,716	3,601	8,317
RMPC (\$)	5,801	4,372	10,173
DDRMPC (\$)	5,604	4,162	9,766

Overall, CEMPC shows the least conservative control profile. Although CEMPC has the lowest overall control cost, which is the sum of heating/cooling cost, CO₂ cost, and water cost, shown in Table 2, it ends up with violating system constraints the most as shown in Table 1. Neglecting the forecast errors makes CEMPC violate constraints, because of the difference between the real weather condition and the weather forecast.

Whenever the weather is colder than predicted, CEMPC would violate the constraint, because the controller neglects the weather forecast errors and usually takes the most aggressive actions. Some violations are so severe that the greenhouse indoor air temperature drops to 14°C, which is 4°C below the minimum required temperature for tomato growth. This is undesirable in a greenhouse because tomatoes are usually sensitive to temperature. Even a 2-3°C temperature drop may cause serious damages to the fruits.

5. Conclusions

In this paper, we developed a DDRMPC framework that was able to simultaneously control multiple system states for indoor climate of a semi-closed greenhouse. To prevent greenhouse climate from becoming harmful to plant and fruit due to harsh weather conditions and inherent uncertainty in weather forecast, uncertainty sets for ambient temperature, solar radiation, and absolute humidity were first constructed by adopting the SVC approach with WGIK on historical weather data. Affine disturbance feedback policy was implemented for solving the optimization problem in proposed DDRMPC. We presented a case study of simultaneously controlling the indoor temperature, CO₂ concentration, and absolute humidity of a semi-closed greenhouse located in Brooklyn, New York for tomato production. The results showed that DDRMPC reduces 14% and 4% control cost compared with rule-based control and CEMPC, respectively, while only violating temperature constraint 0.39% of time. Therefore, DDRMPC showed better control performance compared to other approaches for greenhouse climate control.

References

- J. C. Bakker, G. P. A. Bot, H. Challa, et al. (1995). Greenhouse climate control: an integrated approach. Wageningen Academic Publishers, Netherlands.
- X. Blasco, M. Martínez, J. M. Herrero, C. Ramos, and J. Sanchis, 2007, Model-based predictive control of greenhouse climate for reducing energy and water consumption, *Computers and Electronics in Agriculture*, vol. 55, no. 1, pp. 49-70.
- L. Chen, S. Du, Y. He, M. Liang, and D. Xu, 2018, Robust model predictive control for greenhouse temperature based on particle swarm optimization, *Information Processing in Agriculture*, vol. 5, no. 3, pp. 329-338.
- Y. Chu, F. You, 2015, Model-based integration of control and operations: Overview, challenges, advances, and opportunities. *Computers & Chemical Engineering*, 83, 2-20.
- P. J. Goulart, E. C. Kerrigan, and J. M. Maciejowski, 2006, Optimization over state feedback policies for robust control with constraints, *Automatica*, vol. 42, no. 4, pp. 523-533.
- F. Kasten and G. Czeplak, 1980, Solar and terrestrial-radiation dependent on the amount and type of cloud, *Solar Energy*, vol. 24, no. 2, pp. 177-189.
- H. P. Kläring, C. Hauschild, A. Heißner, and B. Bar-Yosef, 2007, Model-based control of CO₂ concentration in greenhouses at ambient levels increases cucumber yield, *Agricultural and Forest Meteorology*, vol. 143, no. 3, pp. 208-216.
- Meteogram Generator. Dept. of Geological and Atmospheric Sciences, Iowa State University. Available: https://www.meteor.iastate.edu/~ckarsten/bufkit/image_loader.phtml
- C. Ning, F. You, 2018, Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. *Computers & Chemical Engineering*, 112, 190-210.
- C. Ning, F. You, 2019, Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*, 125, 434-448.

- G. Serale, M. Fiorentini, A. Capozzoli, D. Bernardini, and A. Bemporad, 2018, Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: problem formulation, applications and opportunities, *Energies*, vol. 11, no. 3, pp. 631.
- R. R. Shamshiri, F. Kalantari, K. C. Ting, et al., 2018, Advances in greenhouse automation and controlled environment agriculture: a transition to plant factories and urban agriculture, *International Journal of Agricultural and Biological Engineering*, vol. 11, no. 1, pp. 1-22.
- C. Shang, X. Huang, F. You, 2017, Data-driven robust optimization based on kernel learning. *Computers & Chemical Engineering*, 106, 464-479.
- C. Shang and F. You, 2019, A data-driven robust optimization approach to scenario-based stochastic model predictive control, *Journal of Process Control*, vol. 75, pp. 24-39.
- C. Shang, F. You, 2019, Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era. *Engineering*, 5, 1010-1016.
- C. Shang, W.-H. Chen, A. D. Stroock, et al., 2020, Robust Model Predictive Control of Irrigation Systems With Active Uncertainty Learning and Data Analytics. *Ieee Transactions on Control Systems Technology*, 28, 1493-1504.

Initialization strategy for the dynamic optimization of a batch distillation column with a rate-based model

Mayra Margarita May-Vázquez,^a Fernando Israel Gómez-Castro,^{a,*} Edna Soraya Rawlings,^b Vicente Rico-Ramírez,^c Mario Alberto Rodríguez-Ángeles^d

^a *Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Campus Guanajuato, Universidad de Guanajuato, Noria Alta S/N, Col. Noria Alta, Guanajuato, Guanajuato, 36050, México*

^b *Independent Researcher, 664 Hagemann Drive, Livermore, CA, 94551, USA*

^c *Departamento de Ingeniería Química, Tecnológico Nacional de México en Celaya, Av. Tecnológico y García Cubas S/N, Col. FOVISSTE, Celaya, Guanajuato, 38010, México*

^d *Departamento de Ingeniería en Plásticos, Universidad Politécnica de Juventino Rosas, Calle Hidalgo 102, Comunidad de Valencia, Santa Cruz de Juventino Rosas, Guanajuato, 38253, México*
fgomez@ugto.mx

Abstract

Batch distillation is one of the most common separation processes used in the pharmaceutical, agricultural, food and biotechnology industries. The rate-based model is an accurate approach to represent the performance of distillation trays, since it involves the mass and energy transfer equations to describe the interaction between the phases. For batch distillation, such approach results in a set of nonlinear differential and algebraic equations (DAE). Because of the complexity of such systems, good initial values are required for their solution. Thus, a strategy to estimate the initial values for a trayed batch distillation column modeled with a rate-based approach is presented in this work. The column under investigation consists of four sieve trays, a pot, and a total condenser. For the initialization procedure, the model equations are solved under a steady state assumption.

Keywords: rate-based modelling, batch distillation, discretization.

1. Introduction

Distillation is the process of separating of components in the mixture based on the differences among the relative volatilities or boiling points. Distillation can be to continuous or batch, there are some guidelines that can help to decide when a batch process may be favored over a continuous one. Batch distillation the process of separation for liquid mixture that containing high value chemical, such as pharmaceutical, biochemical, food, agricultural industries. Due to their flexibility, batch column can be used for a variety of system with relatively small amounts of components, under various operating conditions and changing compositions and components. In recent years, different models have been proposed to study the performance of batch distillation columns, mainly divided into two different types: equilibrium and rate-based models. The first model requires the solution of mass and energy balances while considering the vapor-

liquid streams leaving the stage to reach thermodynamic equilibrium, but its accuracy depends on the efficiency values assumed for the trays. The rate-based model is more accurate since it involves the addition of the mass and energy transfer equations to describe the interaction between the phases. Correlations of mass and heat transport coefficients are required. The column is described by the MERSHQ equations. May-Vázquez et al. (2020) presented a rate-based model of a batch column to separate methanol and ethanol, in which a novel mass transfer model for sieve trays is incorporated to the rate-based equations. The resulting formulation comprises a set of nonlinear differential and algebraic equations (DAE), whose solution require a significant computational effort. Numerical solution techniques for solving the differential equations have been development: one step methods and multistep methods. Some methods used are Euler's, Runge-Kutta, Adams-Bashforth and Predictor-Corrector. Another classification of the integration techniques depends on whether the method is explicit or not, semi-implicit or implicit: BDF (backward difference formula) and orthogonal collocation method (Diwekar, 2014). The modeling, simulation, and optimization of batch distillation columns is of interest to improve the performance of this dynamic process, involving the solution of the set of DAEs through dynamic optimization. Safdarnejad et al. (2016) mentioned that the optimization column batch can be subdivided into optimal design problems and optimal control problems. There is a difference between the two types of optimization, the first generally deal with column configuration while the second deal with operation. The optimization of the batch column is focused on the solution of optimal control problems, the indices of performed such as maximum distillate, minimum time, and maximum profit. To solve this problem, earlier works (Logsdon and Biegler, 1989) proposed the use of orthogonal collocation over finite elements as a discretization approach to simultaneously simulate and optimize batch distillation columns. Most of the recent literature related to dynamic optimization of such columns use equilibrium models; rate-based approaches are mainly used for packed columns. In any case, to solve this complex set of DAEs using orthogonal collocation over finite elements, it is necessary to provide the initial values and limits of the variables. In this work, a strategy to estimate the initial values for a trayed batch distillation column modeled with a rate-based approach is presented; these values will be used as the initial values for future works to perform the dynamic optimization of the column.

2. Methodology

For the initialization procedure, the model equations are solved under a steady state assumption, approach that has been applied to initialize equilibrium models for batch columns (López-Saucedo et al., 2016). In this work, the use of this technique is extended for the initialization of more complex rate-based models. The methodology is divided into three stages: (i) rate-based model, (ii) initial values using Matlab R2013a, (iii) steady state simulation using GAMS.

The batch distillation column consists four sieve trays, a pot, and a total condenser. Figure 1 shows a diagram of the column. The batch system separates a methanol/ethanol, mixture, with an initial composition of 75% mol of methanol and 25% mol of ethanol. Initial charge of the mixture is 0.163 kmol. The operation is stopped at 80 minutes, with constant thermal duty of 500 W.

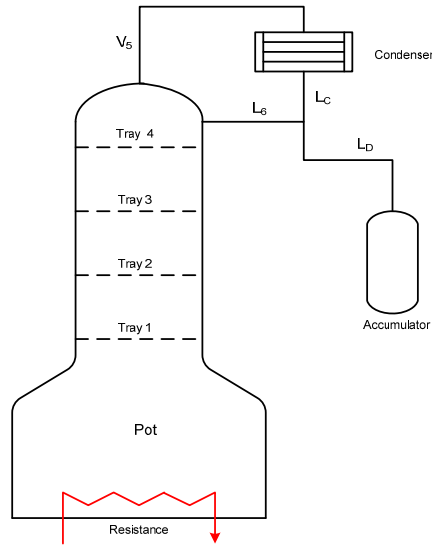


Figure 1.- Simplified representation of the batch column

The rate-based model is based on the MERSHQ equations and the following assumptions: vapor-liquid interface reaches thermodynamic equilibrium, the trays are in mechanical equilibrium, a total condenser is used, the walls of the column are perfectly insulated and the pressure along the column is constant. The mass transfer model is based on the proposal of Taylor and Krishna for sieve trays (Eq.11), with the modifications reported by May-Vazquez et al. (2020). It was also necessary to introduce correlation or methods to calculate physical and transport properties (density, viscosity, diffusivity, heat capacity and thermal conductivity).

The equations used to model the behavior of the batch column are presented below.

Accumulator:

Total material balance

$$\frac{dM_a}{dt} = L_D \quad (1)$$

Component material balance

$$\frac{dx_{i,a}}{dt} = \frac{L_D (x_{i,j} - x_{i,a})}{M_a} \quad (2)$$

Internal Trays, j=1, ...,4; i=1, 2

Total balance material for the vapor and liquid phase

$$\frac{dM_j^V}{dt} = V_{j-i} - V_j + N_{T,j} \quad (3)$$

$$\frac{dM_j^L}{dt} = L_{j+i} - L_j - N_{T,j} \quad (4)$$

Component material balances for component i in vapor and liquid phases on stage j

$$\frac{dy_{i,j}}{dt} = \frac{V_{j-1}}{M_j^V} y_{i,j-1} - \frac{V_j}{M_j^V} y_{i,j} + \frac{N_{i,j}^V}{M_j^V} \quad (5)$$

$$\frac{dx_{i,j}}{dt} = \frac{L_{j+1}}{M_j^L} x_{i,j+1} - \frac{L_j}{M_j^L} x_{i,j} - \frac{N_{i,j}^L}{M_j^L} \quad (6)$$

Energy balances on vapor and liquid phases on stage j

$$0 = V_{j-1}H_{j-1} - V_jH_j + e_j^V \quad (7)$$

$$0 = L_{j+1}h_{j+1} - L_jh_j - e_j^L \quad (8)$$

Rate of mass and energy transfer across interphase

$$e_j^V = e_j^L \quad (9)$$

$$\sum_{i=1}^c N_{i,j}^V = \sum_{i=1}^c N_{i,j}^L = N_{T,j} \quad (10)$$

$$N_i = c_i^V K_{OV} (\Omega^V)^{-1} (y_E - y_i^*) \quad (11)$$

$$N_{i,j}^V = N_{i,j}^L \quad (12)$$

Summation of mol fractions

$$\sum_{i=1}^c y_{i,j}^I = 1.0 \quad (13)$$

$$\sum_{i=1}^c x_{i,j}^I = 1.0 \quad (14)$$

Equilibrium equations

$$y_{i,j}^I = K_{i,j} x_{i,j}^I \quad (15)$$

Reboiler

Total material balance

$$\frac{dM_j}{dt} = L_{j+1} - V_j \quad (16)$$

Component material balances

$$\frac{dx_{i,j}}{dt} = \frac{L_{j+1}x_{i,j+1} - V_j y_{j,i}}{M_a} \quad (17)$$

It was necessary to have initial values for compositions and enthalpies to solve the steady state simulation. To obtain these values, the system of algebraic and differential equations of the rate-based model was written in Matlab R2013 at total reflux for 10 minutes, and the initial composition in the pot was 75% methanol and 25% ethanol. At the beginning of the simulation, the compositions and enthalpies were considered zero at the sieve trays and accumulator. The fourth order Runge-Kutta method with constant step size of 0.001

The differential equations of the rate-based model are equal to zero (Eq.1 to Eq.6 and Eq.16 to Eq.17) and the initial values obtained from MATLAB were used for the initialization of the steady state simulation. The system of algebraic system was solved at GAMS. It was also necessary to introduce upper and lower bounds to the variables.

3. Results

In this section, the obtained experimental data at total reflux is compared with the data obtained by the developed steady-state model.

The system of algebraic equations of the rate-based model at steady state of the binary batch distillation column consists of 301 equations and 28 inequality constraints. CONOPT was used to solve the nonlinear programming.

Figure 1 shows the composition profile when the column operates at steady state. The stage 1 corresponds to the pot and stage 6 to the tray 4. In this figure the composition of methanol in the liquid phase increases in each tray since it is the lightest component. The Figure shows a good agreement between the experimental results and steady state simulation, the average error is 1.63%.

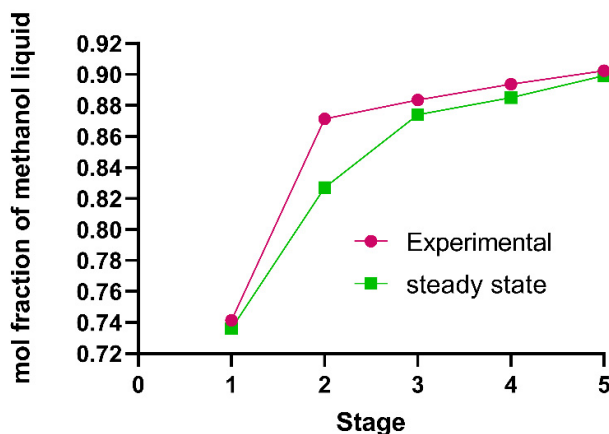


Figure 2.- Composition profile

Figure 2 shows the temperature profile obtained by both, the temperature is lower in the upper stages because the amount of methanol increases. These results indicate that the steady state simulation has a good approximation with the experimental results. The deviations between the experimental data and steady state were $\pm 1.21\%$.

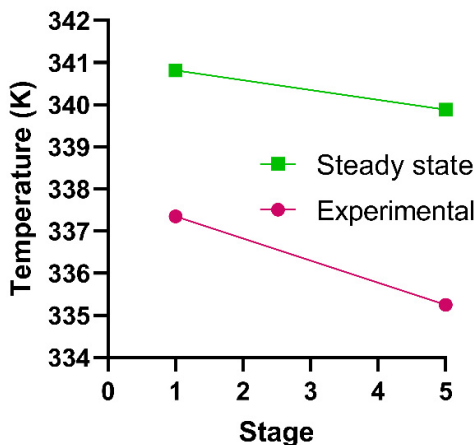


Figure 3.-Temperature profile

4. Conclusions

A strategy to estimate the initial values for a batch distillation column to scale pilot was proposed, the strategy shows a good approximation with the experimental results. These values will be used by the solution of the optimal control. The initial values and limits to the upper and lower bounds on variables are very important for the simulation in GAMS to converge.

Acknowledgements

The authors acknowledge the financial support provided by Universidad de Guanajuato and CONACYT, through the scholarship granted to M.M. May-Vázquez.

References

- L. T. Biegler, 2010. *Nonlinear Programming Concepts, Algorithms, and Applications to Chemical Processes*. Society of Industrial and Applied Mathematics and the Mathematical Optimization Society.
- U. Diwekar, 2014, *Batch Processing: Modeling and Design*, 1st edition, CRC Press.
- J. S. Logsdon, L. T. Biegler, 1989, Accurate Solution of Differential-Algebraic Optimization Problem, *Ind. Eng. Chem. Res.*, 28, 1628-1639.
- E. S. Lopez-Saucedo, I. E. Grossmann, J. G. Segovia-Hernández, S. Hernández, 2016, Rigorous modeling and optimization of a conventional and nonconventional batch reactive distillation column: a comparative study of dynamic optimization approaches, *Chem. Eng. Res. Des.*, 111, 83-99.
- M. M. May-Vazquez, M. A. Rodríguez-Ángeles, F. I. Gómez-Castro, J. Espinoza-Zamora, E. Murriera-Luna, 2020, Development of a mass transfer model for the rate-based simulation of a batch distillation column, *Comput. Chem. Eng.*, 140, 106981.
- S. M. Safdarnejad, J. R. Gallacher, J. D. Hedergren, 2016, Dynamic parameter estimation and optimization for batch distillation, *Comp. Chem. Eng.*, 86, 18-32.p

Adversarial Autoencoder Based Nonlinear Process Monitoring

Kyojin Jang^a, Minsu Kim^a, Hyungjoon Yoon^a, Jonggeol Na^{b,*}, Il Moon^{a,*}

^a*Department of Chemical and Biomolecular Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea*

^b*Department of Chemical Engineering and Materials Science, Ewha Womans University, Seoul 03760, Republic of Korea*

email of the corresponding author: pselab@yonsei.ac.kr

Abstract

As the scale and complexity of the chemical process increase, it is important to detect anomalies in the process at an early stage and respond in real-time. Currently, however, it is difficult for process operators to identify numerous alarms in the factory and to make a consistent and immediate abnormal diagnosis because each has different safety standards. To this end, this study proposed an adversarial autoencoder(AAE) based process monitoring model. AAE uses adversarial training to impose an arbitrary prior distribution on the latent vectors. In other words, the discriminator is trained to distinguish between the samples from the data distribution and the samples from the encoder, and the encoder is trained to match the latent vectors with a prior distribution. In the AAE-based process monitoring model, normal condition samples are used for train data and prior distribution is set up to be Gaussian distribution. T^2 and SPE statistics are constructed in the feature space and residual space respectively to monitor the process. By employing AAE, the model learns a deep generative representation that maps the original data distribution. To demonstrate the performance of the proposed model, a case study using the Tennessee Eastman benchmark process is employed. False alarm rate(FAR) and false detection rate(FDR) are used as the assessment criteria to measure the monitoring performance.

Keywords: Nonlinear process monitoring, Adversarial autoencoder, Fault detection, Gaussian feature learning

1. Introduction

Timely detection of faults in chemical systems is critical to ensuring the safety product quality. However, currently, these methods are relying upon human operators who need to recognize anomalies and make timely corrective decisions. Therefore, there has been a need to develop an intelligent automated fault detection system to assist operators in handling abnormal situations. To this end, various methods have been developed over three decades(N. Olivier-Maget et al, 2008). Data-driven methods possess great potential to be applied to chemical processes since a large amount of data is collected and stored in a distributed control system. Initially, principal component analysis (PCA)(S.Joe Qin, 2003), the most popular feature extraction method, has been widely used for monitoring linear processes. However, the data characteristics of many modern industrial processes are complicated, and the relationships among different variables are highly nonlinear. To overcome this, several nonlinear methods have been proposed. One of them is kernel

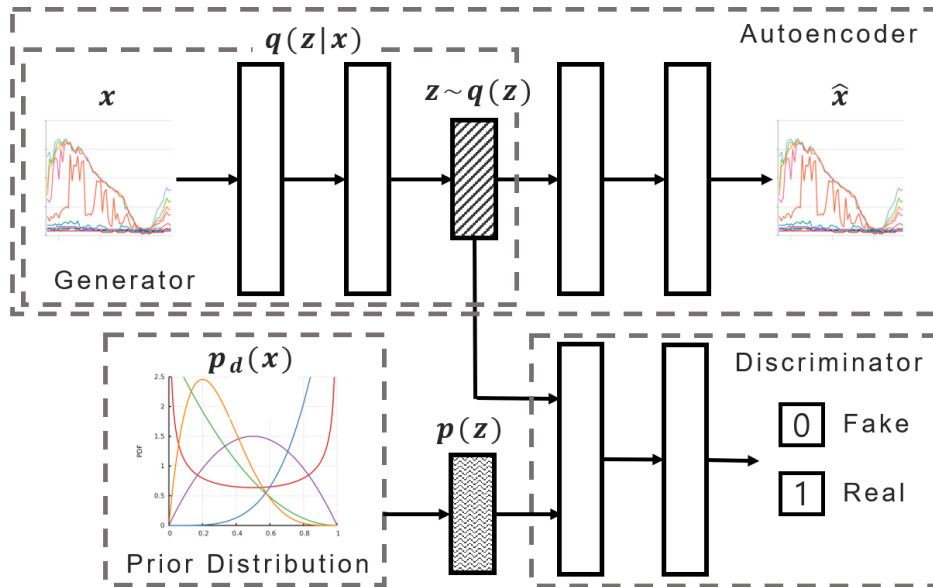


Figure SEQ Figure * ARABIC 1. Structure of AAE

PCA(KPCA)(S. W. Choi et al, 2005) which involves data transformation from the low-dimensional nonlinear observation space into the high-dimensional linear feature space. However, standard kernel functions do not always guarantee good results and the performance is also very sensitive to some hyper-parameters. Recently, deep learning (DL) has received a lot of attention in process monitoring because of its high model flexibility. Among various DL-based methods, autoencoder(AE)(Z. Zhang et al, 2018) provides good performance by achieving automated key latent variables. However, the typical autoencoder based method lacked good model interpretability(Y. Qiu and Y. Dai, 2019). To address these issues, an algorithm for process monitoring based on adversarial autoencoder (AAE) has been developed. AAE(A. Makhzani et al, 2015), which is a probabilistic autoencoder that used the generative adversarial networks, extracts the latent code vector matching the arbitrary prior distribution. Two monitoring indices are proposed and constructed based on the extracted latent code. The proposed method is applied to the Tennessee Eastman (TE) process(J. J. Downs and E. F. Vogel, 1993) to verify its monitoring performance.

2. AAE based fault detection

2.1. AAE based dimension reduction

AAE is a generative autoencoder that uses generative adversarial networks (GAN) to impose an arbitrary prior distribution on the latent code. The structure of AAE is shown in Figure 1. It contains autoencoder and discriminator of GAN. It uses adversarial training (1) to perform variational inference by matching the aggregated posterior distribution $q(z)$ with a prior distribution $p_d(x)p_d(x)$. The encoder network $q(z|x)q(z|x)$ defines

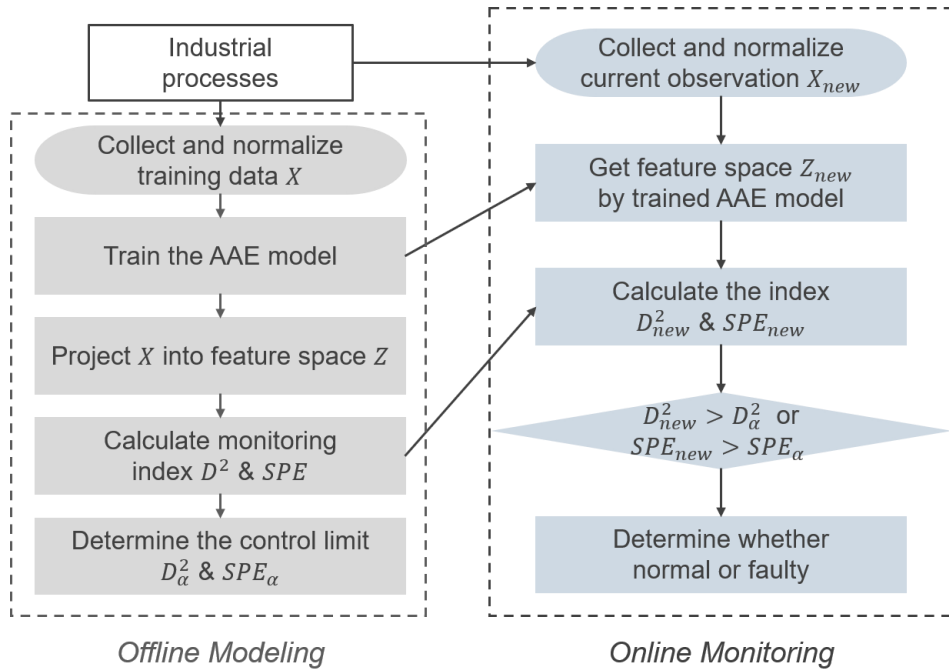


Figure SEQ Figure * ARABIC 2. Flowchart of AAE based monitoring method

an aggregated posterior distribution on the latent vector z as follows:

$$q(z) = \int_x q(z|x)p_d(x)dx$$

Meanwhile, the autoencoder attempts to minimize the reconstruction error as follows:

$$\underset{\theta, \phi}{\operatorname{argmin}} \|x - \hat{x}\|^2 \tag{2}$$

where θ and ϕ are the updating parameters of encoder and decoder respectively and \hat{x} is the reconstructed output of autoencoder. Then the adversarial network differentiates the true samples from the generated samples by updating its discriminator D . The generator G is updated to fool the discriminative network into trusting that the samples from the generator come from the prior distribution. The objective function of an adversarial network is as follows:

$$\begin{aligned} & \underset{\theta, \phi}{\operatorname{argmin}} \|x - \hat{x}\|^2 \\ & \min_{\theta} \max_{\phi} E_{x \sim p_d(x)} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] \\ & \min_{\theta} \max_{\phi} E_{x \sim p_d(x)} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] \end{aligned} \quad (3)$$

Matching the aggregated posterior to the prior samples ensures that generating ones from any part of prior space results in a meaningful latent vector. As a result, the decoder of the AAE learns a deep generative model that maps the imposed prior ones to the data distribution. The extracted latent vector has compressed and enhanced representations.

2.2. Monitoring indices

After AAE based dimension reduction model is developed, latent space and residual space are obtained by the encoder and decoder, respectively. For fault detection, the indices should be constructed to measure how far the variables are out of the normal range.

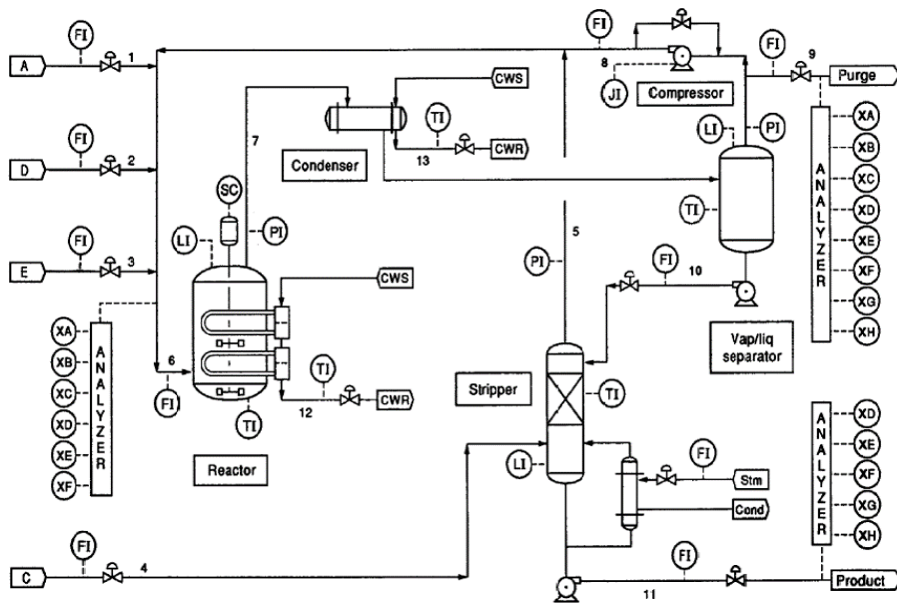


Figure SEQ Figure * ARABIC 3. Layout of Tennessee Eastman process

In the latent space, the index T^2 based on hidden representations is defined as follows:

$$\begin{aligned}
& \underset{\theta, \phi}{\operatorname{argmin}} \|x - \hat{x}\|^2 \\
& \min_G \max_D E_{x \sim p_d(x)} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] T^2 = z^T \Delta^{-1} z \quad (4) \\
& T^2 = z^T \Delta^{-1} z
\end{aligned}$$

where $\Delta\Delta$ is a covariance matrix of a latent vector zz . It describes the comprehensive fluctuation of the model. In the residual space, the index *SPESPE* based on reconstruction error is defined as follows:

$$\begin{aligned}
& \underset{\theta, \phi}{\operatorname{argmin}} \|x - \hat{x}\|^2 \quad \quad \quad SPE = (x - \hat{x})^T (x - \hat{x}) \\
& \min_G \max_D E_{x \sim p_d(x)} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] \quad (5) \\
& T^2 = z^T \Delta^{-1} z
\end{aligned}$$

It reflects the degree of deviation between the process variable and reconstructed variables from the model. In order to detect whether a new sample includes fault information, the control limits of the indices should be determined. As the distribution of real process data is unknown, it is necessary to use a density based approach. Based on

the the confidence level $\alpha\alpha$, control limit is obtained. To sum up, the flowchart with the proposed AAE based fault detection is depicted in Figure 2.

3. Case study

3.1. Process description

The Tennessee Eastman (TE) process is a well-known benchmark process in process control and monitoring. The TE process is shown in Figure 3, which is mainly composed of five operating units: a reactor, a condenser, a separator, a compressor, and a stripper. There are 22 continuously measure variables, 19 composition measurements, and 11 manipulated variables. TE process also has 21 programmed process faults like step, random variation, sticking, and unknowns. Each fault data has 960 samples with a fault introduced from sample 161. Normal data for training the proposed model contains 500 samples.

3.2. Monitoring results

In this case study, the proposed method and several comparative methods (PCA, KPCA, AE) are applied to detect the faults of the process. To demonstrate the performance of the proposed methods, fault detection rates(FDRs) for all fault types and false alarm rates(FARs) of normal data are calculated and shown in Table 1 and Table 2. For large magnitude faults 1, 2, 4, 6, 7, 8, 12, 13, 14, and 17, all four methods show similar detection results. The FDR of the proposed method outperforms that of the other three methods in

some faults (3, 5, 9, 15, 20, and 21) and the average FDR is also higher than the other three methods. Fault detection using AAE model shows strong robustness to all types of faults. In FAR, the proposed method also gives significantly better results than other methods. Extraction of robust features following Gaussian distribution improves the performance of process monitoring.

Table 1. Fault detection rates of PCA, KPCA, AE, and AAE based method

Fault type	PCA	KPCA	AE	AAE
1	1	1	0.997	1
2	0.987	0.987	0.99	0.99
3	0.258	0.135	0.258	0.348
4	0.987	0.973	0.993	0.986
5	0.511	0.336	0.497	0.55
6	1	1	1	1
7	1	1	1	1
8	0.987	0.981	0.99	0.978
9	0.226	0.214	0.255	0.336
10	0.695	0.682	0.678	0.74
11	0.822	0.826	0.843	0.82
12	0.99	0.995	0.993	0.997
13	0.958	0.955	0.963	0.962
14	1	1	1	1
15	0.275	0.252	0.315	0.312
16	0.645	0.645	0.636	0.647
17	0.955	0.958	0.955	0.96
18	0.928	0.93	0.928	0.95
19	0.505	0.498	0.477	0.538
20	0.731	0.697	0.692	0.786
21	0.597	0.568	0.608	0.59

Table 2. False alarm rates of PCA, KPCA, AE, and AAE based method

PCA	KPCA	AE	AAE
0.1	0.0513	0.0813	0.0313

4. Conclusion

The AAE-based process monitoring model is particularly good for nonlinear, non-Gaussian systems because AAE is specialized in extracting Gaussian features that are most desired in these systems. It automatically extracts features through the pre-trained deep neural network and calculates the monitoring statistics using the latent and residual features, so can be easily applied to large-scale data. The proposed method is applied to the benchmark process and shows superior performance in fault detection rate and false alarm rate. Extending the AAE-based model to the multimode process monitoring is an important part of future work.

References

- N. Olivier-Maget, G. Hétreux, J. Le Lann, and M. Le Lann, 2008 “Fault detection and isolation based on the model-based approach: Application on chemical processes”, *Computer Aided Chemical Engineering*, vol. 25, pp. 411-416.
- S. Joe Qin, 2003, “Statistical process monitoring: basics and beyond,” *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 17, no. 8-9, pp. 480–502.
- S. W. Choi, C. Lee, J.-M. Lee, J. H. Park, and I.-B. Lee, 2005, “Fault detection and identification of nonlinear processes based on kernel pca,” *Chemometrics and intelligent laboratory systems*, vol. 75, no. 1, pp. 55–67.
- Z. Zhang, T. Jiang, S. Li, and Y. Yang, 2018, “Automated feature learning for nonlinear process monitoring—an approach using stacked denoising autoencoder and k-nearest neighbor rule,” *Journal of Process Control*, vol. 64, pp. 49–61.
- Y. Qiu, and Y. Dai, 2019, “A stacked auto-encoder based fault diagnosis model for chemical process”, *Computer Aided Chemical Engineering*, vol. 46, pp. 1303-1308.
- A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, 2015, “Adversarial autoencoders,” *arXiv preprint arXiv:1511.05644*.
- J. J. Downs and E. F. Vogel, 1993, “A plant-wide industrial process control problem,” *Computers & chemical engineering*, vol. 17, no. 3, pp. 245–255.

Machine learning for constrained self-optimizing control

Hongxin Su,^{a,b} Chenchen Zhou,^{a,b} Yi Cao,^{a,b,*} Shuanghua Yang^{a,b,*}

^a*College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310027, China*

^b*Institute of Zhejiang University-Quzhou 324000, China*

Abstract

Self-optimizing control (SOC) is a method to select controlled variables (CVs) and keep them constant such that the plant operates optimally. Since the concept of SOC was proposed, some difficult problems in this field have not been solved such as active constraint changes. Previous work either handles the constrained SOC problem with complicated control structure or in the sense of local SOC, and has limitations such as structural complexity and inaccuracy of control. To address the shortcomings of the existing methods, this paper proposed a constrained variable approximation (CVA) method to solve the problem in the global sense using a simple control structure. The constrained variables which may vary between active and inactive are approximated by a nonlinear function of available measurements in the whole operational region using artificial neural network (ANN). Then the CVs are determined as the difference between the nonlinear function and the constrained variables. The system would be near optimal operation when the CVs are controlled at zero. An evaporator process is applied to illustrate the effectiveness of the proposed method.

Keywords: Self-optimizing control, Active constraint changes, Controlled variable selection.

1. Introduction

Operation optimization is difficult to achieve under various disturbances and uncertainties. Real-time optimization (RTO) offers a solution that handles the problem through repeated online optimization, however this would result in large computation burden. Besides, the RTO is an open-loop solution, which means it is not robust when the accuracy of uncertainty estimation is poor. Alternatively, self-optimizing control (SOC) (Skogestad, 2000) focuses on the selection of appropriate controlled variables (CVs), such that the plant operates almost optimally under all circumstances when these CVs are controlled at their constant setpoints through closed feedback control. Since the concept of SOC was proposed, the research associated with CV selection methods has made substantial progress (Jäschke et al., 2017). One of the most interesting problems about SOC is how to control constraints varying between inactive and active. Since there may be different number of unconstrained degrees of freedom in different disturbance regions, Manum and Skogestad (2012) designed a switching structure between different disturbance regions within each of which the active constraints remained unchanged. The extended null space method (Alstad and Skogestad, 2007) was developed to find the CVs for different regions and a special strategy is used to decide when to switch between different regions. Difficulty arises to find the switching method in the presence of measurement noise. Cao (2004) proposed a cascade control structure that the self-

optimizing CVs are in the outer loop while the variables likely to violate the constraints are in the inner loop with a saturation block to restrict the inner loop setpoint so that the variables will not go beyond the constraints. Nevertheless, this scheme requires that the number of CVs is not less than the number of the constraints likely to change between inactive and active because one constraint needs to be controlled by one CV. Hu et al. (2012) considered a simple control structure that aims to find CVs which can ensure all the variables are within their constraints and also within the whole disturbance and uncertainty region. Nonetheless, the CVs are chosen based on the exact local method through minimizing the local average loss.

Existing approaches associated with active constraint changes are either difficult to complement due to the complicated control structure or locally valid. To address this problem, this paper proposes a method to find a CV which makes the average loss acceptable once it is kept zero despite of changes of active constraint. This is achieved by constructing a nonlinear function of available measurements to track the optimal value of the constraints which may vary between active and inactive using artificial neural network (ANN). Then the CV is selected based on the difference between the nonlinear function and the constraint variable. Finally, a simple feedback control structure can achieve constrained global self-optimizing control.

The rest of the paper is organized as follows. Section 2 introduces the constrained global self-optimizing control (gSOC) problem formulation and the CV selection method, then, the proposed approach is demonstrated to be effective through an evaporator case study in Section 3. Finally, the paper is concluded in Section 4.

2. Constrained gSOC formulation and CV selection

Consider a constrained static optimization problem

$$\begin{aligned} \min_u J(u, d) \\ \text{s.t. } g(u, d) \leq 0 \end{aligned} \quad (1)$$

with available measurements

$$y_m = y + n \quad \text{and} \quad y = f(u, d) \quad (2)$$

Where J is a scalar objective (economic cost) function to be minimized, and $u \in \mathbb{R}^{n_u}$, $d \in \mathbb{R}^{n_d}$, $n \in \mathbb{R}^{n_y}$, $y_m \in \mathbb{R}^{n_y}$ and $y \in \mathbb{R}^{n_y}$ are manipulated variables, uncertain disturbances, measurement errors, measured and true output measurements, respectively. $f: \mathbb{R}^{n_u \times n_d} \rightarrow \mathbb{R}^{n_y}$ and $g: \mathbb{R}^{n_u \times n_d} \rightarrow \mathbb{R}^{n_g}$ are the input-output mapping function and operational constraints, respectively. These constraints are often related to operational safety or product qualities.

Suppose c is the self-optimizing CVs. According to the SOC principle, when c are kept as constant, the process operation is optimal or near optimal. This means the desired c at optimal condition, i.e. $c_{opt}(d)$ should be insensitive to the change of d . Therefore, the optimizing problem (1) becomes the problem of choosing appropriate c which remain almost the same under optimal conditions with respect to d . Previous work selects individual measurement y or linear combinations of available measurements Hy as c and most of them are based on the assumption that the active constraints do not change

under different disturbance scenarios. Actually, during the operation, constraints g can be divided into three types: (i) g_1 : always active; (ii) g_2 : always inactive; (iii) g_3 : vary between active and inactive. For case (i), g_1 can always be controlled by u hence consuming the same number of degrees of freedom of g_1 , and for case (ii) g_2 can be neglected because they have no effect on the optimal solution. Therefore, only g_3 deserve careful attention.

It is assumed that the optimal operation data under various disturbances are available. The aim is to select appropriate CVs which are kept at constant setpoints so that the plant is operating near optimally in spite of g_3 switching between active and inactive. Because the optimal values of g_3 are not constant in the whole operation region, they cannot be directly selected as CVs. It is proposed to find a proper function to fit the optimal constraint values of g_3 , and then CVs are selected as the difference between the function and the constraint variables. When the CVs are kept at zero, the optimality of constraint variables in g_3 is maintained by tracking the function in the whole region. Since g_3 sometimes reach the upper bounds, sometimes the lower bounds and sometimes in between, g_3 are strongly nonlinear, hence cannot be fitted well by linear functions. According to the well-known universal approximation theorem of neural network (Hidalgo et al., 2020), that is, a given feedforward neural network with enough hidden nodes, equipped with a linear output layer and activation function of the hidden layer, can approximate any function by any desired error, the ANN method is applied to obtain the nonlinear function which is used to fit the constraint variables in g_3 .

The proposed constrained variable approximation (CVA) method can be described as two parts: one is offline CV selection as shown in Figure 1, and the other is online SOC in Figure 2. As shown in Figure 1, optimal measurements y including controlled variables in g_3 are obtained by solving optimization problem (1) under different disturbances d . Then the controlled variables in g_3 are fitted by a nonlinear function of y , i.e. $Z(y)$ through ANN. Finally, the CV can be selected as $c = Z(y) - g_3$. As shown in Figure 2, the selected CV is applied to the SOC structure. The process can operate near optimally when c is kept at zero through a feedback control.

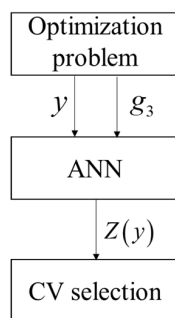


Figure 1: offline CV selection

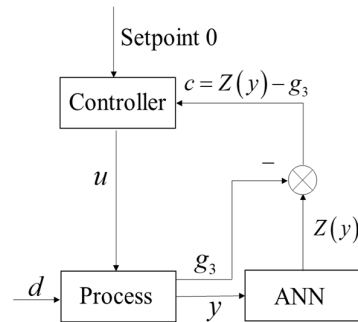


Figure 2: online SOC structure using CVA

3. Case study

A forced-circulation evaporator (Newell and Lee, 1989) is considered as shown in Figure 3. The solute concentration increases through evaporation of the solvent in the forced circulation evaporator. The detailed model equations and the nominally optimal values can both be referred to Ye et al. (2013).

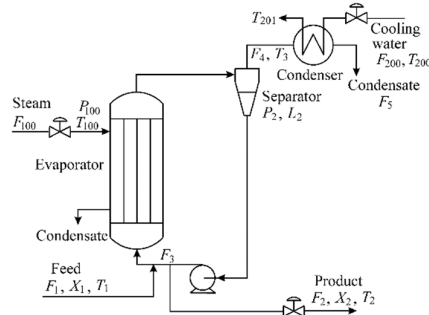


Figure 3: Evaporator system

There are three state variables L_2 , X_2 , P_2 , four manipulated variables F_2 , P_{100} , F_3 , F_{200} and four disturbances F_1 , X_1 , T_1 , T_{200} in the process. The variations of disturbance variables are $\pm 20\%$ of their nominal optimal values. The cost objective function (\$/h) of the whole process is defined as

$$J = 600F_{100} + 0.6F_{200} + 1.009(F_2 + F_3) \quad (3)$$

The process constraints are listed as follows:

$$X_2 \geq 35.5\% \quad (4)$$

$$40 \text{ kPa} \leq P_2 \leq 80 \text{ kPa} \quad (5)$$

$$P_{100} \leq 400 \text{ kPa} \quad (6)$$

$$F_{200} \leq 400 \text{ kg / min} \quad (7)$$

$$0 \text{ kg / min} \leq F_3 \leq 100 \text{ kg / min} \quad (8)$$

The constraints in (7) and (8) are always inactive, hence they can be neglected. Two constraints $P_{100} = 400 \text{ kPa}$ and $X_2 = 35.5\%$ are always active within the whole disturbance region, and the separate level L_2 which has no steady-state effect, but needs to be controlled at its nominal value, hence totally consuming three degrees of freedom (DOF). Hence, only one DOF is left for CV selection to achieve SOC.

Monte Carlo simulation is applied to generate a sequence of 1000 nonuniformly distributed samples within the predefined disturbance set and then optimal measurements are obtained by numerical optimization with respect to different disturbance. Note that P_2 in constraint (5) vary between inactive and active within the whole sampling region, so the number of activeness and inactiveness of P_2 are similar by sampling nonuniformly

to ensure the representativeness of the samples. There are 4 measurements to be considered for CVA method, which are identified as one of the best subsets (Kariwala et al., 2008):

$$y = [F_2, F_{100}, T_{201}, F_3]$$

The ANN is applied to find a nonlinear combination function of optimal y , i.e. $Z(y)$ which can fit the optimal P_2 in the whole operation region. There is one hidden layer with 12 neurons in the ANN structure and the determinant coefficient R^2 is 0.99 after training. Then the self-optimizing controlled variable can be selected as $c = Z(y) - P_2$ and kept at zero through a feedback control. In this way, $Z(y)$ is able to track the optimal setpoint of P_2 with respect to different disturbance, so P_2 can be maintained at optimum through keeping $c = 0$.

To evaluate the performance of the CV selected through the Monte Carlo experiment, 100 random samples are generated within the expected disturbance ranges without measurement noise and the results are shown in Figure 4. It is clear that the values of objective function after controlling $c = 0$ are almost similar with its optimal values at the same disturbance, and the average loss between the two is 0.0864. From Figure 5, it can be seen that P_2 after self-optimizing control is able to track the corresponding optimal P_2 well without violation of its limits. Table 1 compares the proposed constrained variable approximation (CVA) method with the necessary condition of optimality (NCO) approximation using least square method (Ye et al., 2013), which selects CVs as a linear function of the available measurements that approximates the gradient of J with respect to u . To compare fairly, the same measurement set including P_2 is chosen and the constraint equation (5) is ensured when solving the constrained gSOC problem. It is clear that the average loss and maximum loss of CVA method are about two orders of magnitude smaller than the NCO approximation method, which indicates the superiority of the proposed method.

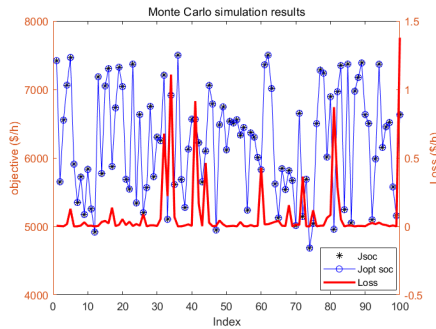


Figure 4: Monte Carlo simulation results

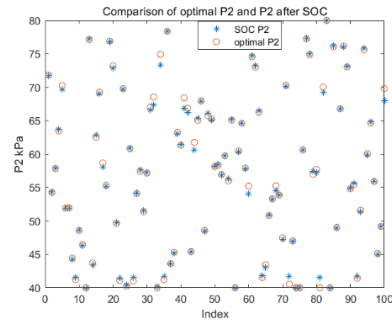


Figure 5: Tracking of optimal P_2

Table 1: Comparison of previous method with this work

Method	Average loss	Maximum loss	Standard deviation
NCO approximation	2.7954	12.9476	3.1219
CVA (our approach)	0.0864	1.3779	0.2302

4. Conclusions

This work proposed a new CV selection approach to solve the constrained gSOC problems. In this approach, the constrained variable which may vary between active and inactive is tracked by a nonlinear function of available measurements using ANN. Then the CV is selected as the difference between the two which are kept zero, leading to good SOC performances. The proposed approach was applied to a forced-circulation evaporator process and was proved to be effective to achieve low loss while satisfying the constrained limits. However, as a preliminary study, only a single degree of freedom (DOF) is considered in this work and how to select CVs when more than one DOF exists will be further investigated.

Acknowledgement

This work was supported in part by Institute of Zhejiang University-Quzhou Science and Technology Project (IZQ2019-KJ-021).

References

- V. Alstad, S. Skogestad, 2007. Null space method for selecting optimal measurement combinations as controlled variables. *Industrial & Engineering Chemistry Research* 46 (3), 846 – 853.
- Y. Cao, 2004. Constrained self-optimizing control via differentiation. *Ifac Proceedings Volumes* 37 (1).
- P. Hidalgo, G. Diaz, M. A. Gutiérrez-Naranjo, 2020. Two-hidden-layer feed-forward networks are universal approximators: A constructive approach - sciencedirect. *Neural Networks* 131, 29 – 36.
- W. Hu, L. M. Umar, G. Xiao, V. Kariwala, 2012. Local self-optimizing control of constrained processes. *Journal of Process Control* 22 (2), 488 – 493.
- J. Jäschke, Y. Cao, V. Kariwala, 2017. Self-optimizing control a survey. *Annual Reviews in Control*.
- V. Kariwala, Y. Cao, S. Janardhanan, 2008. Local self-optimizing control with average loss minimization. *Industrial & Engineering Chemistry Research* 47 (4), 1150 – 1158.
- H. Manum, S. Skogestad, 2012. Self-optimizing control with active set changes. *Journal of Process Control* 22 (5), 873 – 883.
- R. B. Newell, P. L. Lee, 1989. *Applied process control : a case study*.

Driving Force Based Design and Control Performance Analysis of Reactive Distillation Columns

Ashfaq Iftakher,^a Seyed Soheil Mansouri,^b Ahaduzzaman Nahid,^a Anjan K. Tula,^c M. A. A. Shoukat Choudhury,^{a,*} Jay Hyung Lee,^d Rafiqul Gani^{d,e,*}

^a*Department of Chemical Engineering, BUET, Dhaka-1200, Bangladesh*

^b*Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark*

^c*College of Control Science and Engineering, Zhejiang University, China*

^d*Department of Chemical and Biomolecular Engineering, KAIST, South Korea*

^e*PSE for SPEED Company, Skyttemosen 6, DK-3450 Allerød, Denmark*

shoukat@che.buet.ac.bd, rgani2018@gmail.com

Abstract

This paper presents a toolbox for driving force based integrated design and control of reactive distillation (RD) systems. The toolbox provides links to other necessary tools for the design, simulation, and analysis of RD systems. One convenient feature of the toolbox is that it guides the user through the steps of problem definition, including the selection of design-control variables and associated data-flow as well as the generation of the necessary scripts for links to simulation and analysis software. The implementation of this new toolbox is tested through binary/multielement single/double feed RD systems. For each RD system, the needed driving force calculations are done in ProCAPE. Steady state simulation and analysis are performed through ASPEN PLUS, MoT, or ProCADC. Dynamic open and closed-loop PI or MPC simulations are performed in ASPEN PLUS Dynamics, MoT, or MATLAB. The goal of the tests is to confirm the advantages of using the toolbox for studies involving RD systems in terms of fast, reliable, and efficient solution and analysis of integrated design-control problems.

Keywords: MPC, driving force, reactive distillation, ProCADC, toolbox

1. Introduction

Reactive distillation is an attractive intensified process unit because of its many advantages, e.g., the requirement of fewer process units, less heat addition, improved product removal, degradation of azeotropes, etc. (Tian et al., 2018). However, it is inherently difficult to control due to the shrinkage of the controllability region, and propagation of nonlinearity because of non-optimal design decisions (Rafiei and Ricardez-Sandoval, 2020). Hence, it is important to consider design-control issues simultaneously to eliminate potentially promising design alternatives that may be dynamically inoperable. Tian et al. (2020) have proposed a parametric optimization and control approach for simultaneous design-control of RD systems. Hamid et al. (2010) used thermodynamic insights and reverse design for decomposition based integrated design-control of nonreactive and reactive systems. Mansouri et al. (2016) have shown

that by simultaneously considering design and control issues through the driving force concept, better design-control of reactive distillation units can be achieved. Iftakher et al. (2021) extended this concept by showing that various RD systems, designed at the maximum available driving force actually lead to superior controllability under advanced controllers such as MPC. In this work, an integrated design-control toolbox especially suitable for RD systems that provides systematic process modeling, simulation, and verification options based on the extended RD design-control framework of Iftakher et al. (2021), is presented. This paper is organized as follows: first, the extended design-control framework is presented. Next, the work- and data-flow of the developed integrated RD design-control toolbox are described. Then, the calculation steps are highlighted through a case study. Finally, conclusions are stated.

2. Integrated design-control framework for RD systems

The extended integrated design-control framework has 3 main steps: data collection & problem formulation; determination of integrated design-control; and, verification.

2.1. Data collection & problem formulation

Here, an RD system is defined in terms of compounds (and catalysts), feed condition, number of reactions, and a design target. To quantify the design and control issues of the specified RD system, a multiparametric objective function is defined. Iftakher et al. (2021) proposed the use of an objective function in terms of six parameters: total energy consumption by the process, CO₂ footprint per kg feed, integral absolute error, total variation of input, relative gain array (RGA), and the Niederlinski Index. The reactive systems are expressed in terms of elements (difference between the number of compounds and the number of reactions). Therefore, a reactive system with three compounds, and one reaction is expressed in terms of binary elements, whereas a reactive system with multiple elements is represented in terms of equivalent binary elements. At the end of this step, the choice of the elements is verified through an atom-balance of the reactive system.

2.2. Determination of integrated design-control

Here, the integration of design and control is performed through operation at the maximum driving force. First, the reactive phase (vapor-liquid equilibrium) diagram and its translation to a driving force diagram are calculated based on the element reactive system, appropriate thermodynamic models, and the specified operating conditions. The location of the maximum driving force and the operational area are identified to determine the RD column design parameters. Next, the MPC controller parameters are determined (Iftakher et al. 2021).

2.3. Dynamic analysis and verification

Here, at first, the steady state analysis is performed (to verify the designed controller structure from step 2) using the linearized state space model of the process. Next, the dynamic simulation is performed (in open-loop and closed-loop with PI or MPC options). Finally, the objective function parameters are calculated to quantify the design-control performance of the designed RD system.

3. Integrated design-control toolbox for RD systems

The integrated design-control framework has been translated into a toolbox for the study of RD systems. As shown in Figure 1, the toolbox has 5 main steps, which are briefly described below.

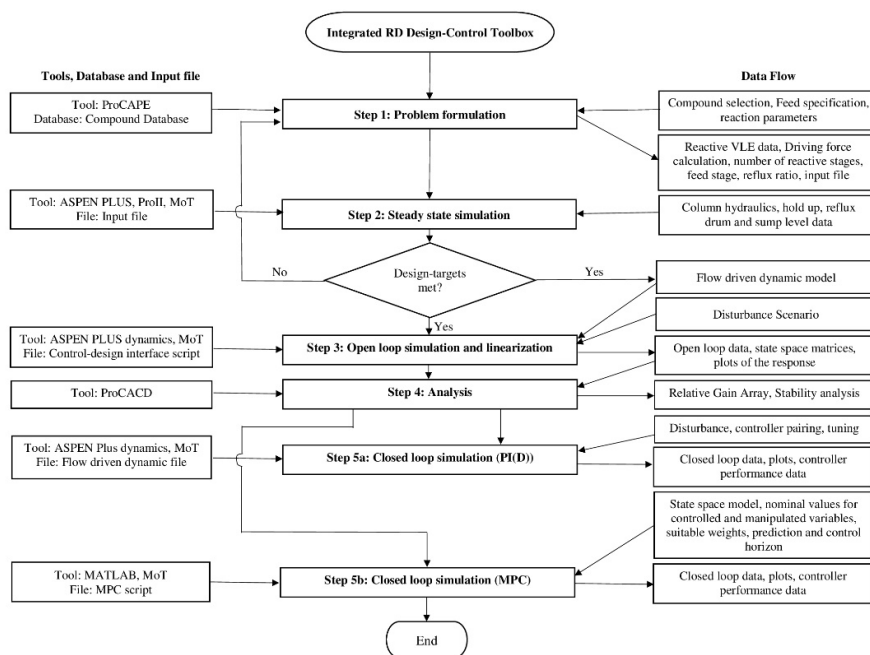


Figure 1: Work-data flow of the Integrated RD design-control Toolbox.

3.1. Step 1: Problem formulation

In this step, the user is guided to enter the following: compounds, feed specification, number of feeds, reaction definition, reaction parameters, design target, choice of elements. The RD toolbox exports these data to ProCAPE where the user selects the needed thermodynamic models to calculate the Reactive VLE data at the specified pressure. The reactive VLE data is then converted to a driving force diagram from which the necessary column design parameters (number of reactive stages, feed stage, reflux ratio) are determined. At the end of this step, a simulator input file is generated that contains all the required flowsheet and design information for steady state simulation in Step 2.

3.2. Step 2: Steady state simulation

The input file generated in Step 1 is loaded to a simulation software (ASPEN PLUS, ProII, or MoT) where all the required process and design variables are defined. Next, the steady state simulation is performed. The product compositions are inspected and if the design targets are satisfied, the required data (column hydraulics, reflux drum level, hold up, etc.) for dynamic simulation are provided for the generation of the corresponding flow driven dynamic model. Otherwise, the user is guided back to Step 1 to refine the problem formulation.

3.3. Step 3: Open-loop simulation and linearization

With the generated dynamic model file, open-loop simulation is performed in ASPEN Plus Dynamics or MoT for a selected disturbance scenario. The open-loop data are used to derive the transfer function model of the process. Also, in this step, a control-design interface script (autogenerated by the integrated RD toolbox), is applied to determine the

state space matrices of the process by linearizing the dynamic model at the nominal steady-state condition.

3.4. Step 4: Analysis

Steady state analysis is performed in ProCADC (Tula et al, 2020). The transfer function models derived from the open-loop data or the state-space matrices are used to determine the relative gain array (RGA), which indicates the degree of loop interaction. Also, the system stability is checked by determining the Niederlinski Index or directly through analysis of the transfer function model.

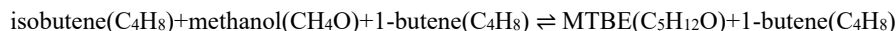
3.5. Step 5: Closed-loop simulation

Closed loop simulation is performed using either the regulatory controller (PI) or the supervisory controller (MPC). For both control structures, the top and bottom compositions of the product(s) of interest are controlled, by varying the reflux rate and reboiler duty, respectively. The selection of the control structure is verified by inspecting the diagonal values of RGA. For MPC closed-loop simulation, the toolbox prompts the user to provide the following inputs: nominal values of the controlled and manipulated variables, weights on the controlled and manipulated variables, prediction horizon, and control horizon. After a check of consistency of data, the integrated RD toolbox links to MATLAB or MoT where a script for MPC closed-loop simulation is autogenerated. After performing the closed-loop simulation, the controller performance is quantified by determining the Integral absolute error (IAE), and the total variation of input (TV).

4. Application example

The application of the integrated design-control toolbox is highlighted through a single feed multielement reactive system (MTBE production with inert) as a case study. Details of application examples for other reactive systems can be obtained from the corresponding authors.

The involved compounds (selected from the in-house database) are isobutene (C₄H₈); methanol (CH₄O); 1-butene (C₄H₈); and MTBE (C₅H₁₂O). The feed condition and the design targets are given in Table 1 (data obtained from Pérez Cisneros et al. 1997). Reversible and exothermic reaction (taking place in acidic catalyst) between methanol and isobutene with the presence of 1-butene (inert) yields MTBE, as follows:



The number of elements required to represent this reactive system is three. SRK equation of state and the Wilson model are used for the generation of reactive VLE data from where the driving force diagram is generated. Using the driving force diagram, the following column design parameters are obtained: total number of stages = 7 (reactive stages from 2 to 6), feed stage = 4, reflux ratio = 2.813).

Table 1: Design targets and feed specifications for MTBE system with inert.

Component	Molar composition		
	Feed	Distillate	Bottom
Isobutene (C ₄ H ₈)	0.590	0.773	0.061
Methanol (CH ₄ O)	0.343	0.0	0.012
1-butene(C ₄ H ₈)	0.067	0.196	0.024
MTBE (C ₅ H ₁₂ O)	0.0	0.031	0.907

Feed flowrate: 100 kmol/h; Feed temperature and Pressure: 320K and 11 atm.

The design-control multi-objective function is defined similarly as proposed by Iftakher et al. 2021. Next, the toolbox creates an input file containing necessary data and the flowsheet description to be loaded to ASPEN PLUS or MoT where the flowsheet is

autogenerated and the steady state simulation is performed. The design targets are satisfied, and the following objective function values (to indicate design issues) are calculated: total energy consumption, $J_1 = 958.34$ KW; total carbon-footprint, $J_6 = 0.0941$ CO₂ eq per kg feed. Next, the flow driven dynamic model is exported to ASPEN PLUS dynamics or MoT for open-loop simulation. The disturbance scenario is as follows: +16.5% step change in methanol flow rate. After linearizing the dynamic model, the following steady state metrics are obtained in ProCADC: relative gain array, $J_4 = \begin{bmatrix} 1.26 & -0.26 \\ -0.26 & 1.26 \end{bmatrix}$; Niederlinski index (N_I), $J_5 = 0.7915$. As the diagonal values are close to unity, the degree of loop interaction is predicted to be minimal. Also, the positive value of N_I along with the positive real parts of the eigenvalues of both the open loop and closed loop dynamic model verifies that the system is stable.

PI closed loop simulation is carried out in ASPEN Plus dynamics or MoT. Based on the RGA values, the MTBE top composition is controlled by varying the reflux rate, and the MTBE bottom composition is controlled by varying the reboiler duty. The controllers are tuned using the IMC rules. The reflux drum and sump level controllers (proportional type) are also included for dynamic model consistency and perfect pressure control is assumed for the RD column.

For the MPC closed-loop simulation, a linear MPC with MIMO control strategy is used. The controller design and corresponding simulation is carried out in MATLAB or MoT where a script is autogenerated by the toolbox. Note that the controller pairings are kept the same as for the PI controllers. The plant inputs are: the condenser duty - Q_c , the reflux mass flow rate - R (kg/hr), the reboiler duty - Q_R , the distillate mass flow rate - D (kg/hr), the bottoms mass flow rate - B (kg/hr), and the feed molar flow rate (kmol/hr). The plant outputs are: the column pressure (stage 1) - P (atm), the mole fraction of the distillate of interest - x^D , the mole fraction of the bottoms of interest - x^B , the reflux drum liquid level - R_{lev} (m), and the sump liquid level - S_{lev} (m). MPC controller is designed in such a way that the nominal set point for all the controlled and manipulated variables is set at 50%. The simulation is run for 25 hours with a sampling time of 30s. The weights on the controlled variables are as follows: $[10 \ 1 \ 1 \ 0.1 \ 0.1]^T$. Note that the relative weights are a measure to assign importance for keeping the controlled variables at the set point. For example, MPC controller assigns 10 times more importance to keep the distillate and bottom product compositions at their set points compared to drum level or sump level. Analogously, maintaining column pressure at the set point is given 100 times more importance than maintaining sump level, since, the column pressure must be regulated tightly for safe operation. Finally, the prediction horizon is set to 30 min which is large enough to make the controller performance insensitive to further increases of the prediction horizon. The control horizon is set to 4 min which is kept small to reduce computational effort. The resulting closed loop dynamic responses for both PI and MPC controllers are shown in Figure 2. It is clear that both PI and MPC controllers are able to reject the feed disturbance very efficiently with minimal process upset. Finally, the control performance parameters (integral absolute error, J_2 and total variation of input, J_3) are calculated for both PI and MPC controllers and listed in

Table 2. Note that the parameters are calculated both for the top and bottom control loops.

5. Conclusions

An integrated toolbox for the design and control of reactive distillation systems using the driving force approach has been presented together with an illustration of an example of application. The toolbox links to other external property prediction tools for problem

formulation and design parameter calculation. It also links to simulation and control tools for steady state simulation, dynamic simulation, and overall quantification of design and control issues. Required scripts for state space model generation and simulation are autogenerated through the integrated toolbox, thereby providing flexible and systematic work-flow through which the study of RD systems can be easily set up for modelling, simulation, and verification with respect to integrated design-control. The application results confirm that RD systems designed at the maximum driving force are able to efficiently reject disturbances under both PI and MPC controllers because the interactions are reduced by the design at the maximum driving force. Future works include the integration of design and control of Petlyuk and dividing wall columns, as well as the extension of the reported toolbox to handle distillation columns with or without reactions.

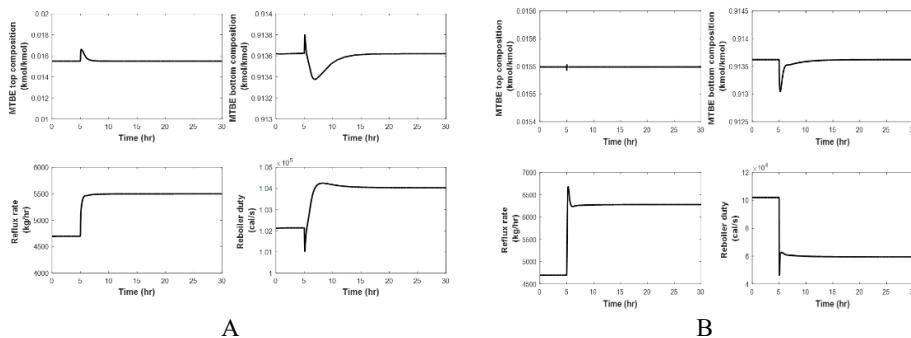


Figure 2: Closed-loop response of MTBE reaction with inert system (designed at the maximum driving force) for step change in feed flow rate; A) PI; B) MPC.

Table 2: Controller performance parameters of MTBE with inert system.

$J_{2,D}$ [hr] (PI)	$J_{2,D}$ [hr] (MPC)	$J_{2,B}$ [hr] (PI)	$J_{2,B}$ [hr] (MPC)	$J_{3,D}$ [-] (PI)	$J_{3,D}$ [-] (MPC)	$J_{3,B}$ [-] (PI)	$J_{3,B}$ [-] (MPC)
0.0010	1.6E-07	9.8E-04	7.5E-04	803.30	2.4E+03	4585	6E+04

References

- A. Iftakher, SS. Mansouri, A. Nahid, AK. Tula, MAAS. Choudhury, JH. Lee, R. Gani, 2020, *AIChE J.*, 2021; e17227. <https://doi.org/10.1002/aic.17227>
- MKA. Hamid, G. Sin, R. Gani, 2010, *Comput. Chem. Eng.*, 34, 683-699.
- SS. Mansouri, JK. Huusom, R. Gani, M. Sales-Cruz, 2016, *AIChE J.*, 62, 3137-3154
- ES. Pérez-Cisneros, R. Gani, ML. Michelsen, 1997, *Chem. Eng. Sci.*, 52, 527-543.
- M. Rafiei, LA. Ricardez-Sandoval, 2020, *Comput. Chem. Eng.*, 132, 106610.
- Y. Tian, SE. Demirel, MF. Hasan, EN. Pistikopoulos, 2018, *Chem. Eng. Process.*, 133, 160-210.
- Y Tian, I Pappas, B Burnak, J Katz, EN Pistikopoulos, 2020, *Chem. Eng. Sci.*, 116232
- AK. Tula, J. Wang, X. Chen, SS. Mansouri, R. Gani, 2020, *Comput. Chem. Eng.*, 136, 106771.

Generating an optimal control policy via apprenticeship and reinforcement learning

Max Mowbray^a, Robin Smith^a, Antonio Del Río Chanona^{b,*}, and Dongda Zhang^{a,*}

^a*University of Manchester, United Kingdom*

^b*Imperial College London, United Kingdom*

**a.del-rio-chanona@imperial.ac.uk and dongda.zhang@manchester.ac.uk*

Abstract

The development of advanced process control schemes is continuing driver of research within process systems engineering. In this work, we propose a framework, which leverages existing process data to automatically learn and update a control policy. This framework is underpinned by machine learning methods, namely, apprenticeship (AL) and reinforcement learning (RL), which compose offline and online learning respectively. In offline learning, we synchronously identify a function descriptive of the control objective and a parameterisation of the existing control policy expressed in data. The parameterised control policy is then updated automatically online via data from the ongoing process and RL. Importantly, the parameterisation learned offline achieves similar performance under the identified control objective to the existing control policy. Ultimately, the framework proposed enables a reduction in the technical intensity of offline RL-based policy learning, and approximates existing controllers in a more robust fashion than by supervised learning. The performance and efficiency of the framework proposed is explored via case study. Future work should focus on increasing sample efficiency in policy learning, satisfaction of constraints and accounting for model uncertainty.

Keywords: Apprenticeship learning, Reinforcement learning, Inverse reinforcement learning, Optimal control, Machine learning

1. Introduction

Recent initiatives for efficiency improvements in industrial process operation has driven interest in the development of high performance, data-driven advanced process control (APC) schemes. Reinforcement learning (RL) has achieved impressive results on benchmark game-based control tasks (Heess et al. (2015)), providing an avenue for research in translation to APC (S. P. K. Spielberg et al. (2017)). In spite of its high and well documented potential (T. Badgwell et al. (2018)), RL has yet to produce any meaningful impact in the (bio)chemical process industry. A number of challenges exist to wider deployment of RL-based controllers. One obstacle is the technical expertise, time and computation demanded in policy training and tuning. This work presents a two-step approach to RL-based policy learning, which leverages process data to parameterise an existing control law and then improves the performance of such control further. The approach promises to increase the learning efficiency of RL-based control policies, by reducing computational and technical investment via the utilisation and synthesis of

existing process knowledge with empirical observations (data) of system response to existing control strategies.

2. Methodology

In this work, we assume that the system concerned displays elements of stochastic behaviour and that process evolution is ultimately Markovian, and adheres to the following description:

$$\begin{aligned} \mathbf{x}_{t+1} &\sim p(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{u}_t) \\ \mathbf{y}_{t+1} &\sim p(\mathbf{y}_{t+1} | \mathbf{x}_{t+1}) \end{aligned} \quad (1)$$

Here, we formally state the process dynamics as a conditional probability density, where the probability of observing some future state $\mathbf{x}_{t+1} \in \mathbb{R}^{n_x}$ at discrete time index $t + 1$ is not conditioned on the process history but rather the current state $\mathbf{x}_t \in \mathbb{R}^{n_x}$ and control $\mathbf{u}_t \in \mathbb{R}^{n_u}$ imparted into the system at discrete time index t . Additionally, we assume the availability of a noisy system observation $\mathbf{y}_{t+1} \in \mathbb{R}^{n_y}$. Such an observation is distributed according to some measurement process $p(\mathbf{y}_{t+1} | \mathbf{x}_{t+1})$, upon which the control policy $\pi(\mathbf{u}_t | \mathbf{y}_t)$ is conditioned. The goal of decision making within an MDP is to maximise the discounted sum of rewards accumulated during a (discrete-time) process trajectory $\boldsymbol{\tau} = (\mathbf{x}_0, \mathbf{y}_0, \mathbf{u}_0, \dots, \mathbf{x}_T, \mathbf{y}_T)$ from some initial state \mathbf{x}_0 over a finite horizon of length T . Given the underlying system adheres to some stochastic dynamics, the probability of observing a given process trajectory is denoted $p(\boldsymbol{\tau} | \pi)$. Hence, the objective of any RL method is to learn a parameterisation $\boldsymbol{\theta}$ upon which the control policy $\pi(\mathbf{u}_t | \mathbf{y}_t, \boldsymbol{\theta})$ is conditioned to maximise the following objective:

$$G(\boldsymbol{\tau}) = \sum_{t=0}^{T-1} \gamma^t R_{t+1} \quad (2)$$

$$J(\boldsymbol{\tau}) = \int p(\boldsymbol{\tau} | \pi(\cdot, \boldsymbol{\theta})) G(\boldsymbol{\tau}) d\boldsymbol{\tau} \quad (3)$$

where $\gamma \in [0, 1]$ is the discount factor and R_{t+1} is the immediate reward as allocated by a reward function $R: Y \times U \rightarrow \mathbb{R}$. The optimal policy $\pi(\cdot, \boldsymbol{\theta}^*)$ maximises Eq. 3, subject to satisfying the set of hard constraints on the available control inputs. In this work, we deploy an 'on-policy', policy optimisation RL algorithm, known as Reinforce. Due to space constraints, little discussion is provided here. For further details see the following works (Petsagkourakis et al. (2020); Sutton et al. (2000)).

Most importantly, apprenticeship learning via inverse reinforcement learning is adopted in this work to formalise an approach to learning a parameterisation of a control policy expressed in process data (Coates et al. (2009)). The general problem statement follows: **given** demonstrations $\mathbf{T} = [\tau_1^E, \dots, \tau_k^E]$ from an existing control scheme; mathematically **abstract** the control objectives guiding the existing policy via the reward function R . The foundational IRL algorithms, construct R as a linear combination of hypothesised control objectives $\boldsymbol{\varphi} = [\varphi_1, \dots, \varphi_D]$ via the following expression:

$$R = \boldsymbol{\alpha}^T \boldsymbol{\varphi} \quad (4)$$

where $\boldsymbol{\alpha} \in \mathbb{R}^D$ denotes a vector of weights and $\varphi_d: Y \times U \rightarrow \mathbb{R}$ denotes a single control hypothesis. Explicitly, each control hypothesis is a hand constructed basis function, which must provide coherent control objective temporally. Given φ_d are of fixed structure, IRL

is concerned with abstracting control objectives via learning the optimal weights α^* , which reproduce the desired characteristics of the demonstrated control response as expressed by \mathbf{T} . These characteristics are quantified by the formulation of trajectory features $\mathbf{v}^Y \in \mathbb{R}^D$, where:

$$v_d^Y = \sum_{t=0}^T \gamma^t \varphi_d(\mathbf{y}_t, \mathbf{u}_t) \quad (5)$$

Ultimately, solution α^* provides control objectives, such that the following is satisfied:

$$\mathbb{E}_{p(\tau)}[\mathbf{v}^{Y,E}] = \mathbb{E}_{p(\tau)}[\mathbf{v}^{Y,\pi}] \quad (6)$$

where $\mathbf{v}^{Y,E} \in \mathbb{R}^D$ are the characteristics of system response τ^E on the control task under the existing control scheme and $\mathbf{v}^{Y,\pi} \in \mathbb{R}^D$ are those under the response τ^π of the RL policy as learned under the weight vector α^* . Specifically, we used maximum entropy IRL (Max-Ent IRL) to learn such a weight vector α^* . From an implementation perspective, it provides explicit mechanisms for matching the condition provided by Eq. 6. More detailed explanation of this algorithm can be found in Ziebart et al. (2013). Explicitly, we propose a framework to: learn an RL-based parameterisation of an existing control scheme by leveraging existing process data and an offline learning procedure (apprenticeship learning); then, to deploy this RL-based parameterisation to the real process to continually learn and update the control policy online via the Reinforce (or other policy optimisation) algorithm. Therefore, the framework naturally mitigates the effects of plant-model mismatch and process variability. Fig. 1 expands this further.

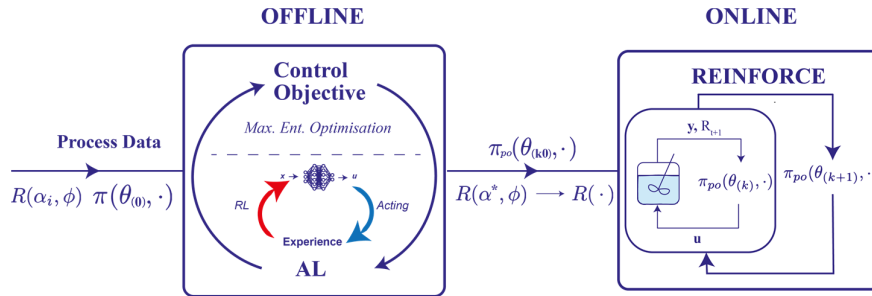


Figure 1: The framework for learning proposed. In offline learning, we leverage maximum entropy inverse RL and existing process data to synchronously abstract the objective function $R(\alpha^*, \phi)$ guiding the existing controller and a parameterisation $\pi_{po}(\theta_{k0}, \cdot)$ of the control law. We then propose to deploy the controller online to continually learn from the ongoing process under a reward function $R(\cdot)$ reflective of the true process objective.

3. Case Study

In the following section, we demonstrate the methodology presented via case study. The investigation is underpinned by a multiple-input, multiple-output (MIMO) control problem descriptive of an endothermic isomerism reaction in a continuous stirred tank reactor (CSTR) i.e. $A \xrightarrow{k_1} B$. We construct the conditional probability functions in Eq. 1 across a continuous state, control and measurement space and subsequently approximate the dynamics of the problem by the following system of stochastic differential equations:

$$\mathbf{x}_{t+1}^* = \mathbf{x}_t^* + \mathbf{h}(\mathbf{x}_t^*, \mathbf{u}_t^*)dt + \delta(\mathbf{x}_t^*)d\mathbf{W}_t \quad (7)$$

$$\mathbf{y}_{t+1}^* = \mathbf{g}(\mathbf{x}_{t+1}^*) \quad (8)$$

where $\mathbf{h}(\cdot): \mathbb{R}^{n_x \times n_u} \rightarrow \mathbb{R}^{n_x}$ represents the underlying process dynamics as derived from first principles; $\delta(\cdot): \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x}$ is process disturbance; \mathbf{W}_t is the Wiener process; and, $\mathbf{g}(\cdot): \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_y}$ is the mechanism of noisy state observation. Explicitly, control is concerned with tracking the set points of the reactor temperature T and concentration of A , C_A in the outlet. The controls available to the scheme are the temperature of a heated jacket TE and the concentration of A in the inlet C_{A0} , respectively. Further details of the process and implementation are provided by Fig. 2. In this work, we constructed a proportional-integral-derivative (PID) control scheme of subjectively near optimal performance. The control scheme consists of two PID controllers, one for each variable control loop. The PID controllers were constructed and tuned via the MATLAB Simulink package. Existing process data \mathbf{T} was generated for the control scheme, by performing Monte Carlo realizations on the approximate process model. Implementation utilised Python 3.7.3. The data generated was then characterised by Eq. 5 in the scope of trajectory feature matching, which underpins MaxEnt IRL and the offline section of the framework.

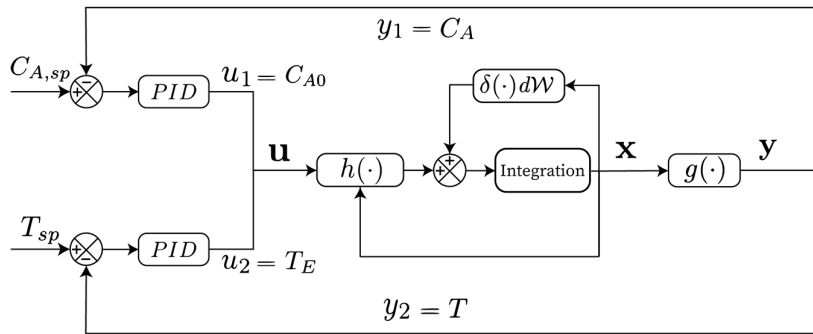


Figure 2: Flow diagram of the existing PID control scheme at implementation level. The underlying process dynamics are described by $\mathbf{h}(\cdot)$ (see Eq 7). Process stochasticity is dependent on the state \mathbf{x} and $\delta(\cdot)$, and modelled via the Wiener process \mathbf{W} . System integration was performed via the Euler-Maruyama method. The state observation is \mathbf{y} as mapped from the true state \mathbf{x} by the function $\mathbf{g}(\cdot)$; subscript sp indicates the variable set point; C_{A0} and T_E represent the control inputs to the system.

4. Results and Discussion

The purpose of this case study is to construct an RL controller which learns from demonstration provided by a near optimal control policy on a single control task and then to improve it further. As such, we demonstrate the full utility of the offline-online framework proposed. The control task itself is a set point change, such that in the concentration control loop, set point is changed by -1 kmol m^{-3} and in the temperature control loop, set point is changed by $+4 \text{ K}$. Firstly, offline learning under MaxEnt IRL is deployed to find a linear combination α^* of control hypotheses ϕ , which describes the control objectives of the existing scheme. Under this reward function a parameterisation of the control policy expressed in process data is learned in order to match the demonstrated process behavior as characterised via $\mathbb{E}_{p(\tau)}[\mathbf{v}^y]$ (see Eq. 6). The learned parameterisation is then improved under the real process objective, which in this case is pure tracking.

Here, the demonstrated control policy is that of a well-tuned PID controller and the learned policy is parameterised by a long-short term memory neural network. In this work, we use the normalized absolute error ϵ to quantify the similarity between the learned policy and that demonstrated, with respect to each v_d^y , which is formalized by Eq. 9:

$$\epsilon = \frac{\mathbb{E}_{p(\tau)}[\mathbf{v}^{y,E}] - \mathbb{E}_{p(\tau)}[\mathbf{v}^{y,\pi}]}{\mathbb{E}_{p(\tau)}[\mathbf{v}^{y,E}]} \quad (9)$$

The results of the offline learning section of the case study described are detailed by Table 1.

Table 1: The expected discounted sum of the state features of the PID ($\mathbf{v}^{y,E}$) and the policy learned through AL ($\mathbf{v}^{y,\pi}$), the normalized absolute error in approximation ϵ and the feature weight (α^*) generated in case study.

	Trajectory Features \mathbf{v}^y					
	v_1^y	v_2^y	v_3^y	v_4^y	v_5^y	v_6^y
$\mathbb{E}_{p(\tau)}[\mathbf{v}^{y,E}]$	21.63	20.68	4.08	7.93	-22.87	-22.43
$\mathbb{E}_{p(\tau)}[\mathbf{v}^{y,\pi}]$	21.41	20.76	4.31	7.03	-22.28	-22.71
ϵ	0.01	0.004	0.06	0.10	0.03	0.01
α^*	0.137	0.652	-0.067	-0.630	-0.194	-0.343

From Table 1, it is concluded that offline learning was able to abstract an RL parameterisation of the demonstrated policy (i.e. PID controller). The quality of solution is indicated by the normalized absolute error ϵ . Table 1 details that the algorithm found a solution, which achieves a maximum $\epsilon = 0.1$ and an average of $\epsilon = 0.035$. This was achieved after just four iterations of the algorithm. It is also worth noting, that the learned weight vector α^* encodes information about the process dynamics into control objectives. Although further discussion is not provided here, this implies that the weight vector itself is both specific to the set point change and the process itself. The policy was then transferred to the “real process”, which was simulated under the assumption of the same process model as in the offline setting but a reward function representative of the true process objective. Through further online learning, it is observed that the parameterisation can achieve a better control strategy compared to the existing PID controller (see Fig. 3). Specifically, RL is able to facilitate a system response, which meets set point faster with less overshoot observed than using the PID controller. For example, the updated parameterisation yields a better temperature response characterised by a fast rise time with no observable overshoot. This performance increase is derived from the ability to account for process stochasticity in a proper closed loop manner, as well as the nonlinearity of the resultant control function.

5. Conclusions

Here, we propose a framework based on apprenticeship learning (AL) to learn a control law based on process data. This approach allows us to synthesize a neural network control policy from a previous controller (e.g. PID, MPC or human controllers) more robustly than with supervised learning. Having learned a parameterisation of the control law, subsequent deployment of RL enables further policy improvement by directly interacting with the real process, thus outperforming the existing control law. Here, AL is

implemented through inverse reinforcement learning (IRL). Given the data-driven nature of AL via IRL, the RL-based policy parameterisation promises to express the action of the control scheme and process knowledge of operators. The RL controller is constructed via a policy optimisation algorithm, although other methods could be applied. Based on the case study, it is concluded that the current framework can effectively extract control information from available process data, resulting in a better optimal control policy efficiently. In future work, we will explore the performance of this framework when handling constrained chemical processes, look to account for uncertainty in the approximate process model and introduce mechanisms to account for process-model mismatch in transfer of the policy to the real process.

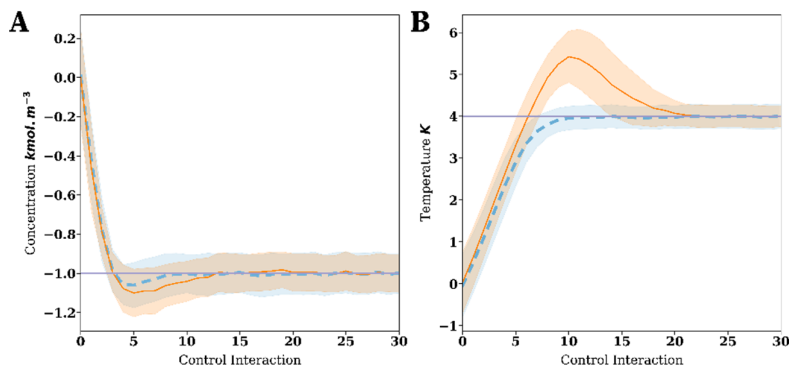


Figure 3: Optimal policy learned in Case study. A and B: System response in the concentration C_A control loop and in the temperature T control loop, respectively. π^A and π^E indicate the learned policy (after online update) and the PID, respectively. Shaded regions indicate the two standard deviations of system response and central line indicates the expected control response. Line properties: dashed line - π^A ; solid line - π^E . Set points are indicated. Only the first 30 control interactions are plotted, given the efficacy of system response

References

- A. Coates and P. Abbeel and A. Ng, 2009, Apprenticeship learning for helicopter control, *Communications of the ACM*, 52, 7, 97–105
- B. Ziebart and J.A. Bagnell, and A.K. Dey, 2013, The Principle of Maximum Causal Entropy for Estimating Interacting Processes, *IEEE Transactions on Information Theory*, 59, 4, 1966–1980
- N. Heess, J.J. Hunt, T.P. Lillicrap, D. Silver, 2015, Memory-based control with recurrent neural networks. Published online.
- P. Petsagkourakis and I.O. Sandoval and E. Bradford and D. Zhang and E.A. del Rio-Chanona, 2020, Reinforcement learning for batch bioprocess optimization, *Comput. Chem. Eng.*, 133.
- R.S. Sutton, D. McAllester, S. Singh, and Y. Mansour, 2000, Policy gradient methods for reinforcement learning with function approximation, *Advances in Neural Information Processing Systems*, 1057–1063
- S. P. K. Spielberg, R. B. Gopaluni and P. D. Loewen, "Deep reinforcement learning approaches for process control," 2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP), Taipei, 2017, pp. 201-206.
- T. Badgwell, J.H. Lee, K. Liu, 2018, "Reinforcement Learning – Overview of Recent Progress and Implications for Process Control", *Comput. Chem. Eng.*, 44.

Systematic model-based dynamic optimization of a combined cooling and antisolvent multistage continuous crystallization process

Jiaxu Liu, Brahim Benyahia*

Department of Chemical Engineering, Loughborough University, Loughborough, LE11 3TU United Kingdom

B.Benyahia@lboro.ac.uk

Abstract

A rigorous and systematic methodology was developed to optimize multistage combined cooling and antisolvent continuous MSMPR (mixed-suspension, mixed-product removal) crystallizers. The crystallization of aspirin (acetylsalicylic acid, ASA) in ethanol (solvent) and water (antisolvent) was used as a case study. A validated mathematical model of the system was firstly used to optimize the steady state performance followed by the development of several optimizations strategies aiming at the minimization of the start-up time using optimal dynamic profiles of antisolvent addition, cooling temperature, seeding flowrate and the combination of these decision variables. The start-up scenarios were also considered in the case of initially prefilled and empty vessels. Using the proposed dynamic optimization strategies, it was predicted that the start-up time can be reduced by up to 70%, which represents a significant gain in time, cost and environmental performance.

Keywords: Continuous crystallization, start-up, antisolvent, optimal seeding, dynamic optimization.

1. Introduction

Continuous manufacturing is increasingly seen as a the most flexible and viable options for the pharmaceutical industry (Mascia et al., 2013). However, despite the strides forward, many technical challenges are still to be overcome. Continuous pharmaceutical campaigns are anticipated to have short operating windows (Benyahia et al., 2012). This makes the impact of start-up and shut-down extremely important on both cost of production and environmental footprint. To address some of these key issues, the next generation pharmaceutical plants require systematic, rigorous and robust optimization strategies for process design, operation and control of single processes and integrated plants (Benyahia et al., 2012; Lakerveld et al., 2015).

Crystallization is a key purification technology widely adopted in the pharmaceutical industry to isolate active pharmaceutical ingredients (API). The critical quality attributes of the drug product, such as solubility, safety and efficacy are commonly determined by the crystal size distribution, shape and purity, which also impact the downstream processability such as flowability, filterability and dyability. A common optimization target in crystallization is to maximise the mean crystal size, lower the coefficient of variation (CV) and improve the aspect ratio (Fysikopoulos et al., 2019;

Hatcher et al., 2020). The fundamental driving force for crystallization is supersaturation. Typically, the supersaturation is generated by cooling, solvent evaporation or addition of antisolvent. Most of crystallization literature focused on the design, optimization and control of batch processes. Various model-based and model-free technique have been extensively implemented and thoroughly discussed in the literature. However only few decision and control variables are available to optimally design and operate a batch crystallization process. The most common approach is to use the temperature or/and antisolvent ratio to optimize the crystal size distribution, CV and batch time.

Over the last decade, continuous crystallization of active pharmaceutical ingredients (API) received growing interest. To date, most literatures focused on three main continuous crystallizers: mixed suspension mixed product removal (MSMPR), plug flow crystallizers and continuous oscillatory baffled crystallizers. The most dominant type of crystallizers in the pharmaceutical industries is based on a stirred tank design which exploits well-established theory and know-how accumulated over decades of batch crystallization. Consequently, many experimental and modelling efforts have been devoted to the continuous MSMPR crystallizers. Several studies focused on the optimization of single, multistage MSMPR and integrated end-to-end continuous pharmaceutical plant with a series of MSMPR crystallizers (Lakerveld et al., 2015; Mascia et al., 2013; Benyahia, 2018; Su et al., 2015). Despite the importance of optimal start-up strategies in continuous pharmaceutical manufacturing, due its significant impact on cost and environmental footprint, only few investigations were reported in the literature which identified mainly the optimal dynamic profile of the cooling temperature and antisolvent flowrate (Yang & Nagy, 2015). However, the MSMPR vessels were assumed to be prefilled with a solution having the steady state composition. This approach is unrealistic and disregards the impact of filling the vessels during the start-up which may increase even further the start-up time. To date, only few papers showed the potential impact of different start-up scenarios, including starting from empty MSMPR vessels, on continuous pharmaceutical campaigns (Benyahia et al., 2012). The later laid the ground to the current investigation, which consists in a systematic optimization approach starting from steady state optimization, aimed at the identification of the attainable mean particle size, followed by different start-up optimization strategies including start-up from empty vessels. The start-up time will be optimized by identifying the optimal discrete piece-wise profiles of antisolvent flowrate, seed flowrate, cooling rates as well as the seed loading in the seed vessel.

2. Materials and Methods

To implement the dynamic and steady state optimization strategies, the crystallization of aspirin (acetylsalicylic acid, ASA) in ethanol (solvent) and water (antisolvent) mixture is considered as a case study. The dynamic mathematical model which represents the population balances, based on the standard method of moment, the energy balance, the total mass balance in each vessel as well as the mass balance of API, solvent and antisolvent was developed based on data available in the literature (Barik et al., 2020; Lindenberg et al., 2008; Liu & Benyahia, 2020).

Prior to the development of the optimal start-up policies, a steady-state optimization problem was solved with the aim of maximizing the mean crystal size (d) using the jacket temperature (T_J), antisolvent flowrate (F_{AS}) and residence time (R) at each

crystallization stage, as the decision variables. This approach delivered the reference attainable crystal size for a 3-stage MSMR system and the optimal design and operating steady state conditions shown in Table 1. However, the steady state optimization, does not consider the dynamic performance of the process and disregards large amounts of off-specification crystals (waste) generated during start-up. To address this issue, several dynamic optimizations strategies were developed to minimize the start-up time by identifying the discrete optimal dynamic profiles (piece-wise constant) of the antisolvent flowrates, cooling rates, seed flowrates and seed loading. In addition, the crystallizers were assumed both empty and prefilled at the beginning of the process.

Table 1 steady state control variable

	T_J (°C)	F_{AS} (g /min)	R (min)
Stage 1	32.75	9.94	5.83
Stage 2	32.10	5.88	10.15
Stage 3	25.00	2.68	14.02

2.1. Start-up optimization using dynamic antisolvent addition

The first series of start-up optimization focused on the manipulation of the antisolvent flowrate profiles at the first, the first two then all three MSMR stages. The approach was developed to investigate whether manipulating the antisolvent flowrates in multiple stages would add any value, despite the increased complexity. As a first attempt, the antisolvent flowrate profiles were divided into n_k equality spaced time intervals. The flowrates of the antisolvent at each interval are the decision variables. The first set of scenarios focussed on prefilled vessels and the second considered empty vessels. The dynamic optimization problems can be expressed using the generic mathematical formulation below:

$$\begin{aligned}
 & \text{Min}_{F_{AS,i,k}, T_{J,i,k}} t_{95} \\
 & \text{s.t.} \\
 & \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{p}, t) \quad \mathbf{x}(t=0) = \mathbf{x}_0 \\
 & 0 = \mathbf{g}(\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{p}, t) \\
 & 0 \leq F_{AS,i,k} \leq 20 \\
 & 20 \leq T_{J,i,k} \leq 40 \\
 & S_i(t) \geq 1 \\
 & i = 1, 2, 3 \\
 & k = 1, 2, \dots, n_k
 \end{aligned}$$

Where t_{95} is the time required for the crystallizer to reach controlled state of operation that has less than 5% variation. $T_{J,i,k}$ (°C) and $F_{AS,i,k}$ (g/min) are the jacket temperature

and antisolvent flow rate at the k th time interval in the i th stage. S_i is the relative supersaturation maintained above 1 in all stages to prevent dissolution of crystals.

2.2. *Start-up optimization using a dynamic seeding profile*

The addition of seeds is commonly adopted in batch systems to generate initial particles without relying on the nucleation, which is difficult to control and requires larger supersaturation levels to reach the boundary of the metastable zone. The addition of seeds in batch systems can provide an effective control strategy over the number of particles and consequently the final particle size distribution. In this work, the seeding approach was used to optimize the start-up time. Instead of one single seeding addition, the optimizer will determine the optimal seed addition profiles as well as the required initial seeding charge in the seeding vessel. The seeds are suspended in a saturated solution having the same composition as the feed solution and the size distribution of seeds was obtained from literature. The seeds were added to the first stage only, where most of the nucleation occurs. The mathematical formulation of the optimization problem is similar to the one described above, with the decision variables being the piece-wise flow rate profile of seed addition alongside the seeding bounds (0: lower bound and 20% of the API load in the feed solution).

2.3. *Start-up optimization using combined seeding, cooling and antisolvent addition*

To maximize the efficiency of the start-up optimization, the optimal seeding policies, which correspond to a piece-wise seeding profile and seed loading ($\mu_{0,seed}$), were combined with the cooling and antisolvent profiles.

3. **Results and Discussion**

The optimization problems were solved using a hybrid method which combines a genetic algorithm and a deterministic method (SQP) available in MATLAB. The start-up optimization results obtained using dynamic antisolvent and cooling (jacket temperature) profiles in the three stages are shown in Figure 1. The predicted optimal start-up time (t_{95}) was 26.1 min in the case of prefilled vessels and 22.2 min in the case empty vessels. Compared to the reference start-up time obtained by the implementation of the steady state operating conditions (69.2 min), an improvement of 62.3% and 67.0% respectively, can be achieved.

Additional start-up optimization problems based on the manipulation of the operating profiles in one stage and two stages starting from prefilled and empty vessels were also addressed. However, the predicted start-up times were larger compared to the cases discussed above. These results are not presented here for the sake of brevity. Despite the increased complexity, manipulating the antisolvent flowrates and cooling profiles in three stages enhances the flexibility and degrees of freedom and help achieve shorter start-up times and increased on-specification production.

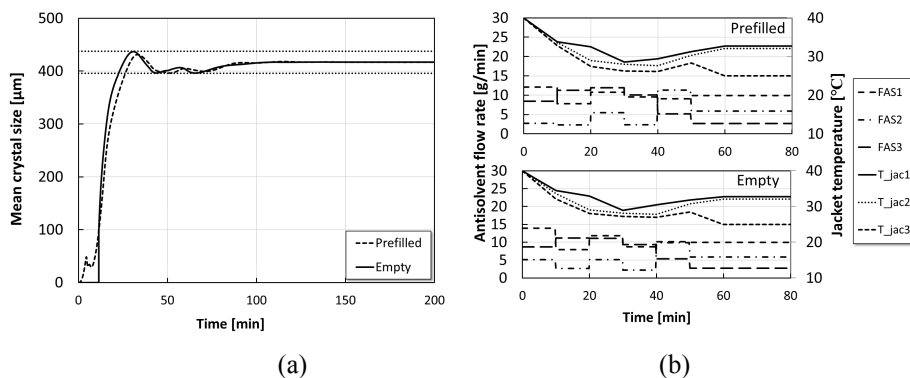


Figure 1. Optimization results obtained by manipulating jacket temperatures and antisolvent flowrates in the 3 MSMPR stages (a) Mean crystal size profiles (b) optimal cooling and antisolvent addition profiles.

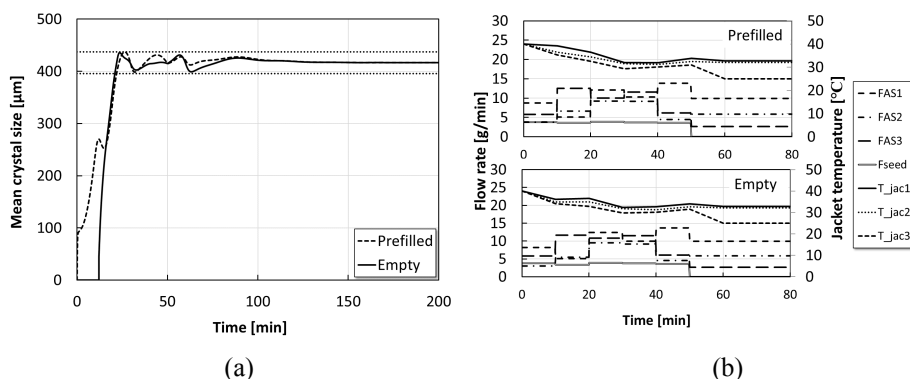


Figure 2. Optimization results obtained by manipulating the jacket temperatures and antisolvent and seeding flowrates in the 3 MSMPR stages (a) Mean crystal size profiles (b) optimal cooling, seeding and antisolvent addition profiles.

Figure 2 shows the dynamic optimization results obtained by manipulating the jacket temperatures as well as the antisolvent and seeding flowrates in the 3 MSMPR stages. With the addition of the seeding policies (seeding flow rates and initial seed load) to the set of decision variable comprising the jacket temperature and antisolvent flowrate profiles, the start-up time can be shortened to 21.8 min (in the case of prefilled vessels) and 20.9 min (in the case of initially empty vessel). This is equivalent to a 67.8% and 69.8% improvement respectively compared to the reference steady state settings. Moreover, a start-up time of 28.5 min was obtained using solely the optimal seeding policies (not presented here). It becomes clear that the combination of optimal seeding policies, cooling and antisolvent addition enhances the degrees of freedom and optimization flexibility and therefore delivers the best start-up performance.

4. Conclusion

The start-up performance of a three-stage MSMPR crystallizer was optimized based on a rigorous and systematic model-based approach. The attainable steady state particle

size and start-up time, obtained using steady state optimization, were used as references for the start-up optimization. To minimize the start-up time, several strategies were adopted including starting from empty and prefilled vessels. The dynamic optimization of the start-up was achieved using discretised piece-wise antisolvent flowrates, seed addition and the combination of both. The discretisation policy was based on equally spaced time intervals which can be refined further using optimized variable time intervals. The seed charge in the seeding vessel was also optimized in order to identify the optimal seeding policies. Overall, the start-up cases associated with the empty vessels seem to outperform the cases starting from prefilled vessels. The start from prefilled vessels may indicate that the nucleation requires longer time compared to starting from empty vessels where nucleation may happen very quickly. The number of time intervals was also increased from 5 to 10 and 20. However, these results were not presented for the sake of brevity.

It was shown that the implementation of optimal piece-wise continuous cooling and piece-wise antisolvent flowrates in the three stages can help reduce the start-up time by 67%. However, the best performance was obtained by combining the optimal cooling, antisolvent and seeding policies which predicted a reduction of the start-up time by nearly 70%. It is worth mentioning that the optimization strategies developed in this work are in essence open-loop and their implementation may require an advanced control strategy such as model predictive control to ensure effective trajectory tracking and the predicted minimization of the start-up time.

References

- Barik, K., Prusti, P., & Mohapatra, S. S. 2020. Single- and multi-objective optimisation for a combined cooling and antisolvent semi-batch crystallisation process with an ACADO toolkit. *Indian Chemical Engineer*, 62(3), 287–300.
- Benyahia, B. 2018. Applications of a plant-wide dynamic model of an integrated continuous pharmaceutical plant: Design of the recycle in the case of multiple impurities. In *Computer Aided Chemical Engineering* (1st ed., Vol. 41). Elsevier B.V.
- Benyahia, B., Lakerveld, R., & Barton, P. I. 2012. A plant-wide dynamic model of a continuous pharmaceutical process. *Industrial and Engineering Chemistry Research*, 51(47), 15393–15412.
- Fysikopoulos, D., Benyahia, B., Borsos, A., Nagy, Z. K., & Rielly, C. D. 2019. A framework for model reliability and estimability analysis of crystallization processes with multi-impurity multi-dimensional population balance models. *Computers and Chemical Engineering*, 122, 275–292.
- Hatcher, L. E., Li, W., Payne, P., Benyahia, B., Rielly, C. D., & Wilson, C. C. 2020. Tuning morphology in active pharmaceutical ingredients: Controlling the crystal habit of lovastatin through solvent choice and non-size-matched polymer additives. *Crystal Growth and Design*, 20(9), 5854–5862.
- Lakerveld, R., Benyahia, B., Heider, P. L., Zhang, H., Wolfe, A., Testa, C. J., Ogden, S., Hersey, D. R., Mascia, S., Evans, J. M. B., Braatz, R. D., & Barton, P. I. 2015. The Application of an Automated Control Strategy for an Integrated Continuous Pharmaceutical Pilot Plant. *Organic Process Research and Development*, 19(9), 1088–1100.
- Lindenberg, C., Krättli, M., Cornel, J., Mazzotti, M., & Brozio, J. 2008. Design and Optimization of a Combined Cooling/Antisolvent Crystallization Process. *Crystal Growth & Design*, 9(2), 1124–1136.

- Liu, J., & Benyahia, B. 2020. Systematic Model-Based Steady State and Dynamic Optimization of Combined Cooling and Antisolvent Multistage Continuous Crystallization Processes. *Proceedings*, 62(1), 7.
- Mascia, S., Heider, P. L., Zhang, H., Lakerveld, R., Benyahia, B., Barton, P. I., Braatz, R. D., Cooney, C. L., Evans, J. M. B., Jamison, T. F., Jensen, K. F., Myerson, A. S., & Trout, B. L. 2013. End-to-end continuous manufacturing of pharmaceuticals: Integrated synthesis, purification, and final dosage formation. *Angewandte Chemie - International Edition*, 52(47), 12359–12363.
- Su, Q., Benyahia, B., Nagy, Z. K., & Rielly, C. D. 2015. Mathematical Modeling, Design, and Optimization of a Multisegment Multiaddition Plug-Flow Crystallizer for Antisolvent Crystallizations. *Organic Process Research and Development*, 19(12), 1859–1870.
- Yang, Y., & Nagy, Z. K. 2015. Combined cooling and antisolvent crystallization in continuous mixed suspension, mixed product removal cascade crystallizers: Steady-state and startup optimization. *Industrial and Engineering Chemistry Research*, 54(21), 5673–5682.

Predicting Microclimate of a Closed Greenhouse Using Support Vector Machine Regression

Farhat Mahmood^a, Rajesh Govindan^b, Tareq Al-Ansari^{a,b*}

^a *Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation. Doha, Qatar*

^b *Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation. Doha, Qatar*
talansari@hbku.edu.qa

Abstract

A semi-closed greenhouse has been studied in this research. The greenhouse is designed to keep the inside temperature and relative humidity within the optimum growing range throughout the year while maximizing the utilization of solar energy for the plants. The Venlo shaped greenhouse has 4 mm tempered glass as the covering material and three air handling units which control the microclimate. Using air handling units to manage the microclimate of a greenhouse is becoming increasingly popular in arid climates. The plant yield productivity and quality depend on the accurate monitoring and control of the greenhouse microclimate. This study aims to develop a dynamic model that predicts the temperature and relative humidity to improve climate monitoring control accuracy. The data-driven model determines the temperature and relative humidity by incorporating factors affecting the microclimate, such as solar radiation, ambient temperature, relative humidity, fan speed, etc. The available greenhouse data spans from April to June (3 months). Results illustrate that the model predicted accurate values for temperature and relative humidity with an R^2 value of 0.930 and 0.911 and an RMSE value of 0.826 and 1.740, respectively. By accurately predicting the temperature and relative humidity inside the greenhouse, the crop yield can be increased, while minimizing the energy consumption by the air handling units, making greenhouses in an arid climate a more economically feasible option.

Keywords: Greenhouse, Dynamic modeling, Support vector machine, Climate control, microclimate.

1. Introduction

As the world population is expected to rise to 9.7 billion by the year 2050, achieving food security is one of the most challenging problems of the 21st century (UN, 2019). With the expected increase in the population coupled with the demand for food, it is essentially required to develop clean and sustainable food production systems. Agriculture greenhouse presents a viable solution as they minimize resource consumption while increasing the yield by 10-20% (De Gelder et al., 2012). Greenhouses offer a closed and controlled environment to the plants while maintaining optimum conditions for plant development. Maintaining these optimum growing conditions throughout the year is a significant concern as various factors affect the greenhouse microclimate. Greenhouse microclimate variables are complex and nonlinear and depend on environmental factors such as solar radiation, ambient air

temperature, outside relative humidity, etc. Therefore, developing an accurate model to predict the microclimate parameters is an essential part of greenhouse management.

Microclimate prediction models can be based on thermodynamics laws that incorporate the heat and mass exchange taking place between the different driving factors or data-driven models using data from the input-output of the process occurring inside the greenhouse. The microclimate models are highly complex, resulting in inconsistent results when physical models are used to predict the microclimate (Van Henten, 1994). On the other hand, data-driven models are increasingly being applied to the ongoing development of new computational methods. The data-driven approach provides accurate and quick results for agricultural applications such as predicting the temperature and relative humidity inside a greenhouse (Chen et al., 2016).

Neural networks have been extensively used to model nonlinear systems such as the greenhouse microclimate (Ribeiro et al., 2019). However, the application of neural networks is limited due to certain disadvantages such as overfitting and stopping at local minima (Vapnik, 1995). Support vector machines can model nonlinear systems without the mentioned disadvantages (Engel et al., 2002). Wang et al. (2009) used support vector regression (SVR) to develop a greenhouse microclimate model to predict the temperature inside the greenhouse based on external and internal factors. Results demonstrated that the SVR successfully predicted the temperature with a mean square value of 0.351. Liu et al., (2016) developed a wireless sensor network-based prediction model to predict the greenhouse microclimate. Results illustrated that SVM had a value of R^2 of 0.9865 and 0.9605 for the temperature and relative humidity inside the greenhouse, respectively. Similarly, Taki et al., (2018) developed a model to estimate the temperature of air, plant, and soil inside a greenhouse using SVM and artificial neural networks. The model included factors such as inside temperature, outside temperature, solar radiation, and wind speed. Results illustrated the SVM had an R^2 value of 0.83, 0.99, and 0.98 for the air, plant, and soil temperature, respectively.

Most of the studies existing in literature have developed models for greenhouses with natural ventilation in tropical or cold environments. Semi-closed greenhouses depend on the cooling system to maintain the optimum growing conditions inside the greenhouse. Maintaining optimum temperature is a challenging task during the summer season as the solar radiation and ambient temperature rise, increasing the cooling load of the greenhouse. In this study, a microclimate model is developed to predict the temperature and relative humidity inside a closed greenhouse with no ventilation in an arid climate to improve greenhouse management.

2. Greenhouse data

The experiments were conducted in a Venlo shaped greenhouse (800 m²) with 4 mm tempered glass as the covering material located on a greenhouse farm in Qatar. The data was collected from April 1, 2020, to June 30, 2020, using tomatoes as the plants. The temperature and relative humidity inside the greenhouse were measured using a sensor module shown in Figure 1. The data was recorded after every five minutes and saved in the data logger. The external conditions such as solar radiation, ambient temperature, and outside relative humidity were all measured and recorded by external sensors.



Figure 1: Temperature and relative humidity monitoring sensor.

3. Methodology

Support vector machines (SVM) were initially developed to recognize patterns and classify objects. A small subset from the training sample is used to define the decision boundary called support vectors. The concept of support vectors originated from the optimum separation of the hyperplane which could separate the points without error and maximize the distance between the closest point and the hyperplane. SVM is similar to perceptron in neural networks as the output is a combination of the nodes present in the middle, as illustrated in Figure 2.

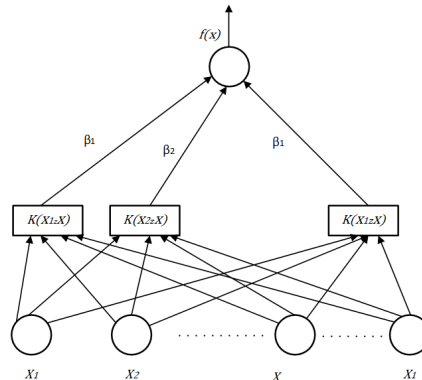


Figure 2: Support vector machine structure.

Support vector machine regression (SVMR) calculates the real value function by generalizing the support vectors. Support vector machine regression (SMVR) is also called support vector regression (SVR) to differentiate it from support vector machine (SVM). SVR follows the structural risk minimization (SRM), which has been shown to give better results than empirical risk minimization (ERM) (Wang et al., 2009). SVR maps data from a vector x into an F , which is a high dimensional feature space, and the linear regression is performed as with b as the threshold value:

$$f(x) = (w \cdot \varphi(x)) + b(\varphi : R^n \rightarrow F, w \in F) \quad (1)$$

The resulting linear regression corresponds to nonlinear regression in a low dimension space. The value of w can be calculated by the following:

$$R(w) = R_{emp} + \lambda \|w\|^2 = \sum_{i=1}^l e(f(x_i) - y_i) + \lambda \|w\|^2 \quad (2)$$

Where e is the cost function, l is the number of samples, R_{emp} is the sum of empirical risk, and λ is the constant of regularization. The cost function can be represented in a linear form given as:

$$e(f(x) - y) = \max(0, |f(x) - y| - \epsilon) \quad (3)$$

Quadratic cost function:

$$e(f(x) - y) = (f(x) - y)^2 \quad (4)$$

Huber cost function:

$$e(f(x) - y) = \begin{cases} \mu |f(x) - y| - \frac{\mu^2}{2}, & \text{if } |f(x) - y| > \mu \\ \frac{1}{2} |f(x) - y|^2, & \text{other} \end{cases} \quad (5)$$

The function $R(w)$ is minimized by first determining $(\alpha - \alpha_i^*)$

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varphi(x_i) \quad (6)$$

4. Results and Discussion

The data set has outside temperature, solar radiation, fan speed, cooling temperature setpoint, and humidity deficit as the input factors, while the greenhouse temperature and relative humidity are the outputs. The dataset is split into training and testing data, with 70% dedicated for training and 30% for the testing set. The model's performance is evaluated at different time steps such as 5, 10, 15, and 20 minutes (SVM-5, SVM-10, SVM-15, SVM-20) by increasing the prediction time step. The SVM model algorithm is implemented in Python 3.7.

Table 1 shows the temperature and relative humidity predicted results using the SVR model. The results have been obtained by increasing the time step of the data from 5 up to 20 minutes.

Table 1: SVM prediction accuracy for temperature and relative humidity.

	SVM-5		SVM-10	
	Temperature	Relative Humidity	Temperature	Relative Humidity
RMSE	0.826	1.740	0.909	2.207
R²	0.930	0.911	0.917	0.857
	SVM-15		SVM-20	

	Temperature	Relative Humidity	Temperature	Relative Humidity
RMSE	0.982	2.517	0.998	2.598
R²	0.899	0.818	0.897	0.807

The temperature has an RMSE and R^2 value of 0.826 and 0.93, respectively and relative humidity has an RSME and R^2 value of 1.740 and 0.911 at time step 5. The accuracy of the model decreases by increasing the time step as the number of data points decreases and has the lowest accuracy at time step of 20 minutes, which illustrates for better prediction performance, a more extensive data set should be used. Figure 3 shows the prediction of the trained model over an unseen data set of 3 days. The graph on the left side is the actual vs. predicted temperature of the greenhouse, and the graph on the right shows the actual vs. predicted data for the relative humidity. The red shows the actual data, and the yellow lines indicate the predicted temperature and relative humidity values, respectively. The model gives accurate predictions for the temperature and relative humidity, as shown in the graphs. Also, the actual vs. predicted graph illustrates that the trained model is not overfitted as it produces accurate results for a new set of values for both the temperature and relative humidity inside the greenhouse.

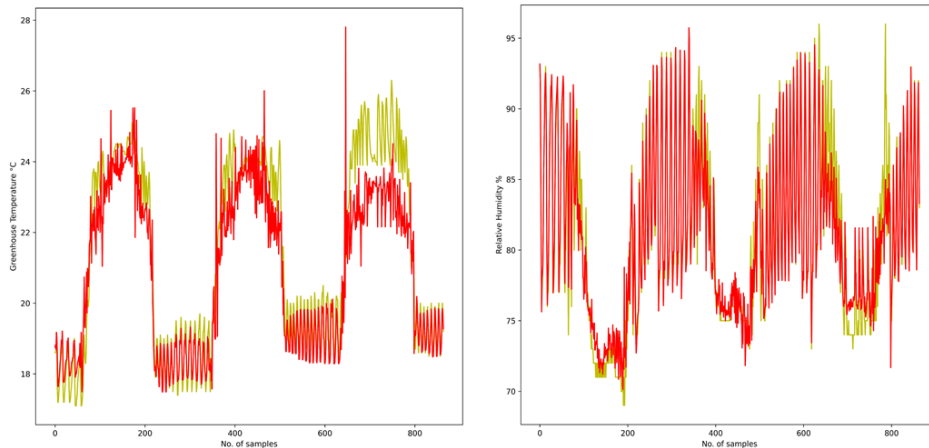


Figure 3: Actual vs. Predicted for a 3-day data set; (left) Temperature (Right) Relative humidity.

To improve the prediction accuracy of the relative humidity, data regarding the irrigation history and soil moisture can be included by using the relevant sensors as the greenhouse uses a hydroponic system to provide water and nutrients to the plants.

5. Conclusion

The microclimate of a greenhouse is a dynamic system where various factors such as solar radiation, ambient temperature, relative humidity, etc., affect the temperature and relative on the inside of the greenhouse. The study shows the possibility of predicting the greenhouse temperature and relative humidity using support vector machines. Support vector machines are feasible for modeling dynamic environments where the variables are constantly changing, such as the environmental factors. The developed model has an R^2 and RMSE value of 0.930 and 0.826 for temperature and 0.911 and

1.740 for relative humidity. The accuracy of the model for temperature and relative humidity dropped by increasing the data time step as the training data points were reduced. The model had the lowest prediction accuracy at a twenty-minute interval with an R^2 value of 0.897 and 0.807 for temperature and relative humidity. By predicting the temperature and relative humidity based on the factors, the water and energy consumption can be optimized while providing optimum growing conditions to the plants. The results of this study show the potential of applying machine learning for greenhouse management.

Acknowledgment

The authors would like to thank Yara Fertilizer for their help and technical support to carry out this research.

References

- .N.Chen, L., Zhang, B., Yao, F., Cui, L., 2016. Modeling and simulation of a solar greenhouse with natural ventilation based on error optimization using fuzzy controller. Chinese Control Conf. CCC 2016-Augus, 2097–2102. <https://doi.org/10.1109/ChiCC.2016.7553676>
- De Gelder, A., Dieleman, J.A., Bot, G.P.A., Marcelis, L.F.M., 2012. An overview of climate and crop yield in closed greenhouses. *J. Hortic. Sci. Biotechnol.* 87, 193–202. <https://doi.org/10.1080/14620316.2012.11512852>
- Engel, Y., Mannor, S., Meir, R., 2002. Sparse online greedy support vector regression. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 2430, 84–96. https://doi.org/10.1007/3-540-36755-1_8
- Liu, Q., Jin, D., Shen, J., Fu, Z., Linge, N., 2016. A WSN-based prediction model of microclimate in a greenhouse using extreme learning approaches. *Int. Conf. Adv. Commun. Technol. ICACT 2016-March*, 730–735. <https://doi.org/10.1109/ICACT.2016.7423609>
- Ribeiro, R., Casanova, D., Teixeira, M., Wirth, A., Gomes, H.M., Borges, A.P., Enembreck, F., 2019. Generating action plans for poultry management using artificial neural networks. *Comput. Electron. Agric.* 161, 131–140. <https://doi.org/10.1016/j.compag.2018.02.017>
- Taki, M., Abdanan Mehdizadeh, S., Rohani, A., Rahnama, M., Rahmati-Joneidabad, M., 2018. Applied machine learning in greenhouse simulation; new application and analysis. *Inf. Process. Agric.* 5, 253–268. <https://doi.org/10.1016/j.inpa.2018.01.003>
- UN, 2019. *World Population Prospects 2019: Highlights*. United Nations Publ. 1.
- Van Henten, E.J., 1994. Greenhouse climate management: an optimal control approach. *Agric. Syst.* 45, 342.
- Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*. *Nat. Stat. Learn. Theory*. <https://doi.org/10.1007/978-1-4757-2440-0>
- Wang, D., Wang, M., Qiao, X., 2009. Support vector machines regression and modeling of greenhouse environment. *Comput. Electron. Agric.* 66, 46–52. <https://doi.org/10.1016/j.compag.2008.12.004>

Adjustable Robust Optimal Control for Nonlinear Chemical Process under Uncertainty

Enzhi Liang,^a Zhihong Yuan^{a,*}

^a*Department of Chemical Engineering, Tsinghua University, Beijing, 10084, China*

**Corresponding author's E-mail: zhihongyuan@mail.tsinghua.edu.cn*

Abstract

This paper has proposed an adjustable robust dynamic optimization (ARDO) scheme for nonlinear chemical process under time-dependent uncertainties. A novel continuous affine decision rule extended from multistage affine decision rule is developed to approximate the causal dependence of wait-and-see decision variables on the infinite dimensional uncertainty factors. Through applying the state-of-the-art first-order Taylor expansion method, the adjustable robust counterpart can be formulated. The effectiveness and applicability of the proposed ARDO scheme is demonstrated by a classic Williams-Otto semi-batch problem. Compared with traditional robust optimization for uncertain dynamic systems, the proposed framework can adjust the decision variables based on true values of the revealed uncertainties, leading to more flexible control profiles and less conservative solutions.

Keywords: adjustable robust dynamic optimization, time-dependent uncertainty, continuous affine decision rule, nonlinear semi-batch reactor

1. Introduction

Dynamic optimization has received extensive attention during the past decades to deal with batch optimal operation problems. Unfortunately, usually employing deterministic optimal control will force systems running on safety margins, potentiate constraints violation in the presence of uncertainties. Although traditional static robust optimization (SRO) methods provide effective frameworks to overcome uncertainties, in general, their inherent conservatism cannot be completely avoided, because the manipulated variables in SRO schemes, which are referred to as here-and-now decision variables, are optimized without knowledge of the true values of the uncertainties (Diehl et al, 2006). To reduce the conservativeness of the solution, lessons can be learned from the adjustable robust optimization (ARO) method for two-stage linear problems (Ben-Tal et al, 2004). By introducing wait-and-see or adjustable decision variables, part of the manipulated variables in ARO scheme can correspond to gradually revealing information about the uncertainty, thus leading to lower conservative system performance and better objective values. However, owing to the intrinsic difficulties of nonlinear problems, research on adjustable robust nonlinear optimization with only algebraic constraint or differential-algebraic constraints is scarce.

Triggered by the progresses of ARO on multistage planning and scheduling (Zhang et al, 2016), we propose an adjustable robust optimal control scheme for nonlinear chemical processes under time-dependent uncertainties on continuous time horizon. A novel continuous affine decision rule is developed to approximate the causal dependence of wait-and-see decision variables on the infinite dimensional uncertainty factors to overcome potential computational intractability. Then the classical first-order Taylor

expansion method is applied to derive the robust counterpart which is finally solved numerically by IPOPT in Pyomo. The effectiveness and applicability of the proposed ARDO scheme is demonstrated by a classic Williams-Otto semi-batch problem.

2. Model formulation

2.1. General formulation of deterministic dynamic optimization

General deterministic dynamic optimization problem defined on $t \in [0, t_f]$ can be formulated as follows:

$$\min_u \Phi(x(t_f), t_f) \quad (1)$$

$$\frac{dx}{dt} = f(x, u, t, p); \quad x(0) = x_0, \quad (2)$$

$$0 \geq G(x, u, t, p) \quad (3)$$

where $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$ and $p \in \mathbb{R}^{n_p}$ represent state, controls and model parameters respectively. Besides, $\Phi(\cdot)$ is a scalar Mayer-type cost function, $f(\cdot)$ represents differential equations with initial conditions x_0 while $G(\cdot)$ denotes inequalities.

2.2. General formulation of traditional static robust dynamic optimization

Regarding the problem under uncertainty, model parameters are divided into $p = [\zeta; \bar{p}]$ where $\zeta \in \mathbb{R}^{n_\zeta}$ and $\bar{p} \in \mathbb{R}^{n_p}$ denote uncertain and known parameters respectively. Furthermore, time-dependent uncertainties are denoted as $\zeta(\tau) \in \mathbb{R}^{n_\zeta}$. Time horizon for uncertainty realization is denoted as $\tau \in [0, t_f]$ while horizon for states and controls is $t \in [0, t_f]$. Formulation of traditional SRO defined on $\tau, t \in [0, t_f]$ is:

$$\underset{u(t)}{\text{minimize}} \quad \max_{\zeta(\tau) \in Z} \Phi(x(t_f), t_f) \quad (4)$$

$$\frac{dx}{dt} = f(x, u, t, \bar{p}, \zeta); \quad x(0) = x_0; \quad \forall \zeta(\tau) \in Z \quad (5)$$

$$0 \geq G(x, u, t, \bar{p}, \zeta); \quad \forall \zeta(\tau) \in Z \quad (6)$$

$$Z := \{\zeta(\tau) \in \mathbb{R}^{n_\zeta} \mid -\Delta\zeta_j \leq \zeta_j(\tau) - \zeta_j^{nom} \leq \Delta\zeta_j, \quad j \in I_\zeta\} \quad (7)$$

Where Z is a predefined box uncertainty set and I_ζ is the subscript set of uncertainties.

2.3. General scheme of proposed adjustable robust dynamic optimization

The proposed adjustable robust dynamic optimization is proposed by applying wait-and-see strategy to nonlinear dynamic optimization problems under time-dependent uncertainty. In the proposed ARDO scheme, the wait-and-see decision variables are determined based on a general decision policy $d_t(\cdot)$ defined as:

$$u(t) = d_t(\hat{\zeta}(t)), \quad d_t(\cdot): \prod_{\tau=t_0}^{\tau=t} Z_\tau \rightarrow \mathbb{R}^{n_u} \quad (8)$$

where $\hat{\zeta}(t)$ is the trajectory of uncertainties revealing in time interval $\tau \in [0, t]$.

Proposed ARDO scheme is defined as Eqs.(9)-(13),

$$\underset{d_t(\cdot)}{\text{minimize}} \quad \max_{\zeta(\tau) \in Z_\tau} \Phi(x(t_f), t_f) \quad (9)$$

$$\frac{dx}{dt} = f(x, u, t, \bar{p}, \zeta); \quad x(0) = x_0; \quad \forall \zeta(\tau) \in Z \quad (10)$$

$$0 \geq G(x, u, t, \bar{p}, \zeta); \quad \forall \zeta(\tau) \in Z \quad (11)$$

$$Z := \{\zeta(\tau) \in \mathbb{R}^{n_\zeta} \mid -\Delta\zeta_j \leq \zeta_j(\tau) - \zeta_j^{nom} \leq \Delta\zeta_j, j \in I_\zeta\} \quad (12)$$

$$u(t) = d_t(\hat{\zeta}(t)) \quad (13)$$

The above formulation is generally computationally intractable due to undermentioned reasons. On one hand, functional space of possible $d_t(\cdot)$ is infinite. On the other hand,

requirement of satisfying constraints over uncertainty set leads to a semi-infinite problem. Therefore, this general ARDO model will be reformulated to gain tractability in next section.

3. ARDO reformulations

3.1. Continuous affine decision rule

As a preliminary, multistage affine decision rule is defined as Eq.(14), which restricts decision-making at t_k depend on observed uncertainty till t_k , since the undesired effect of uncertainty at t_1 may not be fully compensated only by decision $u(t_1)$, which is bounded within operation limits.

$$\mathbf{u}(t_k) = \mathbf{u}^0(t_k) + \sum_{i=0}^k \mathbf{u}^1(t_k, \tau_i) \zeta(\tau_i), \quad k = 0, 1, \dots, N \quad (14)$$

Obviously, Eq.(14) cannot be directly employed for differential-algebraic systems due to its discrete nature. Therefore, a novel continuous affine decision policy is proposed to approximate $\mathcal{d}_t(\cdot)$ for dynamic optimization problem on continuous time horizon.

The slope parameter in Eq.(14) can be written as $\mathbf{u}^1(t_k, \tau_i) = w(t_k, \tau_i) \Delta \tau_i$ where the time span is defined as $\Delta \tau_i = \tau_i - \tau_{i-1}$. In such way, Eq.(14) is converted to,

$$\mathbf{u}(t_k) = \mathbf{u}^0(t_k) + \sum_{i=0}^k w(t_k, \tau_i) \zeta(\tau_i) \Delta \tau_i, \quad k = 0, 1, \dots, N \quad (15)$$

Naturally, if the number of time intervals N approaches infinity, time spans $\Delta \tau_i$ become infinitesimals, denoted as $d\tau$. Thus Eq.(15) is extended to continuous time horizon as,

$$\mathbf{u}(t_k) = \mathbf{u}^0(t_k) + \int_{\tau=0}^{\tau=t_k} w(t_k, \tau) \zeta(\tau) d\tau, \quad k = 0, 1, \dots, N \quad (16)$$

If t_k is replaced by t in Eq.(16), the continuous affine decision rule can be formulated as,

$$\mathbf{u}(t) = \mathbf{u}^0(t) + \int_0^{\tau=t} w(t, \tau) \zeta(\tau) d\tau, \quad t \in [0, t_f] \quad (17)$$

3.2. Adjustable robust counterpart formulation

The uncertain problem given by Eqs. (9)-(12) is a semi-infinite program. Therefore, the state-of-the-art first-order Taylor expansion is applied at the nominal uncertainty values to guarantee constraints be satisfied on uncertainty set. The robust counterparts of inequality constraints Eq.(11) are written as:

$$0 \geq G(x, u, t, \bar{p}, \zeta^{nom}) + \|\nabla_{\zeta} G(x, u, t, \bar{p}, \zeta) \Delta \zeta\|_1 \quad (18)$$

$$\nabla_{\zeta} G(x, u, t, \bar{p}, \zeta) = \left(\frac{\partial G}{\partial \zeta} \right) + \left(\frac{\partial G}{\partial x} \right) D(t) + \left(\frac{\partial G}{\partial u} \right) \left(\frac{\partial u}{\partial \zeta} \right) \quad (19)$$

The inner level of original objective function is approximated as:

$$\left(\max_{\zeta(\tau) \in Z_{\tau}} \Phi(x(t_f), t_f) \right) = \Phi(x(t_f), t_f) + \left\| \left(\frac{\partial \Phi}{\partial x} \right) D(t_f) \Delta \zeta \right\|_1 \quad (20)$$

To summarize, the adjustable robust counterpart of ARDO model in Eqs. (9)-(13) can be described as Eqs. (21)-(27).

$$\underset{w(t, \tau), u^0(t)}{\text{minimize}} \quad \Phi(x(t_f), t_f) + \left\| \left(\frac{\partial \Phi}{\partial x} \right) D(t_f) \Delta \zeta \right\|_1 \quad (21)$$

$$\frac{dx}{dt} = f(x, u, t, \bar{p}, \zeta^{nom}); \quad x(0) = x_0; \quad (22)$$

$$u(t) = u^0(t) + \int_0^{\tau=t} w(t, \tau) \zeta(\tau) d\tau, \quad t \in [0, t_f] \quad (23)$$

$$\frac{dD(t)}{dt} = \left(\frac{\partial f}{\partial \zeta} \right) + \left(\frac{\partial f}{\partial x} \right) D(t) + \left(\frac{\partial f}{\partial u} \right) \left(\frac{\partial u}{\partial \zeta} \right) \quad (24)$$

$$0 \geq G(x, u, t, \bar{p}, \zeta^{nom}) + \|\nabla_{\zeta} G(x, u, t, \bar{p}, \zeta) \Delta \zeta\|_1 \quad (25)$$

$$\nabla_{\zeta} G(x, u, t, \bar{p}, \zeta) = \left(\frac{\partial G}{\partial \zeta} \right) + \left(\frac{\partial G}{\partial x} \right) D(t) + \left(\frac{\partial G}{\partial u} \right) \left(\frac{\partial u}{\partial \zeta} \right) \quad (26)$$

$$D(0) = 0 \quad (27)$$

4. Case study: Williams-Otto semi-batch

We apply the adjustable robust dynamic optimization scheme to the Williams Otto process with reactions $A + B \rightarrow C$, $B + C \rightarrow P + E$ and $C + P \rightarrow G$. Reactant A is the only prefilled substance and reactant B is continuously fed in. Corresponding adjustable robust dynamic optimization formulation under uncertain inlet temperature $T_{in}(\tau)$ is:

$$\underset{F_{B,in}(t), T_j(t)}{\text{minimize}} \quad \max_{T_{in}(\tau)} \Phi(t_f) \quad (28)$$

$$\frac{dx_A}{dt} = -\frac{F_{B,in}x_A}{1000V} - k_1\eta_1x_Ax_B \quad (29)$$

$$\frac{dx_B}{dt} = -\frac{F_{B,in}(x_B - 1)}{1000V} - k_1\eta_1x_Ax_B - k_2\eta_2x_Bx_C \quad (30)$$

$$\frac{dx_C}{dt} = -\frac{F_{B,in}x_C}{1000V} + 2k_1\eta_1x_Ax_B - 2k_2\eta_2x_Bx_C - k_3\eta_3x_Cx_P \quad (31)$$

$$\frac{dx_P}{dt} = -\frac{F_{B,in}x_P}{1000V} + k_2\eta_2x_Bx_C - \frac{1}{2}k_3\eta_3x_Cx_P \quad (32)$$

$$\frac{dx_E}{dt} = -\frac{F_{B,in}x_E}{1000V} + 2k_2\eta_2x_Bx_C \quad (33)$$

$$\frac{dx_G}{dt} = -\frac{F_{B,in}x_G}{1000V} + \frac{3}{2}k_3\eta_3x_Cx_P \quad (34)$$

$$\frac{dT}{dt} = -\frac{R_1 + R_2 + R_3}{C_p} - L_1T + L_2T_j + \frac{F_{B,in}(T_{in} - T)}{1000V} \quad (35)$$

$$\frac{dV}{dt} = \frac{F_{B,in}}{1000V} \quad (36)$$

$$\frac{d\Phi}{dt} = -5554.1 \left(k_2\eta_2x_Bx_C - \frac{1}{2}k_3\eta_3x_Cx_P \right) V - 251.82k_2\eta_2x_Bx_CV \quad (37)$$

$$0 \leq V(t_f) \leq 5; \quad 60 \leq T(t) \leq 90 \quad (38)$$

$$0 \leq F_{B,in}(t) \leq 5.784; \quad 0.02 \leq T_j(t) \leq 0.1 \quad (39)$$

$$Z = \{T_{in}(\tau) \in \mathbb{R} \mid 32 \leq T_{in}(\tau) \leq 38\} \quad (40)$$

$$x(0) = [1, 0, 0, 0, 0, 0, 65, 2, 0]^T \quad (41)$$

where manipulated variables $F_{B,in}(t)$ and $T_j(t)$ are feeding flowrate and dimensionless jacket temperature limited by Eq.(38). Other necessary parameters can be found in the paper (Assassa and Marquardt, 2014). In this problem, the continuous affine decision rules for $F_{B,in}(t)$ and $T_j(t)$ are defined as:

$$F_{B,in}(t) = F_{B,in}^0(t) + \int_{\tau=0}^{\tau=t} w_1(t, \tau) T_{in}(\tau) d\tau \quad (42)$$

$$T_j(t) = T_j^0(t) + \int_{\tau=0}^{\tau=t} w_2(t, \tau) T_{in}(\tau) d\tau \quad (43)$$

4.1. Optimal continuous affine decision rule

Figure 2 depicts the optimized $w_1(t, \tau)$, $w_2(t, \tau)$, $F_{B,in}^0(t)$ and $T_j^0(t)$, respectively. Slope parameters $w_1(t, \tau)$ and $w_2(t, \tau)$ indicate the sensitivities of manipulated $F_{B,in}(t)$ and $T_j(t)$ to the realized uncertainty. The intercepts $F_{B,in}^0(t)$ and $T_j^0(t)$ represent the values of wait-and-see decisions when uncertain T_{in} equals to zero.

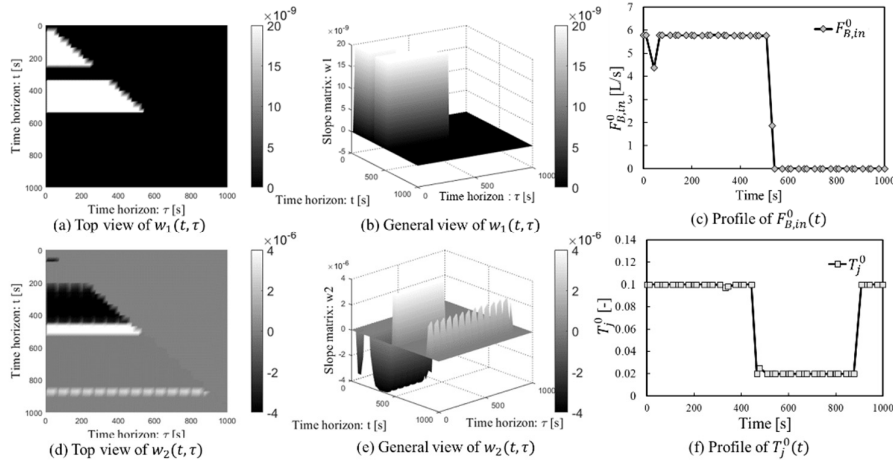


Figure 2. Optimal parameters of continuous affine decision rule

In Figure 2.(a)(b), we have $\forall t, \tau, w_1(t, \tau) \leq 2 \times 10^{-8}$, which suggests that the optimal feeding rate should take fixed actions with respect to uncertain T_{in} . Such phenomenon can be explained that the objective values monotonously decrease with the increment of volume, decided only by feeding rate $F_{B,in}(t)$, thus the reactor should always be fulfilled at terminal time regardless of uncertain feeding temperature profile.

On the other hand, the optimal $w_2(t, \tau)$ shown in Figure 2.(d)(e) can lead to a maximum adjustment at 0.02 for $T_j(t)$ applying Eq.(40). Slope parameter $w_2(t, \tau) \leq 0$ when $t \in [200, 450]$, indicating value of scaled jacket temperature should increase to meet to decrement of uncertainty in these time intervals. The choice of $w_2(t, \tau)$ is consistent with reactor temperature profiles in Figure 3.(a), where reactor temperature T is close to its lower bound in $t \in [200, 450]$. In this time period, decrement of feeding temperature may cause violation to constraint $T \geq 60$, therefore manipulated jacket temperature $T_j(t)$ should be raised to compensate the undesired effect by T_{in} falling. Moreover, the gray upper-triangular zone reflects that current manipulated action is only related to preceding uncertainty realization, while the gray areas in the lower-triangular zone indicates no adjustments are applied in these time periods.

4.2. Benefits evaluation of ARDO solution

In order to examine the benefits of ARDO solution, 200 random scenarios are generated via Eq.(44), where $r(\tau) \in [0, 1]$ is time-varying primitive uncertainty under uniform distribution.

$$T_{in}(\tau) = T_{in}^{LB} + r(\tau)(T_{in}^{UB} - T_{in}^{LB}) \quad (44)$$

The overall computational results are provided in Table 1. Conclusion can be drawn that the proposed ARDO method reaches lower average objective cost than SRO. No violation is observed in 200 random scenarios, demonstrating the robust feasibilities of proposed method.

Table 1. Benefit evaluation results of ARDO framework

Method	SRO	ARDO
Average objective value [€]	-4734.26	-4746.36
Average T_{min} [°C]	61.06	61.03
Constraint violations	0	0

Figure 3.(a) provides detailed profiles of reactor temperature and Figure 3.(b) shows the objective values of 200 uncertain scenarios, where each point stands for a single scenario. In Figure 3.(a), it is indicated that both proposed ARDO method can indeed prevent constraint violation under uncertainty as the reactor temperature profile T^{ARDO} is always between predefined lower bound T^{LB} and upper bound T^{UB} . Furthermore, in Figure 3.(b), significant improvement of objective performance by ARDO method can be seen. Objective costs Φ^{ARDO} is lower than Φ^{SRO} in every scenario, hinting the superiority of ARDO to SRO in achieving better objective performances.

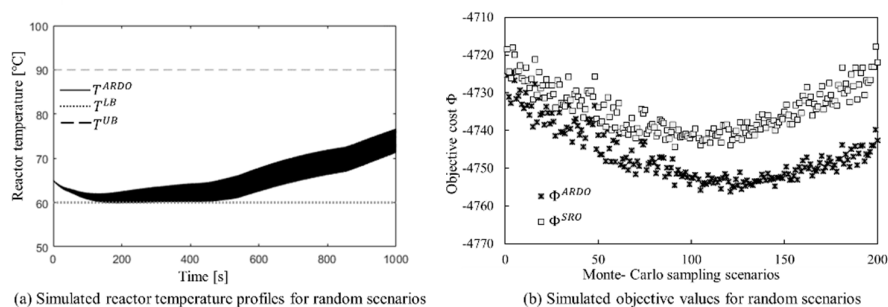


Figure 3. Robust solutions under 200 random scenarios.

5. Conclusion

The adjustable robust dynamic optimization scheme was proposed in this work. A novel continuous affine decision rule was developed to approximate the causal dependence of wait-and-see decision variables on the infinite dimensional uncertainty factors. To the best knowledge, it was the first attempt to employ adjustable robust optimization approach to addressing dynamic optimization of nonlinear chemical processes under uncertainty. Through the application to the classical nonlinear chemical process, when compared to traditional robust optimization approach, the proposed framework led to more flexible control profiles and less conservative solutions.

Acknowledge

We appreciate National Nature Science Foundation of China (Grant No. 21978150) for the financial support for this work.

References

- A. Ben-Tal, A. Goryashko, E. Guslitzer, A. Nemirovski, Adjustable robust solutions of uncertain linear programs, *Mathematica Programming: Series A*, 2004, 99:351-376
- F. Assassa and W. Marquardt, 2014, Dynamic optimization using adaptive direct multiple shooting, *Computers and Chemical Engineering*, 60:242-259.
- M. Diehl, H. G. Bock, E. Kostina, 2006 An approximation technique for robust nonlinear optimization, *Mathematica Programming: Series B*, 2006, 107, 213-230.
- Q. Zhang, M. Morari, I. E. Grossmann, A Sundaramoorthy, J. M. Pinto., 2016, An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load. *Computers & Chemical Engineering*, 2016, 86: 106-119.

Adaptive Optimal Control of Baker's Yeast Fermentation Process with Extreme Learning Machine and Recursive Least Square Technique

Kazeem Alli, Jie Zhang

*Newcastle University, School of Engineering, Merz Court, Newcastle upon Tyne NE1
7RU, UK.*

K.alli2@ncl.ac.uk

Abstract

This paper presents combining extreme learning machine (ELM) and recursive least square (RLS) technique in modelling and optimisation of a fed-batch fermentation process. ELM has some characteristic features of fast training together with better generalisation capability. In order to cope with batch-to-batch variations due to unknown disturbances such as unknown process condition drift, the RLS algorithm is integrated with the ELM to update the output layer weights recursively from batch to batch. The offline trained output layer weights of the ELM are used as the initial parameter estimation in RLS. After updating the ELM model, optimisation is carried out to update the feeding policy for the next batch. The proposed method is applied to a simulated baker's yeast fermentation process and the results obtained shows that the proposed method can cope with unknown disturbance and improve process operation from batch to batch.

Keywords Extreme Learning Machine, Fed-batch Process, Optimization, Recursive Least Square, Optimization Control.

1. Introduction

Mostly in biotechnology and food industries, fermentation process plays vital role in the production of baker's yeast also known as *Saccharomyces cerevisiae*, alcoholic beverages, organic solvents and antibiotics or biopolymers. The importance of these products in food industries cannot be over-emphasised which leads to their highly competitive market value demands and hence, the industrial production of baker's yeast.

Fed-batch processes operation are commonly seen in many fermentation industries. This mode of operation allows the substrate concentration and all other operating variables (such as the feed-rate, temperature, pressure, and agitation rate) to be varied and monitored during the progress of production to obtain maximum and desired biomass yield at the end of production. According to (Xiong and Zhang 2005), the significant requirement of batch process modelling lies in the accurate mathematical representation of the process.

The new trending concept of machine learning coupled with big data analytics has become a widely acceptable concept used in modelling, monitoring and optimisation of

many industrial process operations. Thus, the data-driven modelling concept utilizes statistical theories in establishing process models, monitoring the progress of pre-set conditions of operation for optimization purpose and generalization capabilities in predicting the unforeseen circumstances. In this concept, detailed knowledge of the process operation is not necessary but past historical data of the process operation is required. As much as the historical data is large and representative enough, data-driven modelling and process optimization can be established with machine learning and statistical theories conceptualization (Liu *et al.*, 2019).

In extreme learning machine (ELM), the hidden layer-weights are arbitrarily assigned and fixed without repeatedly adjustment unlike the traditional training approaches for single hidden layer feedforward networks (SLFNs). Parameters to be learnt in ELM are the connections (weights) between the hidden layer and the output layer, which are determined with a one-step regression type approach using Moore-Penrose (MP) generalized inverse operation of the hidden layer output matrices. ELM has been successfully applied to batch process modelling (Alli and Zhang, 2020).

This paper proposes integrating ELM with RLS algorithm in obtaining reliable modelling for batch-to-batch optimal control of the final biomass concentration of a baker's yeast fermentation process. An ELM model is initially developed from historical process operation data which is used to calculate the optimal control policy for the baker's yeast fermentation process. After the completion of each batch, the output layer weights of the ELM model are updated using the data from the newly completed batch through the RLS algorithm. The updated ELM model is then used to optimise the next batch and the procedure is repeated from batch to batch. By using this proposed modelling approach, the ELM model can keep track of any variations that may arise during the progress of the batch process operation.

The rest of this article is organized as follows. Batch-to-batch modelling with recursive updated ELM model is given in Section 2. Section 3 presents application to a fed-batch fermentation process and Section 4 gives the final concluding remarks.

2. Batch-to-Batch Modelling with an Updated ELM Model

2.1. Extreme Learning Machine

An ELM is a single-hidden layer feedforward networks (SLFNs) where the hidden layer weights and bias are assigned randomly (Huang *et al.*, 2006). The output layer usually uses the linear activation function, and the output is calculated as:

$$f_l(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta \quad (1)$$

where $\beta = [\beta_1, \dots, \beta_L]^T$ is a vector of the output layer weights, L is the number of hidden neurons, and $h(x) = [h_1(x), \dots, h_L(x)]$ is a vector of the hidden layer outputs calculated as:

$$h_i(x) = G(a_i \cdot x_j + b_i) \quad (2)$$

where $j = 1, \dots, N$, $a_i = [a_{i1}, a_{i2}, \dots, a_{iN}]^T$ is a vector of weights connecting the i th hidden node to the inputs, b_i is the bias of the i th hidden nodes, x_j is the j th input sample, and G is the hidden layer neuron activation function typically taken as the sigmoid function.

The output layer weights, denoted by β , are obtained by solving a regression type problem as follows:

$$\|H\beta - T\|^2 \quad (3)$$

where H is the hidden layer output matrix and T is the training target matrix.

The optimal solution of Eq. (3) is given by:

$$\hat{\beta} = H^\dagger T \quad (4)$$

where H^\dagger denotes the Moore-Penrose generalised inverse of matrix H .

2.2. Batch-wise Updated ELM Model with RLS

Batch to batch variations exist due to the presence of unknown disturbances. Any variations in the batch process could make the ELM model invalid. To cope with operating condition changes, the ELM can be updated from batch to batch by using the RLS algorithm.

In RLS, the parameter estimation at the current batch is obtained by using the previous parameter estimation and data from the current batch. The correctional term applied to the output layer weights is proportional to the prediction error at the current batch. The updating output layer weights by the RLS algorithm is given as:

$$\beta(k+1) = \beta(k) + F(k).E(k+1) \quad (5)$$

$$F(k) = P(k)H_{k+1}^T [\lambda + H_{k+1}P(k)H_{k+1}^T]^{-1} \quad (6)$$

$$P(k+1) = [I - F(k).H_{k+1}] P(k) / \lambda \quad (7)$$

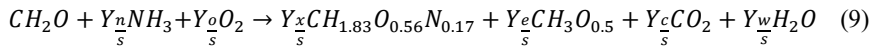
$$P(k) = P(0) = \sum_{i=i_1}^0 [\lambda^{k-i} H(i).H(i)^T]^{-1} \quad (8)$$

where $E(k+1) = y_{k+1} - H_{k+1}\beta(k)$ is the ELM model prediction error at the current batch, $\beta(k)$ is the current parameter estimate (the ELM output layer weights), $F(k)$ is the Kalman gain, H_{k+1} is a vector of the hidden neuron outputs at the current batch, and λ is a forgetting factor.

3. Application to a Fed-Batch Fermentation Process

3.1. Baker's Yeast Fermentation Process

Baker's yeast also known as *Saccharomyces cerevisiae* is one of the various products of culturing yeast which is being produced from glucose as substrate to grow microorganism. The main source of energy in growing this culture is the glucose in molasses for the growth reaction of the yeast cells and energy production. The chemical reaction representing this description is given by (Karakuzu et al., 2006) as:



where CH_2O , $CH_{1.83}O_{0.56}N_{0.17}$, and $CH_3O_{0.5}$ are glucose, biomass, and ethanol concentrations in (g/L) respectively, the coefficient Y_s^i represents the stoichiometric yield coefficients where $i = n, o, x, e, c, w$ and their values can be found in (Karakuzu et al., 2006). The main objective of production is to maximise the biomass production thereby limiting the formation of ethanol as a by-product in the fermentation process.

Ethanol formation can be limited either through limiting the oxygen concentration in the reaction process or overfeeding the substrate concentration to completely oxidise the yeast cells in the fermentor for maximum productivity during fermentation. To limit the amount of ethanol formation, the feed rate needs to be controlled and monitored during the progress of reaction. Detailed cell kinetics and reactor dynamic equations describing the baker's yeast fermentation process are presented in (Yüzgeç *et al.*, 2009; Baron and Zhang, 2017).

Based on the kinetics, cell dynamics and the model parameters from (Baron and Zhang, 2017), simulation algorithm codes were developed in MATLAB to generate process operational data for modelling the fed-batch fermentation process with ELM+RLS algorithms and also determine the optimal final biomass concentration of the process operation through optimal control policies.

3.2. Batch-to-Batch Modelling Using ELM and RLS

The mechanistic model-based simulation is used to produce 100 batches of historical operation data. The batch time is 16.5h and is divided into 10 equal intervals. The model inputs are the substrate feed flow rates randomly generated in the range of 0 to 2500 L/h. The ELM model is of the following form:

$$Y_i(t_f) = f(x_i) \quad (10)$$

where $Y_i(t_f)$ and $x_i = [x_1, x_2 \dots x_{10}]$ are the final biomass concentration and the substrate feed rates of the i th batch respectively. The simulated data were scaled to zero mean and unit variance before parting the data set into training and testing (70% of the data) while the remaining data were used as validation data (30% of the data).

After obtaining the ELM model with the first 20 batches, the output weight and bias obtained were used to initialize the parameter estimation of the RLS algorithm. These weights and biases are updated through the RLS algorithm based on the amount of data in the simulated training data sets.

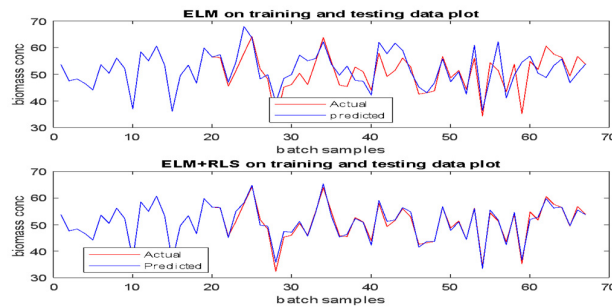


Figure 1. Model prediction performance on training and testing data.

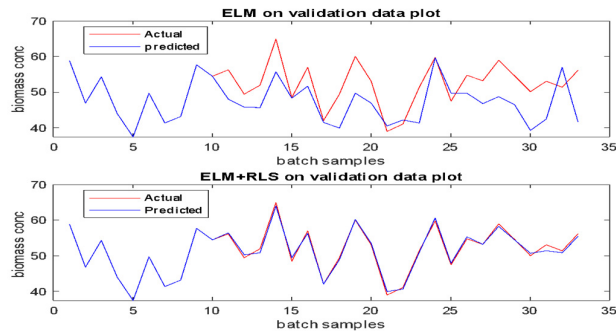


Figure 2. Model prediction performance on validation data

Figure 1 shows the model prediction performance on the training and testing data for both ELM and the proposed method (ELM+RLS). Figure 2 shows the model prediction performance on the validation data for both ELM and the proposed method (ELM+RLS). It can be seen from Figures 1 and 2 how the proposed method tracked the actual data accurately on both training and validation plots. Table 1 shows the mean square error (MSE) of the predictions shown in Figures 1 and 2.

Table 1. Mean Squared Error Values

Models	MSE (Training and Testing)	MSE (Validation)
ELM	0.4783	0.4234
ELM+RLS	0.0257	0.0039

3.3. Batch-to-Batch Optimization Control

The proposed adaptive optimal control of the baker's yeast fermentation is based on the ELM+RLS model prediction given by Eq. (10). The objective is to maximise the final biomass concentration by modifying the substrate feed rate subject to process constraints given below where the optimisation problem can be represented as:

$$J = -f_{ELM+RLS}(F)$$

$$\text{subject to: } \{0 \leq F_i \leq 3000, i = 1, 2, \dots, 10, V_f \leq 10000\} \quad (11)$$

where $f_{ELM+RLS}(F)$ is the ELM+RLS predicted final biomass concentration and F is the vector of substrate feeding rates. The constrained optimisation problem in Eq. (11) was solved with the "fmincon" function in the MATLAB 2019a Optimization Toolbox using interior-point method and random selection of a batch from 100 initial substrate feeding rate in modelling simulation as the starting point.

Table 2. Optimal control results on final biomass concentration

Models	Model final biomass concentration predicted (g/L)	Actual final biomass concentration (g/L)
--------	---	--

ELM	58.13	40.79
ELM+RLS	65.39	53.55

The optimisation results given in Table 2 substantiate the improvement in model accuracy and reliability of the proposed method. The optimal control policy obtained using the ELM+RLS model gives much better performance than that using the ELM model, especially when applied to the actual process (i.e., mechanistic model simulation).

4. Conclusions

An adaptive optimal control strategy of a fed-batch fermentation process is proposed in this paper through integrating ELM and RLS technique. Historical process operation data is used to establish an initial ELM model and in order to cope with the presence of unknown disturbances, the ELM model is updated after each batch using RLS. The updated ELM with RLS is used to find the optimal control policy for the next batch. The proposed batch-wise modelling and optimisation control strategy is demonstrated on a simulated fed-batch fermentation process for producing baker's yeast. It has been shown that the proposed adaptive optimal control strategy can effectively overcome the effect of unknown disturbances and achieve improved product output operation from batch to batch.

Acknowledgement

The work is supported by Petroleum Technology Development Fund (PTDF-Nigeria) through a PhD scholarship.

References

- Alli, K. and J. Zhang (2020). Adaptive Modelling of Fed-batch Processes with Extreme Learning Machine and Recursive Least Square Technique. *ICAART* (2).
- Baron, C. M. C. and J. Zhang (2017). Reliable on-line re-optimization control of a fed-batch fermentation process using bootstrap aggregated extreme learning machine. *Informatics in Control, Automation and Robotics*, (eds.) O. Gusikhin and K. Madani, Springer, 2020, 272–294.
- Huang, G. B., Q. Y. Zhu, and C. K. Siew (2006). Extreme learning machine: Theory and applications, *Neurocomputing*, 70(1), pp. 489-501.
- Karakuzu, C., M. Türker, S. Öztürk (2006). Modelling, on-line state estimation and fuzzy control of production scale fed-batch baker's yeast fermentation. *Control Engineering Practice*, 14(8): 959-974.
- Liu, X., L. Liu, L. Wang, Q. Guo, and X. Peng (2019). Performance sensing data prediction for an aircraft auxiliary power unit using the optimized extreme learning machine. *Sensors*, 19(18): 3935.
- Xiong, Z. and J. Zhang (2005). A batch-to-batch iterative optimal control strategy based on recurrent neural network models. *Journal of Process Control*, 15(1): 11-21.
- Yüzgeç, U., M. Türker, and A. Hocalar (2009). On-line evolutionary optimization of an industrial fed-batch yeast fermentation process. *ISA Transactions*, 48(1): 79-92.

Data-based Industrial Soft-sensor Design via Optimal Subset Selection

Martin Mojto^a, Karol Lubušký^b, Miroslav Fikar^a and Radoslav Paulen^a

^a*Slovak University of Technology in Bratislava, Bratislava 81237, Slovakia*

^b*Slovnaft, a.s., Bratislava 82412, Slovakia*

Abstract

Inferential (or soft) sensors are used in industry to infer the values of imprecisely and rarely measured (or completely unmeasured) variables from variables measured online (e.g., pressures, temperatures). The main challenge, akin to classical model overfitting, in designing an effective inferential sensor is to select a correct structure represented by the number of sensor inputs. This work is focused on the design of an inferential sensor for bottom product composition of an industrial distillation column. We study effectiveness of various subset selection methods that regard different model-overfitting criteria. Our results show that the subset selection is a viable methodology to sensor design and that we are able to improve accuracy of the current refinery sensor by around 15 %.

Keywords: Inferential (Soft) Sensors, Process Monitoring, Subset Selection

1. Introduction

The accuracy and reliability of industrial measurements have a huge impact on the effectiveness of industrial process control (Khatibisepehr et al., 2013). Especially, the control performance of advanced process controllers (Qin and Badgwell, 2003) is highly related to the indication quality of controlled variables (CVs). It is often the case that the crucial CVs (e.g. distillate purity) are too expensive or impossible to measure at the frequency required for an effective feedback control. This gave rise to a use of so-called inferential (or soft) sensors (Mejdell and Skogestad, 1991; Kordon et al., 2003; Curreri et al., 2020).

The purpose of an inferential sensor is to infer the CV value (output) using the data from other measured variables (inputs). The design procedure aims at a) identifying a sensor structure and b) at estimating the sensor parameters. While the latter problem can be solved relatively easily, the former issue of structure selection can be much more challenging in practice.

The focus of the paper is on a class of subset selection (SS) methods. These methods use mixed-integer programming to determine the optimal structure of an inferential sensor using various model-overfitting criteria such as adjusted R^2 (R_{adj}^2), corrected Akaike information criterion (AIC_c), Bayesian information criterion (BIC), and cross-validation. We make a comparison of effectiveness of the SS methods investigating a

linear soft-sensor design for a depropanizer column in an industrial fluid catalytic cracking (FCC) unit.

2. Problem Description

Our goal is to identify models of inferential sensors of the following linear form:

$$y = m \left(a_1, a_2, \dots, a_{n_p} \right)^T = ma, \quad (1)$$

where y stands for the desired CV inferred by the sensor, m is the vector of available input variables, and $a \in \sim^{n_p}$ represents the vector of sensor parameters.

2.1. Industrial FCC unit

We study a depropanizer that is a part of an FCC unit of the refinery Slovnaft, a.s. in Bratislava, Slovakia. The column separates a seven-component feed to a C3-fraction-rich distillate and to a C4/C5-fraction-rich bottom product. Plant description is given in Mojto et al. (2020). The candidate input vector for inferring the bottom impurity is:

$$m = \left(F, R, Q_B, p_D, p_B, T_D, T_B, T_{10}, T_{37}, \frac{R}{F}, \frac{Q_B}{F} \right)^T, \quad (2)$$

with feed flowrate F , reflux flowrate R , reboiler heat duty Q_B , pressure at the top of the column p_D , pressure at the bottom of the column p_B , and temperatures of distillate T_D , at the 10th tray T_{10} , at the 37th tray T_{37} and at the bottom T_B . This set ($n_p = 11$) involves all variables measured directly at the column and their commonly used fractions.

Any use of a thermodynamic model to monitor top/bottom stream compositions is prohibitive here, even under some ideality assumptions. This occurs as there are too many degrees of freedom for a seven-component mixture that cannot be inferred from online data. Current inferential sensor (denoted as ref) in use in the refinery is designed according to King (2011) and uses p_B , T_{37} , and Q_B/F as inputs.

3. Soft-sensor Design by Optimal Subset Selection

This section introduces the optimal SS methods for soft-sensor design. An effective design procedure usually requires splitting the available dataset with n measurement points ($M := (m_1^T, m_2^T, \dots, m_n^T)^T, Y := (y_1, y_2, \dots, y_n)^T$) into the following subsets: dataset for sensor design that contains training data ($M(T), Y(T)$) and dataset used for the performance evaluation of designed sensors that contains testing data ($M(S), Y(S)$). Here T and S denote the corresponding row-selection operators.

3.1. Optimal Subset Selection with Model-overfitting Criteria

Subset selection denotes a class of methods that explicitly seek for the simplest possible sensor structures such that some model-overfitting criterion $J(a, z)$ is minimized (Miyashiro and Takano, 2015). Here the variable z denotes a vector with binary entries $z \in \{0, 1\}^{n_p}$ signifying selection of i^{th} input into sensor structure. Correspondingly, the sum of the vector entries $\sum_{i=1}^{n_p} z_i = 1^T z$ denotes the sensor complexity.

Optimal subset selection solves the following bi-level program (Bertsimas et al., 2016):

$$\min_{a, z \in \{0,1\}^{n_p}} J(a, z) \tag{3a}$$

$$\text{s.t. } a \in \arg \min_{\tilde{a}} \| Y(T) - M(T)\tilde{a} \|_2^2 \quad \text{s.t. } -\bar{a}z_i \leq \tilde{a}_i \leq \bar{a}z_i, \forall i \in \{1, \dots, n_p\}. \tag{3b}$$

where \bar{a} represents an upper bound on $\| a \|_\infty$ to be tuned and the optimization criterion $J(\cdot)$ might take the form (RSS := $\| Y(T) - M(T)a \|_2^2$):

$$J_{R_{\text{adj}}}^2 = \frac{\text{RSS}}{n-1^T z-1} \quad \text{or} \quad J_{\text{AICc}} = 2^T z + \log^n \frac{\text{RSS}}{n} \quad \text{or} \quad J_{\text{BIC}} = \log^{Tz}(n) + \log^n \frac{\text{RSS}}{n}. \tag{4}$$

The bi-level program (3) can be effectively solved by standard MIQP solvers using big-M reformulation as shown in Takano and Miyashiro (2020).

3.2. Optimal Subset Selection with Cross-Validation Criterion

The principle of this method is to mimic a standard cross-validation procedure within the training dataset. Let us divide the training data into K smaller subsets N_k , such that:

$$T = \bigcup_{\forall k \in \{1, \dots, K\}} N_k, \quad N_k \cap N_{k'} = \emptyset, \quad \forall k \neq k', \quad K \geq 2. \tag{5}$$

The data is distributed into training (T_k) and validation (V_k) sets as follows:

$$V_k := N_k, \quad T_k := T \setminus N_k, \quad \text{card}(T_k) \geq n_p, \quad \forall k \in \{1, \dots, K\}. \tag{6}$$

where V_k sets contain unique data, while the different T_k sets involve recurring measurements. The SS with cross-validation solves (Takano and Miyashiro, 2020):

$$\min_{a^{(k)}, \forall k \in \{1, \dots, K\}, z \in \{0,1\}^{n_p}} \sum_{k=1}^K \| Y(V_k) - M(V_k)a^{(k)} \|_2^2 \tag{7a}$$

$$\text{s.t. } \forall k \in \{1, \dots, K\}; a^{(k)} \in \arg \min_{\tilde{a}} \| Y(T_k) - M(T_k)\tilde{a} \|_2^2 \tag{7b}$$

$$\text{s.t. } -\bar{a}z_i \leq \tilde{a}_i \leq \bar{a}z_i, \forall i \in \{1, \dots, n_p\}. \tag{7c}$$

The problem (7) can be solved for several values of K — considering constraints on parameter identifiability, i.e., the cardinality condition in Eq. (6) — and for different randomly generated distributions of data into T_k and V_k sets. The structure of the resulting sensor is then given by the most frequent inputs occurring in the calculated sensors. Once the optimal sensor structure is calculated, a least-squares fitting of such model is used with the entire training dataset to determine the parameters of designed soft-sensor. Similarly to problem (3), the problem (7) can be effectively resolved by standard MIQP solvers.

4. Results

Industrial data available from the refinery represents more than two years of production. We possess 177 lab measurements of the bottom product composition.

We use MATLAB, Yalmip (Löfberg, 2004), and Gurobi (Gurobi Optimization LLC, 2020) to solve various instances of the problems (3) and (7). When determining the best sensor according to SS with cross-validation, we take a median of $1^T z$ from the results

Table 1: Comparison of the number of inputs (n_p^*) and sensor accuracy (RMSE) for the soft sensors designed using time series data.

	R_{adj}^2	AIC_c	BIC	Cross-validation	ref
n_p^*	9	4	4	4	3
RMSE	0.110	0.106	0.106	0.106	0.128

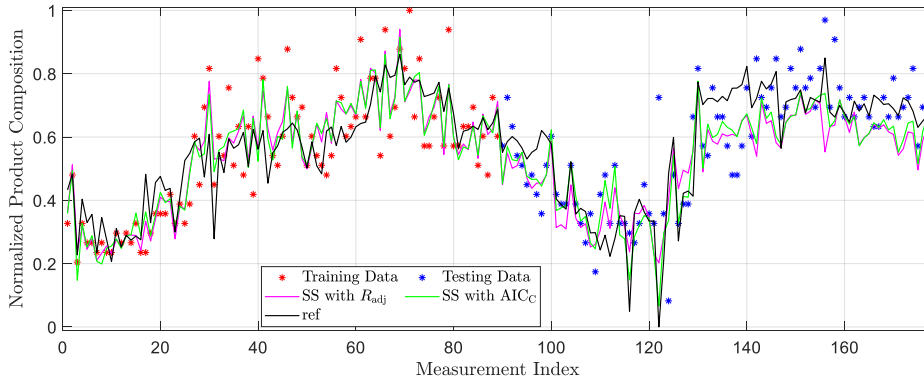


Figure 1: Comparison of the soft sensors designed using time series data.

of different runs (different $K \leq 6$ and data distribution) to obtain the $n_p^* \leq n_p$, i.e., the number of inputs of the final sensor. We then select the n_p^* most frequent inputs from the results of different runs to finalize the sensor structure.

4.1. Design of an Inferential Sensor using Time Series Data

We chronologically assign first 50 % of the data to the training set and the last 50 % of the data to the testing set. The accuracy of the designed sensors is assessed by root mean squared error (RMSE) evaluated on the testing dataset.

Table 1 shows the obtained results. The SS with R_{adj}^2 suggests to include almost all available inputs (except T_D and Q_B) and the accuracy of this sensor is slightly decreased compared to other designed soft sensors. Therefore we conclude that this criterion is not appropriate option for structure selection of an soft sensor. The performance of SS using AIC_c , BIC and cross-validation is the same. These methods suggest to include four common variables (T_B , T_{10} , T_{37} and Q_B/F) into the structure of the soft sensor. Suitable candidates for a good quality soft-sensor thus seem to be temperatures in the column (T_{10} and T_{37}) and variables measured near to the inferred variable (T_B and Q_B/F).

The inferential sensors designed via SS with AIC_c , BIC, and cross-validation show better performance compared to the reference inferential sensor. We can thus conclude that an improvement of the current inferential sensor is possible with only slight modifications, i.e., at least one extra variable in the inferential sensor is required. The accuracy

improvement achieved by the inferential sensor of SS with AIC_c , BIC and cross-validation is more than 15 % compared to the reference inferential sensor.

Table 2: Comparison of the number of inputs (n_p^*) and sensor accuracy (RMSE) for the soft sensors designed using randomly distributed data.

	R_{adj}^2	AIC_c	BIC	Cross-validation	ref
n_p^*	8	6	5	5	3
RMSE	0.107	0.107	0.109	0.111	0.127

Figure 1 shows a comparison of the measured data with the output of the sensors. We plot the predictions of the reference sensor and of the sensors designed by SS with R_{adj}^2 and SS with AIC_c (the same as the rest of SS-based sensors). The performance of the designed sensors on the training data is good as can be expected, despite we can clearly observe problems of the reference sensor in fitting the data. This already suggests its inappropriate structure. This is further documented when looking at the testing data, where the quality of the reference sensor rapidly deteriorates once leaving the training-data window. The last period of the testing data (measurements 130–177) shows a significant discrepancy between the measurements and values inferred by all the designed sensors. This might be caused by a major change in the operating conditions of the FCC unit. A possible remedy could be to design a new sensor (with different structure) or a simple bias correction, which seems to be more appropriate in this case. The bias correction strategy is actually used at the refinery to improve the reference sensor.

4.2. Design of an Inferential Sensor using Randomly Distributed Data

To further investigate the problem, we study the impact of the training/testing data distribution on the sensors performance. Therefore, we randomly assign 50 % of the available data to the training set and leave the rest of the data for testing.

Table 2 shows the results averaged over fifty different randomly generated training/testing data distributions. These results show slightly increased complexity of the inferential sensors designed by SS with R_{adj}^2 and AIC_c compared to inferential sensors designed by SS with BIC and cross-validation. Nevertheless, these SS approaches suggest five common variables (Q_B , T_B , T_{10} , T_{37} and Q_B/F) into the structure of the soft sensor.

The sensors accuracy (see Table 2) confirms better performance of the inferential sensors designed via SS compared to the current inferential sensor. However, only SS with R_{adj}^2 improved its accuracy compared to the sensors in Sec. 4.1. Therefore, the overall improvement of the designed soft sensors is comparable with Sec. 4.1 (around 15 %).

In comparison to the results shown in Fig. 1, the performance of the designed inferential sensors using random distributed data (Fig. 2) is slightly improved in the section represented by the measurements 130–177. Nevertheless, the performance of the reference inferential sensor is almost the same as in Sec. 4.1. Therefore, we can conclude that the simple structure of the reference inferential sensor provides the robustness and constant accuracy at the whole time interval. On the other hand, the designed inferential sensors with enhanced structure (more input variables) are more accurate than the reference inferential sensor, but only within a short time horizon. We thus conclude that

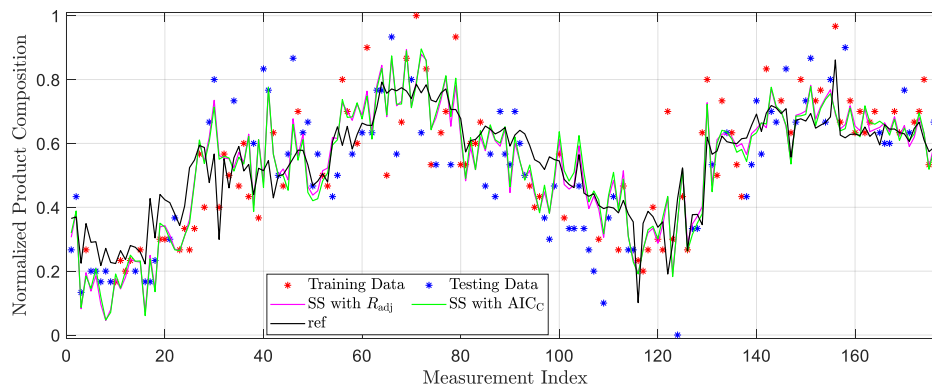


Figure 2: Comparison of the soft sensors designed using randomly distributed data.

a design of one efficient global sensor at the whole time interval is too complicated or impossible and one should better design a family of switching sensors or an appropriate mechanism for update of sensor parameters (beyond the simple bias correction). These are the directions for our further research.

Conclusions

We analyzed effectiveness of optimal subset selection to design a soft sensor. We used several variants of the SS method with different model-overfitting criteria and with cross-validation. According to the time series data, the use of AIC_C , BIC and cross-validation results in better performing sensors than if R_{adj}^2 is used. The designed soft sensors via SS could improve the current soft sensor by around 15 %. Further investigations revealed that any further improvements would be possible using a set of switching sensors.

References

- D. Bertsimas, A. King, R. Mazumder, 2016. Best subset selection via a modern optimization lens. *Ann. Statist.* 44 (2), 813–852.
- F. Curreli, S. Graziani, M. G. Xibilia, 2020. Input selection methods for data-driven soft sensors design: Application to an industrial process. *Information Sciences* 537, 1–17.
- Gurobi Optimization LLC, 2020. Gurobi optimizer reference manual. URL <http://www.gurobi.com>.
- S. Khatibisepehr, B. Huang, S. Khare, 2013. Design of inferential sensors in the process industry: A review of Bayesian methods. *Journal of Process Control* 23, 1575–1596.
- M. King, 2011. *Process Control: A Practical Approach*. John Wiley & Sons Ltd.
- A. Kordon, G. Smits, A. N. Kalos, E. Jordaan, 2003. Robust soft sensor development using genetic programming. *Data Handling in Science and Technology* 23, 69–108.
- J. Löfberg, 2004. Yalmip: A toolbox for modeling and optimization in MATLAB. In: *Proceedings of the CACSD Conference*. Taipei, Taiwan.
- T. Mejdell, S. Skogestad, 1991. Composition estimator in a pilot-plant distillation column using multiple temperatures. *Industrial & Engineering Chemistry Research* 30 (12), 2555–2564.
- R. Miyashiro, Y. Takano, 2015. Mixed integer second-order cone programming formulations for variable selection in linear regression. *European Journal of Operational Research* 247, 721–731.
- M. Mojto, K. Ľubušký, M. Fikar, R. Paulen, 2020. Advanced process control of an industrial depropanizer column using data-based inferential sensors. Vol. 48 of *Computer Aided Chemical Engineering*. Elsevier, pp. 1213–1218.
- S. J. Qin, T. A. Badgwell, 2003. A survey of industrial model predictive control technology. *Control Engineering Practice* 11, 733–764.
- Y. Takano, R. Miyashiro, 2020. Best subset selection via cross-validation criterion. *TOP* 28, 475–488.

Economic MPC with Modifier Adaptation using Transient Measurements

Erika Oliveira-Silva^{a,b,*}, Cesar de Prada^{a,b}, Daniel Navia^c

^a*Department of Systems Engineering and Automatic Control, School of Industrial Engineering, University of Valladolid, Dr. Mergelina s/n, 47011, Valladolid, Spain*

^b*Institute of Sustainable Process, Dr. Mergelina s/n, 47011, Valladolid, Spain*

^c*Dpto. Ingeniería Química y Ambiental, Universidad Técnica Federico Santa María, Avd. Vicuña Mackenna, Campus San Joaquín, Santiago, Chile*

erika.oliveira@autom.uva.es, prada@autom.uva.es, daniel.navia@usm.cl

Abstract

This paper presents an eMPC controller that integrates Modifier Adaptation (MA) methods in order to achieve in closed loop the real optimum in spite of the presence of model-process structural mismatch. The main difficulty associated with MA is the correct process gradient estimation that usually requires waiting for several process steady states. Then in a large settling time process, the time required for computing corrective actions with MA can be so large that the method becomes impractical. To avoid these problems, this work presents an eMPC+MA that estimates process gradients using transient data and an on-line identification algorithm (TMA). The integrated eMPC+TMA method is applied to the Williams-Otto reactor. The results show that eMPC+TMA can approach the plant's real steady state optimum despite process-model mismatch without using steady state measurements in a sensible period of time.

Keywords: Real-time Optimization, Modifier Adaptation, Uncertainty, Transient Measurements, MPC.

1. Introduction

Real-Time Optimization (RTO) is a largely used technique to increase the industry competitiveness, fulfilling quality, environmental, and security demands. RTO is formulated as an optimization problem with economic objectives that use explicit process models to calculate the optimal decision variables to be used as setpoints of low-level controllers, commonly MPC (Model Predictive Control). RTO generally optimizes an economic cost function ϕ with respect to the decision variables or inputs \mathbf{u} and constraints \mathbf{g} using a steady state model as in (1).

$$\begin{aligned} & \text{Min}_{\mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U} \phi(\mathbf{u}, \mathbf{y}) \\ & \text{s.t. } \mathbf{g}(\mathbf{u}, \mathbf{y}) \leq 0 \end{aligned} \quad (1)$$

In case of parametric and/or structural model-process mismatch, (1) will not provide the correct process optimum. So modifier adaptation methodology (MA) can be applied to (1) in order to match model and real plant optimum (Marchetti et al., 2009). MA adds modifiers to the cost function and constraints, resulting in problem (2).

$$\begin{aligned} \min_{\mathbf{u}^L \leq \mathbf{u} \leq \mathbf{u}^U} \phi_{MA} &= \phi(\mathbf{u}) + \boldsymbol{\lambda}_k^T (\mathbf{u} - \mathbf{u}_{k-1}^*) \\ \text{s.t. } \mathbf{g}_{MA}(\mathbf{u}) &= \mathbf{g}(\mathbf{u}) + \boldsymbol{\gamma}_{k,i}^T (\mathbf{u} - \mathbf{u}_{k-1}^*) + \boldsymbol{\varepsilon}_k \leq 0 \end{aligned} \quad (2)$$

Where ϕ_{MA} is the modified cost function and \mathbf{u}_{k-1}^* are the optimum values of the decision variables in the previous steady state. $\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\varepsilon}$ are the modifiers computed as in **Error! Reference source not found.**, using the information of the plant (subscript P) and model.

$$\boldsymbol{\lambda}_k^T = \left. \frac{\partial \phi_P}{\partial \mathbf{u}} \right|_{\mathbf{u}_{k-1}^*} - \left. \frac{\partial \phi}{\partial \mathbf{u}} \right|_{\mathbf{u}_{k-1}^*}, \boldsymbol{\gamma}_k^T = \left. \frac{\partial \mathbf{g}_P}{\partial \mathbf{u}} \right|_{\mathbf{u}_{k-1}^*} - \left. \frac{\partial \mathbf{g}}{\partial \mathbf{u}} \right|_{\mathbf{u}_{k-1}^*}, \boldsymbol{\varepsilon}_k = \mathbf{g}_P(\mathbf{u}_{k-1}^*) - \mathbf{g}(\mathbf{u}_{k-1}^*) \quad (3)$$

The RTO+MA problem (2) is a static optimization that uses a different model than the model predictive controller (MPC). To avoid compatibility problems and benefit from considering dynamics, it is logical to think of an integration of these two layers where the economic problem of the RTO+MA could be used as an objective function of an economic MPC (eMPC). Another point to have in mind in this integration is the process gradient estimation. In traditional MA, process gradients in **Error! Reference source not found.** are estimated using steady state measurements, requiring quite a few steady states to the problem (2) reach the optimum. Therefore, when dealing with slow dynamics processes, as often happens in practice, the time required for computing corrective actions with RTO+MA can be so large that the method becomes impractical. Another problem is that MPC controllers usually operates in a time scale of seconds or minutes and they constantly interact with the system making a traditional MA impractical. So, an estimation of process gradients using transient measurements is required.

Recent works have presented proposals to solve the RTO+MA+MPC problem as in Vaccari and Pannocchia (2016) eMPC, Pannocchia (2018), Vaccari and Pannocchia (2018), Hernández and Engell (2019), Faulwasser and Pannocchia (2019), and Vaccari et al. (2020). Most of these papers maintain two layers inside the controller: a static economic optimization problem and a target optimization problem. Faulwasser and Pannocchia (2019) is an exception. All the authors mentioned show different frameworks to the unification eMPC+MA and some of them apply identification algorithms or other methods to estimate the processes gradients using steady or transient measurements.

To contribute to the practical applications of MA, the present paper proposes a different method, using transient measurements and an identification framework, to estimate process gradients to an eMPC+MA controller. This new architecture will be called eMPC+TMA. The paper is organized as follows. Section 2 describes the algorithm to estimate gradients with transient data. Then, the structure of the eMPC is presented and eMPC+TMA is applied to the benchmark example of William Otto reactor in Section 4. Finally, in the last section some conclusions are presented.

2. Modifier Adaptation using transient measurements (TMA)

One possible approach to estimate process gradient using transient information is to approximate the variation of the process cost function using a second order Taylor's expansion with an adaptative estimation technique (Navia et al., 2017; Rodríguez-Blanco et al., 2017). In this paper, this idea is expanded and integrated into the context of eMPC assuming that the process cost depends not only on previous decisions but also on its time dynamics. Then, a second order Taylor's expansion will lead to:

$$\begin{aligned} \Delta\phi_{P,k-1} &= \frac{\partial\phi_P}{\partial u_{k-1}}\Delta u_{k-1}^T + \frac{\partial\phi_P}{\partial u_{k-2}}\Delta u_{k-2}^T + \frac{\partial^2\phi_P}{\partial u_{k-1}^2}\frac{1}{2}\Delta u_{k-1}\Delta u_{k-1}^T \\ &+ \frac{\partial^2\phi_P}{\partial u_{k-1}u_{k-2}}\Delta u_{k-1}\Delta u_{k-2} + \frac{\partial^2\phi_P}{\partial u_{k-2}^2}\frac{1}{2}\Delta u_{k-2}\Delta u_{k-2}^T + \eta\frac{\partial J}{\partial t}\Delta t \end{aligned} \quad (4)$$

Now, if we define two vectors φ and θ as in (5), (6), then, (4) can be written as (7). Notice that (5) is a vector of the changes in the decision variables and the derivative of the cost function in the past iterations. The derivative with respect to time can be approximated using past data and Nordsieck's vector (Butcher, 2002). In the same way, (6) is a vector of unknown values that contains the gradients of the process cost function w.r.t. the decision variables in the two first positions. So (7) has the form of a typical model used in parameter identification problems.

$$\begin{aligned} \varphi_{k-1}^T &= \left[\Delta u_{k-1}^T \quad \Delta u_{k-2}^T \quad \frac{1}{2}\Delta u_{k-1}\Delta u_{k-1}^T \quad \Delta u_{k-1}\Delta u_{k-2}^T \quad \frac{1}{2}\Delta u_{k-2}\Delta u_{k-2}^T \quad \frac{\partial\phi_P}{\partial t}\Delta t \right] \end{aligned} \quad (5)$$

$$\hat{\theta}_{k-1}^T = \left[\frac{\partial\phi_P}{\partial u_{k-1}} \quad \frac{\partial\phi_P}{\partial u_{k-2}} \quad \frac{\partial^2\phi_P}{\partial u_{k-1}^2} \quad \frac{\partial^2\phi_P}{\partial u_{k-1}u_{k-2}} \quad \frac{\partial^2\phi_P}{\partial u_{k-2}^2} \quad \eta \right] \quad (6)$$

$$\Delta\hat{\phi}_{k-1} = \varphi_{k-1}^T\hat{\theta}_{k-1} \quad (7)$$

In our case, the normalized least mean square algorithm, NLMS is used to minimize the module of difference of the current estimate of \square and the real one (Isermann and Münchhof, 2011). The NLMS algorithm is easy to implement and it is computationally less expensive than other recursive methods (8).

$$\hat{\theta}_k = \hat{\theta}_{k-1} + \mu \frac{[\Delta\phi_k - \varphi_k^T\hat{\theta}_{k-1}]}{\sigma + \|\varphi_k^T\|^2}, 0 < \mu < 2 \quad (8)$$

3. eMPC+TMA formulation

The eMPC+MA is formulated in a single layer using an economic objective function composed of three terms (9). The first term, $\phi(t_{pred})$, corresponds to an economic target computed from the continuous dynamic model $F(\dot{x}, x, u, v_k)$ at the end of the prediction horizon has a typical RTO cost function. This implies that $\phi(t_{pred})$ depends on the expected dynamic of the system, control moves, and disturbances. The second term in (9) is a MA type modifier term that modifies the objective function to match the NCO of the real plant when it reaches steady state. The term is equivalent to the first order modifier in the traditional MA formulations (2). The term \square_k is computed every sampling time k using the process gradient estimated using transient measurements and model gradients as in **Error! Reference source not found.** and then filtered. Similar formulations apply to the constraints modifiers, \mathbf{y}_k and $\boldsymbol{\varepsilon}_k$. Finally, the third term of (9), with a matrix weighting factor \mathbf{Q} , penalizes changes in the manipulated variables (move suppression parameter) over the control horizon, increasing stability, and contributes to model adequacy and convexification (François and Bonvin, 2013).

An overall view of the proposed eMPC+TMA architecture is presented in Figure 1. The MHE module is a moving horizon estimator responsible for estimating states and disturbances for the controller (Rawlings et al., 2019). In each iteration k that could correspond to each sample time of the controller, the measurements \mathbf{y}_k and the value of manipulated variables \mathbf{u}_{k-1} are used to estimate the process states, $\hat{\mathbf{x}}_k$, and disturbance

\mathbf{v}_k in MHE module. These values, together with process measurements $\phi_{p,k}$, $\mathbf{g}_{p,k}$ are used to estimate the process gradients that besides model gradient are used to calculate the modifiers as in Section 2. The modifiers calculated in TMA are filtered and used in eMPC+MA module, problem (9),(10) resulting in the vector of control moves to be applied in the process.

$$\min_{\mathbf{u}_{k+i}} \phi(t_{pred}) + \mathbf{Q}_k^T(\bar{\mathbf{u}} - \mathbf{u}_{k-1}) + \sum_{i=0}^{n_u-1} \mathbf{Q}_{k+i}^T \mathbf{Q}_{k+i} \mathbf{u}_{k+i} \quad (9)$$

$$\begin{aligned} & s. t. \mathbf{g}(\mathbf{u}) + \mathbf{Y}_k^T(\bar{\mathbf{u}} - \mathbf{u}_{k-1}) + \mathbf{Q}_k \leq \mathbf{0}, \forall t \in [t_k, t_{pred}] \\ & \mathbf{u}_{k+i} = \mathbf{u}_{k+i-1} + \Delta \mathbf{u}_{k+i}, \quad i = 0, 1, \dots, n_{pred} - 1 \\ & \mathbf{u}^L \leq \mathbf{u}_{k+i} \leq \mathbf{u}^U, \quad i = 0, 1, \dots, n_{pred} \\ & \mathbf{u}(t) = \mathbf{u}_{k+i}, \quad t \in [t_{k+i}, t_{k+i+1}], \quad i = 0, 1, \dots, n_{pred} - 1 \\ & \Delta \mathbf{u}_{k+i} = \mathbf{0}, \quad i = n_u, \dots, n_{pred} - 1 \\ & \mathbf{x}(t_k) = \hat{\mathbf{x}}_k \\ & \mathbf{F}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, \mathbf{v}_k) = \mathbf{0}, \forall t \in [t_k, t_{pred}] \end{aligned} \quad (10)$$

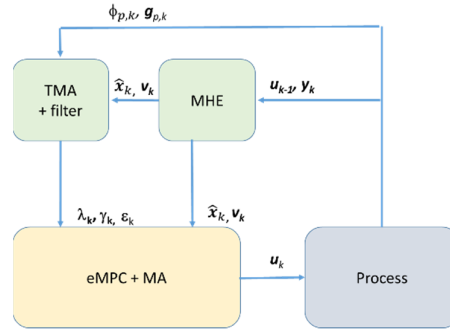


Figure 1: eMPC+TMA architecture.

4. Williams-Otto Case study

The benchmark example (Williams and Otto, 1960) is used to validate the proposed eMPC+TMA approach. The process consists of a continuous stirred tank reactor CSTR where the reactants A and B combine to generate four species C, E, G, P in three different reactions. The eMPC uses a simplified dynamic model that differs in structure and parameters from the process. Only two reactions and 5 components A, B, E, G, P are considered. The nonlinear dynamic models used to represent the real process and the controller dynamic model consist of mass balance equations for each component (Faulwasser and Pannocchia, 2019). The eMPC+TMA problem to be solved is (11). The initial states and disturbances in k are provided by the MHE module. The modifiers' values are calculated by the TMA module and then filtered with a first order filter.

The reactor simulation and the dynamic optimization problem in MHE are formulated in a continuous domain in MATLAB. The eMPC was implemented using the *nlmpc* object with the custom cost function (11). The optimization problems are solved using the NLP solver *fmincon*. The eMPC+TMA framework executes every 2 minutes. The control/prediction horizons are $n_u = 3$, $n_{pred} = 30$, MA filter $K = 0.7$ and move suppression $\beta = (0.5, 0.5)$. The NLMS parameters, $\mu = 0.1$ and σ is a small value to avoid division by zero ($\sigma = 0.0001$ for example). The first 8 minutes of data are

collected to be used in the identification algorithm. Then the controller starts at minute 10.

$$\begin{aligned}
 \max_{u=[F_B, T]} \phi &= \bar{F}_R(\bar{X}_P p_P + \bar{X}_E p_E) - F_A X_{A0} p_A - \bar{F}_B X_{B0} p_B + \lambda_{1,k} \sum_{i=0}^{n_u-1} \Delta F_{B,k+i} \\
 &+ \lambda_{2,k} \sum_{i=0}^{n_u-1} \Delta T_{k+i} + \beta_1 \sum_{i=1}^{n_u-1} (\Delta F_{B,k+i})^2 + \beta_2 \sum_{i=1}^{n_u-1} (\Delta T_{k+i})^2 \\
 \text{s.t. nonlinear dynamic model} \\
 \square F_{B,k+i} &= F_{B,k+i} - F_{B,k+i-1}, \quad i = 0 \dots n_{pred} - 1 \\
 \square T_{k+i} &= T_{k+i} - T_{k+i-1}, \quad i = 0 \dots n_{pred} - 1 \\
 F_B(t) &= F_{B_{k+n_u}}, \quad t \in [t_{k+n_u}, t_{pred}] \\
 T(t) &= T_{k+n_u}, \quad t \in [t_{k+n_u}, t_{pred}] \\
 180 \frac{l}{\text{min}} &\leq F_{B,k+i} \leq 360 \frac{l}{\text{min}}, \quad i = 0 \dots n_{pred} - 1 \\
 75^\circ\text{C} &\leq T_{k+i} \leq 100^\circ\text{C}, \quad i = 0 \dots n_{pred} - 1
 \end{aligned} \tag{11}$$

Figure 2 shows the results of the eMPC+TMA in the case of study. The closed loop system stabilizes around 50 min, which is, approximately, twice the time of the reactor open loop response (30 min). For comparison, the cost function obtained without the TMA module is shown in the right bottom. It means, if there is no MA correction in the cost function ($\lambda_1 = \lambda_2 = 0$ in (11)) as happens in traditional eMPC or RTO problems, the cost obtained could be far away from the real optimum. The small offset in the temperature results may be caused by problems in the optimizer (the cost function region next to the optimum is relatively flat), besides the process gradients are estimated using transient data and not steady-state data.

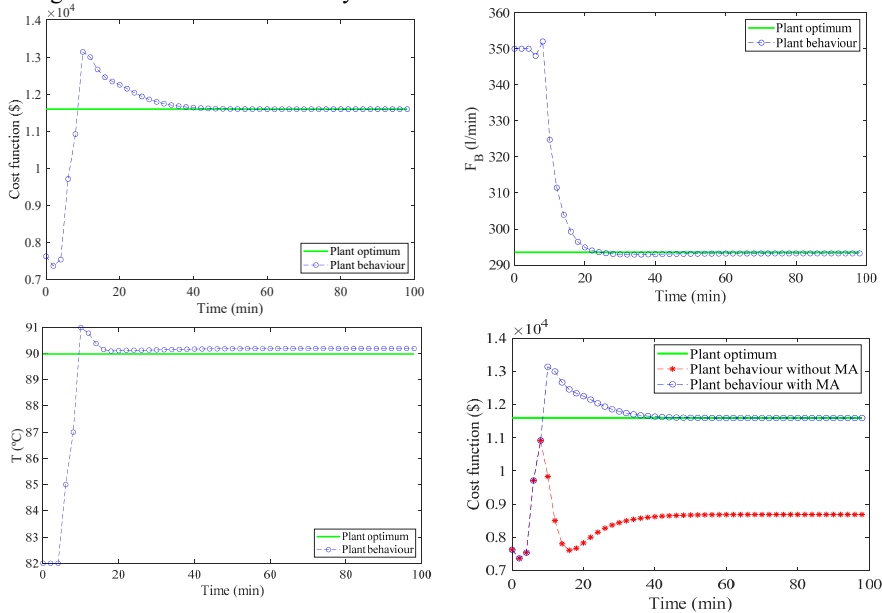


Figure 2: Cost function and manipulated variables over time.

5. Conclusions

The implementation of eMPC+TMA shows that it is possible to achieve an operating point close to the real optimum, despite the parametric and structural mismatch between model and real process using transient measurements in a shorter time than the one required in the RTO+MA+MPC traditional architecture, avoiding without waiting for multiple process steady states.

6. Acknowledgments

This work was supported by JCYL under programs CLU 2017-09 and UIC 233, as well as by project InCO4In, from Spanish AEI under project PGC2018-099312-B-C31, both with FEDER funds. The authors thank the European Social Fund and the “Conserjería de Educación de la Junta de Castilla y León”.

References

- Butcher, J.C., 2002. Nordsieck methods with an off-step point 1, 87–101.
- Faulwasser, T., Pannocchia, G., 2019. Toward a Unifying Framework Blending Real-Time Optimization and Economic Model Predictive Control. *Ind. Eng. Chem. Res.* 58, 13583–13598. <https://doi.org/10.1021/acs.iecr.9b00782>
- François, G., Bonvin, D., 2013. Use of Convex Model Approximations for Real-Time Optimization via Modifier Adaptation. *Ind. Eng. Chem. Res.* 52, 11614–11625. <https://doi.org/10.1021/ie3032372>
- Hernández, R., Engell, S., 2019. Economics optimizing control with model mismatch based on modifier adaptation. *IFAC-PapersOnLine* 52, 46–51. <https://doi.org/10.1016/j.ifacol.2019.06.035>
- Isermann, R., Münchhof, M., 2011. Identification of Dynamic Systems: An Introduction with Applications, *Identification of Dynamic Systems*. https://doi.org/10.1007/978-3-540-78879-9_9
- Marchetti, A., Chachuat, B., Bonvin, D., 2009. Modifier-Adaptation Methodology for Real-Time Optimization. *Ind. Eng. Chem. Res.* 48, 6022–6033. <https://doi.org/10.1021/ie801352x>
- Navia, D., Puen, A., Bergh, L., Rodriguez-Blanco, T., Sarabia, D., de Prada, C., 2017. Modifier-Adaptation Based on Transient Measurements Applied to a Laboratory-Scale Flotation Column. *Comput. Aided Chem. Eng.* 40, 1729–1734. <https://doi.org/10.1016/B978-0-444-63965-3.50290-7>
- Pannocchia, G., 2018. An economic MPC formulation with offset-free asymptotic performance. *IFAC-PapersOnLine* 51, 393–398. <https://doi.org/10.1016/j.ifacol.2018.09.332>
- Rawlings, J.B., Mayne, D.Q., Diehl, M.M., 2019. *Model predictive control: Theory, Computation, and Design*, 2nd ed, *Studies in Systems, Decision and Control*. Nob Hill Publishing, Santa Barbara.
- Rodríguez-Blanco, T., Sarabia, D., Pitarch, J.L., de Prada, C., 2017. Modifier Adaptation methodology based on transient and static measurements for RTO to cope with structural uncertainty. *Comput. Chem. Eng.* 106, 480–500. <https://doi.org/10.1016/j.compchemeng.2017.07.001>
- Vaccari, M., Pannocchia, G., 2018. Implementation of an economic MPC with robustly optimal steady-state behavior. *IFAC-PapersOnLine* 51, 92–97. <https://doi.org/10.1016/j.ifacol.2018.10.180>
- Vaccari, M., Pannocchia, G., 2016. A Modifier-Adaptation Strategy towards Offset-Free Economic MPC. *Processes* 5, 2. <https://doi.org/10.3390/pr5010002>
- Vaccari, M., Pelagagge, F., Bonvin, D., Pannocchia, G., 2020. Estimation technique for offset-free economic MPC based on modifier adaptation, in: *IFAC 2020 - 21st IFAC World Congress*.
- Williams, T.J., Otto, R.E., 1960. A generalized chemical processing model for the investigation of computer control. *Trans. Am. Inst. Electr. Eng. Part I Commun. Electron.* 79, 458–473. <https://doi.org/10.1109/tce.1960.6367296>

Attack Detection Using Unsupervised Learning Algorithms in Cyber-Physical Systems

Divyang Deep Tiwari, Sourjya Naskar, Arasavelli Siva Sai, Venkata Reddy Palleti

*Department of Chemical Engineering, Indian Institute of Petroleum and Energy,
Visakhapatnam, India*

venkat_palleti.che@iipe.ac.in

Abstract

Cyber-Physical Systems (CPS) are collections of physical and computer components that are integrated with each other to operate a process safely and efficiently. Examples of CPS include industrial control systems, water systems, robotics systems, smart grid, etc. However, the security aspect of CPS is still a concern that makes them vulnerable to cyber attacks on the control elements, network or physical systems. The work reported here is an attempt towards detecting cyber attacks and improving process monitoring in CPS; using unsupervised machine learning anomaly detection algorithms such as one-class SVM, isolation forest, elliptic envelope. These algorithms are evaluated using the dataset of a real Water Distribution Plant (WADI) built at the iTrust centre at Singapore University of Technology and Design for cyber security research. For modelling purposes, process 1 and 2 of the aforementioned plant were taken into consideration because the implemented attacks were closely related to only these sub-processes. The result of the experiment shows that one-class SVM is found to be the most effective algorithm in determining anomalies for this particular dataset.

Keywords: Cyber-Physical System, Machine Learning, Unsupervised Learning, Anomaly Detection, One-Class SVM, Isolation Forest, Elliptic Envelope, Principal Component Analysis (PCA)

1. Introduction

The automation in industries is increasing proportionally with the advancement of technology. Although, security infrastructure of Industrial Control Systems (ICS) is continuously ongoing development, yet there is an abundant possibility that artificial intelligence and domain knowledge can be merged to help us identify deviation in processes more effectively, in order to increase the safety of industries. Machine learning allows us to take into account all inputs at once and return the status of the process. Therefore, it is possible that these trained models can be installed in process industries and could potentially act as an additional layer of security and monitoring. A number of papers are available that discuss the use of various supervised learning algorithms for attack detection. Observations made by Gharibian and Ghorbani (2007) demonstrated that Decision Trees are very sensitive to the training data and don't learn quite well from imbalanced data. Furthermore, they found that Decision Trees and Random Forests (ensemble of Decision Trees) are very sensitive to the training data and their performance can vary significantly based on it. Naive Bayes Classifier implements Bayes theorem for classification problems. In comparison to MLP, Naive Bayes

classifiers can be trained within a short period of time (Fleizach and Fukushima, 1998). According to Domingos and Pazzani (1996) & Langley and Sage (1994), Naive Bayes can perform quite well when there exists a reasonable dependency in the data. It has been also observed that the performance of Naive Bayes classifier improves when redundant features are removed. The observations from Huy Anh and Deokjai (2008), where they applied a wide variety of algorithms: Bayesian approach, Decision Trees, Rule-based models, etc., states that no single algorithm is capable of detecting all kinds of attack with high detection and low false alarm rate. Muda Z, et al. (2011) performed a hybrid learning approach by combining Naive Bayes and K-means Clustering. The training dataset was divided into k-clusters based on an initial value known as the seed points into each cluster centre. The results showed that this hybrid approach performed better as compared to only Naive Bayes Classifier. Wang H. et al. (2011) made an attempt to improvise SVM by combining Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO). PCA is quite effective in reducing the dimensions of data. Thereafter, PSO was used to optimize the kernel parameters. The experimental results showed that the performance of PCA and PSO combined SVM was higher than those of PSO-SVM and standard SVM. Another work is conducted by Cheng Feng et al. (2019), proposed an approach which combined several machine learning and data mining techniques to generate a significant number of invariant rules (defined as a physical condition that must be satisfied for any given state of an ICS) (S. Adepu and A. Mathur, 2016). They observed that generated invariant rules can achieve high anomaly detection performance by demonstrating on two real-world ICS case studies. Their results outperform commonly used residual error-based anomaly detection models. Another advantage of this methodology is that it can be applied to diverse ICS frameworks as it is dependent only on the general control dynamics of ICS. The approach followed in the present work is slightly different and includes the outlier detection methods using unsupervised learning.

2. Methodology

The testbed under study for the scope of this work is Water Distribution Plant (WADI), built at the iTrust centre at Singapore University of Technology and Design for cyber security research (C. M. Ahmed et al., 2017). The entire plant is divided into 3 sub-processes namely process 1 (primary supply and analysis), process 2 (domestic grid with booster pump) and process 3 (return water system) as shown in Figure 1.

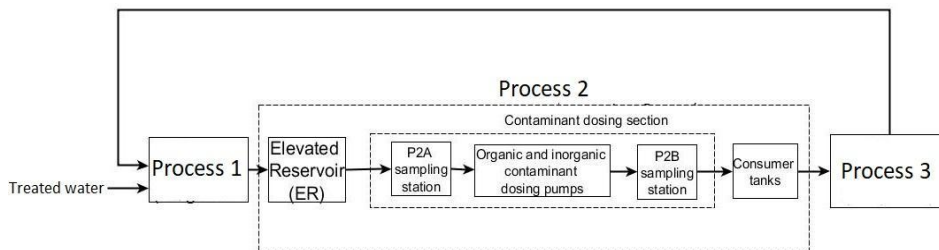


Figure 1: Overview of WADI testbed

The plant was operated for 14 days and the dataset¹ is divided into two parts. The initial 12 days correspond to normal operation data, and in the last 2 days, at different time

¹ https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/

segments, a variety of 15 attacks were performed on the testbed. The stated attack detection task is tackled as an anomaly detection problem. Outlier and novelty detection classification algorithms, such as One-Class SVM, Isolation Forest and Elliptic Envelope were trained on cleaned and transformed normal operation data. Then, the trained models were used to predict attacks/anomalies in the remaining 2 days data, which consists of both attack and normal instances. Then, based on the prediction result for each model, their hyperparameters are manually adjusted in order to obtain the most optimum performance. Since the last 2 days data belongs to the category of imbalance dataset, using only accuracy as the performance indicator is ineffective. Therefore, the confusion matrix is used for evaluating the models' performance. The major steps involved in the present study are as follows: (I) Slicing process variables from the normal operation data (II) Data preprocessing (III) Implementing Principal Component Analysis (PCA) for dimensionality reduction. (IV) Modelling the data with all three methods followed by manual hyperparameter optimization. The above steps were executed for both process 1 and process 2 respectively.

3. Procedure and Results

3.1. Process 1 (Primary Supply and Analysis)

Process 1 data is sliced from the initial 12 days normal operation dataset. It consists of 19 process variables which include different sort of equipment, such as motorized valve, level indicator, transfer pumps, physical properties sensors (eg. turbidity, conductivity, pH, ORP, TRC). Then, this data is cleaned and preprocessed in order to prepare it for PCA implementation. After performing PCA, 19 variables were reduced to 3 principal components which explained 99.8 % variance of the original dataset. Similarly, PCA is applied to the variables of process 1 in the last 2 days data, to obtain 3 principal components. After implementing PCA, process 1 normal operation dataset is then trained for each algorithm; i.e., one-class SVM, isolation forest and elliptic envelope. After training, models were tested with last 2 days data to examine their performance. The results are shown in Table 1.

Table 1: Performance of different algorithms modelled on Process 1 data and tested on the last 2 days dataset to detect attacks.

	One-Class SVM		Isolation Forest		Elliptic Envelope	
	Attack	Normal	Attack	Normal	Attack	Normal
Precision	0.25	0.97	0.39	0.96	0.21	0.96
Recall	0.48	0.91	0.33	0.97	0.32	0.93
F-1 score	0.33	0.94	0.36	0.96	0.26	0.94

It can be observed from Table 1 that the isolation forest performed slightly better than one-class SVM in terms of F-1 score. However, considering the ability to detect maximum number of attacks, it is found that one-class SVM detected 11 attacks out of

15. Whereas, isolation forest detected 10 attacks out of 15. Figure 2 highlights the timestamps that are detected as attacks (anomalies) in the last 2 days dataset by the one-class SVM model discussed in Table 1. The model classifies inlier (normal) as 1 and outlier (anomaly) as -1. As shown in Figure 2, the duration of the last 2 days data starts at 18:00 on Day 1 and ends at 18:00 on Day 3. Similar kinds of plots were obtained for isolation forest and elliptic envelope algorithms.

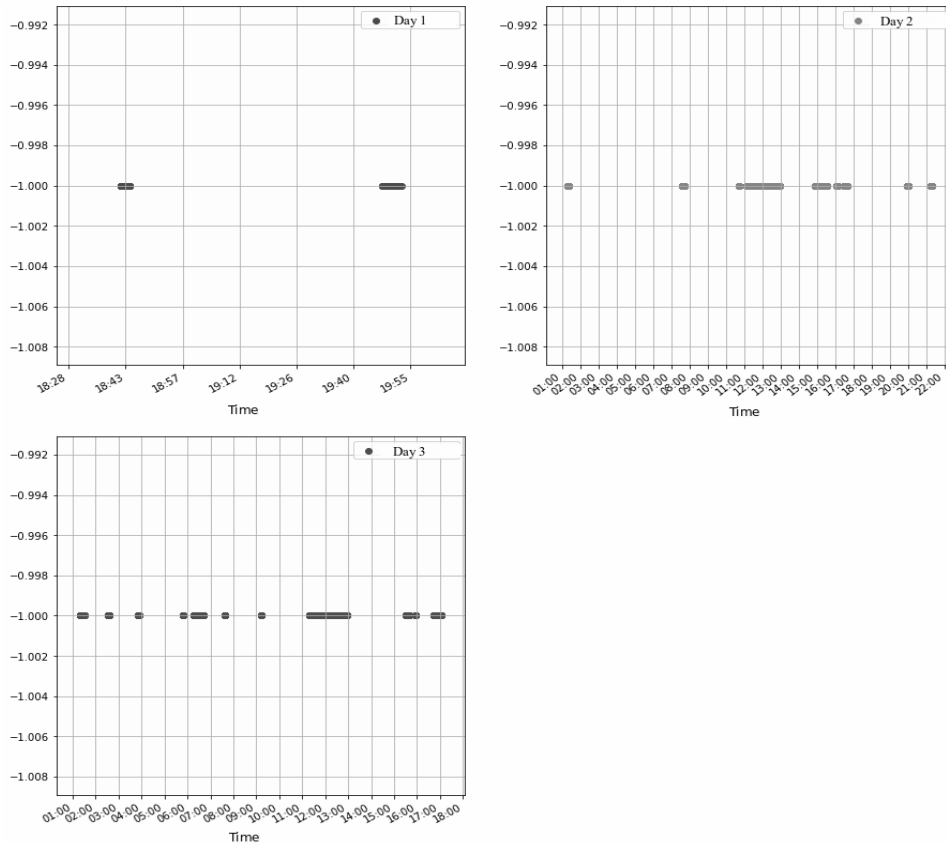


Figure 2: Predicting attacks in the last 2 days data using one-class SVM trained on process 1.

3.2. Process 2 (Domestic Grid with Booster Pump)

Process 2 is the largest among all the 3 processes. Initially, when sliced from the 12 days normal operation dataset, it consists of 86 variables. After removal of some redundant features, a total of 75 variables were extracted followed by the PCA transformation step. After performing PCA, these 75 variables were reduced to 7 principal components which explained 95 % variance of the original dataset. Similarly, PCA is applied to the variables of process 2 in the last 2 days data, to obtain 7 principal components. After implementing PCA, process 1 normal operation dataset is then trained for each algorithm; i.e., one-class SVM, isolation forest and elliptic envelope. After training, models were tested with last 2 days data to examine their performance. The results are shown in Table 2 which indicates that one-class SVM performed better

in detecting attacks as compared to isolation forest and elliptic envelope. In fact, one-class SVM detected 14 attacks out of 15 when modelled on process 2.

Table 2: Performance of different algorithms modelled on Process 2 data and tested on the last 2 days dataset to detect attacks.

	One-Class SVM		Isolation Forest		Elliptic Envelope	
	Attack	Normal	Attack	Normal	Attack	Normal
Precision	0.32	0.97	0.34	0.96	0.07	0.94
Recall	0.48	0.94	0.26	0.97	0.13	0.89
F-1 score	0.38	0.95	0.30	0.96	0.09	0.91

Figure 3 highlights the timestamps that are detected as attacks (anomalies) in the last 2 days dataset by the one-class SVM model discussed in Table 2.

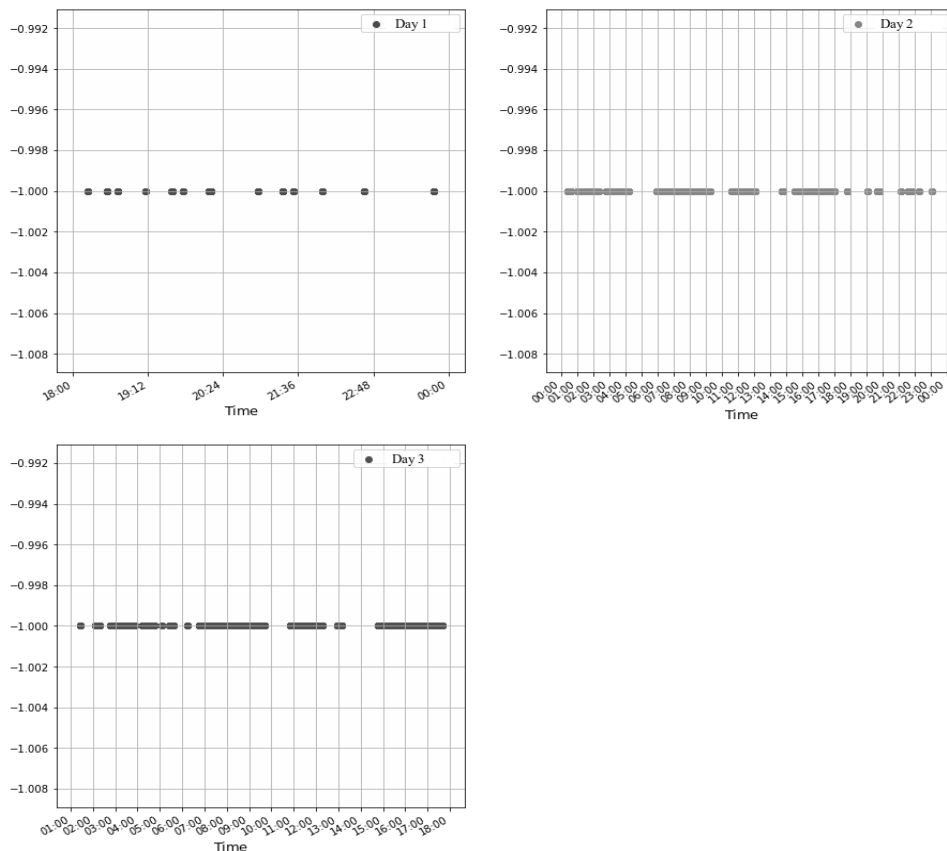


Figure 3: Predicting attacks in the last 2 days data using one-class SVM trained on process 2.

4. Conclusion

This work explored the possibility of using unsupervised machine learning anomaly detection algorithms for cyber attack detection in cyber-physical systems. The task is considered as an anomaly detection problem, where each attack instance is expected to be classified as an anomaly. Three anomaly detection techniques; i.e., one-class SVM, isolation forest and elliptic envelope were taken into consideration for the scope of this work. Modelling of all the 3 algorithms was carried out on both process 1 and process 2 of the Water Distribution Plant (WADI); i.e., a total of 6 models were trained. Based on the distribution of training data used in this work, it is found that one-class SVM performed reasonably well for detecting attacks in both the processes. However, few challenges that hold the possibility of further improvements are: (I) Reducing the number of false alarms that are generated by the models (II) Reducing the time delay in identification of attacks. It would be interesting to address these issues in the future work.

References

- Gharibian F, Ghorbani A.A, Comparative Study of Supervised Machine Learning Techniques for Intrusion Detection, Proc. of the Fifth Annual Conference on Communication Networks and Services Research, 2007, pp. 350–358.
- C. Fleizach and S. Fukushima, “A naive bayes classifier on 1998 kdd cup,” 1998.
- Domingos P. and Pazzani M., Beyond Independence: Conditions for the optimality of the simple Bayesian Classifier, In proceedings of the 13 th International Conference on Machine Learning, 1996, pp.105-110.
- Langley P, Sage S, Induction of selective Bayesian classifiers, Proc. of the Tenth Conference on Uncertainty in Artificial Intelligence, 1994,pp. 399-406, Seattle, WA: Morgan Kaufmann
- Huy Anh Nguyen, Deokjai Choi, Application of Data Mining to Network Intrusion Detection: Classifier Selection Model, 2008, pp.399-408, Springer-Verlag.
- Muda Z, Yassin W, Sulaiman M.N, Udzir N.I , Intrusion Detection based on k-means clustering and Naive Bayes classification, Proc. of 7 th Intl. Conference on IT in Asia, 2011, pp.1-6
- Wang, H., Zhang, G., Mingjie, E. et al. A novel intrusion detection method based on improved SVM by combining PCA and PSO. Wuhan Univ. J. Nat. Sci. 16, 409 (2011)
- Cheng Feng, Venkata Reddy Palleti, Aditya Mathur and Deeph Chana: A Systematic Framework to Generate Invariants for Anomaly Detection in Industrial Control Systems, 26th Annual Network and Distributed System Security Symposium (NDSS) (2019)
- S. Adepu and A. Mathur, “Using process invariants to detect cyber attacks on a water treatment system,” in IFIP International Information Security and Privacy Conference. Springer, 2016, pp. 91–104
- C. M. Ahmed, V. R. Palleti, and A. P. Mathur, “WADI: A water distribution testbed for research in the design of secure cyber physical systems,” in Proceedings of the 3rd International Workshop on CyberPhysical Systems for Smart Water Networks, ser. CySWATER '17, 2017, pp. 25–28

Application of Self-Optimizing Control in a Real Heat Exchanger Network

Ahmet Coşgun,^{a,*} Oğuz Uzun,^a Murathan Bağdat,^a

^a*Turkish Petroleum Refinery, 41790 Körfez, Kocaeli, Turkey*

ahmet.cosgun@tupras.com.tr

Abstract

The optimal operation of heat exchanger networks are important for increasing energy efficiency and reducing the environmental impact. In this paper control method called self-optimizing control is analysed and applied to a network with stream splits in a real refinery environment. By using only the temperature measurements, new control variables are defined. Optimization is done by manipulating the split ratios in the network. The optimal operation is defined by the economics of the process, which is maximizing the exit temperature of the exchanger network. Optimal control of the exchanger networks, whose operational purpose is to maximize the total heat transfer, are achieved by determining the amount of flow to be distributed to the individual branches at all times. Results showed that this method achieves an important increase in network exit temperature, which means a decrease in energy consumption compared to traditional strategy used in the refinery.

Keywords: optimization, self-optimizing control, heat exchanger networks

1. Introduction

Heat Exchanger Networks (HENs) are often used in refinery processes to recover as much thermal energy as possible from hot streams using thermal integration methods before sending the hot product or intermediate to the tank or to the next process (Biyanto et al., 2016).

The first main stage of oil refining is the distillation process. A pretreatment step is required for the distillation process before the crude oil stream is fed to the distillation column; preheating of crude oil to around 380 °C. For this purpose, about 60-70% of the required thermal energy is supplied by a network of heat exchangers called the crude preheat train (CPT) (Assis et al., 2013). The remainder of the thermal energy is provided by furnaces. It is important to gain as much as thermal energy from CPT units to decrease the need for energy supplied from furnaces, which uses external energy sources and increases the energy consumption of the refinery.

In CPTs, cold flow is divided into two or more branches and heated with different hot streams in different heat exchangers. The ideal operation is to maximize the total heat transfer or, in other words, to reach the highest cold flow temperature at the junction point at all times. This optimization goal is tried to be accomplished by adjusting the split ratio of the cold streams into the branches of the network. In literature, real-time optimization (RTO) approach is well studied, where the potential savings are very high but

computational needs are also very complicated. Other approach for optimization of CPTs is a practical method used in the refinery as current strategy, which is adjusting the split ratio by equalizing each cold stream's outlet temperature (Lid and Skogestad, 2001). Although this method has no computational need, it can easily diverge from optimal operation point.

Another method is called self-optimizing control (SOC), where, a cost function is continuously optimized to update the set points of the control variables to study the optimal solution. The control variable defined with its self-optimizing control concept aims to keep the process at an economically optimum level. The newly designed control variables are translated into the design of a simple control structure that keeps the process at high efficiency levels despite the changing conditions. This simplification in the control structure can result in suboptimal levels, but in most cases the benefits of a simple program outweigh the increased 'goodness' of complex models due to the high costs of implementation and maintenance. The control variables defined for CPT in self-optimizing control is called "Jäschke temperature" in the literature. For each stream, Jäschke temperature is calculated, where the objective for optimum operation is to equalize these control variables in each stream (Jäschke et al., 2017).

This study aims to optimize the heat exchanger networks of CPTs by maximizing the final CPT outlet temperature to reduce furnace fuel consumption and carbon emissions by using self-optimizing control system for flow distribution in a real refinery environment and comparing the results with current strategy used in the process.

2. Theory

The theoretical background for self-optimizing control of heat exchanger networks are well defined in the literature (Jäschke and Skogestad, 2014a). Optimization is done by assuming arithmetic mean temperature difference. This assumption causes little divergence from the real optimum but it is an acceptable loss compared to computational work needed. The objective function is defined to be the economically best scenario, which is the maximum outlet temperature in the network exit. The manipulated variable for the network optimization that maximizes the total heat transfer is the split ratio. The self-optimization control variable called Jäschke temperature is calculated for achieving this goal for each branch of the network as described in the literature (Jäschke and Skogestad, 2014b). The computation is well-defined and highly adaptable to different kinds of network configurations. For the calculations, only temperature values are used and other variables such as flow rates of the streams, heat transfer properties, or any other information regarding the process is not needed. The optimal controlled variable is also defined in the literature to be equal values of Jäschke temperatures in all branches. Jäschke temperatures can be regarded as the representation of the available potential for heat transfer. In example, a branch with zero Jäschke temperature stem from equal temperatures in cold stream outlet temperature and hot stream inlet temperature. In this case, increasing the cold stream flow rate would not result with an increase in heat transfer rate. Therefore, it can be said that all the potential in that branch is already utilized. In a network with two or more branches, for optimum operation, streams should be adjusted to increase the flow on the branch with higher Jäschke temperature. By doing that the potential of that stream would be utilized more, and corresponding Jäschke temperature will decrease. Keeping Jäschke temperature difference between branches at zero is the defined condition for optimal operation. The varying operating conditions and

disturbances, such as the change oil characteristics, fouling in heat exchangers, or the changes in stream flow rates and temperatures, are not needed to be monitored in terms of optimal operation, as the defined optimal control condition is valid in all cases (Girei et al., 2014).

3. Case Study

The study of self-optimizing control of HEX networks is performed in a vacuum distillation unit in Tüpraş İzmit refinery. The mathematical equations for the calculation of self-optimizing control variable are well defined in the literature and given in related papers (Jäschke and Skogestad, 2014a). The adaptation of the standardized calculations in the literature is carried out for the selected CPT unit by evaluating the network configuration. In Fig. 1, the network configuration of the studied CPT unit is given. The network consists of two branches with different hot stream passing through from each branch. The number and the type of the heat exchangers are different in each branch as well as the hot stream characteristics. Only temperature values are needed for the calculation of self-optimizing variables for each branch, and as shown in the Fig 1, the boxed variables are the ones used in the calculations.

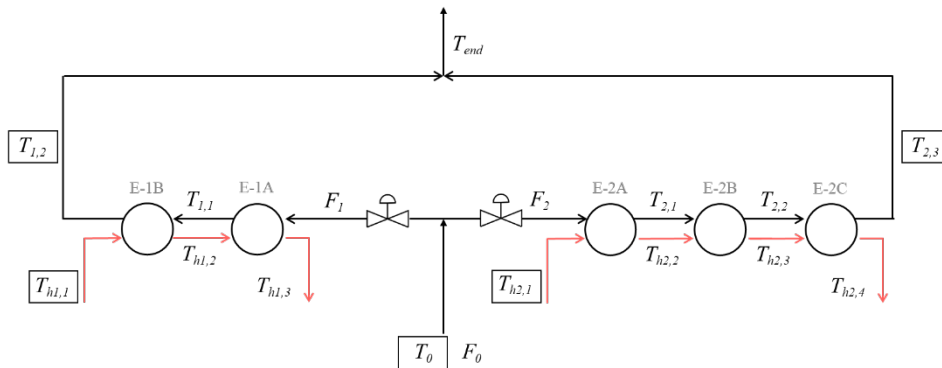


Figure 1: The heat exchanger network studied for self-optimizing control

The Jäschke temperatures for each branch is calculated. As defined in the literature equalizing these control variables in each branch results in optimum operation.

Before the implementation of this method into real operation, the historical data is collected and examined. In steady operational conditions, Jäschke temperature difference between branches must directly effects the exit temperature of the network stream as stated. To identify this behaviour, the historical data of the selected heat exchanger network is filtered by selecting a subset where the operational variables can be regarded as constant. The analysis of Jäschke temperature differences versus exit temperature of the network stream describes a negative correlation between those two as shown in Figure 2. The straight line in the figure shows the linear relationship between Jäschke difference between branches with exit temperature. Although the relationship is not a linear relationship as discussed in the literature, it is used as it is for illustration purposes, which

is a fair approximation locally. As backed up with the literature, it is also apparent in the historical data that minimizing the Jäschke temperature difference between branches positively effects the exit temperature.

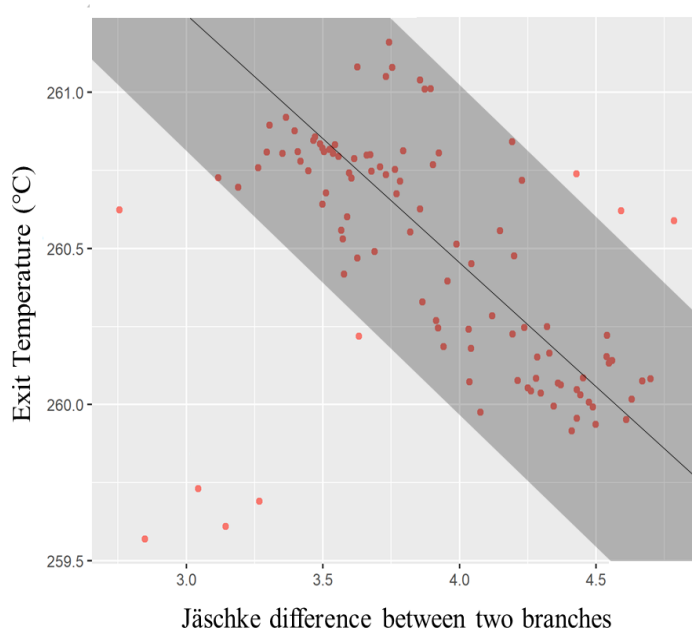


Figure 2: Jäschke temperature difference between branches versus exit temperature of the network stream

4. Results

To understand the effect of self-optimizing control variables in the exit temperature of the network stream in real time, a test run is constructed. Two scenarios are tested, while the operational variables are kept constant. The duration of the scenarios were selected to be twelve hours where the starting hours are same for eliminating the effect of day and night temperature differences. Two scenarios were tested in real operation by only manipulating the control variables defined by self-optimizing control by keeping the Jäschke temperature difference between branches at 2 and 10 as shown in Figure 3. This values of Jäschke temperature difference between branches are selected to approximately compare the best use case of self-optimizing control with current strategy. The results showed that, in the region of testing, an 8 °C decrease in the differences in Jäschke temperatures, resulted in 3.6 °C increase in exit temperature of the network stream and by considering the downstream furnace operation; a 0.2 decrease in energy intensity index (EII).

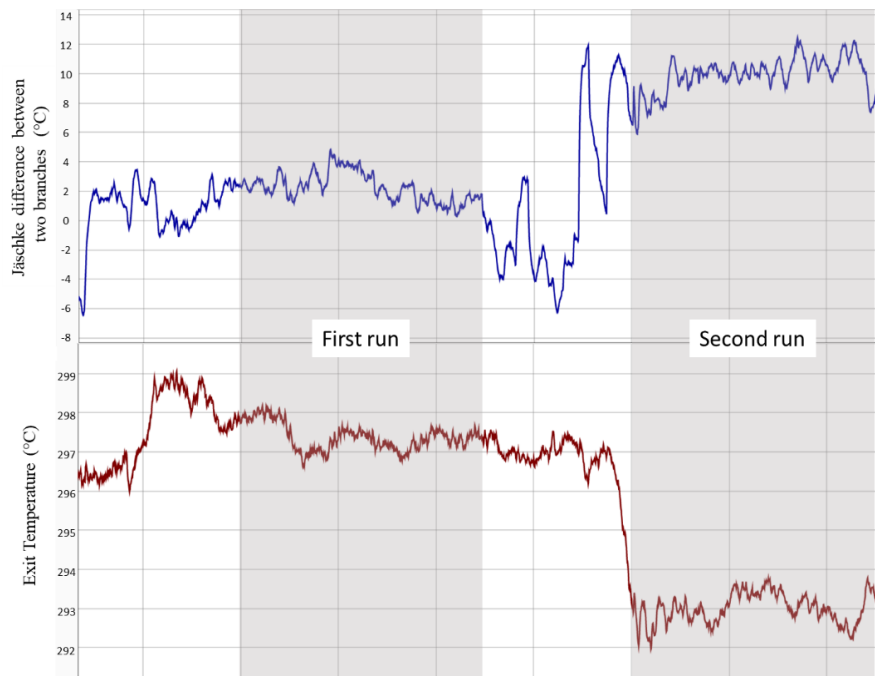


Figure 3: Test run results, Jäschke difference between branches and corresponding exit temperature over time

As shown in Figure 4, the average value of the Jäschke temperature difference between branches in the heat exchanger unit in current strategy is 3.7 °C. After verification of the effect of self-optimizing control method in test run, the method is integrated into DCS screens (open loop implementation). The manipulation of the stream splits are done manually by the operators by controlling the Jäschke temperature difference between branches. By this manual operation, the average Jäschke temperature difference between branches is decreased into 1.9 °C. However, because of the need for constant action to manipulate the stream splits stemming from operational changes, an optimal operation did not sustained. For automation purposes, implementation into Automatic Performance Control (APC) system is done (closed loop implementation). As shown in Figure 4, Jäschke temperature difference between branches is decreased into 0.5 °C by automation of the self-optimizing control method.

Heat exchanger networks in operational environment is prone to the changes in the system variables. Hence, test runs are particularly important for exploring the relation between control variable and exit temperature, where all the variables are kept almost constant. After the construction of test runs, the results are linearly interpolated and, it is found that open loop implementation resulted in 0.8 °C increase in exit temperature of the network stream and 0.045 decrease in EII, and closed loop implementation resulted in 1.4 °C increase in exit temperature of the network stream and 0.08 decrease in EII.

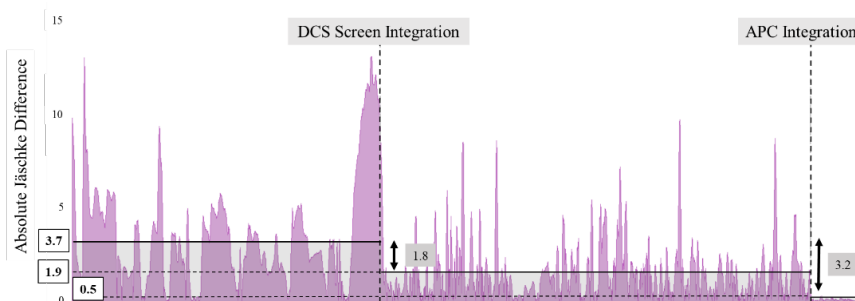


Figure 4: Absolute Jäschke difference over time

5. Conclusions

Self-optimizing control method for heat exchanger networks is well defined in the literature. Previous works in the literature also proved the usefulness of the method by simulations. This paper focused on real use case of the method and results showed that compared to past strategy used for maximizing the exit temperature of the network, self-optimizing control resulted in more optimized solution in real refinery environment. The economic and environmental effects are investigated, although a major assumption of linear interpolation is done for estimating the effects from the test run results. Because of the simplicity of the method compared to RTO methods, the implementation and the maintenance of the self-optimizing control method is done relatively easily without any need for investment. However there are also limitations that may prevent the optimal operation stemming from the operational safety limits, minimum and maximum flows permitted for each branch and, maximum allowable level of pressure drop in the heat exchanger network.

References

- B.C.G Assis, C.O. Gonçalves, F.S. Liporace, S.G. Oliveira, E.M. Queiroz, F.L.P Pessoa, A.L.H Costa, 2013, Constrained thermohydraulic optimization of the flow rate distribution in crude preheat trains, *Chemical Engineering Research and Design*, 91, 1517–1526.
- T.R Biyanto, M. Ramasamy, A.B. Jameran, H.Y. Fibrianto, 2016, Thermal and hydraulic impacts consideration in refinery crude preheat train cleaning scheduling using recent stochastic optimization methods, *Applied Thermal Engineering*, 108, 1436-1450.
- S.A. Girei, Y. Cao, A.S. Grema, L. Ye, V. Kariwala, 2014, Data-driven Self-optimizing Control, *Computer Aided Chemical Engineering*, 33, 649-654.
- J.Jaschke, S. Skogestad, 2014a, Optimal operation of heat exchanger networks with stream split: Only temperature measurements are required, *Computers and Chemical Engineering*, 70, 35-49.
- J.Jaschke, S. Skogestad, 2014b, A self-optimizing strategy for optimal operation of a preheating Train for a Crude Oil Unit, *Computer Aided Chemical Engineering*, 33, 607-612.
- J. Jaschke, Y. Cao, V. Kariwala, 2017, Self-optimizing control – A survey, *Annual Reviews in Control*, 43, 199-223.
- T. Lid, S. Skogestad, 2001, Implementation issues for real-time optimization of a crude unit heat exchanger network, *Computer aided chemical engineering*, 9, 1041-1046.

Data-Driven Control Strategies for the Autonomous Operation of the Pharmaceutical Crystallization Process

Merve Öner,^a Gürkan Sin^a

^aProcess and Systems Engineering Center (PROSYS), Technical University of Denmark, Søtofts Plads B228A, Kgs. Lyngby 2800, Denmark

Abstract

In this contribution, we studied the deep neural network (DNN) for the control of the cooling crystallization of a model compound system. To this end, firstly the performance of the different neural network architectures in conjunction with the various combination of the time-series process data was tested for the training of the data-based model in order to assess the best fit model-training data architecture. The identified network model, which was trained with the offline process data, was utilized in a predictive control strategy. The objective of the control strategy was to optimize the supersaturation generating/decaying variable in the crystallizer to achieve a target crystal-state property profile throughout the process. The performance of the proposed control strategy was tested in the presence of the process disturbance and benchmarked against a radial basis function (RBF) based control strategy. The results showed that the DNN model was able to approximate the crystallization process input-output relation with R^2 ranging between 0.767 and 0.990 and achieve the target profile at the end of the operation with a 22.3 % offset.

Keywords: crystallization, data-based models, control, neural networks.

1. Introduction

In pharmaceutical manufacturing, crystallization is an essential unit in the downstream processing to recover and purify high-purity active pharmaceutical ingredients (API) and fine chemicals. The design, operation and control of a pharmaceutical crystallization process is a great challenge due to the highly nonlinear and multivariable nature of the crystallization process in conjunction with the strict and versatile requirements of the end-product. The current industrial practice for the crystallization process design and optimization relies on the quality-by-design framework based on the factorial design of the resource-intensive open-loop crystallization experiments (Szilagyi et al., 2020). Despite the significant progress in academia in the area of model-based (based on high fidelity population balance equations) and model-free (e.g. direct nucleation control, supersaturation control) techniques for the control of the crystallization process, establishing control over the crystallization process is still an issue in pharmaceutical manufacturing that can be reasoned due to strict dichotomy of the developed control strategies (Griffin et al., 2016). A third category, which is different from model-based and model-free techniques, can be attributed to the data-driven (based on process data) control strategies. The main advantage of the data-driven approaches is their ability to learn from the provided multidimensional process data with high accuracy. Accordingly, they became very useful and quickly applicable particularly for the optimization and

control of the complex processes that are troublesome, and have nonlinear and stochastic characteristics (Qin and Chiang, 2019; Venkatasubramanian, 2019) such as the crystallization process. A review of literature studies focusing on the data-driven control applied to the crystallization process can be found elsewhere (Öner et al., 2020a). Although the development of efficient training algorithms, availability of more powerful computers, easy-to-use software, and access to the online crystallization process data through the process analytical (PAT) tools enable overcoming of past challenges that hindered the expanding application of the data-driven strategies, the number of the studies focusing on the crystallization process is still limited. Therefore, in this contribution, we studied a data-driven control strategy based on the deep neural network (DNN) for a pharmaceutical crystallization process. The feasibility of the different neural network structures (models, training data) was explored, and the control performance was tested and benchmarked in the presence of the process disturbance.

2. Methods

2.1. Crystallization modeling

A mechanistic model incorporating population, mass, and energy balance equations for the crystallization process was developed in our previous work (Öner et al., 2020b). This model was employed firstly to generate data to be used to develop/train neural network models for the crystallization process and secondly to mimic a physical crystallization process under the data-based feedback control. As a case study, a seeded cooling crystallization (from 40 °C down to 20 °C) of paracetamol in ethanol solvent (400 mL) was demonstrated. The kinetic parameters of paracetamol crystallization were adopted from the literature. A total of 20 batches of data was generated by varying the process parameters of initial relative supersaturation (1.0 – 1.2), seed crystal size distribution (mean (90 – 250 μm) and standard deviation (5 – 10 μm) and shear rate created by the impeller in terms of secondary nucleation parameters (coefficient (50 % of the reported value) and exponent (5 % of the reported value)) based on the Latin Hypercube Sampling strategy from parameter space similar to the workflow described in our previous study (Öner et al., 2020a). The remaining process parameters were kept constant between the batches.

2.2. Deep neural network: structure, training and validation for process modeling

As one of the machine learning techniques, neural networks are structured in terms of processing units of input, hidden and output layers to resemble the human nervous system and the structure of the brain, by which they were inspired. Each layer contains nodes or units, which have connections to the other nodes and layer in adjacent layers with a weight value. Weighted summation of the inputs at each unit undergoes a transformation based on an activation function that is then fed to the respective unit in the next layer. Consequently, the outcome of the final output layer yields the solution for the problem (Shrestha and Mahmood, 2019). Various algorithms of the deep neural networks have been developed and used with varying purposes such as long short-term memory for forecasting models, convolutional neural networks in computer vision and image recognition, and recurrent neural networks for time series problems and forecasting (Hwangbo and Sin, 2020; Shrestha and Mahmood, 2019). In this study, the aim is to design a deep neural network model based on the ground of supervised learning for the crystallization process using the process data generated from the crystallization mechanistic model simulations. For this purpose, a general deep learning algorithm is adopted following the previously published framework (Hwangbo and Sin, 2020). To this end, a list of deep neural network model candidates are generated by manipulating the

hyperparameters in the structure of the deep neural network. The explored hyperparameters are the number of hidden layers and hidden neurons. A key step in the deep network architecture is to solve a regression problem embodying activation functions, connections and a loss function (Hwangbo and Sin, 2020). In this network architecture, the connections between the neurons in a hidden layer (h), weights and bias are linear as shown in Eq. (1). $W^{[h]}$ is a weight matrix and $b^{[h]}$ is a bias matrix having the shape of $(n^{[h]}, n^{[h-1]})$ and $(n^{[h]}, 1)$, respectively. The number of hidden neurons is represented by $n^{[h]}$ ($1 \leq h < H$) where $n^{[0]}$ and $n^{[H]}$ are equal to the numbers of input and target (output) variables. $A^{[h]}$ is the activation matrix (Eq. (2)) having a shape of $(n^{[h]}, 1)$, where $A^{[0]}$ represents the input matrix $[x_1, x_2, \dots]^T$ and $A^{[H]}$ contains one single value from the output layer. The activation function is indicated by $g(z)$. A commonly used function, which is the hyperbolic tangent sigmoid function, is used as the activation function in both hidden layers and the output layer (Lau and Lim, 2017; Hwangbo and Sin, 2020). As the loss function, mean squared error difference (MSE) between the target value and the estimated value from the neural network comprising all training data is used (Wang et al., 2016):

$$Z^{[h]} = W^{[h]}A^{[h-1]} + b^{[h]} \quad (1 \leq h \leq H) \quad (1)$$

$$A^{[h]} = g(Z^{[h]}) \quad (1 \leq h \leq H), \quad g(z) = \frac{2}{1 + e^{-2z}} - 1 \quad (2)$$

The machine learning toolbox in Matlab 2018b was used for the training of the DNN. The process variables such as processing time (t), bulk temperature in the crystallizer (T_b), solute concentration (C) and relative supersaturation (SS) were regarded as inputs, while the sum of square weighted crystal counts, mean crystal size and yield were used as the outputs. The choice of the parameters can be attributed to the fact that the main driving force for the crystallization process is the difference between the solute concentration and the saturation concentration (simplification of the chemical potential) that is often defined as supersaturation, and it is created by the manipulation of the temperature in a cooling crystallization process. On the other hand, crystal size distribution related measures (e.g. mean crystal size) and yield are some of the parameters that are commonly regarded as the critical product quality attributes or key process performance indicators. Therefore, a total of 165 different combinations (11 model structure, 5 input data structure, 3 output data structure) were tested (Table 1). A simple min-max data normalization was applied to both input and output data so that the data take the value between 0 and 1. To support the model selection criteria, two estimators of information criterion have been consulted. These criteria are the Akaike information criterion (AIC) and Bayesian information criterion (BIC). These information criteria are calculated from the sum of squared of errors (SSE) between the target value and the details can be found elsewhere (Hwangbo and Sin, 2020). The optimal model candidate is identified by the minimum AIC and BIC values.

2.2.1. Crystallization control

The ultimate goal of the crystallization process operation is to have a robust strategy that achieves a desired (or a reference) profile of a solid-state attribute in the presence or absence of any process disturbances. To achieve this goal, the particular control strategy

Table 1: DNN architecture candidates based on model, input and output structures.

Model	Network structure	Input	Output
1	[4 2]	[t T _b]	$\sum_{10}^{500} NL^2$ (Sum of square weighted counts)
2	[8 4]		
3	[16 4]	[T _b C]	$\frac{\sum NL}{\sum N}$ (Mean crystal size)
4	[32 4]		
5	[16 4 4]	[t T _b C]	
6	[32 8 4]		$\frac{c_o - c_f}{c_o - c_{sat(@T_f)}}$ (Yield)
7	[16 8 4 4]	[T _b SS]	
8	[32 16 4 4]		
9	[32 16 8 4]	[t T _b SS]	
10	[16 16 8 4 4]		
11	[32 16 8 4 4]		

utilizes the DNN model trained offline in a predictive control to optimize the cooling water (jacket) temperature setpoint for a specific time horizon in near future. The optimization of the temperature setpoint was realized using `fmincon` function in Matlab 2018b with respect to a cost function (minimization of the square of the normalized difference between the reference state and the DNN-predicted state) under imposed constraints (temperature and heating/cooling rate). The control strategy implemented in this study follows the same strategy published in our previous work (Öner et al., 2020a), where a radial basis network (RBF) model was used as a data-driven model trained with a very limited process data (up to 2 batches). The main difference between the DNN based strategy compared to RBF based strategy was i) the usage of a higher number of batch data and ii) training of the data-driven model offline. Considering these differences, the DNN based control was benchmarked against the RBF based control as well.

3. Results and Discussion

3.1. Deep neural network model for the crystallization process

As mentioned previously, 5 input data and 3 output data structures in conjunction with 11 neural network model structures were evaluated with respect to the best performance of the deep neural network. The datasets from 20 crystallization batches (shown in Figure 1) consisting of approximately 42000 data points were fed into all these neural network models to train and evaluate the performance. The ratio between training/validation/test data was 0.8/0.1/0.1. Levenberg-Marquardt back-propagation was used as the network training function with respect to the objective function of MSE. The maximum number of validation failure and number of epochs to train were 10 and 10,000, respectively. The training of the DNN on a computer with an Intel Core i5-6200U 2.30 GHz processor took between 30 minutes and 12 hours. The optimum combination was found considering the minimum information criteria (AIC and BIC). In the cases, in which AIC and BIC criteria showed different neural network structures, an additional criterion of the coefficient of determination (R^2) was also considered to discriminate the best fit. The obtained results can be seen in Table 2. For the outputs of the sum of the square weighted crystal counts (output 1) and mean crystal size (output 2), the same input data type [t T_b SS] yielded the best fit incorporated in the structure of the neural network Model 9 and Model 11, respectively. For the output of yield (output 3), in the discriminated structure of Model 9, the replacement of supersaturation data with the concentration data in the input showed

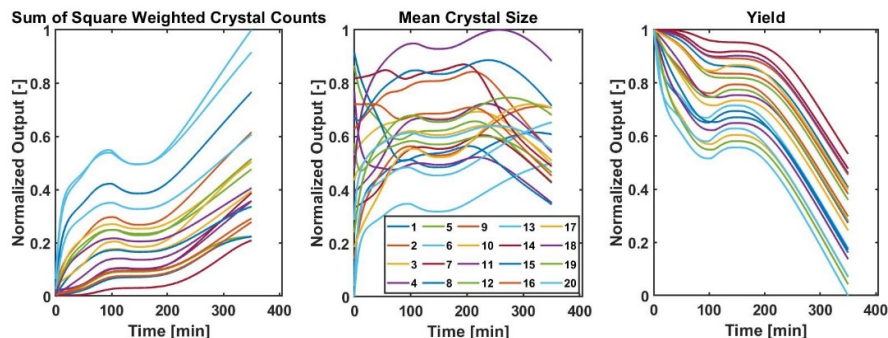


Figure 1: The process output data of 20 crystallization batches.

Table 2: Selected DNN structures based on the information criteria.

Output	Input	Best structure	R^2_{Training}	R^2_{Test}
Sum of sqr. wt. counts	[t T_b SS]	[32 16 8 4]	0.960-0.981	0.956-0.990
Mean crystal size	[t T_b SS]	[32 16 8 4 4]	0.675-0.853	0.653-0.864
Yield	[t T_b C]	[32 16 8 4]	0.874-0.952	0.921-0.951

the best fit. This can be rationalized with the direct relation between concentration and yield terms.

3.2. Crystallization control

The DNN based control was studied in comparison with the RBF based control in the presence of the process disturbance on the initial supersaturation. During the first 90 minutes, both data-based controls were inactivate, while a pre-defined temperature profile was applied to the system. Corresponding data was collected during this period. After this data collection period, the data-based controls were activated, and they took the control of the crystallization operation, run the system autonomously and taking decisions of heating or cooling the system to achieve the target solid-state profile. Compared to the RBF based control (final offset was 2.4 %), the DNN based control behaved more conservatively and achieved the final target profile with an offset up to 22.3 %. The manipulated variable set points predicted by the RBF based control could eliminate the disturbance-sourced offsets better in the earlier stages of the operation and consequently could follow smoothly the target profile. While these initial results showed DNN approach does not have a superior performance against RBF based data-driven approach, which employs a simpler network configuration, future work will examine the ability of DNN to extract patterns from heterogeneous sources of data including image analysis.

4. Conclusions

The results of this study demonstrated the feasibility of the data-driven models substituted for the knowledge-driven models for the crystallization process. The regarded performance metrics made evident that the DNN performs quite well with high prediction quality even though training data was still limited to 20 batches in contrast to the common applications involving the big data. However, system-dependent studies are still required

to discriminate the best network architecture for the specific input/output information and a generalization is not possible. With respect to control, the utilization of the DNN as a predictive model in a control strategy performed acceptably compared to RBF based control. As a future work, further comprehensive analyses will be performed to assess and improve the performance of the DNN based control. It can be concluded that the data driven control has a great potential due to its quick and simple implementation compared to time, effort and knowledge needed for developing a knowledge-driven model-based optimization and control for the highly nonlinear and stochastic crystallization process.

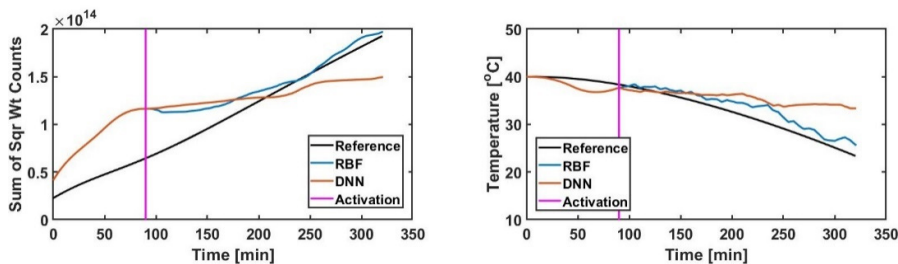


Figure 2: The performance of the DNN and RBF based predictive controls in the presence of positive disturbance on the initial supersaturation.

References

- D. J. Griffin, M. A. Grover, Y. Kawajiri, R.W. Rousseau, 2016, Data-driven modeling and dynamic programming applied to batch cooling crystallization, *Ind. Eng. Chem. Res.* 55, 1361–1372.
- S. Hwangbo, G. Sin, 2020, An integrated framework for plant data-driven process modeling using deep learning with Monte-Carlo simulations, *Computers and Chemical Engineering* 143, 107071.
- M. M. Lau, K. H. Lim, 2017, Investigation of activation functions in deep belief networks, In: 2nd International Conference on Control and Robotics Engineering (ICCRE) 55, 201–206.
- M. Öner, F. C. C. Montes, T. Ståhlberg, S. M. Stocks, J. E. Bajtner, G. Sin, 2020a, Comprehensive evaluation of a data driven control strategy: Experimental application to a pharmaceutical crystallization process, *Chemical Engineering Research and Design* 163, 248–261.
- M. Öner, S. M. Stocks, G. Sin, 2020b, Comprehensive sensitivity analysis and process risk assessment of large scale pharmaceutical crystallization processes, *Computers & Chemical Engineering* 135, 106746.
- S. J. Qin, L. H. Chiang, 2019. Advances and opportunities in machine learning for process data analytics, *Computers and Chemical Engineering* 126, 465–473.
- A. Shrestha, A. Mahmood, 2019, Review of deep learning algorithms and architectures, *IEEE Access* 7, 53040–53065.
- B. Szilagy, A. Eren, J. L. Quon, C. D. Papageorgiou, Z. K. Nagy, 2020, Application of model-free and model-based quality-by-control (QbC) for the efficient design of pharmaceutical crystallization process, *Crystal Growth & Design* 20, 3979–3996.
- V. Venkatasubramanian, 2019, The promise of artificial intelligence in chemical engineering: Is it here, finally?, *AIChE Journal* 65,2, 466–478.
- S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, P. J. Kennedy, 2016, Training deep neural networks on imbalanced data sets, In *International Joint Conference on Neural Networks (IJCNN)*, IEE 4368–4374.

Dynamic optimization of a stand-alone alkaline water electrolyzer for future operation using a predicted solar forecast

Haider Niaz, J.Jay Liu*

Department of Chemical Engineering, Pukyong National University, Busan 48513, South Korea

jayliu@pknu.ac.kr

Abstract

Renewable energy is a promising candidate as a green and sustainable energy source. Nevertheless, it is also well known for its dynamic and unpredictable nature. These characteristics make it difficult to decide whether to choose it as an energy source or not. On the other hand, ample renewable energy availability creates insecurity when it comes to storing this energy. A battery can be a short-term energy storage viability. However, in the long run, hydrogen is considered a promising candidate. In this study, standalone solar powered 4.5 MW alkaline water electrolyzer (AWE) operation is optimized for one-year predicted solar data via a dynamic optimization approach. The purpose of optimization was to find (1) optimal power split fraction for the battery and electrolyzer and (2) discharge power value when subjected to a predicted solar profile. These optimal points later will be used to run the model on actual weather data to analyze the possibility of future operation and control action plan for enhanced plant performance. The optimization results show a 24.6 % increase in annual hydrogen production with a 39.14% reduction of startups compared to the base case. Conversely, the optimal values of A and P_D at 0.9784 and 111 kW, respectively, ensured 26.53% reduction in startups for the actual profile too in comparison to the base case, and with a 3.45% reduction in annual hydrogen production (kg) than the optimized case for the predicted solar profile. This study will serve as a supporting tool for future investors to analyze renewable hydrogen's operational and economic potential in any region by just inputting solar-generated power profile.

Keywords: Electrolyzer, dynamic operation, optimization, solar, hydrogen production.

Nomenclature

PV	Photovoltaic	$nH_{2_{STORED}}$	Total hydrogen (kg)
AWE	Alkaline water electrolyzer	P_A	Power to AWE
BESS	Battery energy storage system	P_C	Power to BESS

SAM	System advisor model	P_D	Power from BESS
SOC	State of charge (%)	P_X	Power for sale
nH_2^{GEN}	Hydrogen produced (kg/hr)	$P_{C_{MIN}}, P_{C_{MAX}}$	Minimum and maximum charging power
η_{Ele}	Efficiency of electrolyzer (%)	$P_{D_{MIN}}, P_{D_{MAX}}$	Minimum and maximum discharging power
P	Power to AWE	SOC_{MIN}, SOC_{MAX}	Minimum and maximum state of charge
LHV	Lower heating value of H ₂	η^+, η^-	Charging and discharging efficiencies

1. Introduction

Sooner or later, fossil fuels will run out, but their impact on our environment would take years to recover. Currently, the threshold for atmospheric CO₂ levels is at 415.26 ppm, which is the highest recorded up to date and has not been witnessed in millions of years (Moreno-Benito et al., 2016; Hannah Ritchie and Max Roser, 2020). The increase in greenhouse emissions has created an urge to implement renewable energy as an alternative to fossil fuels, but renewable energy's unpredictability is still persistent. This behavior makes renewable energy a less reliable resource when it comes to a continuous production system. However, storing renewable energy in batteries is a viable option but not the optimal one. For long-term storage, renewable hydrogen is considered the most viable means of achieving long-term energy storage at a lower cost and greater volumetric energy density than other fuels (Hirscher et al., 2020).

The principal method for converting renewable energy into hydrogen is through an electrolyzer. Among electrolyzer types, the alkaline water electrolyzer (AWE) is the most common and commercially accepted technology for hydrogen production (Brauns and Turek, 2020). However, a hydrogen production system will not be efficient if operated with an interrupted energy supply. With a proper set of controls and energy backup such as BESS, renewable energy fluctuations can be mitigated effectively (Kwon et al., 2016; Pascuzzi et al., 2016). Ulleberg presented a dynamic model of a PV-hydrogen with a battery and grid support in Germany (Ulleberg, 2003). The possibility of combining two or more renewable energy sources such as solar, wind, and hydropower to produce hydrogen has also been studied (Hoste et al., 2020). For a smooth operation, there should be a minimum interruption in power supplied to electrolyzers. To overcome that, a battery energy storage system (BESS) plays a prominent role in providing power at times with no solar-based energy or during power outages to operate the AWE at a minimum operational capacity (Eichman et al., 2014).

Therefore in this study, a 4.5 MW AWE dynamic operation for one year is optimized based on a predicted solar forecast for possible future operation based on the optimal split value of power to the AWE and BESS for hydrogen production and BESS charging, respectively. Also, an optimal discharge power value from BESS to AWE is selected for safe BESS operation. When subjected to the actual solar data-based operation, these optimized split fractions and discharge power values resulted in a continuous hydrogen production with no shutdowns, however a bit lower hydrogen production. Results suggested a substantial insight into operation planning and control for enhanced performance and optimal operation via optimizing predicted solar-based

operation for actual operation. This study will guide as a supporting tool in estimating the power distribution to the AWE and BESS based on the solar potential for any region.

2. Method

The solar irradiation data for the year 2018, both predicted, and actual weather of Incheon Korea was obtained from the System Advisor Model (SAM) (Ezeanya et al., 2018). SAM was used to simulate the power profile of a 6.5 MW PV system. The solar power production model by SAM takes the following assumptions:

- a) Solar PV modules are operated in a shadow-free environment.
- b) The PV system is standalone with no connection to the grid.

The general equations representing the dynamic operation of AWE and moles of hydrogen generated (nH_2^{GEN}) can be represented, as shown in Eq. (1).

$$nH_2^{GEN}(t) = \frac{\eta_{Ele} \times P}{LHV} \quad (1)$$

Where η_{Ele} is the efficiency of electrolyzer (68%), P the power to the electrolyzer, and LHV is lower heating value of hydrogen. The amount of hydrogen being stored ($nH_{2_{STORED}}$) can be represented, as shown in Eq. (2), where t is equal to 8760 hrs.

$$nH_{2_{STORED}}(t) = \int_0^t nH_2^{GEN}(t') dt' \quad (2)$$

To ensure a safe operation, some constraints must be considered for BESS operation and are enlisted below in Eq. (3) – (6):

$$I. \quad P_C(t) \times P_D = 0 \quad (3)$$

$$II. \quad P_{C_{MIN}} \leq P_C(t) \leq P_{C_{MAX}} \quad (4)$$

$$III. \quad P_{D_{MIN}} \leq P_D \leq P_{D_{MAX}} \quad (5)$$

$$IV. \quad SOC_{MIN} \leq SOC(t) \leq SOC_{MAX} \quad (6)$$

P_C is the power to BESS, $P_{C_{MAX}}$ is the maximum charging power for BESS, P_D is the power discharged from the BESS, $P_{D_{MIN}}$ is the minimum discharge power from the BESS to ensure hydrogen production during the night at lowest possible power. $P_{D_{MAX}}$ is

the maximum discharge power from the BESS, SOC is the state of charge of the BESS, and SOC_{MIN}, SOC_{MAX} represent the minimum and maximum SOC level respectively. In this study, SOC_{MIN}, SOC_{MAX} are 30% and 100%, respectively. Overall SOC can be represented, as shown in Eq.(7).

$$SOC(t) = SOC(t - \Delta t) + \left((\eta^+ \times P_C(t)) - (\eta^- \times P_D(t)) \right) \Delta t \quad (7)$$

Where η^+, η^- are the charging and discharging efficiencies. Charging and discharging efficiencies consider 5% loss.

Figure SEQ Figure * ARABIC 1. AWE with BESS and hydrogen storage.

The data used for modeling the electrolyzer and values of the model's parameters are taken from (Ulleberg, 2003; Mayyas and Mann, 2018). gPROMS platform was used to model the dynamic equations for the alkaline water electrolyzer. An illustration for the overall model can be seen in Figure 1. P_s represents the solar power, A/B represent the split fractions for the power to AWE ($P_A = P_s \times A$) and power to BESS ($P_B = P_s \times B$). P_x represents the surplus power when neither the AWE and neither the BESS needs power, i.e., either the BESS is fully charged, or the AWE is already operating at maximum load. P_x is used as a product and is assumed to be sold to neighboring industries. P represents the total power and is the sum of $P_A + P_D$.

Solar power (P_s) is split into two fractions (A) and (B), as shown in Figure 1. These fractions are set at different ratios. In this model, (A) is 0.7, then (B) will be 0.3, provided their sum always counts to 1.0. P_A is the power split going to the electrolyzer, and P_C is the power split going to the BESS for charging the batteries. In general, during day time (06:00 AM to 06:00 PM), power (P) is primarily the power

coming from P_A . However, during the initial morning hours, due to lesser solar energy production, the BESS supports maintaining a smooth power profile for the AWE until it is greater than the discharge power (P_D). Two data sets are considered in this study for P_s , one is based on the predicted solar data, and the second one comprises the actual weather data, both for the year 2018. Firstly the base model with predicted solar data set was run with A as 0.7, B as 0.3, and P_D as 700. Secondly, for Case 1(Optimized model), the base model was optimized to get the optimal values for A , B , and P_D . Thirdly for Case 3 (Base model with actual solar data), was run with optimal values evaluated in Case 2. The objective of the optimization was to maximize the hydrogen production, keeping the SOC within the operational constraints i.e. $SOC_{MIN} \leq SOC(t) \leq SOC_{MAX}$. The optimization formulation is shown below with the constraints mentioned in Eq. (8) – (11):

$$MAX_{A,P_D} (nH_{2_{STORED}}) \text{ s.t.}$$

$$0 \leq A \leq 1 \tag{8}$$

$$A + B = 1 \tag{9}$$

$$P_{D_{MIN}} \leq P_D \leq P_{D_{MAX}} \tag{10}$$

Figure SEQ Figure * ARABIC 2. and for base Case, Case 1 and Case 2.

$$SOC_{MIN} \leq SOC(t) \leq SOC_{MAX} \tag{11}$$

3. Results

As shown in Figure 1, the base model was run with A as 0.7, B as 0.3, and P_D as 700. The value P_D was kept higher to maximize BESS's utilization during night hours. The cumulative hydrogen produced for Base Case, Case 1, and Case 2 can be seen in Figure 2. Each graph in Figure 2 on the top right corner shows a greater resolution of

time from 70 to 100 h to better understand the main graph. Figure 2 shows that the Base case had the lowest hydrogen produced among all the discussed cases with 157,280 kg. Whereas the optimized case i.e., Case 1, had the highest production of 196,024 kg. Throughout the year, Base case faced a total of 3683 startups due to the lack of solar or not having enough power in the BESS to support the AWE operation. These startups were reduced by 39.14% for Case 1 (optimized case). Results show that an optimized operation not only produced higher production but also reduced the number of startups. Optimized decision variables from Case 1 were then given as an input for Case 2. Case 2 was powered by an actual solar profile to analyze the possibility of operation based on optimized operation for a predicted solar forecast. Hydrogen produced for Case 2 was 20.34% higher than the Base case. Furthermore, a reduction of 26.53 % in the startups throughout the year were reported. Results demonstrate that there is a possibility to pre-plan a year ahead of operations and manage any unforeseen situations to enable the plant's maximum potential based on predicted solar data and available resources. Furthermore, a year ahead of planning can help in monitoring the annual performance of the system. However, if this time is reduced from a year ahead to a week ahead of planning, the results would be different and may even result in improved control and performance. Currently, the model operates on time-invariant values. However, a huge potential lies if the model is operated with control variables that are optimized every hour based on the plant's current state and available renewable energy.

4. Conclusion

In this study, a 4.5 MW AWE dynamic operation was analysed for a year operation. The model was tested on a predicted solar profile for the upcoming year. Later the model was optimized to find optimal values for the control variables to ensure maximum production while observing model constraints. Later, those optimal values were used to run the model with the actual weather data to see if there is a possibility to plan a year before for plant operation. Results demonstrated that planning based on the predicted solar data and optimized control variables could help in better plant operation and plan for any unforeseen situation in the future operation, keeping in mind the unpredictable nature of renewable energy. Furthermore, in future work, the control variables will be optimized for piecewise control, i.e., for each interval, an optimal control variable value will be found to ensure a more robust control and reliable plan for future operation based on the predicted solar profile.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT (2019M3E6A1064422, 2019M3E6A1064422).

References

- Brauns, J., and Turek, T. (2020). Alkaline Water Electrolysis Powered by Renewable Energy: A Review. *Processes* 8, 248. doi:10.3390/pr8020248.
- Eichman, J., Harrison, K., Peters, M., Eichman, J., Harrison, K., and Peters, M. (2014). Novel Electrolyzer Applications : Providing More Than Just Hydrogen Novel Electrolyzer Applications : Providing More Than Just Hydrogen. *NREL Rep.*, 1–24.
- Ezeanya, E. K., Massiha, G. H., Simon, W. E., Raush, J. R., and Chambers, T. L. (2018). System

- advisor model (SAM) simulation modelling of a concentrating solar thermal power plant with comparison to actual performance data. *Cogent Eng.* 5. doi:10.1080/23311916.2018.1524051.
- Hannah Ritchie, and Max Roser (2020). CO₂ and Greenhouse Gas Emissions. *Publ. online OurWorldInData.org*. Available at: <https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions#citation> [Accessed April 14, 2020].
- Hirscher, M., Yartys, V. A., Baricco, M., Bellosta von Colbe, J., Blanchard, D., Bowman, R. C., et al. (2020). Materials for hydrogen-based energy storage – past, recent progress and future outlook. *J. Alloys Compd.* 827, 153548. doi:10.1016/j.jallcom.2019.153548.
- Hoste, G. R. G., Dvorak, M. J., and Jacobson, M. Z. (2020). Matching Hourly and Peak Demand by Combining Different Renewable Energy Sources 1 . *Electric Grid Basics*.
- Kwon, S., Won, W., and Kim, J. (2016). A superstructure model of an isolated power supply system using renewable energy: Development and application to Jeju Island, Korea. *Renew. Energy* 97, 177–188. doi:10.1016/j.renene.2016.05.074.
- Mayyas, A., and Mann, M. (2018). Manufacturing Competitiveness - Analysis for Hydrogen Refueling Stations and Electrolyzers. *DOE Hydrog. Fuel Cells Progr.* June.
- Moreno-Benito, Marta, Agnolucci, P., Will McDowall, A., and Papageorgiou., L. G. (2016). Towards a sustainable hydrogen economy: role of carbon price for achieving GHG emission targets.No Title. *Comput. Aided Chem. Eng.* 38, 1015–1020.
- Pascuzzi, S., Anifantis, A. S., Blanco, I., and Mugnozza, G. S. (2016). Electrolyzer performance analysis of an integrated hydrogen power system for greenhouse heating a case study. *Sustain.* 8, 1–15. doi:10.3390/su8070629.
- Ulleberg, Ø. (2003). Modeling of advanced alkaline electrolyzers: A system simulation approach. *Int. J. Hydrogen Energy* 28, 21–33. doi:10.1016/S0360-3199(02)00033-2.

A Mixed Integer Dynamic Optimization Approach for a Hybrid-Stand Alone Solar and Wind Powered Alkaline Water Electrolyser for Renewable Hydrogen

Haider Niaz, J.Jay Liu*

Department of Chemical Engineering, Pukyong National University, Busan 48513, South Korea.

jayliu@pknu.ac.kr

Abstract

Renewable energy sources are sustainable, cost-effective, and environment-friendly alternatives for fossil fuels. However, renewable energy's intermittency and storage remain a challenge, leading researchers to study hybrid energy systems. In general, batteries can be considered as a short-time energy storage option. For sustainable and long term energy storage, hydrogen can be a suitable candidate. Among hydrogen production pathways, water electrolysis is considered a sustainable and widely accepted energy conversion route when powered by a hybrid renewable energy source (i.e., solar and wind energy). However, essential improvements are required for renewable hydrogen to be competitive enough with fossil fuels derived hydrogen energy. In this work, a mixed integer dynamic optimization (MIDO) approach is presented. The problem's objective is to minimize the Levelized cost of hydrogen (LCOH) (\$/kg) by determining the optimal size for a battery energy storage system (BESS). The hybrid system is connected with a BESS to ensure the electrolyzer's operation with no support from the grid electricity. The model is simulated over one year based on an actual solar and wind profile to include seasonal variation and its performance effect. With the inclusion of economic studies, a mixed integer dynamic optimization problem is constructed and presented a LCOH of 10.33 \$/kg and BESS capacity of 6575 kW. This study can be extended into simulation-based design tools for estimating the optimal size and approximate cost estimates for stand-alone renewable energy powered electrolyzer for hydrogen production for any given location.

Keywords: Alkaline water electrolyzer, Mixed integer dynamic optimization, hydrogen production, levelized cost of hydrogen

Nomenclature

PV	Photovoltaic	$nH_{2_{STORED}}$	Total hydrogen (kg)
AWE	Alkaline water electrolyzer	P_A	Power to AWE
BESS	Battery energy storage system	P_C	Power to BESS

SAM	System advisor model	P_D	Power from BESS
SOC	State of charge (%)	P_X	Power for sale
nH_2^{GEN}	Hydrogen produced (kg/hr)	$P_{C_{MIN}}, P_{C_{MAX}}$	Minimum and maximum charging power
η_{Ele}	Efficiency of electrolyzer (%)	$P_{D_{MIN}}, P_{D_{MAX}}$	Minimum and maximum discharging power
P	Power to AWE	SOC_{MIN}, SOC_{MAX}	Minimum and maximum state of charge
LHV	Lower heating value of H ₂	η^+, η^-	Charging and discharging efficiencies
P_S, P_W	Power from solar and wind	$LCOH$	Levelized cost of Hydrogen

1. Introduction

Fossil fuel extinction is a real threat to the upcoming generations. Their excessive use has done enough damage to our environment and will take years to recover from it. A radical step must be taken to control the ever-increasing emissions and replace fossil fuels with environment-friendly energy, providing renewable energy resources (Tokarska and Gillett, 2018). Renewable energy being dynamic creates a void at being a reliable energy source. But storing this energy in batteries can be a good remedy. Storing all energy in batteries can cost us several million dollars, but storing renewable energy in the form of hydrogen can help us overcome this huge capital investment as hydrogen has the properties to store the largest energy content among fuels (Nicita et al., 2020).

electrolyzers are recognized as green hydrogen generators when powered by a renewable energy source. The possibility of combining two or more renewable energy sources such as solar, wind, and hydropower to produce hydrogen has also been studied (Gökçek and Kale, 2018; Hoste et al., 2020). A battery energy storage system (BESS) can play an important role in ensuring a consistent energy supply to the electrolyzers. BESS can help minimize the power fluctuations by supplying the power when needed and store when there is enough renewable generation. In this context, a combination of a commercially accepted electrolyzer technology, i.e., alkaline water electrolyzer (AWE), a BESS, and solar and wind power for hydrogen generation, will be a perfect and promising candidate.

In this study, a 4.5 MW AWE is powered by a combination of 4.5 MW solar, 2.0 MW wind, and a 1.0 MW BESS. The study focuses on solving a mixed integer dynamic optimization problem to find the optimal size for a BESS, maximize the hydrogen production, and minimize the Levelized cost of hydrogen (LCOH) (\$/kg). The model is simulated over one year with one hour time interval to accommodate the solar and wind generation's seasonal variations. This study's dynamic and economic perspective can help evaluate the overall plant costs at minimum capital expenditure (CAPEX) by evaluating the optimal size for the BESS. Future work will consider minimizing the operational cost

and CAPEX based on the available hydrogen demand and required optimal size for AWE, solar, and wind capacity.

2. Method

The solar irradiation and wind data for the year 2018 was taken from Phoenix, Arizona, USA, was obtained from the System Advisor Model (SAM) (Ezeanya et al., 2018). SAM was used to simulate a 4.5 MW PV and Wind system's voltage and power profile. An overall block diagram for the solar and wind-powered AWE & BESS system with hydrogen storage can be seen in Figure 1.

Figure . Solar and Wind Powered 4.5 MW AWE with BESS and hydrogen storage.

The general equations representing the dynamic operation of AWE and moles of hydrogen generated (nH_2^{GEN}) can be represented, as shown in Eq. (1).

$$nH_2^{GEN}(t) = \frac{\eta_{Ele} \times P}{LHV} \quad (1)$$

Where η_{Ele} is the efficiency of electrolyzer (68%), P the power to the electrolyzer, and LHV is lower heating value of hydrogen. The amount of hydrogen being stored ($nH_{2_{STORED}}$) can be represented, as shown in Eq. (2), where t is equal to 8760 hrs.

$$nH_{2_{STORED}}(t) = \int_0^t nH_2^{GEN}(t') dt' \quad (2)$$

To ensure a safe operation, some constraints must be considered for BESS operation and are enlisted below in Eq. (3) – (6):

$$I. \quad P_C(t) \times P_D = 0 \quad (3)$$

$$II. \quad P_{C_{MIN}} \leq P_C(t) \leq P_{C_{MAX}} \quad (4)$$

$$III. \quad P_{D_{MIN}} \leq P_D \leq P_{D_{MAX}} \quad (5)$$

$$IV. \quad SOC_{MIN} \leq SOC(t) \leq SOC_{MAX} \quad (6)$$

P_C is the power to BESS, $P_{C_{MAX}}$ is the maximum charging power for BESS, P_D is the power discharged from the BESS, $P_{D_{MIN}}$ is the minimum discharge power from the BESS to ensure hydrogen production during the night at lowest possible power. $P_{D_{MAX}}$ is the maximum discharge power from the BESS, SOC is the state of charge of the BESS, and SOC_{MIN}, SOC_{MAX} represent the minimum and maximum SOC level respectively. In this study, SOC_{MIN}, SOC_{MAX} are 30% and 100%, respectively. Overall SOC can be represented, as shown in Eq.(7).

$$SOC(t) = SOC(t - \Delta t) + \left((\eta^+ \times P_C(t)) - (\eta^- \times P_D(t)) \right) \Delta t \quad (7)$$

Where η^+, η^- are the charging and discharging efficiencies. Charging and discharging efficiencies consider 5% loss.

The data used for modeling the electrolyzer and values of the model's parameters are taken from (Ulleberg, 2003; Mayyas and Mann, 2018). gPROMS platform was used to model the dynamic equations for the alkaline water electrolyzer. An illustration for the overall model can be seen in Figure 1. P_S represents the sum of solar and wind power, A/B represent the split fractions for the power to AWE ($P_A = P_S \times A$) and power to BESS ($P_B = P_S \times B$). P_X represents the surplus power when neither the AWE and neither the BESS needs power, i.e., either the BESS is fully charged, or the AWE is already operating at maximum load. P_X is used as a product and is assumed to be sold to neighboring industries. P represents the total power and is the sum of $P_A + P_D$.

Solar and wind power (P_S) is split into two fractions (A) and (B), as shown in Figure 1. These fractions are set at different ratios. In this model, (A) is 0.7, then (B) will be 0.3, provided their sum always counts to 1.0. P_A is the power split going to the electrolyzer, and P_C is the power split going to the BESS for charging the batteries. In general, during day time (06:00 AM to 06:00 PM), power (P) is primarily the power coming from P_A . However, at times with lesser solar and wind energy production, the BESS supports maintaining a smooth power profile for the AWE until it is greater than the discharge power (P_D).

Technoeconomic model and the method to calculate the LCOH used in this study were similar to the National Renewable Energy Laboratory (NREL) (Short et al., 1995; Dutta et al., 2011). Other raw materials (i.e., steam and KOH) consumption per kg of hydrogen produced for hydrogen production were estimated using the ones quoted in the NREL

(Ivy, 2004). The cost of the Electrolyser is taken from Gim and Yoon (Gim and Yoon, 2012). The cost of storage and compression equipment was taken from NREL and other sources. The plant life was assumed to be 30 years with an IRR of 4.5%. Our study's startup time was taken as 0.25 years with a working capital equivalent of 5% of the fixed capital investment (FCI). LCOH was calculated using NREL's equation, as shown in Eq.8 (Short et al., 1995).

$$LCOH = (TLCC - NPV_{OTHER-PRODUCTS}) / \sum_0^L (nH_2^{GEN} / (1 + D)^n) \tag{8}$$

Where $TLCC$ is the total life cycle cost of the plant, $NPV_{OTHER-PRODUCTS}$ is the net present value of the other products, i.e., P_x , D is the discount rate and n is the analysis year, and L is the analysis period in years i.e., 30.

Figure . and for a 4.5 MW AWE powered by 4.5 MW solar and 2.0 MW wind power.

The objective of the optimization was to minimize the $LCOH$ keeping the SOC within the operational constraints i.e. $SOC_{MIN} \leq SOC(t) \leq SOC_{MAX}$. $CAPEX_{BESS}$ can be represented as shown in Eq. (9).

$$CAPEX_{BESS} = CAP_{BESS} * COST_{kW} \tag{9}$$

Where $COST_{kW}$ represents the \$/kW cost for the battery. For $CAPEX_{BESS}$, SAM was used to evaluate the total cost for the BESS (Ezeanya et al., 2018). SAM incorporates the number of cells required for a specific size of a battery and can present the \$/kW for any BESS capacity. The optimization formulation is shown below with the constraints mentioned in Eq. (9) – (13), where A, B, P_D are continuous values and CAP_{BESS} is an integer value:

$$MIN_{A, P_D, CAP_{BESS}} (LCOH) \text{ s.t.}$$

$$0 \leq A \leq 1 \tag{9}$$

$$A + B = 1 \tag{10}$$

$$P_{D_{MIN}} \leq P_D \leq P_{D_{MAX}} \quad (11)$$

$$SOC_{MIN} \leq SOC(t) \leq SOC_{MAX} \quad (12)$$

$$CAP_{BESS}^{MIN} \leq CAP_{BESS} \leq CAP_{BESS}^{MAX} \quad (13)$$

gPROMS platform was used to solve the problem. Solvers used for dynamic optimization was CVP_SS (control vector parameterization), and for MINLP, OAERAP solver was employed, which uses an outer approximation (OA) algorithm for the solution of MINLP.

3. Results

Figure 2 demonstrates the power (P) to power the AWE. Followed by the hydrogen produced in a year. In Figure 2 the top right corner graph shows a greater time resolution for the main power graph. The optimization resulted in a cumulative nH_2^{GEN} of 223 tons of hydrogen a year. Keeping in mind the other objectives of the optimization was to evaluate the operational variables, i.e., A and P_D , for an optimal operation. Results demonstrated and proposed a value of 1 for A and 8 kW for P_D . It is assumed that 8 kW is the minimum possible power at which the AWE can run. Based on these values, the optimal size for the BESS was taken as 6575 kW.

Apart from operational results, the techno-economic model evaluated the LCOH, operational, and the CAPEX for the plant. Results demonstrated that for a 4.5 MW AWE powered by 4.5 solar and 2.0 MW wind turbine with a 6575 kW of BESS will have an equipment cost of \$10.81 MM. Major equipment costs include \$3.6 MM for a solar plant, \$2.83 for AWE, \$1.57 MM for BESS and \$ 2.8 MM for the wind turbine, and other costs, including compression storage and cost for inverters. The project resulted in an LCOH of \$10.33 /kg.

4. Conclusion

In this study, a mixed integer dynamic optimization was performed for a 4.5 MW AWE for a year operation using solar and wind power coupled with a BESS. The objective of the optimization was to maximize the hydrogen and minimize the LCOH and CAPEX for BESS. Results demonstrated that a dynamic operation and economic optimization could overcome and consider the seasonal fluctuations for a renewable energy powered project and provide realistic cost estimates. Furthermore, optimal sizing for the BESS can reduce the oversizing of the plant and save cost. In future work, this study will be extended into a simulation-based design tool for estimating the optimal size and approximate cost estimates for stand-alone renewable energy powered electrolyzer for hydrogen production for any given location.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT (2019M3E6A1064422, 2019M3E6A1064422).

References

- Dutta, A., Talmadge, M., Nrel, J. H., Worley, M., Harris, D. D., Barton, D., et al. (2011). Process Design and Economics for Conversion of Lignocellulosic Biomass to Ethanol Thermochemical Pathway by Indirect gasification and mixed alcohol synthesis. *Renew. Energy*. doi:10.2172/1015885.
- Ezeanya, E. K., Massiha, G. H., Simon, W. E., Raush, J. R., and Chambers, T. L. (2018). System advisor model (SAM) simulation modelling of a concentrating solar thermal power plant with comparison to actual performance data. *Cogent Eng.* 5. doi:10.1080/23311916.2018.1524051.
- Gim, B., and Yoon, W. L. (2012). Analysis of the economy of scale and estimation of the future hydrogen production costs at on-site hydrogen refueling stations in Korea. *Int. J. Hydrogen Energy* 37, 19138–19145. doi:10.1016/j.ijhydene.2012.09.163.
- Gökçek, M., and Kale, C. (2018). Techno-economical evaluation of a hydrogen refuelling station powered by Wind-PV hybrid power system: A case study for İzmir-çeşme. *Int. J. Hydrogen Energy* 43, 10615–10625. doi:10.1016/j.ijhydene.2018.01.082.
- Hoste, G. R. G., Dvorak, M. J., and Jacobson, M. Z. (2020). Matching Hourly and Peak Demand by Combining Different Renewable Energy Sources 1 . *Electric Grid Basics*.
- Ivy, J. (2004). Summary of electrolytic hydrogen production. *Small*, 27.
- Mayyas, A., and Mann, M. (2018). Manufacturing Competitiveness - Analysis for Hydrogen Refueling Stations and Electrolyzers. *DOE Hydrog. Fuel Cells Progr.* June.
- Nicita, A., Maggio, G., Andaloro, A. P. F., and Squadrito, G. (2020). Green hydrogen as feedstock: Financial analysis of a photovoltaic-powered electrolysis plant. *Int. J. Hydrogen Energy*, 1–20. doi:10.1016/j.ijhydene.2020.02.062.
- Short, W., Packey, D., and Holt, T. (1995). A manual for the economic evaluation of energy efficiency and renewable energy technologies. *Renew. Energy* 95, 73–81. doi:NREL/TP-462-5173.
- Tokarska, K. B., and Gillett, N. P. (2018). Cumulative carbon emissions budgets consistent with 1.5 °c global warming. *Nat. Clim. Chang.* 8, 296–299. doi:10.1038/s41558-018-0118-9.
- Ulleberg, Ø. (2003). Modeling of advanced alkaline electrolyzers: A system simulation approach. *Int. J. Hydrogen Energy* 28, 21–33. doi:10.1016/S0360-3199(02)00033-2.

A framework for adaptive online model-based redesign of experiments in dynamic systems

Arun Pankajakshan,^a Panagiotis Petsagkourakis,^a Federico Galvanin^{a,*}

*^aDepartment of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK
f.galvanin@ucl.ac.uk*

Abstract

Online model-based redesign of experiments (OMBRE) methods are advanced methods to optimize input controls and sampling times to attain data with the highest information for precisely estimating the parameters of a dynamic process model. The way OMBRE methods are implemented is to first design a dynamic experiment, but only implement it over a fixed time interval, then perform parameter estimation with data collected in that interval to update the process model as the process is running and, finally, optimizing the experiment repeatedly until the experiment is concluded. In this work, we propose an adaptive version of OMBRE (A-OMBRE) where an algorithm determines when to re-optimize/update the current experiment based on parameter estimation results. The proposed method does not require any user input regarding the updating policy, instead it uses the real time parametric uncertainty to determine the updating times. The method is tested on a simulated case study of fermentation kinetics and is compared with i) a conventional offline design; ii) an OMBRE approach with user-defined redesign strategy.

Keywords: model-based redesign of experiments, adaptive design, collocation

1. Introduction

Identification of reliable mathematical models for dynamic systems is crucial as dynamic simulation mimics the behaviour of real plant and can be used to represent plant's real time operation, control and optimization. In several systems, mathematical models with a set of unobservable parameters are constructed from the physical laws governing the process and identification of those models needs precise estimation of the model parameters. Model-based design of experiments (MBDoE) methods are applied to dynamic systems with the purpose of designing a set of experiments yielding the most informative process data to be used for the estimation of the process model parameters (Franceschini and Macchietto, 2008). In conventional MBDoE methods, the experiment design phase, the actual experimental phase and the parameter estimation phase are carried out in a strictly sequential manner. This means that information is exploited only when the experiment is concluded. In contrast, in online MBDoE methods (Galvanin et al., 2009) a dynamic experiment is initially designed by optimizing input control configurations, allocation of sampling times, initial values of states and experiment duration, but the experiment is then redesigned within predefined time intervals ("updating intervals"). In each updating interval the information regarding model parameters is updated using parameter estimation techniques using past data and the updated model is used for designing the remaining part of the experiment repeatedly, until the model parameters are determined with minimal variance (Galvanin et al., 2009). Thus, in the online model-based redesign of experiments (OMBRE) method, information

content in the data is used during the experiment. The idea of online redesign was first proposed in Mehra (1974), applied to linear systems and later extended to nonlinear systems (Stigter et al., 2006). Compared to the conventional offline design, online experimental redesign methods offer great potential for the real time parameter estimation and model identification. However, a challenging aspect of the online redesign methods is the systematic choice of the number of redesigns and duration of each updating interval which are usually given as user-defined inputs. An improper choice of the above updating policy without system knowledge may lead to process data with insufficient information for parameter estimation in each updating interval. In De-Luca et al. (2016), authors proposed a method to minimize the duration of updating intervals while ensuring a sufficient reduction in parametric uncertainty in each updating interval. However, the method seems to be feasible for systems with nearly continuous measurements. More recently, a method based on auto-updating of sampling time redesign (Wang and Yue, 2019) was proposed to determine the length of updating intervals to assure identifiability for parameter estimation. To tackle the issue of properly determining the updating times, a framework for adaptive online model-based redesign of experiments for the identification of dynamic models is proposed in this work. The framework introduces a rational criterion based on the real time parametric uncertainty assessed through likelihood function of model parameters for the redesign of experiments. In the framework, orthogonal collocation method (Biegler, 2010) is used to discretize the dynamic model and the optimization is formulated as a mixed integer nonlinear programming (MINLP) problem to facilitate the optimal allocation of sampling points. The proposed redesign framework is tested on a simulated case study and compared with offline design and online redesign with user-defined redesign strategy.

2. Problem statement

The optimal design of dynamic experiments can be formulated as a dynamic optimization problem of the form of Eq. (1). The problem seeks the optimal design vector $\boldsymbol{\phi}$ that excites the dynamic system in order to optimize the function ψ towards the end of the experiment t_f . For convenience, the set of optimization variables are aggregated to form the design vector $\boldsymbol{\phi}$, which usually contains the vector of initial conditions \mathbf{y}_0 , the experiment duration t_f , the vector of input trajectories $\mathbf{u}(t)$ and the N_{sp} -dimensional set of sampling points \mathbf{t}_{sp} i.e., $\mathbf{t}_{sp} = [\mathbf{y}_0, t_f, \mathbf{u}(t), \mathbf{t}_{sp}]$. The basis for such a problem is the dynamic model \mathbf{f} that describes the evolution of the state variables with time $\mathbf{x}(t)$, starting from the initial condition $\mathbf{x}(0)$.

$$\begin{aligned}
 & \underset{\boldsymbol{\phi}}{\text{minimize}} && \psi(\mathbf{V}_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \boldsymbol{\theta})) \\
 & \text{subject to} && \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t)), \hat{\mathbf{y}}(t) = \mathbf{g}(\mathbf{x}(t)) \quad \forall t \in [0, t_f] \\
 & && \mathbf{x}(t) \in \mathcal{X}, \mathbf{u}(t) \in \mathcal{U}, \quad \forall t \in [0, t_f] \\
 & && \mathbf{x}(0) \in \mathcal{X}_0, \boldsymbol{\theta} \in \Theta, \mathbf{t}_{sp} \in [0, t_f]
 \end{aligned} \tag{1}$$

In Eq. (1), $\mathbf{x}(\cdot) \in \mathbb{R}^{N_x}$ is the vector of differential states, $\hat{\mathbf{y}}(\cdot) \in \mathbb{R}^{N_y}$ is the vector of model predictions for the response variables $\mathbf{y}(t)$ (those state variables which are actually measured), $\mathbf{u}(\cdot) \in \mathbb{R}^{N_u}$ is the vector of input controls and $\boldsymbol{\theta} \in \mathbb{R}^{N_{\theta}}$ is the vector of model parameters. The differential states $\mathbf{x}(\cdot)$, controls $\mathbf{u}(\cdot)$ and parameters $\boldsymbol{\theta}$ are bounded within the admissible sets \mathcal{X} , \mathcal{U} and Θ respectively; $\mathbf{f}: \mathbb{R}^{N_x} \times \mathbb{R}^{N_{\theta}} \times \mathbb{R}^{N_u} \mapsto \mathbb{R}^{N_x}$ and $\mathbf{g}: \mathbb{R}^{N_x} \mapsto \mathbb{R}^{N_y}$

are differentiable functions. In Eq. (1), the predicted parameter covariance matrix $\mathbf{V}_\theta(\boldsymbol{\varphi}, \boldsymbol{\theta}, t_f)$ toward the end of the experiment is defined as

$$\mathbf{V}_\theta(\boldsymbol{\varphi}, \boldsymbol{\theta}, t_f) = \left\{ \left[\mathbf{V}_\theta^0 \right]^{-1} + \sum_{t=t_1}^{t_{N_{sp}}} \mathbf{Q}^T(t) \boldsymbol{\Sigma}_Y^{-1} \mathbf{Q}(t) \right\}^{-1} \quad (2)$$

In Eq. (2), $\boldsymbol{\Sigma}_Y$ denotes the covariance matrix of measurement errors which is assumed as a diagonal matrix with variances along the diagonal, \mathbf{V}_θ^0 denotes the prior uncertainty of the model parameters and $\mathbf{Q}(t)$ is the dynamic sensitivity matrix (Franceschini and Macchietto, 2008).

3. Methodology

In the online experimental redesign framework proposed in this work called as adaptive online model-based redesign of experiments (A-OMBRE), dynamic experiment design problem in Eq. (1) is solved using the simultaneous discretization approach using orthogonal collocation (Biegler, 2010). The method discretizes the continuous domains in the model and approximates the differential equations using algebraic equations defined at the discretization points. To incorporate the constraint on the number of measurements, we modify the optimization problem of Eq. (1) (discretized version) into a mixed integer non-linear programming (MINLP) problem in which the objective function is evaluated as

$$\mathbf{V}_\theta(\boldsymbol{\varphi}, \boldsymbol{\theta}, t_f) = \left\{ \left[\mathbf{V}_\theta^0 \right]^{-1} + \sum_{i=1}^N b_i \cdot \left(\mathbf{Q}^T(t_i) \boldsymbol{\Sigma}_Y^{-1} \mathbf{Q}(t_i) \right) \right\}^{-1} \quad (3)$$

where, $t_i \in [0, t_f]$, $b_i \in \{0, 1\}$, $\sum_{i=1}^N b_i = N_{sp}$

In Eq. (3), the binary variables b_i are used to choose the N_{sp} number of sampling points from the set of discretization points. Given the experimental budget in terms of number of samples allowed (N_{sp}) and the maximum duration of the experiment (t_f^{ub}), Eq. (1) is solved using the prior estimate $\boldsymbol{\theta}_0$ and the prior uncertainty \mathbf{V}_θ^0 to obtain the optimal experimental settings $\boldsymbol{\varphi}$ at which the dynamic experiment is started. This is shown in the first three blocks from top of the flow sheet in Figure 1 and in the first line of Algorithm 1. The first parameter estimation (PE) is carried out after sufficient data (at least as many as twice the number of model parameters; an empirical rule used in this work) is obtained, which is necessary for a proper estimation of the parameters and the variances. In cases where all parameters are not estimated well, which is most likely after the first parameter estimation, the likelihood ratio (LR) test (Quaglio et al., 2019) is used for an additional decision making which involves whether to ‘CONTINUE’ the running experiment or to ‘REDESIGN’ the experiment. At this purpose, the likelihood ratio (LR) test is carried out against a null hypothesis which states there is no significant difference between the likelihood functions at the prior $\boldsymbol{\theta}_0$ and at the new estimate $\hat{\boldsymbol{\theta}}$ obtained from the parameter estimation. The LR is evaluated by computing the statistic according to Eq. (4):

$$LR = -2 \left(\log L(\boldsymbol{\theta}_0 | \mathbf{y}(t_1, \dots, t_{n_{sp}})) - \log L(\hat{\boldsymbol{\theta}} | \mathbf{y}(t_1, \dots, t_{n_{sp}})) \right) \sim \chi^2(N_\theta, \alpha) \quad (4)$$

If the LR test appears to be statistically significant (p-value < 0.05), the current experiment which is optimized using the prior $\boldsymbol{\theta}_0$ might become sub-optimal and the remaining part of

the experiment is then redesigned using the new estimate $\hat{\theta}$ (updated prior) and modified constraints. If the LR test is not significant, there is no evidence to disprove the null hypothesis and the running experiment remains optimal and hence continues. This concept is explored to decide whether the experiment needs to be redesigned after a new sample has been collected. The above redesign strategy is shown in the second decision block of the flow sheet in Figure 1 and in lines 4 to 11 of Algorithm 1.

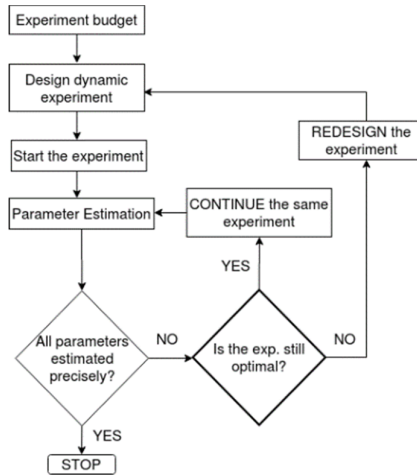


Figure 1. Flowsheet for the proposed adaptive ombre framework

Algorithm 1 Adaptive OMBRE

- 1: Input: N_{sp} , t_f^{ub} , \mathbf{V}_0^0 ; solve Eq. (1), return ϕ
- 2: Start exp; wait until $n_{sp} = 2 \cdot N_0 / N_y$
- 3: Solve PE using $y(t_1, \dots, t_{n_{sp}})$, return $\hat{\theta}$, \mathbf{V}_θ
- 4: **if** \mathbf{V}_θ is satisfactory **then** STOP
- 5: **else** compute LR statistic using Eq. (4)
- 6: **if** LR test is significant **then** modify and solve Eq. (1) and replace running exp.
- 7: **else** Continue the same experiment
- 8: **end if**
- 9: $n_{sp} = n_{sp} + 1$, solve PE and go to step 4
- 10: Continue step 9 until $n_{sp} = N_{sp}$
- 11: **end if**

4. Case study

A process for the fermentation of biomass by Baker's yeast presented in (Cooney and McDonald, 1995) is chosen as the case study to test the performance of the proposed algorithm. A Monod-type kinetic model which is described by following the set of DAEs is assumed to characterize the real system.

$$\begin{aligned}
 \frac{dx_1}{dt} &= (r - u_1 - \theta_4) x_1 \\
 \frac{dx_2}{dt} &= -\frac{r x_1}{\theta_3} + u_1 (u_2 - x_2) \\
 r &= \frac{\theta_1 x_2}{\theta_2 + x_2}
 \end{aligned} \tag{5}$$

In Eq. (5), the state variables x_1 and x_2 represent the biomass concentration [g L^{-1}] and the substrate concentration [g L^{-1}] respectively. The control variables in the model include the dilution factor [h^{-1}] and the substrate concentration in the feed [g L^{-1}], which are respectively denoted as $u_1(t)$ and $u_2(t)$. The experimental design vector ϕ contains the initial biomass concentration x_1^0 (range: 1-10 g L^{-1}), the experiment duration with an upper bound t_f^{ub} of 18 h, the dilution factor $u_1(t)$ (range: 0.05-0.20 h^{-1}), the substrate concentration in the feed $u_2(t)$ (range: 5-35 g L^{-1}) and the set of N_{sp} sampling points at which the state variables x_1 and x_2 are measured. The assumed true values for model parameters are given in Table 1 which are used to generate in-silico measurements by

adding random noise (Σ_V with 0.01 and 0.05 along the diagonal) to the simulated values. The assumed initial values of model parameters (with 90 % relative error with respect to the true values) are also provided in Table 1. The entire time horizon is divided into 18 equal control intervals or finite elements. The 18 finite element points except the starting point was allowed as potential sampling points in the MINLP design formulation. The maximum number of sampling points allowed (N_{sp}) was assumed as 12. An A-optimal criterion was used to design the experiments in all the design configurations: 1) a conventional offline method (MBDoE); 2) an online redesign method with pre-defined updating times (OMBRE); 3) the proposed method (A-OMBRE). The comparison is shown in Figure 2. In all the three cases, the experimental design involved four-point Legendre collocation, 18 finite elements and the controls have been discretized as piecewise constant profiles. The same initialisation and variable bounds discussed above were used in all the three cases.

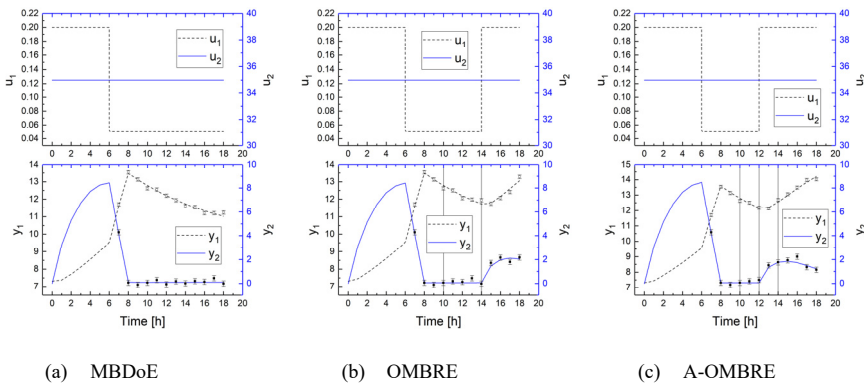


Figure 2. Experimental design results using three different design strategies (MBDoE, OMBRE, A-OMBRE). Figure shows the optimal control trajectories $\mathbf{u}(t)$ and the predicted state trajectories $\mathbf{y}(t)$. Measured states within the error bar are shown at the optimal allocation of sampling points. The vertical lines on panel b and c indicate the updating instances

In MBDoE, Eq. (1) was solved to obtain the optimal controls and sampling times shown in Figure 2(a). In OMBRE method, Eq. (1) was solved initially to obtain the design vector ϕ . Later, the running experiments were redesigned twice, each time after collecting new $2 \cdot N_\theta$ measurements or $2 \cdot N_\theta / N_y$ sampling points. The proposed A-OMBRE method followed the methodology discussed in section 3 and involved three redesigns. The optimal control settings, sampling times and the updating times in OMBRE and A-OMBRE are shown in Figure 2(b) and Figure 2(c) respectively. The comparison of parameter estimation results using the three different design strategies (Table 1) indicates that slightly more precise parameter estimation was achieved in the OMBRE method compared to the proposed adaptive OMBRE method. However, the significant advantage of the proposed method is that it allows to track the parametric uncertainty with respect to the optimal experimental design in real time with the help of a hypothesis testing (LR test) problem. The details of hypothesis testing problem (discussed in section 3) which is used to correlate the updating decision with parametric uncertainty in the proposed A-OMBRE method is illustrated in Table 2.

Table 1. Parameter estimation results including estimates and 95 % confidence intervals at the end of three different design strategies

Model parameter	True value	Initial value	Estimates from MBD _{oE}	Estimates from OMBRE	Estimates from A-OMBRE
θ_1	0.31	0.589	0.313 ± 0.021	0.307 ± 0.007	0.312 ± 0.013
θ_2	0.18	0.018	0.257 ± 0.328	0.153 ± 0.039	0.153 ± 0.057
θ_3	0.55	1.045	0.549 ± 0.012	0.549 ± 0.013	0.561 ± 0.028
θ_4	0.05	0.005	0.049 ± 0.003	0.050 ± 0.005	0.054 ± 0.012

Table 2. Redesign decision criteria in the adaptive ombre framework

Parameter estimation	Time [h]	LR statistic	p value	Decision
1	11	4.54	0.33	CONTINUE
2	12	13.17	0.01	REDESIGN
3	13	0.92	0.92	CONTINUE
4	14	11.65	0.02	REDESIGN
5	15	0.07	0.99	CONTINUE
6	16	0.77	0.94	CONTINUE
7	17	1.23	0.87	CONTINUE
8	18	0.49	0.97	CONTINUE

5. Conclusion

In this work, a novel algorithm is proposed to automatically redesign a dynamic experiment in the online model-based redesign of experiments method which is used for precisely estimating the parameters of a dynamic process model. The proposed algorithm uses the updated parametric uncertainty to redesign a dynamic experiment. The effectiveness of the algorithm is tested on a simulated case study by comparing the proposed algorithm to a standard offline MBD_{oE} and an OMBRE method. The algorithm appears to be promising for online identification of dynamic models due to its ability to automate the calculation of updating times in redesign policies based on a rational criteria, an approach lacking in current state of the art algorithms.

References

- Biegler LT. Nonlinear programming: concepts, algorithms, and applications to chemical processes. Siam; 2010.
- Cooney MJ, McDonald KA. Optimal dynamic experiments for bioreactor model discrimination. *Appl Microbiol Biotechnol* 1995.
- De-Luca R, Galvanin F, Bezzo F. A methodology for direct exploitation of available information in the online model-based redesign of experiments. *Comput Chem Eng* 2016;91:195–205.
- Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sci* 2008;63:4846–72.
- Galvanin F, Barolo M, Bezzo F. Online model-based redesign of experiments for parameter estimation in dynamic systems. *Ind Eng Chem Res* 2009;48:4415–27.
- Mehra RK. Optimal Input signals for Parameter Estimation in Dynamic Systems—Survey and New Results. *IEEE Trans Automat Contr* 1974;19:753–68.
- Quaglio M, Fraga ES, Galvanin F. Statistical diagnosis of process-model mismatch by means of the Lagrange multiplier test. *Comput. Aided Chem. Eng.*, 2019.
- Stigter JD, Vries D, Keesman KJ. On adaptive optimal input design: A bioreactor case study. *AIChE J* 2006;52:3290–6.
- Wang K, Yue H. Auto-updating of sampling time redesign for system identification under parameter uncertainty. *ICAC 2019 - 2019 25th IEEE Int. Conf. Autom. Comput.*, 2019.

Investigation of the use of transient process data for steady-state Real-Time Optimization in presence of complex dynamics

Rodrigo Curvelo^{a,*}, Pedro de A. Delou^a, Maurício B. de Souza Jr.^{a,b} and Argimiro R. Secchi^a

^a*Chemical Engineering Program/COPPE, Federal University of Rio de Janeiro, Rio de Janeiro - Brazil*

^b*Chemical Engineering Department/School of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro – Brazil*
curvelo@peq.coppe.ufrj.br

Abstract

Hybrid Real-Time Optimization (HRTO) approaches consist of the economical steady-state optimization performed after the dynamic adaptation of the process model, usually carried out by an Extended Kalman Filter. Despite the increasing number of works concerning this technique, the literature still lacks a larger number of case studies supporting these alternatives, including cases that present nonlinear behavior at certain operating ranges and model uncertainties. This work aims to analyze the HRTO applied to a wide range of examples, exploring systems presenting different transient patterns, including slow (or fast) non-minimum (or minimum) phase behavior. Different RTO frameworks were evaluated, and nonlinear model predictive control was implemented as the advanced control layer for all architectures. As expected, under no structural uncertainties, the reached steady-state is the same for all approaches. Concerning transient operation, the area under the objective function curve was calculated by numerical integration and applied as a quantitative performance indicator, with the DRTO curve set as a reference since it outperforms its counterparts. The better performance of the HRTO for several systems with complex dynamics gives this approach more possibilities for industrial application, but care should be taken for systems with long time delays.

Keywords: Steady-state optimization, Dynamic optimization, Hybrid Real-time optimization, Kalman Filter, Van de Vusse Reactor.

1. Introduction

Real-Time Optimization (RTO) algorithms have a large participation in the industry, especially in petrochemical processes (Ruiz, 2009). Traditionally, steady-state model parameters are updated through real-time steady-state process measurements. After this step, an economic optimization is carried out with the updated steady-state model, generating the optimal setpoint for the controlled variables (Melo et al., 2009). Regarding the parameter estimation and steady-state optimization steps, the RTO is commonly referred to as a *two-step approach* (Trierweiler, 2014). One of the main challenges involving traditional RTO algorithms is the steady-state detection step, which usually uses statistics or heuristics to assess whether the system is sufficiently close to steady-state operation (Cao and Rhinehart, 1995). This matter becomes more relevant when facing real large-scale processes, with frequent transitions and long

transient dynamics (Bousbia-Salah et al., 2017).

Dynamic RTO (DRTO), unlike traditional RTO, performs the calculation of the optimal trajectory of the decision variables of the plant through a dynamic model of the process, making the steady-state detection step no longer needed. Despite this advantage, a major problem arises: the computational cost associated with the dynamic optimization. Solving a dynamic optimization problem for nonlinear systems can be a highly costly task, especially for large-scale systems. Given this fact alone, the industry in general still opts to apply the traditional RTO (Krishnamoorthy et al., 2018).

Driven by these limitations, some authors (Krishnamoorthy et al., 2018; Matias and Le Roux, 2018) have suggested an alternative hybrid approach for the real-time optimization that combines the strengths of traditional RTO and DRTO: the static optimization combined with the dynamic adaptation of the model. As this is a very recent technique, there are not many applications reported in the literature. Thus, the main goal of this work is to evaluate how hybrid RTO performs in processes presenting nonlinear and non-minimum phase behaviors and to evaluate its robustness under diverse scenarios, comparing different real-time optimization architectures.

2. Real-time optimization frameworks

Model-based real-time optimization techniques are imperative to deal with parametric plant-model mismatch (Serralunga et al., 2012). Structural divergences can make the system erroneously reach operating points that differ from the real optimum, highlighting the importance of a well-designed estimator. Thus, four RTO architectures were considered:

- **RTO-LS:** a traditional two-step RTO. First, the plant runs until its stationarity is addressed by a steady-state detection step. Once the process is at steady-state, its model is updated by a nonlinear least-squares estimator in order to perform the steady-state optimization. This optimizer returns the optimal controlled variables setpoints and the optimal manipulated variables targets that are fed into the advanced control layer with the updated model parameters. It is noteworthy that the controller only initiates when the first pair setpoint-target is calculated, i.e., when the plant is at steady-state for the first time;
- **RTO-EKF:** this framework differs from the previous RTO scheme when it comes to the model update phase. An extended Kalman filter was designed and acts as a state-parameter estimator at all sampling steps. This comes with the reasoning that, since non-measured disturbances are only estimated at steady-state, setpoint control is compromised when the process is at transient operation;
- **DRTO:** traditional dynamic real-time optimization, i.e, the optimization problem and estimation step consider the dynamic model of the process;
- **HRTO:** the original approach designed by Krishnamoorthy et al. (2018) was applied. Similarly to the mentioned RTO-EKF, the steady-state detector was removed. Thus, the dynamic model of the process is updated combined with the optimization of the steady-state model.

Concerning the mathematical formulation, more details can be found at Krishnamoorthy et al. (2018), where the optimization and estimation problems are stated for traditional

RTO (RTO-LS in the present work), DRTO, and HRTO.

3. Illustrative example: the Van de Vusse CSTR

The Van de Vusse reactor (Van de Vusse, 1962) is a continuous stirred-tank reactor containing a reaction system in which cyclopentadiene is used to synthesize cyclopentanol by electrophilic addition of water. The reaction is catalyzed by an acid solution (Diels-Alder reaction or cycloaddition), with the generation of cyclopentenediol and dicyclopentadiene as side products due to the high reactivity of reagents and products. This system is a well-known benchmark for multivariable control algorithms and nonlinear optimization. This process combines nonlinear behavior in several operational ranges with a relatively simple modelling.

Summarizing, this work considers the measurement vector $y = [C_A, C_B, T]^T$, the input vector $u = \left[\frac{F}{V}, T_k \right]^T$, $\theta = [C_{A,in}, T_{in}]^T$ as unmeasured disturbances, and C_B and T are the controlled variables. The remainder of the physico-chemical parameters are considered known.

Concerning the optimization problem for this system, the yield of B (C_B) is usually implemented as the objective function, with another common formulation being the sum of selectivity and conversion for the same component ($C_B / (C_{A,in} - C_A) + C_B / C_{A,in}$). A study of the system's steady-state profiles shows that both objective functions always raise with the increase of the jacket temperature. Thus, in order to prevent the saturation of the manipulated variables throughout RTO operation, a new objective function to be maximized that takes into account the energy cost related to higher jacket temperatures was proposed:

$$J = \frac{C_B}{C_{A,in} - C_A} + \frac{C_B}{C_{A,in}} - 6 \times 10^{-4} T_k \quad (1)$$

Regarding complex dynamics, different models were considered:

- **Base Van de Vusse:** standard model formulation with equations and parameters available in Engell and Klatt (1993);
- **Van de Vusse with dead time:** in this case, fixed time delays were inserted into the measured and input variables.

The NLP problems were implemented with CasADi v3.5.5 (Anderson, 2013) using Python v3.8.3 programming environment and solved using IPOPT v3.12.2 (Wächter and Biegler, 2006) and IDAS (Hindmarsh, 2005) was the integrator applied to develop the plant simulator with relative and absolute accuracies of 10^{-6} for both solvers. The sampling time and total simulation time were 9s and 2h for all cases and architectures. A common practice in industrial applications is to run the optimizer in a fixed frequency; in this case the optimization cycle was of 10 sampling times. For all frameworks, nonlinear model predictive control (NMPC) was considered for the advanced control layer with control and prediction horizons of 10 and 40 sampling times, respectively. The steady-state detection uses the last five plant measurements at all sampling times.

3.1. Base Van de Vusse

Figure 1 shows the results for all architectures considering step perturbations in the disturbances at different simulation times. Regarding the objective function performance and computational effort, Table 1 summarizes the average CPU time per optimization cycle for each approach.

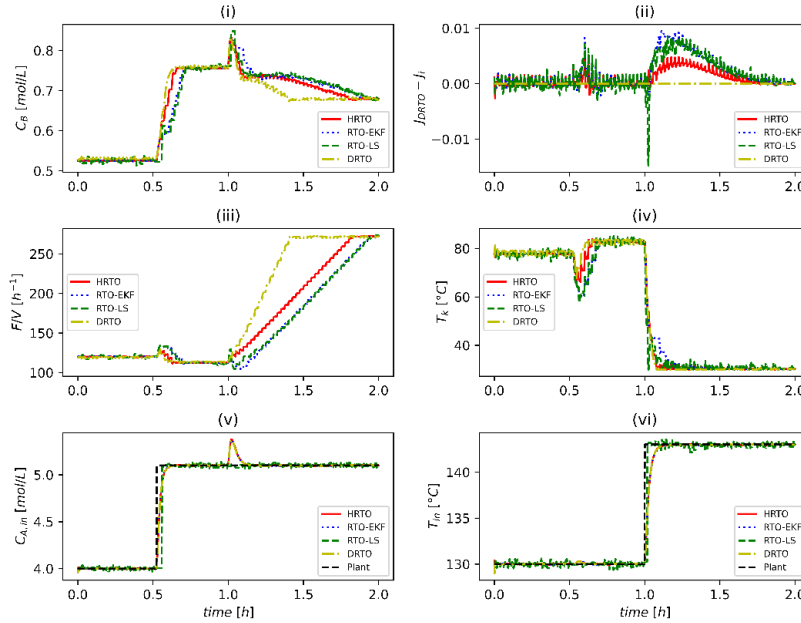


Figure 1: Performance of the RTO frameworks: (i) estimates of C_B ; (ii) difference of the objective function plant value to DRTO; (iii) input values for $\frac{F}{V}$; (iv) input values for T_k ; (v) estimates of $C_{A,in}$; and (vi) estimates of T_{in} .

Table 1: Integrated objective function and average CPU time per optimization cycle.

Approach	Integrated objective function	Avg. time per optimization cycle [s]
DRTO	518.32	2.62
HRTO	517.70	2.11
RTO-EKF	516.92	2.11
RTO-LS	517.06	2.13

It is clear that the NMPC acts faster with the DRTO, and the hybrid approach performs better when compared to its steady-state counterparts, which are held back by the steady-state detection step. The steady-state optimum is expected to be the same for all cases, and the delay in RTO-EKF and RTO-LS to reach the new optimum is evident. It is noteworthy that all optimization approaches are limited to a fixed input variation constraint of 5 units, explaining why DRTO can move faster than the other strategies since it is able to perform 3 actions during the 10 sampling times cycle. In addition, HRTO architecture behaved as fast as its steady-state counterparts while performing slightly better, as shown in Table 1 and Figure 1.

It is also important to state the robustness of the HRTO approach to disturbances of different types and magnitudes, considering that only parametric uncertainty is present and that the plant is open-loop stable. Different disturbance scenarios should only affect the estimation convergence rate since the EKF was tuned in order to converge reasonably fast for this specific perturbation scenario.

3.2. Van de Vusse with dead time

Regarding this study, the implemented delays were [10, 10] sampling times for the inputs and [15, 15, 6] for the plant measurements. Since the results for the base case study showed how HRTO overperformed in comparison to RTO-EKF and RTO-LS, this example has the purpose of evaluating its behavior in a process presenting dead time. Figure 2 illustrates performances for HRTO in four different cases: base case, delay only on the inputs, delay only on the measurements, and delay on both.

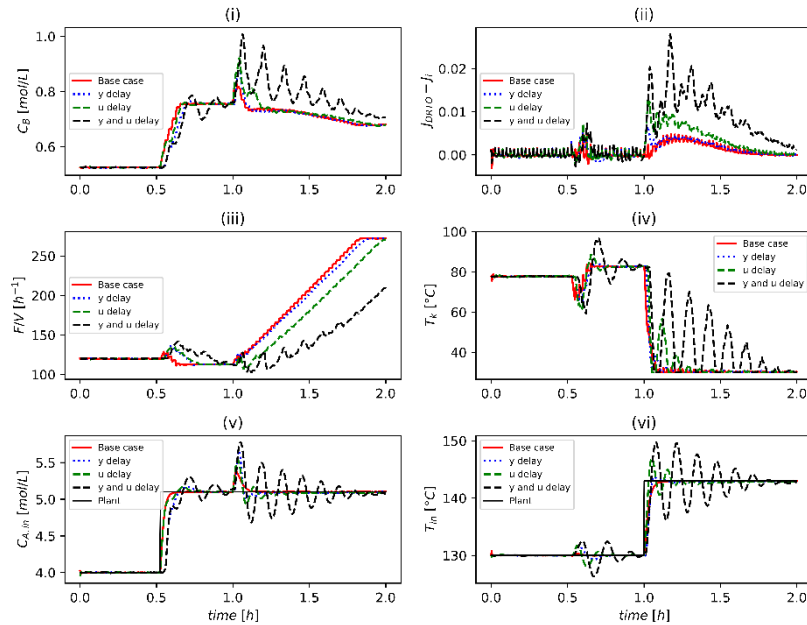


Figure 2: HRTO performance in the presence of dead-time: (i) estimates of C_B ; (ii) difference of the objective function plant value to DRTO; (iii) input values for F/V ; (iv) input values for T_k ; (v) estimates of $C_{A,in}$; and (vi) estimates of T_{in} .

Figure 2 clearly shows that the dead time's inclusion affects HRTO's performance. It is noteworthy that the estimation becomes especially compromised with the inclusion of dead times, which translates into poorer performances from the optimizer and controller. Concerning C_B and the objective function profiles, the system struggles to reach the steady-state, and the HRTO base curve stabilizes faster. As expected, the insertion of dead time in both inputs and measurements proved to be the worst-case scenario as it comes with the presence of oscillatory modes. Simulations with greater delays resulted in the system becoming unstable. It is noteworthy that the DRTO and RTO-EKF simulation with the same delays did not present oscillatory behavior, unlike HRTO and RTO-LS.

Table 2 shows the objective function performance and CPU effort for DRTO, HRTO and RTO-EKF considering dead-time in the inputs and outputs, confirming that the existence of time delays in the inputs and outputs compromises the overall performance of the real-time optimization. RTO-LS was not considered for this comparison since the LS estimator relies on steady-state operation and the system becomes highly unstable, as seen in Figure 2.

Table 2: Van de Vusse with dead-time - Integrated objective function and average CPU time per optimization cycle.

Approach	Integrated objective function	Avg. time per optimization cycle [s]
DRTO	517.84	2.58
HRTO	514.52	1.99
RTO-EKF	514.35	2.17

4. Conclusion

In this work, the hybrid RTO approach was applied to the Van de Vusse CSTR. The DRTO is still expected to outperform its counterparts. This result is showed by numerically integrating the objective function curve while in transient operation and using this value for the DRTO as reference. By comparing the objective function profile for the RTO and HRTO, a similar behavior was observed during most of the steady-state operation, with close performances. The HRTO stands out especially thanks to the absence of the steady-state detection step, which makes standard RTO have a delayed performance at some operational ranges. HRTO's behavior in the presence of dead time was also evaluated, and it was observed that it does affect its performance, especially in the estimation step. The presence of dead time arose oscillatory modes, and it is even possible to destabilize the system depending on its amplitude. However, this issue can possibly be overpassed by the use of dead-time compensators.

References

- J. Andersson, 2013. A general-purpose software framework for dynamic optimization. Ph.D. thesis, Arenberg Doctoral School, KU Leuven, Department of Electrical.
- R. Bousbia-Salah, F. Lesage, G.-H. Hu, A. Latifi, 2017. Dynamic real-time optimization of a batch polymerization process. In: *Computer Aided Chemical Engineering*. Vol. 40. Elsevier, pp. 1741–1746.
- S. Cao, R. R. Rhinehart, 1995. An efficient method for on-line identification of steady state. *J. Process Control* 5, 363–374.
- S. Engell, K.-U. Klatt, 1993. Nonlinear control of a non-minimum-phase cstr. In: *1993 American Control Conference*. IEEE, pp. 2941–2945.
- A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, C. S. Woodward, 2005. Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)* 31 (3), 363–396.
- D. Krishnamoorthy, B. Foss, S. Skogestad, 2018. Steady-state real-time optimization using transient measurements. *Computers & Chemical Engineering* 115, 34–45.
- J. O. Matias, G. A. Le Roux, 2018. Real-time optimization with persistent parameter adaptation using online parameter estimation. *Journal of Process Control* 68, 195–204.
- D. N. Melo, A. P. Mariano, E. C. V. de Toledo, C. B. Costa, R. Maciel Filho, 2009. An evaluation of a multimethod tool for real-time implementation of two-layer optimization. In: *Computer Aided Chemical Engineering*. Vol. 26. Elsevier, pp. 537–541.
- C. A. Ruiz, 2009. Real time industrial process systems: Experiences from the field. In: *Computer Aided Chemical Engineering*. Vol. 27. Elsevier, pp. 133–138.

- F. Serralunga, M. C. Mussati, P. A. Aguirre, 2012. Real-time optimization of energy systems in sugar and ethanol facilities: a modifier adaptation approach. In: *Computer Aided Chemical Engineering*. Vol. 31. Elsevier, pp. 375–379.
- J. O. Trierweiler, 2014. Real-time optimization of industrial processes. *Encyclopedia of Systems and Control*, 1132–1141.
- J. G. Van de Vusse, 1962. A new model for the stirred tank reactor. *Chemical Engineering Science* 17 (7), 507–521.
- A. Wächter, L. T. Biegler, 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming* 106 (1), 25–57.

Evaluating Control Room Operator Training Outcomes Through Eye Gaze Augmented Multi-Scale Data

Mohammed Aatif Shahab,^a Babji Srinivasan,^{a*} Rajagopalan Srinivasan^{a*}

^a*Indian Institute of Technology Madras, Tamil Nadu, 600036, India*

* raj@iitm.ac.in, babji.srinivasan@iitm.ac.in

Abstract

The significance of operator training has dramatically increased due to complex automation strategies in modern process plants. It is reported that human errors account for 70% of the accidents in process industries, with inadequate training cited as one of the most common reasons for these incidents. Our previous work has shown the potential of eye-tracking to infer the mental state of control room operators. In this work, we propose a methodology that combines multi-scale data from the process simulator, control actions performed, and eye gaze data of the operators to evaluate their training outcomes. Specifically, we use fixation transition entropy, an eye-tracking metric, which can help infer the mental models of the process abnormalities developed by the operators during repeated control room tasks. Results indicate that the fixation transition entropy decreases on account of development of correct mental models of process while it remain at higher values when operator fails to update their mental models during plant abnormalities. Thus, the proposed metric can be used to gauge the development of operator's mental models during training to understand the transition from novice to becoming experts.

Keywords: Operator Training, Eye-tracking, Transition Entropy, Mental Model, Process Safety

1. Introduction

Operator training is an important part of industrial safety, during which novice operators acquire skills and develop mental models of the process. After training, the operators are deployed to a real control room setting to monitor the automated system and intervene when the automation underperforms, especially during abnormalities. When dealing with such abnormalities, an incorrect mental model of the abnormal situation developed by the operator increases the likelihood of committing. Thus, it is crucial to have skilled operators with the correct mental models of the process. However, the complex nature of modern plants, with sophisticated and complex automation strategies, makes it challenging for the operators to develop correct mental models of the process (Kludge et al., 2014). As a result, assessment of mental models during training is important to evaluate their understanding of process dynamics.

Typically, process industries rely on Operator Training Simulators (OTS) to impart skills to the operators. Several studies have been conducted to evaluate operator performance during these training trials. Most of these studies either focus on developing comprehensive OTS that can accurately mimic plant behavior (Patle et al., 2014) or

assessing operator's ability to follow standard guidelines based on metrics derived from process and operator interactions. For instance, Manca et al. (2014) evaluated the performance based on the time taken and the sequence of actions by the operator against a standard protocol in a virtual polymerization plant. Nevertheless, these studies do not explicitly focus on the evolution of the operator's knowledge about the dynamics of the process during the training trials. In addition, there are hardly a few works that focus on understanding the cognitive behavior of operators, crucial to enhancing operators' skills and abilities (Bullemer and Nimmo, 1994).

Eye-tracking is one of the widely used approaches to analyze the cognitive behavior of human operators in several safety-critical domains such as aviation, health care, and nuclear industries (Srinivasan et al., 2019). In the past, our research group has utilized physiological data obtained from eye tracking to understand the cognitive behavior of control room operators during abnormal situations (Bhavsar et al., 2017, Srinivasan et al., 2019). Our studies indicated that eye gaze analysis could identify the orientation, diagnosis, and execution tasks performed by the operator during an experiment (Kodapully et al., 2016). Statistical studies revealed that gaze-based entropy measures could be used to understand the situational awareness of the operators while handling plant abnormalities (Bhavsar et al., 2017). With the knowledge acquired from these studies, in this work, we develop a methodology to understand the evolution of operator's mental models and expertise level during training tasks using metrics derived from process dynamics, operator actions, and eye-tracking data. The details of the proposed methodology, along with experimental studies are discussed in the following section.

2. Experimental setup and methodology

The experimental study consisted of participants interacting with the Human Machine Interface (HMI) of a simulated ethanol production plant. Participants were asked to monitor and perform necessary control actions (using sliders provided on the HMI) during plant abnormalities. There are six different scenarios that result in abnormal situations that need to be handled by the participant. If the participant fails to handle the abnormality, they can use emergency shutdown button else the plant will shut down automatically in 2 mins after the occurrence of alarms. The reader is referred to Bhavsar et al. (2017) for more details on the simulation. Ten participants were involved in the study. Each participant carried out repetitive tasks (repetition of a scenario) in different trials. Overall, these ten participants performed 81 trials for each scenario leading to a total of 486 tasks. We recorded process data, alarm information, and operator action data during all these tasks. In addition to these measurements, we also recorded eye-gaze data obtained from Tobii TX 300 eye tracker.

For assessing the effect of training (repetition of scenarios), in every trial, we analyzed the number of alarms triggered (N_t), the number of alarms cleared (N_c), and the number of slider actions (N_s). In addition to these process and action-based metrics, we also quantified operators' cognitive behavior using eye gazed fixation transition entropy (\tilde{H}_F).

To calculate \tilde{H}_F , we divided the HMI into L rectangular-shaped areas of interest (AOI) based on the location of variables tag, alarm panel, and trend panel. In this work, we have defined 19 (L) AOIs. We obtained the fixation distribution on these AOIs and generated a sequence of fixations based on the AOI they belong. Next, we computed the transition matrix (T' of size $L \times L$) which represents the frequency of fixation transitions between A_i ($i = 1, 2, \dots, L$) to A_j ($j = 1, 2, \dots, L$).

$$T' = \begin{pmatrix} A'_{11} & \cdots & A'_{1L} \\ \vdots & \ddots & \vdots \\ A'_{L1} & \cdots & A'_{LL} \end{pmatrix} \quad (1)$$

Here A'_{Lj} denotes the number of transitions from L^{th} AOI to j^{th} AOI and likewise. We then weighted T' by a weight matrix (W) to avoid the effect of unintentional fixation transitions between nearby AOI's. The weight matrix (W) is computed based on the Euclidian distance between the center of the AOIs. The coordinates of the center of i^{th} AOI (A_i) is given by (x_i, y_i) . An element of W is given by:

$$W_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}, i = 1, 2, \dots, L, j = 1, 2, \dots, L \quad (2)$$

where W_{ij} represents Euclidian distance between A_i and A_j . We then rescale W between 0 and 1. For the sake of simplicity, we call the rescaled weight matrix as W . The weighted transition matrix (T) is then obtained as:

$$T' = \begin{pmatrix} A'_{11}W_{11} & \cdots & A'_{1L}W_{1L} \\ \vdots & \ddots & \vdots \\ A'_{L1}W_{L1} & \cdots & A'_{LL}W_{LL} \end{pmatrix} = \begin{pmatrix} 0 & \cdots & A_{1L} \\ \vdots & \ddots & \vdots \\ A_{L1} & \cdots & 0 \end{pmatrix} \quad (3)$$

The diagonal elements in T' (self-transitions) are zero as W_{ii} (self-distance) is zero. After assigning weights to the transition matrix, we calculated \tilde{H}_F as per equation (4) and (5):

$$H_F = -\sum_{i=1}^L \sum_{j=1, j \neq i}^L P(A_{ij}|A_i) \log_2 P(A_{ij}|A_i) P(A_i) \quad (4)$$

$$\tilde{H}_F = \frac{H_F}{H_F^{max}} \quad (5)$$

where H_F is the weighted fixation transition entropy. Here $P(A_{ij}|A_i)$ denotes the probability of transition of fixation from AOI A_i to AOI A_j , and $P(A_i)$ is the probability of fixation on AOI A_i . \tilde{H}_F represents the relative weighted fixation transition entropy, while H_F^{max} is the maximum value of H_F given by $H_F^{max} = \log_2(L-1)$. The larger the \tilde{H}_F , the more disordered fixation pattern on HMI and vice-versa. It is expected that with repetition of trials, the operator's fixation pattern will move towards lower entropy values as the transition takes place only between those regions which give information of disturbance or help in removing disturbance.

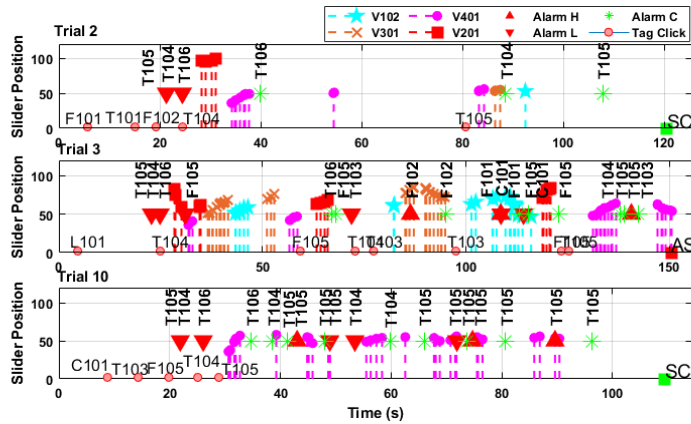


Figure 1: Operator actions with alarm and slider action details in key trials.

3. Results and Discussion

In this section, we demonstrate the potential of the proposed methodology in assessing operator's ability as well as inability in developing correct mental models of the process dynamics.

3.1. Illustrative Example 1

In this illustration, we demonstrate the potential of the proposed methodology to assess the development of a correct mental model of the process dynamics under a given abnormality with the training of novice operators. Consider an operator who performed ten trials of a scenario that involves a disturbance in the reflux ratio of the distillation column, and which can be rectified by manipulating V401. Operator actions in key trials are shown in Fig. 1. Consider trial 2 (Fig. 1), at around 22 s alarms T104 and T105 occur, followed by T106. Operator clicks on the tag of T104 to observe the trend of the process variable T104 and takes first control action at around 29 s by manipulating V201. At around 35 s, the operator manipulates V401 (correct control action). In the same trial, the operator also manipulates V301 later, depicting that the operator keeps on exploring the possible actions which can clear the disturbance in the process. In trial 2 and 3, the operator has used multiple control actions as shown by different markers of stem plot (Fig. 1). The use of multiple control actions indicates that the operator is unaware of the root cause of the disturbance. While manipulating different sliders, operator somehow manages to clear all the alarms in trial 2, but it is not necessary that the operator understands the root cause of the disturbance. Trial 3 ends in automatic shutdown because of operator's inability to deal with the disturbance. With the repetition of trials, the operator develops an understanding of the underlying principle of the process and is reflected in the trial 10 where the operator only manipulates V401.

The alarm information (N_t & N_c) and operator actions (N_s) with the trial number is shown in Fig. 2 (a) and 2 (b), respectively. We observe a decrease in N_t towards the end of the trial (Fig. 2 (a)), implying that the operator has taken the right control actions because of which few additional alarms are triggered. Also, the operator successfully cleared all the triggered alarms ($N_t - N_c = 0$). The operator action data (Fig. 2 (b)) shows a decrease in N_s from trial 8 onwards. This implies that with repetition of trials, an understanding of the scenario (less slider manipulation) is developed by the operator.

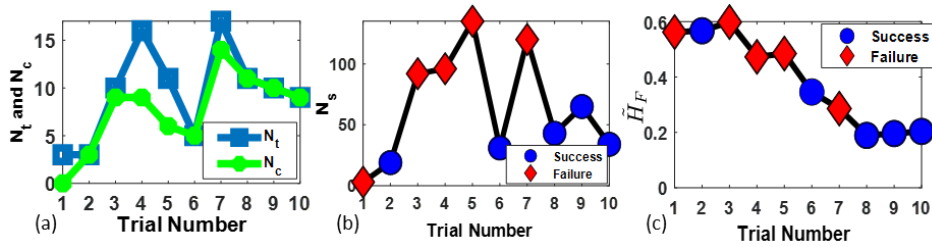


Figure 2: (a) Number of alarms triggered (N_t) and cleared (N_c) with trial number (b) Number of slider actions (N_s) with trial number (c) Weighted Fixation Transition Entropy (\tilde{H}_F) with trial number

Trials 1 and 2 need special mention as both N_t and N_s stay at lower values. In trial 1, the operator puts minimal effort ($N_s = 3$) to clear the alarms, which leads to the shutdown of the plant. During trial 2, there is a possibility that the operator manages to clear the disturbance by chance. This is supported by the fact that the next three trials (trials 3, 4 & 5) are not completed successfully. From Fig. 2 (c), it can be observed that \tilde{H}_F decreases (from 0.562 in the first trial to 0.203 in the last trial) with repetition of trials and remains at lower values towards the end of the trial. This signifies that with increasing trials, the attention is increasingly directed to the relevant AOIs. During trials 1 and 2, \tilde{H}_F remains at a higher value in contrast to N_t and N_s . This indicates that the operator is unaware of the dynamic response of the process because of which distributed attention on HMI is observed. Thus \tilde{H}_F is able to track the mental model of the operator under the given abnormality correctly and therefore has the potential to gauge the learning progress of novice operators.

3.2. Illustrative Example 2

In this illustrative example, we demonstrate the potential of \tilde{H}_F to capture the instances when operators struggle to update their mental models of the process, and as such, do not learn with repetition of trials. Consider an operator who performed seven trials of scenario 5 in which feed flow to the CSTR gets disturbed, resulting in alarms F101, F105 and C101. The number of alarms triggered and cleared (N_t & N_c), and number of slider actions (N_s) are shown in Fig. 3 (a) and 3 (b) respectively. The operator was unsuccessful in bringing the plant to the normal operating range till the last trial ($N_t - N_c > 0$), except during trials 2 and 4. The inability to learn is captured by the inability to clear all the alarms within the stipulated amount of time as shown in Fig. 3(a), and as also reflected by N_s which does not show a decreasing trend with trial number. Even though trials 2 and 4 were a success, with low values in all the process and action-based metrics, there was a very little decrease in \tilde{H}_F as depicted by Fig. 3 (c). Nevertheless, \tilde{H}_F remained more or less constant, during the other trials as well, even towards the end of the trials, indicating that the operator did not update their mental model of the given abnormality with repetitions. Failure to update the mental model is manifested in the inability to improve the fixation pattern from a disordered one to a more directed pattern of information acquisition from the HMI. Thus, the proposed eye-tracking based metric has the potential to capture whether an operator is learning or not and can complement all other process and action-based metrics by providing critical insights into the operator cognitive behaviour.

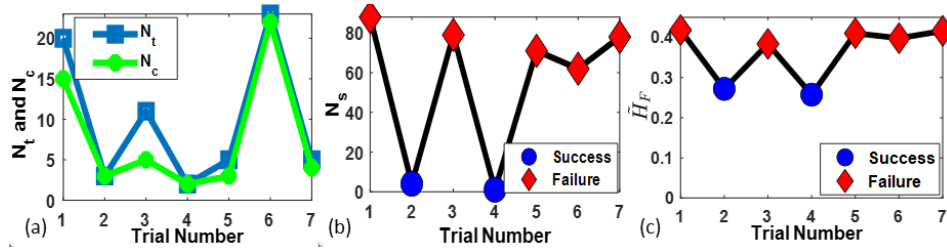


Figure 3: (a) Number of alarms triggered (N_t) and cleared (N_c) with trial number (b) Number of slider actions (N_s) with trial number (c) Weighted Fixation Transition Entropy (\tilde{H}_F) with trial number

4. Conclusions

The present work proposed a multi-scale data-driven to gauge the development of expertise level of control room operators during training. When the operators develop the correct mental models of the process, the process-based measures (N_t & N_c), operator actions (N_s), and fixation transition entropy (\tilde{H}_F) decrease and remain at a lower value. The lower values of \tilde{H}_F signifies that operators have arranged their fixation pattern on only a few regions on HMI. Further, our results demonstrate that the proposed cognitive measure \tilde{H}_F help infer the mental models of the operators even when the process and operation action based measures fail. Our future work will focus on developing robust multivariate measures using various physiological sensors (such as EEG, ECG, audio and video analytics etc.) for expertise level assessment of control room operators.

References

- Bhavsar, P., Srinivasan, B., & Srinivasan, R. (2017). Quantifying situation awareness of control room operators using eye-gaze behavior. *Computers & chemical engineering*, 106, 191-201.
- Bullemer, P. T., & Nimmo, I. (1994, October). Understanding and supporting abnormal situation management in Industrial process control environments: a new approach to training. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics* (Vol. 1, pp. 391-396). IEEE.
- Kluge, A., Nazir, S., & Manca, D. (2014). Advanced applications in process control and training needs of field and control room operators. *IIE Transactions on Occupational Ergonomics and Human Factors*, 2(3-4), 121-136.
- Kodappully, M., Srinivasan, B., & Srinivasan, R. (2016). Towards predicting human error: Eye gaze analysis for identification of cognitive steps performed by control room operators. *Journal of Loss Prevention in the Process Industries*, 42, 35-46.
- Manca, D., Nazir, S., Colombo, S., & Kluge, A. (2014). Procedure for automated assessment of Industrial operators. *Chemical Engineering Transactions*, 36, 391-396.
- Patle, D. S., Ahmad, Z., & Rangaiah, G. P. (2014). Operator training simulators in the chemical industry: review, issues, and future directions. *Reviews in Chemical Engineering*, 30(2), 199-216.
- Srinivasan, R., Srinivasan, B., Iqbal, M. U., Nemet, A., & Kravanja, Z. (2019). Recent developments towards enhancing process safety: Inherent safety and cognitive engineering. *Computers & Chemical Engineering*, 128, 364-383.

Multi-layer Multivariate Statistical Analysis for Process Monitoring

Simin Li,^a Shuanghua Yang,^{a,b} Yi Cao^{a,b}

^a*Zhejiang University, 38 Zheda Road, Lingyin Street, Hangzhou 310058, China*

^b*Institute of Zhejiang University-Quzhou, 78 Jiuhua North Avenue, Baiyun Street, Quzhou 324002, China*

Abstract

In order to ensure the safety and stability of industrial processes, it is necessary to research more efficient and accurate fault detection technology. Methods based on multivariate statistics analysis are very suitable for complex industrial processes. Based on the pros and cons of current fault detection methods, deep principle component analysis (DePCA) was developed by Deng previously. In this work, by combining DePCA with dynamic process, the deep dynamic principle component analysis (DeDPCA) is proposed, where optimizing variables for extended time series is used to reduce computational complexity. Case study in the benchmark Tennessee Eastman process demonstrates the superior performance of the proposed DeDPCA method over the DePCA method.

Keywords: fault detection, DPCA, KPCA, deep learning

1. Introduction

With the development of modern science and technology and the intensification of market competition, the production level of the chemical process continues to improve. Traditional detection systems lack reliability, and data-driven methods are getting more and more attention^[1].

Deep learning, as one of the most currently remarkable machine learning techniques, has achieved great success in many applications such as image analysis, speech recognition and text understanding. Deng^[2] took the advantages of deep learning strategy and designed a hierarchical statistical model structure to extract multilayer data features, including both the linear and nonlinear principal components, which is called deep principle component analysis (DePCA). Although DePCA's ability to detect faults is better than kernel principle component analysis (KPCA)^[3] and serial principal component analysis (SPCA)^[4], it is not sensitive to dynamic data for most industrial processes.

When combine DePCA with dynamic process^[5], deep dynamic principle component analysis(DeDPCA) is established. Then variables for extended time series is optimized using autocorrelation analysis to reduce computational complexity. Which layer to extract dynamic feature was conducted by comparing two models with different structure, one extract dynamic feature in first linear layer and another in second nonlinear layer.

The remainder of this paper is structured as follows. In Section 2, our proposed DeDPCA method for nonlinear dynamic process monitoring is given. In Section 3, case study of TE process is used to validate the proposed method, and our conclusions are drawn in Section 4.

2. Construction of Proposed Model

2.1. Proposed DeDPCA model

The proposed DeDPCA model extracts linear and nonlinear features by integrating dynamic principle component analysis(DPCA) and KPCA into a 2-layer structure. In this deep feature extraction structure, first the training data matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$ with N samples of M variables expands to an augmented matrix $\mathbf{X}(k) \in \mathbb{R}^{(N-k) \times M(k+1)}$ with a lag of k . Then the linear principal components are extracted from the augmented matrix $\mathbf{X}(k)$ to obtain the first layer feature $\mathbf{T}^{(1)} \in \mathbb{R}^{(N-k) \times \tilde{N}^{(1)}}$, where $\tilde{N}^{(1)} = M(k+1)$. Then KPCA is applied to $\mathbf{T}^{(1)}$ to extract nonlinear principal components to obtain the second layer feature $\mathbf{T}^{(2)} \in \mathbb{R}^{(N-k) \times \tilde{N}^{(2)}}$, where $\tilde{N}^{(2)}$ is the number of nonlinear non-zero features extracted in the second layer.

More specifically, in the first feature layer, the linear optimization task is designed as

$$\begin{aligned} & \max_{\mathbf{p}^{(1)}} \frac{1}{N-1} (\mathbf{p}^{(1)})^T \mathbf{X}(k)^T \mathbf{X}(k) \mathbf{p}^{(1)} \\ & \text{s. t. } (\mathbf{p}^{(1)})^T \mathbf{p}^{(1)} = 1 \end{aligned} \quad (1)$$

to obtain $M(k+1)$ projection vector $\mathbf{p}_i^{(1)} \in \mathbb{R}^{M(k+1)}$ and score vector $\mathbf{t}_i^{(1)} = \mathbf{X}(k) \mathbf{p}_i^{(1)}$, $1 \leq i \leq M(k+1)$. The score vector constitutes the first layer feature $\mathbf{T}^{(1)} = [\mathbf{t}_1^{(1)} \mathbf{t}_2^{(1)} \dots \mathbf{t}_{M(k+1)}^{(1)}]$. In the second feature layer, we perform nonlinear optimization task

$$\begin{aligned} & \max_{\mathbf{p}^{(2)}} \frac{1}{N-1} (\mathbf{p}^{(2)})^T \boldsymbol{\Phi}(\mathbf{T}^{(1)})^T \boldsymbol{\Phi}(\mathbf{T}^{(1)}) \mathbf{p}^{(2)} \\ & \text{s. t. } (\mathbf{p}^{(2)})^T \mathbf{p}^{(2)} = 1 \end{aligned} \quad (2)$$

where $\mathbf{p}^{(2)}$ is the nonlinear projection vector that can be obtained by

$$\mathbf{p}^{(2)} = \sum_{j=1}^{N-k} \boldsymbol{\Phi}^T(\bar{\mathbf{t}}_j^{(1)}) \boldsymbol{\alpha}_j^{(2)} = \boldsymbol{\Phi}^T(\mathbf{T}^{(1)}) \boldsymbol{\alpha}^{(2)} \quad (3)$$

where $\bar{\mathbf{t}}_j^{(1)} \in \mathbb{R}^{\tilde{N}^{(1)}}$ is the j th column of $(\mathbf{T}^{(1)})^T$, $\boldsymbol{\alpha}^{(2)} = [\alpha_1^{(2)} \alpha_2^{(2)} \dots \alpha_N^{(2)}]^T$. Besides, denote $\mathbf{K}^{(1)} = \boldsymbol{\Phi}(\mathbf{T}^{(1)}) \boldsymbol{\Phi}^T(\mathbf{T}^{(1)})$, The nonlinear optimization formula can be written as

$$\begin{aligned} & \max_{\boldsymbol{\alpha}^{(2)}} \frac{1}{N-k-1} (\boldsymbol{\alpha}^{(2)})^T \mathbf{K}^{(1)} \mathbf{K}^{(1)} \boldsymbol{\alpha}^{(2)} \\ & \text{s. t. } (\boldsymbol{\alpha}^{(2)})^T \mathbf{K}^{(1)} \boldsymbol{\alpha}^{(2)} = 1 \end{aligned} \quad (4)$$

resulting in the projection vector $\boldsymbol{\alpha}_i^{(2)} \in \mathbb{R}^{N-k}$ and the corresponding nonlinear score vector $\mathbf{t}_i^{(2)} = \mathbf{K}^{(1)} \boldsymbol{\alpha}_i^{(2)} \in \mathbb{R}^{N-k}$, $1 \leq i \leq \tilde{N}^2$. Therefore, the second feature layer is $\mathbf{T}^{(2)} = [\mathbf{t}_1^{(2)} \mathbf{t}_2^{(2)} \dots \mathbf{t}_{\tilde{N}^2}^{(2)}] \in \mathbb{R}^{(N-k) \times \tilde{N}^2}$.

For the testing vector $\mathbf{y}_t = [\mathbf{x}_t \mathbf{x}_{t-1} \dots \mathbf{x}_{t-k}]$, the feature vector of the first layer $\mathbf{t}_t^{(1)} = [\mathbf{t}_{t,1}^{(1)} \mathbf{t}_{t,2}^{(1)} \dots \mathbf{t}_{t,M(k+1)}^{(1)}]^T$ can be calculated as

$$\mathbf{t}_{t,i}^{(1)} = \mathbf{y}_t^T \mathbf{p}_i^{(1)}, 1 \leq i \leq M(k+1) \quad (5)$$

Second layer feature vector $\mathbf{t}_t^{(2)} = [t_{t,1}^{(2)} \ t_{t,2}^{(2)} \ \dots \ t_{t,\bar{N}^2}^{(2)}]^T$, can be expressed as

$$\mathbf{t}_{t,i}^{(2)} = \mathbf{k}_t^{(1)T} \boldsymbol{\alpha}_i^{(2)} \quad (6)$$

where the kernel vector $\mathbf{k}_t^{(1)} = \boldsymbol{\Phi}(\mathbf{T}^{(1)})\boldsymbol{\Phi}(\mathbf{t}_t^{(1)}) = [k_{t,1}^{(1)} \ k_{t,2}^{(1)} \ \dots \ k_{t,N-k}^{(1)}]^T \in \mathbb{R}^{N-k}$. As can be emphasized that, like the DePCA method, the DeDPCA method also does not require complex network optimization to achieve layer-by-layer feature extraction.

For the first layer, $T^{2(1)}$ and $Q^{(1)}$ monitoring statistics can be obtained as

$$T^{2(1)} = \mathbf{t}_t^{(1)T} \Lambda_1^{-1} \mathbf{t}_t^{(1)}$$

$$Q^{(1)} = \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|^2 = (\mathbf{y}_t - \mathbf{P}_P \mathbf{t}_t^{(1)})^T (\mathbf{y}_t - \mathbf{P}_P \mathbf{t}_t^{(1)}) \quad (7)$$

where Λ_1 is $K_1 \times K_1$ -dimensional matrix. Eigenvalues $\lambda_i, 1 \leq i \leq K_1$ are their diagonal elements. $\hat{\mathbf{y}}_t = \mathbf{P}_P \mathbf{t}_t^{(1)}$ is the reconstruction of the test vector \mathbf{y}_t .

For the second layer, $T^{2(2)}$ and $Q^{(2)}$ monitoring statistics can be obtained as

$$T^{2(2)} = \mathbf{t}_t^{(2)T} \Lambda_2^{-1} \mathbf{t}_t^{(2)}$$

$$Q^{(2)} = \|\boldsymbol{\Phi}(\mathbf{y}_t) - \hat{\boldsymbol{\Phi}}(\mathbf{y}_t)\|^2 = \sum_{i=1}^{\bar{N}^2} \mathbf{t}_t^{(2)}{}^2 - \sum_{i=1}^{K_2} \mathbf{t}_t^{(2)}{}^2 \quad (8)$$

where Λ_2 is $K_2 \times K_2$ -dimensional matrix. Eigenvalues $\lambda_i, 1 \leq i \leq K_2$ are their diagonal elements. $\hat{\boldsymbol{\Phi}}(\mathbf{y}_t)$ is the reconstruction of the test vector $\boldsymbol{\Phi}(\mathbf{y}_t)$.

2.2. Improved DeDPCA

The purpose of extending the time series matrix to the original matrix is to eliminate the autocorrelation of variables. If the autocorrelation of a variable is low, not only little information can be included, but also it will increase the unnecessary calculation complexity. Therefore, the improved DeDPCA algorithm analyzes the autocorrelation α to set an appropriate lag k for each variable. We denote a strong autocorrelation standard α_{\min} . α_{\min} was set to 0.5 empirically. Taking variable x_1 for example, if $\alpha(n) > \alpha_{\min}$ when $k = n$, and $\alpha(n+1) < \alpha_{\min}$ when $k = n+1$, $[x_{1,t-1}, x_{1,t-2} \dots x_{1,t-n}]$ will be added into expansion matrix at time t , where $n \leq k_{\max}$, k_{\max} is the maximum time lag calculated above. In this way, variables with strong autocorrelation are retained and variables with weak autocorrelation are eliminated, which can effectively retain the time series characteristics of data, but also simplify the data set and reduce unnecessary calculations.

2.3. Control limit calculation

Kernel density estimation(KDE)^[6] is an effective method for estimating probably density function(PDF), especially for univariate random processes. Therefore, it is very suitable for univariate T^2 and Q index estimation.

The probability density function $\hat{p}(x)$ at point x estimated according to the kernel function $K(\cdot)$ is defined as follows

$$\hat{p}(x) = \frac{1}{Mh} \sum_{k=1}^M K\left(\frac{x - x_k}{h}\right) \quad (9)$$

Where x_k , $k=1,2,\dots,M$ is the sample of x , and h is the bandwidth.

Through the probability density function, the control limits T_{lim}^2 and Q_{lim} under a given confidence interval α can be obtained.

2.4. Fault information fusion

For DeDPCA-based online sample monitoring, each layer provides two monitoring statistics $T^{2(l)}$ and $Q^{(l)}$, where $l = 1,2$. In order to integrate the monitoring statistics of all feature layers, a fusion strategy based on Bayesian inference is proposed to conduct the overall decision.

Taking T^2 metric for example, we express the failure probability of sample \mathbf{x}_t under the fault condition C_f as $P_{T^2}^{(l)}(\mathbf{x}_t|C_f)$, under the normal condition C_n as $P_{T^2}^{(l)}(\mathbf{x}_t|C_n)$

$$P_{T^2}^{(l)}(\mathbf{x}_t|C_f) = \exp(-\gamma T_{\text{lim}}^{2(l)}/T^{2(l)}) \quad (10)$$

$$P_{T^2}^{(l)}(\mathbf{x}_t|C_n) = \exp(-\gamma T^{2(l)}/T_{\text{lim}}^{2(l)}) \quad (11)$$

where γ is a parameter and was set to 0.2 empirically.

According to Bayesian inference, monitoring statistics $T^{2(l)}$ can be transformed to posterior probability $P_{T^2}^{(l)}(C_f|\mathbf{x}_t)$ by

$$P_{T^2}^{(l)}(C_f|\mathbf{x}_t) = \frac{P_{T^2}^{(l)}(\mathbf{x}_t|C_f)P_{T^2}^{(l)}(C_f)}{P_{T^2}^{(l)}(\mathbf{x}_t)}, \quad 1 \leq l \leq 2 \quad (12)$$

where $P_{T^2}^{(l)}(C_f)$ is prior failure probability equal to significance level δ . $P_{T^2}^{(l)}(\mathbf{x}_t)$ is the occurrence probability of sample \mathbf{x}_t

$$P_{T^2}^{(l)}(\mathbf{x}_t) = P_{T^2}^{(l)}(\mathbf{x}_t|C_f)P_{T^2}^{(l)}(C_f) + P_{T^2}^{(l)}(\mathbf{x}_t|C_n)P_{T^2}^{(l)}(C_n) \quad (13)$$

where $P_{T^2}^{(l)}(C_n)$ is prior normal probability equal to confidence level $1 - \delta$.

By weighting the posterior failure probability of each layer, an overall monitoring statistic data PT^2 is constructed

$$PT^2 = \frac{\sum_{l=1}^L \omega_{T^2}^{(l)} P_{T^2}^{(l)}(C_f|\mathbf{x}_t)}{\sum_{l=1}^L \omega_{T^2}^{(l)}} = \sum_{l=1}^2 \bar{\omega}_{T^2}^{(l)} P_{T^2}^{(l)}(C_f|\mathbf{x}_t) \quad (14)$$

where $\omega_{T^2}^{(l)}$ and $\omega_Q^{(l)}$ are weighting factor, $\bar{\omega}_{T^2}^{(l)} = \omega_{T^2}^{(l)}/\sum_{l=1}^2 \omega_{T^2}^{(l)}$.

PQ can be obtained by the same way. The process is in normal condition if $PT^2 \leq PT_{\text{lim}}^2$ and $PQ \leq PQ_{\text{lim}}$, otherwise a fault is detected. PT_{lim}^2 and PQ_{lim} was set to δ .

3. Case study

The TE process^{[7],[8]}, detailed by Downs and Vogel, has been a benchmark process for evaluating the process monitoring and fault diagnosis methods. For process monitoring

modelling, we select 52 monitored variables and 21 typical faults. The corresponding simulation data can be downloaded from <http://web.mit.edu/braatzgroup/links.html>. The normal operation dataset includes 500 samples. Each fault data set involves 960 samples and the faults are introduced after the 160th sample.

For the DePCA, and DeDPCA, the Gaussian kernel function is used with a kernel width of $\sigma = 50N$ found empirically. The number of the retained linear and nonlinear PCs is determined by the average method, which keep PCs greater than the average variance. The control limits of the monitoring statistics are computed by the KDE method, and the 95% confidence limits are computed for each statistic. Bayesian inference is used to integrate the monitoring statistics of each feature layer to generate an overall decision. Three performance indices are used to evaluate the monitoring results, which are the fault detection rate (FDR), the fault detection time (FDT) and the false alarming rate (FAR).

The detection results of the DePCA and DeDPCA are demonstrated in Table I. The average FDR of DeDPCA method is higher than DePCA for about 2.35%, indicating that DeDPCA is more likely to detect faults. The average FDT of DeDPCA method is earlier than DePCA for about 7 samples, indicating that DeDPCA can discover the fault more sensitive than DePCA. The average FARs of two methods are both low. Through the comparison of the three detection performance indices, the superiority of the proposed DeDPCA method can be clearly seen.

Table I. Average FDRs(%), FDTs(sample) and FARs(%) of the 21 TE process faults obtained by DePCA and DeDPCA methods

Method	Average FDR	Average FDT	Average FAR
DePCA	81.57	17	3.21
DeDPCA	83.92	10	3.52

In order to determine which layer to extract dynamic feature could have better monitoring performance, two models with different structure have been tested. One model extracting the dynamic feature at first linear layer is called DeDPCA-1, another extracting the dynamic feature at second nonlinear layer is called DeDPCA-2. The detection results are listed in Table II. Obviously, DeDPCA-1 gets higher FDR and lower FAR, denoting that extracting the dynamic feature at first linear layer is more suitable. Extracting the dynamic feature at second layer may loss some dynamic information at the first linear feature extracting process.

Table II. Average FDRs(%), FDTs(sample) and FARs(%) of the 21 TE process faults obtained by DeDPCA-1 and DeDPCA-2 methods

Method	Average FDR	Average FDT	Average FAR
DeDPCA-1	83.92	10	3.52
DeDPCA-2	79.47	6	14.01

Although extending the time series matrix could improve the monitoring performance in TE process, it also increases the computational complexity. Thus improved DeDPCA method was proposed to reduce the computational complexity and keep effective monitoring performance at the same time. Table III shows the detection results of improved DeDPCA. Online monitoring programs of each method was runned 5 times for

average, configured with Intel Core i5-6200U CPU(2.4 GHz) and 8G RAM memory. It shows that the average time to compute one sample of improved DeDPCA reduced by 34% compared to the former. The average FDRs, FDTs and FARs are about the same.

Table III. Average FDRs(%), FDTs(sample), FARs(%) and C-time(s) of the 21 TE process faults obtained by DeDPCA and improved DeDPCA methods

Method	Average FDR	Average FDT	Average FAR	Average C-time per sample
DeDPCA	83.92	10	3.52	1.3×10^{-2}
Improved DeDPCA	81.55	7	3.25	8.6×10^{-3}

4. Conclusions

This paper takes the Tennessee-Eastman process as the test object to improve the DePCA method based on deep learning. Aiming at the problem that it does not consider the autocorrelation of process variables, a dynamic process is added to construct a DeDPCA model and is optimized. The simulation results show that DeDPCA is conducive to early detection of faults. Besides, extracting the dynamic feature at first linear layer can get best detection results. Finally, improved DeDPCA reduces the computational complexity and meanwhile keeps excellent monitoring performance.

There are still many issues worth discussing and improving. The current deep model only includes one linear layer and one non-linear layer. Whether it is necessary to extend several layers to increase the depth of potential data feature mining is a question. Besides, how to calculate the contribution of variables in the deep model to monitoring statistics requires further investigating.

References

- [1] Y. Shen, S. X. Ding, A. Haghani, et al, 2012, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process.," J. Process Contr., vol. 22, no. 9, pp. 1567-1581.
- [2] X. Deng, X. Tian, and S. Chen, 2019, "Deep Principal Component Analysis Based on Layerwise Feature Extraction and Its Application to Nonlinear Process Monitoring," IEEE T. Contr. Syst. T., vol. 27, no. 6, pp. 2526-2540.
- [3] J.-M. Lee, C. K. Yoo, S. W. Choi et al, 2003, "Nonlinear process monitoring using kernel principal component analysis," Chem. Eng. Sci., vol. 59, no. 1, pp. 223-234.
- [4] X. Deng, X. Tian, S. Chen, et al, 2018, "Nonlinear process fault diagnosis based on serial principal component analysis," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 3, pp. 560-572.
- [5] W. Ku, R. H. Storer, and C. Georgakis, 1995, "Disturbance detection and isolation by dynamic principal component analysis," Chemometrics Intell. Lab. Syst., vol. 30, no. 1, pp. 179-196.
- [6] P.-E. P. Odiowei, Y. Cao, 2009, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations," IEEE T. Ind. Inform., vol. 6, no. 1, pp. 36-45.
- [7] P.R. Lyman and C. Georgakis, 1995, "Plant-wide control of the Tennessee Eastman problem," Comput. Chem. Eng., vol. 19, no. 3, pp. 321-331.
- [8] C. K. Lau, K. Ghosh, M. A. Hussain, et al, 2013, "Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS," Chemometr. Intell. Lab., vol. 120, no.15, pp. 1-14.

Digitalization, Control and Optimization for Cement Plants

Zhanhao Zhang,^{a,c} Marcus K. Nielsen,^{a,c} Steen Hørsholt,^{a,c}
Guruprasath Muralidharan,^b John B. Jørgensen^a

^a*DTU Compute, Technical University of Denmark, Kgs. Lyngby 2800, Denmark*

^b*Smarta-Opti Solutions and DALAPC, India;* ^c*2-control ApS, Denmark*

Abstract

In this paper, we present digitalization, control and optimization for entire cement plants based on model predictive control (MPC) technologies. We provide an overview of the employed MPC technologies and report statistics from several industrial implementations. This demonstrates that MPC technology can simultaneously increase the production by 5-10%, reduce the fuel consumption for the cement kiln by 2-4%, reduce the power consumption by 3-8%, and significantly improve the consistency and uniformity of the final cement powder. The cement production industry contributes approximately 8% of the global CO₂ emissions. We outline oxy fuel combustion and carbon capture technologies for cement plants operated along with water electrolysis plants and other power-2-X facilities. These technologies can decarbonize the cement production completely and rely on MPC technologies for their coordinated and optimized operation.

Keywords: Model Predictive Control, Cement plants, CO₂ capture, Power-2-X

1. Introduction

World cement production has increased around 20% in the past decade and reached a high of 4100 million tonnes in 2019. The electrical energy consumed in the cement production is approximately 110 kWh/tonne. 30% of the electrical energy is used for raw material crushing and grinding while around 40% of this energy is consumed for grinding clinker to cement powder. Hence, global cement production uses 18.7 TWh which is approximately 2% of the world's primary energy consumption and 5% of the total industrial energy consumption. Furthermore, manufacturing of cement emits 3 Gton CO₂, which is 45% of industrial's CO₂ emission. (IEA, 2020). Figure 1 describes the cement manufacturing process. It consists of blending, raw meal grinding, pre-calcining, clinker burning and cement grinding. In the crusher, mixing bed, and raw mill, limestone and other materials containing calcium, silicon, aluminums, and iron oxides are crushed, blended, and milled into a raw meal of a certain chemical composition and size distribution. This raw meal is blended and heated in the pre-heating system (cyclones) to start the dissociation of calcium carbonate into calcium oxide. In the kiln, the material is heated, kept at a temperature of 1200-1450°C, and reacts to form calcium silicates and calcium aluminates. The reaction products leave the kiln as a nodular material called clinker. The clinker is cooled in the clinker coolers before being stored in the clinker silos. The clinker is ground with gypsum and other materials such as fly ash in a cement mill to form Portland cement. In this paper, we present two cement related industrial MPC applications. A cement raw-mix blending application and a cement mill grinding MPC application. In applications, we apply a linear model predictive control (LMPC) algorithm with soft constraints. The proposed LMPC in both applications is able to optimally control the system operation, reduce power consumption, and improve production.

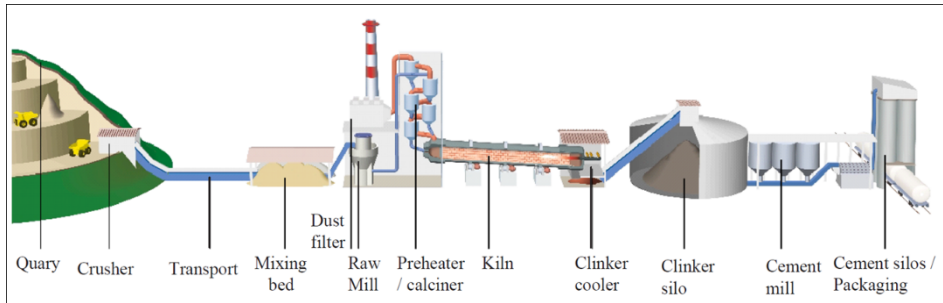


Figure 1. Cement manufacturing process

Carbon capture, storage, and utilization (CCUS) technologies can reduce CO₂ emissions of cement plants significantly. Voldsund et al. evaluated and compared several CO₂ capture technologies, e.g., oxyfuel combustion, chilled ammonia process, membrane-assisted CO₂ liquefaction and calcium looping. Oxyfuel combustion provides the best energy performance and may significantly reduce CO₂ emissions of cement kilns (Voldsund et al. 2019). Power-2-X refers to a series of technologies which can turn electricity into hydrogen and liquid fuels. We apply Power-2-X facilities to convert captured CO₂ into added-value products by using renewable energies. The combination of CCUS and power-2-X technologies provides possibilities for reaching zero-emission cement plants and has the potential to be implemented in the cement manufacturing process of the future.

2. Cement related Industrial MPC Applications

2.1. Cement Raw-Mix Blending MPC Application

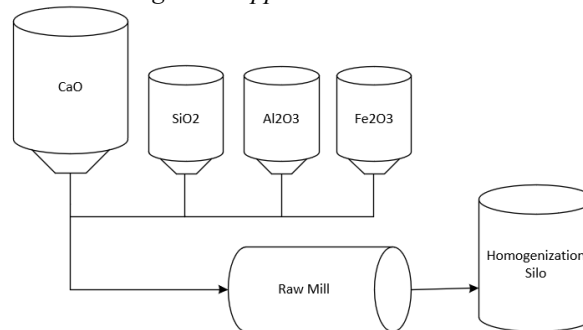


Figure 2. The cement raw-mix blending process.

Figure 2 describes the cement raw-mix blending process. The quarried materials such as limestone, clay, sand, and iron are crushed and sorted into piles in the mixing bed. The chemical compositions of sorted piles are assumed known and described in terms of the oxides CaO, SiO₂, Al₂O₃ and Fe₂O₃. These piles are transported to a raw mill by a conveyor with a certain mixture proportion. In the raw mill, the raw material mixtures are blended and ground to the raw mill with a certain size. The output raw mill then is transported and stored in the homogenization silos. The cement raw-mix blending process is important since it will affect the quality of the clinker and the quality of the final product.

Two coupled input-output models describe the blending process

$$Z_1(s) = G_1(s) \sum_{m \in M} P_m U_m(s) + H_1(s) W_1(s), \quad (1a)$$

$$Z_2(s) = G_2(s) Z_1(s) + H_2(s) W_2(s). \quad (1b)$$

Z_1 are the compositions of the mixture measured before the raw mill process and Z_2 is the compositions measured in the raw meal. $G_1(s)$, $G_2(s)$ are deterministic models and $H_1(s)$, $H_2(s)$ are stochastic models. P_m is a constant parameter indicating the compositions of pile m and U_m is the input from the pile m , W_1 and W_2 are stochastic inputs.

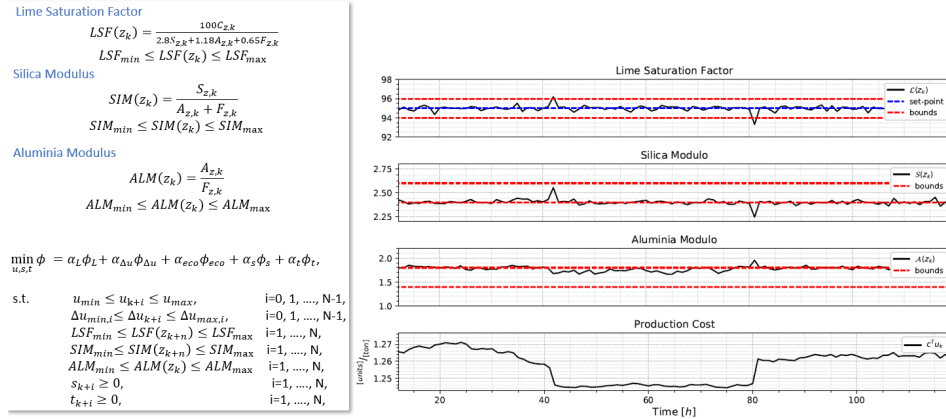


Figure 3. The objective function of the soft constrained LMPC with three compositional parameter constraints and the simulation result.

The raw meal must fulfil a set of quality requirements determined by three nonlinear compositional parameters: lime saturation parameter (LSP), silica modulus (SIM), and alumina modulus (ALM). Figure 3 shows the objective function. The objective function consists of a reference tracking objective ϕ_L , an input rate of movement objective $\phi_{\Delta u}$, an economic objective ϕ_{eco} and two penalty function terms ϕ_s , ϕ_t . For the reference tracking objective, the LSF parameter tracks the given setpoint for the optimal quality. The economic term of the LMPC objective function reduces the system input consumption. To avoid the infeasibility problem caused by hard constraints, we express the compositional parameter boundaries, LSF, ALM, and SIM as soft output constraints.

Figure 3 shows the simulation results obtained by applying the soft constrained LMPC on a cement raw-mix blending model. The soft constrained linear model predictive controller is able to control the raw meal within the compositional parameter bounds. Furthermore, the produced clinker of cement kiln is increased by 3-5% and the energy consumption of the cement raw-mix blending process is reduced by 2-4%. The product quality improvement is up to 30%.

2.2. Cement Mill Grinding MPC Application

Figure 4 describes the cement mill grinding process. Clinker, gypsum, and fly ash are fed to a ball mill by a feed conveyor with a certain proportion. The ball mill has two chambers. In the first chamber, the material is crushed roughly with large steel balls. In the second chamber, the roughly crushed material is finely ground to a fine grey powder. The grey powder then is transported by an elevator to an air separator. In the air separator, particles are suspended by the stream from the bottom and rotated with a specific speed around the

wall. Coarse particles will hit the wall and fall into the recycle pipe. Finally, classified fine particles (cement powder) will be stored in cement silos.

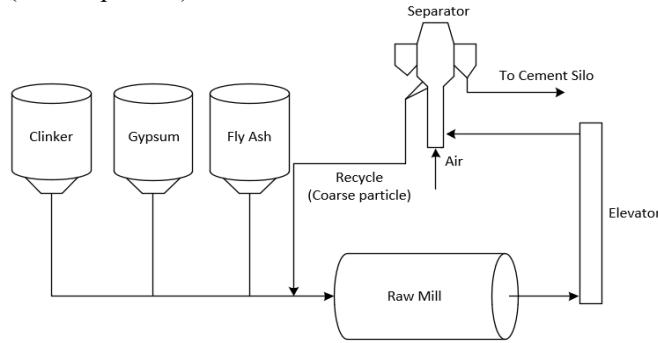


Figure 4. The cement mill grinding process

A MIMO continuous-time input-output model describes the cement mill grinding process

$$Z(s) = G(s)U(s) + G_d(s)(D(s) + W(s)), \quad (2a)$$

$$y(t_k) = z(t_k) + v(t_k). \quad (2b)$$

The system input is $U(s) = [\text{feed rate}; \text{separator speed}]$, $D(s) + W(s)$ are stochastic inputs. $G(s)$ is the deterministic model and $G_d(s)$ is the disturbance model. The measurement output $Y(s) = [\text{elevator load}; \text{fineness}]$ is obtained at discrete time and $v(t_k)$ is the measurement noise. In the grinding process, the unmeasurable variations in material properties will cause significant nonlinearities and uncertainties. These nonlinearities and uncertainties lead to a model mismatch problem. Furthermore, the loading material level of the ball mill can also influence the system operation. Too much loading material will cause a plugging problem and too less loading material will cause energy waste (Prasath et al, 2013).

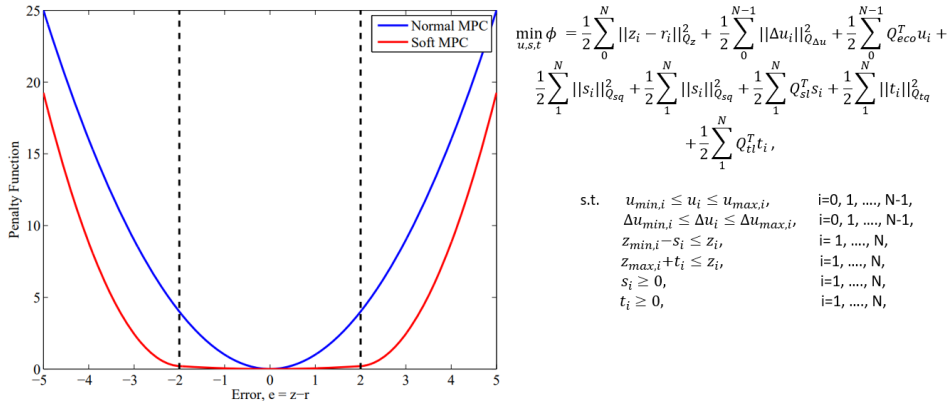


Figure 5. The principle of MPC with soft output constraints and normal MPC (Prasath et al, 2013).

To solve the described problems, we apply a soft constrained LMPC algorithm. The soft constrained LMPC can indirectly control the loading material level of the ball mill to reach the maximum capacity. The model mismatch problem can be solved by soft output constraints. The purpose of applying soft constraints in the cement mill grinding application is to make the controlled system less sensitive to model uncertainties. Figure 5 describes the principle of normal constrained LMPC and soft constrained LMPC. A dead zone is created by the soft output constraints penalty function (red line). In this dead

zone, the penalty function is not sensitive to the change of error $e = z - r$. The value of the penalty function is extremely small and close to 0 in the dead zone and increase rapidly when the error is out of the dead zone. The range of the dead zone depends on the covariance analysis of the nominal system (Prasath et al. 2010). In this way, the controlled system is more robust to the stochastic noise and model uncertainties.

We test the soft constrained LMPC algorithm on a simulated cement mill grinding circuit. The simulation results imply that the soft constrained LMPC can control and stabilize the cement mill grinding circuit. Furthermore, it improves the quality of the final cement powder. The production is increased by 3-8% and the power consumption is reduced by 3-8%. The standard deviation of the final cement powder fineness is decreased significantly, in the range of 10-20%.

3. CCUS & Power-2-X Technologies

3.1. Oxyfuel Combustion CO₂ Capture

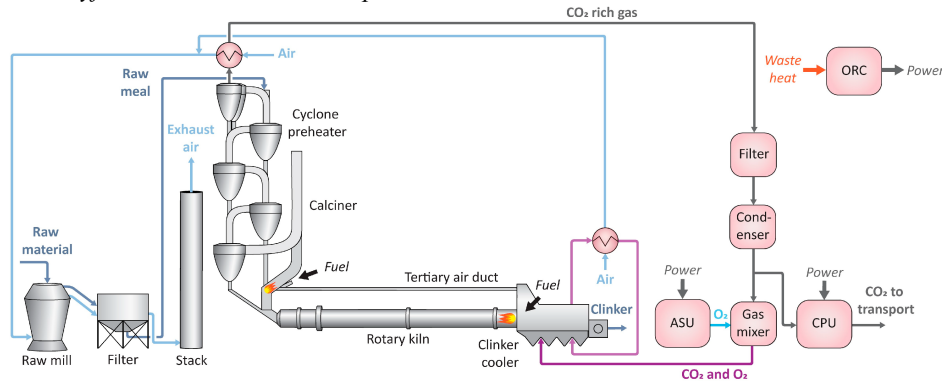


Figure 6. Reference clinker burning line with oxyfuel CO₂ capture.

Figure 6 shows a schematic illustration of a cement kiln with an oxyfuel combustion CO₂ capture system. The oxyfuel combustion process mixes recycled CO₂ and air to produce a CO₂ rich gas. Some of the CO₂ rich gas is purified by a CO₂ processing unit (CPU) and other is mixed with O₂ in a gas mixer. O₂ is produced by an air separation unit (ASU). Both ASU and CPU require power to work and some of the power requirement may be provided by an organic Rankine cycle (ORC). The mixed gas of CO₂ and O₂ is used in the rotary kiln process (Gardarsdottir et al. 2019).

We can apply MPC technology to optimal control the oxyfuel combustion CO₂ capture process. The operation of ASU and CPU requires additional electrical energy and economic MPC can help reduce system energy consumption and improve efficiency. In this way, the combination of the MPC algorithm and CO₂ capture technology can significantly reduce CO₂ emissions and power consumption.

3.2. Power-2-X Advanced Process Control

Figure 7 shows a network diagram that illustrates how Power-2-X facilities interact with CO₂-emitting industry (cement plants). Application of CCUS technologies to capture CO₂ emissions and Power-2-X facilities can convert captured CO₂ into added-value products, e.g. methanol, ammonia, fertilizer, etc. The energy consumption of these facilities is covered by renewable wind and solar energy. Furthermore, integrated electrolytic hydrogen plants can produce H₂ and SNG plants can, fuel and SNG+ by using

renewable energy. These added-value products can be used in other industries, e.g. chemical industries, farms, mobility, power and gas consumers.

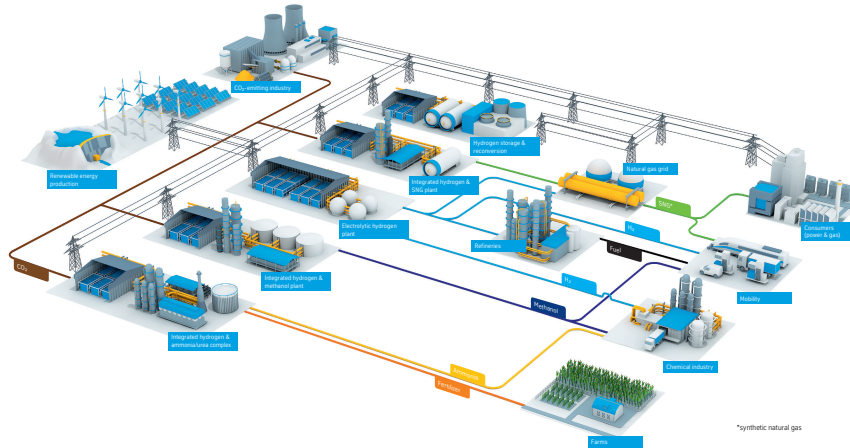


Figure 7. Carbon Capture, storage, and utilization (CCUS) & Power-to-X technologies.

Coordination and optimization of such a system presented in Figure 7 require advanced process control system; and MPC technologies are in development and ideal to optimally control and coordinate the facilities in such an integrated system. In this way, we may achieve zero-emission cement plants that only rely on renewable energy.

4. Conclusion

This paper describes two cement related industrial MPC applications and outlines the oxyfuel combustion CO₂ capture process for cement kilns as well as Power-2-X technologies. A soft constrained LMPC algorithm is implemented on a cement raw-mix blending process and a cement mill grinding process. The proposed soft constrained LMPC technology can increase the production by 5-10%, reduce the energy consumption for cement kilns by 2-4%, and improve product quality significantly. For the cement mill grinding process, the production is increased by 3-8% and the power consumption is reduced by 3-8%. The standard deviation of the final cement powder fineness is decreased between 10-20%. The combination of the MPC algorithms, CCUS as well as Power-2-X technologies have the potential to achieve zero-emission cement plants.

References

- IEA, 2020, World Energy Outlook, <https://www.iea.org/reports/world-energy-outlook-2020>.
- Prasath, G., Reche, B., Chidambaram, M., Jørgensen, J. B., 2013, Soft constrained based MPC for robust control of a cement grinding circuit, IFAC proceedings volumes, Volume 10, Issue 1, pp. 475-480.
- Prasath, G., Chidambaram, M., Reche, B., Jørgensen, J. B., 2010, Soft constrained MPC Applied to an Industrial Cement Mill Grinding Circuit, 9th International Symposium on Dynamics and Control of Process Systems, Volume 9, Issue 1, pp. 288-293.
- Voldsund, M., et al, 2019, Comparison of Technologies for CO₂ Capture from Cement Part 1: Technical Evaluation, Energies, Volume 12, Issue 3.
- Gardarsdottir, S. O., 2019, Comparison of Technologies for CO₂ Capture from Cement Production Part 2: Cost Analysis, Energies, Volume 12, Issue 3.

Process Monitoring with Advanced Analytics for Improved Plant Efficiency in the Pharmaceutical Industry

Deniz Koç,^a Daniel Castro-Rodriguez,^b Dimitrios I. Gerogiorgis^{a*}

^a*School of Engineering (IMP), University of Edinburgh, Edinburgh, EH9 3FB, UK*

^b*GlaxoSmithKline (GSK), Montrose, Angus, DD10 8EA, UK*

D.Gerogiorgis@ed.ac.uk

Abstract

Pharmaceutical manufacturing plants consist of a wide range of complex unit operations, from upstream processing aimed at synthesis of Active Pharmaceutical Ingredients (APIs) to intermediate purifications and downstream processing for final product formulation. To ensure high pharma plant performance at minimised total cost and solvent waste, operators continuously assess key units, frequently with only input-output data available. First-principles modelling is desirable, yet parameterisation is often arduous and elusive. Statistical techniques, e.g. Latent Variable (LV) methods such as Principal Component Analysis (PCA) are widely used to explore industrial data and discover causality patterns. The present paper uses a PCA model to evaluate multicomponent adsorption of Volatile Organic Compound (VOC) gas emissions from salbutamol synthesis, via industrial FTIR data for mixtures of dichloromethane, chloroform, toluene, methanol and ethanol (IMS). Two industrial case studies are considered to analyse the interplay of these solvents, with strong evidence of preferential toluene adsorption on the activated carbon bed discussed.

Keywords: Principal Component Analysis (PCA), Volatile Organic Compounds (VOC).

1. Introduction

Primary (upstream) pharmaceutical manufacturing is characterized by the consumption of large solvent volumes, key in reactions and separations. Volatile Organic Compounds (VOCs) are prominent but also problematic, as their high volatility at standard conditions has catastrophic effects on human health and environment, with extensive research on mitigation (Constable et al., 2007; Balasubramanian et al., 2012; Perez-Vega et al., 2013). Adsorption is a very effective abatement method, as it exploits the strong potential of porous materials to selectively adsorb VOCs from pharma effluents (Yang et al., 2019). Widely employed to treat high waste stream volumetric flows of low VOC concentrations (Das et al., 2004), it can achieve high process efficiency at relatively low energy demand. Nevertheless, adsorbent material procurement and/or regeneration costs are a challenge.

The present paper describes the development and use of a novel *Latent Variable, LV* (specifically a *Principal Component Analysis, PCA*) model to analyse multicomponent VOC adsorption from pharma effluent mixtures onto an industrial activated carbon bed. This ongoing project has three goals. First, to develop and validate LV (PCA) models. Second, to explore causality and improve process understanding. Third, to improve VOC abatement performance by preventing undesirable (e.g. selective adsorption) phenomena.

2. Salbutamol Manufacturing Process Description

Salbutamol is a proven bronchodilation medication: reliably expanding lung airways to comfort patients with respiratory conditions, it is commonly prescribed against Chronic Obstructive Pulmonary Disease (COPD), bronchitis, emphysema and other ailments. Manufacturing this API requires 4 feedstock compounds (*diacetate*, *glycyl*, *benzyl*, *base*), each produced in the plant as part of the salbutamol batch production schedule (Fig. 1). Prior to this project, the first two (*diacetate* and *glycyl*) subsystems were conclusively identified as the two major VOC sources, thus forming the exclusive focus of our study. The root cause of VOC emissions are the distillation units needed for solvent separation. Mass balance analysis for the majority of VOC emissions has previously indicated that: (1) for *diacetate production*, VOCs come from dichloromethane (DCM) atmospheric and vacuum distillation (occurring in vessel 3/V3), and toluene distillation (in vessel 4/V4). (2) for *glycyl production*, VOCs originate from chloroform and toluene distillation (V3).

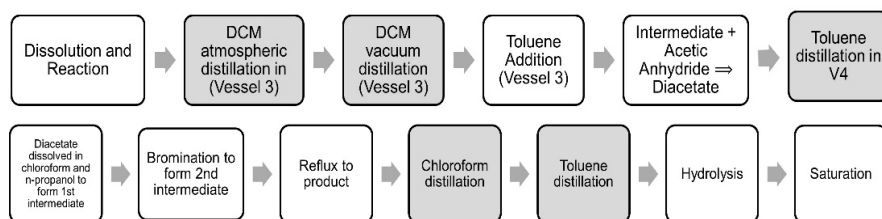


Figure 1: Diacetate (top) and glycyl (bottom) production schedules (grey denotes separation steps).

Fig. 2 describes the VOC abatement system: all solvent emissions due to salbutamol production are simultaneously fed to the same treatment (fixed activated carbon bed) unit. Gas effluents from all vessels mainly comprise four solvents: toluene, dichloromethane (DCM), trichloromethane (TCM, chloroform), and Industrial Methylated Spirit (IMS), a mixture of methanol (MeOH) and ethanol (EtOH) which is widely used in pharma plants. Incoming fumes first enter a heat exchanger, which reduces the Relative Humidity (RH) to 50%, as lower water content is known to improve VOC adsorption by the carbon bed. Once treated, the purified gas stream is subsequently discharged into the atmosphere, with the FTIR sensor collecting VOC concentration data within the exhaust stack (Fig. 2/star).

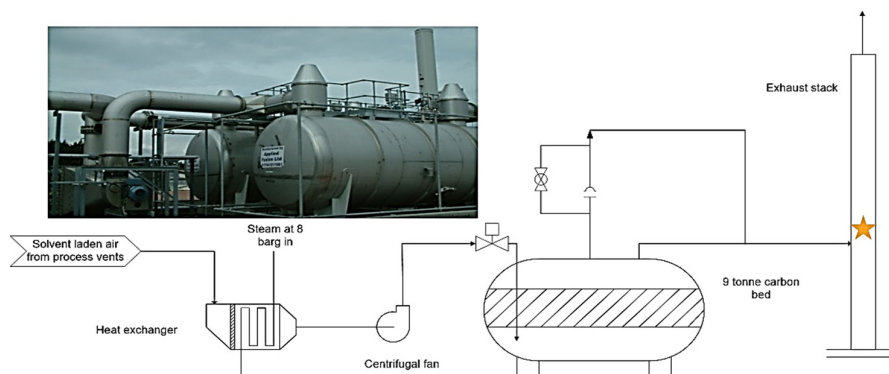


Figure 2: The VOC abatement system (a star denotes the FTIR sensor sampling point of the stack).

3. Principal Component Analysis (PCA) Methodology

Principal Component Analysis (PCA) is an established statistical methodology we use here for VOC adsorption evaluation, to extract process features and elucidate key trends (MacGregor et al., 2005; Rajalahiti & Kvalheim, 2011; Venkatasubramanian et al., 2013). Industrial (GSK) data on VOC emissions production conditions have been compiled from an effluent gas monitoring station for the salbutamol API plant (Montrose, Scotland, UK). The VOC emissions discharged from the API plant exhaust stack (Fig. 2) are continuously monitored via Fourier Transform Infrared Spectroscopy (FTIR), and the key compound concentrations, process conditions (P, T) and vessel numbers, are fed to a nearby station.

For the purpose of this project, two distinct phases and case studies have been considered, corresponding to VOC-critical production stages encompassing distillation separations. The raw FTIR data has been pre-processed and split into a training and a testing dataset. For each dataset, observations have been mean-centred and auto-scaled to reduce bias, and then pre-processed to remove obvious outliers due to sensor shutoffs. The Hotelling T^2 values have been subsequently computed in order to identify other potential outliers.

After essential pre-processing, PCA models are validated and used for two case studies. An important PCA modelling decision is the selection of the number of latent variables (Principal Components, PCs) required, for which we have considered three (3) criteria:

(1) The *first criterion* is the level of variance explained by Principal Components: we have opted for our PCA model to explain a minimum of **80% of the variance** in the data, (2) The *second criterion* is the pair of R^2 and Q^2 : the **coefficient of determination (R^2)** is the proportion of the dependent variable variance that is predictable from independent variables, Eq (1); the **predictive squared correlation coefficient (Q^2)** is the residual variation after applying it to samples held out (the portion of such a sample not explicable by the model), Eq (2). The combination of R^2 and Q^2 metrics helps selecting the number of PCs at the point where R^2 and Q^2 are about equal and as the Q^2 value begins to decline.

$$R^2 = 1 - \frac{\text{Var}(\mathbf{X} - \hat{\mathbf{X}})}{\text{Var}(\mathbf{X})} \quad (1)$$

$$Q^2 = 1 - \frac{\text{Var}(\mathbf{X} - \hat{\mathbf{X}}_{CV})}{\text{Var}(\mathbf{X})} \quad (2)$$

where Var denotes variance of the data sample (R^2) and cross-validation sample (for Q^2). Their calculation is almost identical: the only difference is that R^2 is computed from data used for PCA model training, while Q^2 is computed from held-out (cross-validation) data.

(3) The *third criterion (Root Mean Square Error of Cross-Validation, RMSECV)*, is calculated using Eq. (3). To select the number of PCs, we use the point where RMSECV values increase and then begin to decline, reflecting Q^2 trends from the second criterion.

$$RMSECV = \sqrt{\frac{1}{(IJ)} \sum_{i=1}^I \sum_{j=1}^J (X_{ij}^{(r)})^2} \quad (3)$$

where $X_{ij}^{(r)}$ is the residual of sample i (total: I) and variable j (total: J) after r components.

4. First Case Study: Results and Discussion

The first PCA model addresses the simultaneous operation of toluene distillation in V4 (during the first batch of the diacetate stage), the DCM atmospheric distillation in V3 (during the second batch of diacetate stage) and the chloroform and toluene distillation in V15 during the glycol stage. For ease of reference, this dataset is referred to as the simultaneous operation of V3 and V4, as per the colour-coded process diagram (Fig. 3). The *loadings scatter plot* (left) shows the use of Principal Components (PCs) to explain each variable: here, we note PC2 separates toluene emissions from the rest of the VOCs. The *score plot* (middle) quantifies how PCs explain each observation in a dataset, with data point colours corresponding to time periods in which vessels operate simultaneously. For example, toluene distillation (in V4) and DCM atmospheric distillation (in V3) occur together between 06:00-09:00; data point swarms clearly shift with time (orange arrows). The *biplot* (right) synthesises information from the said two plots, illustrating a combination of all PC scores and loadings which helps us draw quantitative conclusions. A key observation is that the amount of toluene recorded declines with time (data subsets).

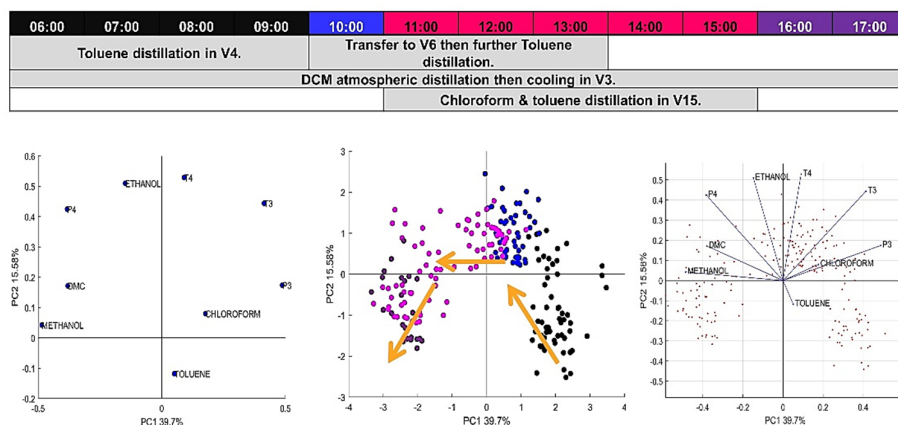


Figure 3. Case study 1 schedule (top); loadings scatter plot, score plot and biplot (L to R, bottom).

The contribution plots are valuable in portraying variance captured vs. PCs used (Fig. 4). The first two PCs explain 55.37% but four (4) PCs can explain 80.81% of data variance. The R^2 and Q^2 plots are also key in deciding the number of PC employed: for the case of 4 PCs, we note that both R^2 and Q^2 values are similar and the Q^2 value starts to decline. The RMSECV plot confirms that 4 PCs suffice here, as it mirrors the trend of the Q^2 plot.

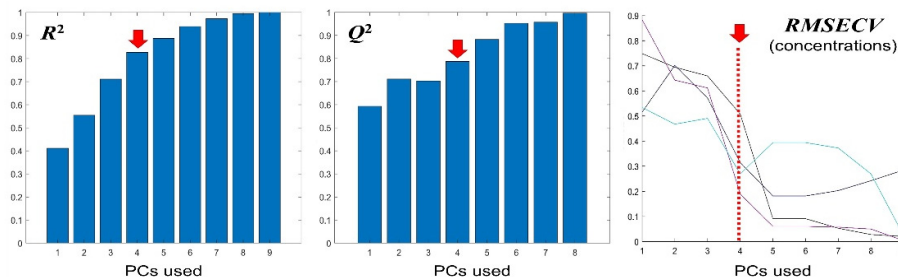


Figure 4. Contribution plots, R^2 and Q^2 and RMSECV plots for variable PC number - case study 1.

5. Second Study: Results and Discussion

The second PCA model addresses the simultaneous DCM vacuum distillation (in V3) and chloroform distillation (in V15), which is followed by toluene distillation therein (Fig. 5). The PCA model construction and analysis methodology is identical to the one described and followed in the first case study, but the situation and observations are quite different. The *loadings scatter plot* shows that PC1 separates toluene from other VOC emissions; the latter change considerably as time progresses, from toluene, to ethanol, to chloroform, to DCM and methanol. The biplot corroborates these indications of industrial importance.

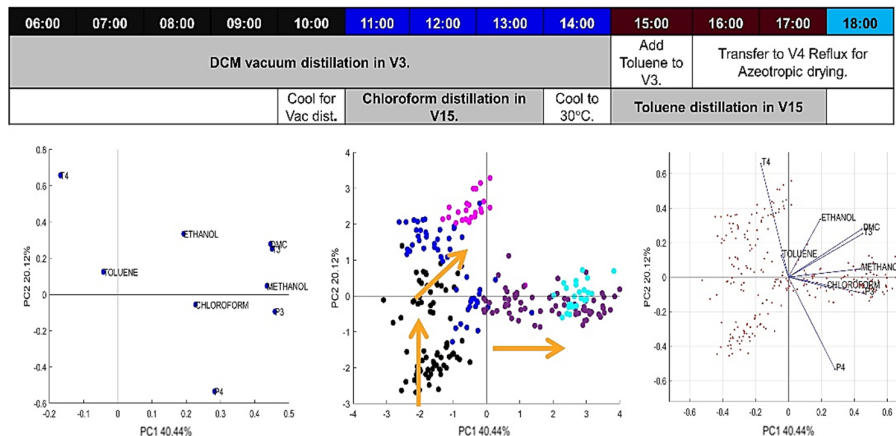


Figure 5. Case study 2 schedule (top); loadings scatter plot, score plot and biplot (L to R, bottom).

From the contribution plots (Fig. 6), we see the first 2 PCs explain 60.57% of the data, while the first 4 PCs explain 82.76% of the data, thus being adequate for our PCA model. The R^2 and Q^2 plots, and the RMSECV plot, concur that 4 PCs suffice to explain the data.

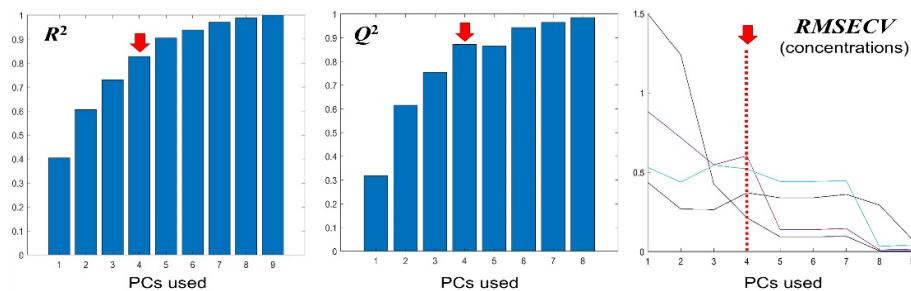


Figure 6. Contribution plots, R^2 and Q^2 and RMSECV plots for variable PC number - case study 2.

The PCA analysis plots for both case studies clearly indicate that toluene emissions progressively decline over time, with concurrent increase of others (DCM, MeOH/IMS), despite the existence of toluene in the raw stream fed to the carbon bed at the same time. This observation suggests the *selective adsorption of toluene* onto the carbon bed, a trait that differs from what happens for other VOC compounds coexistent in the VOC mixture. *Adsorption Theory* informs us that vapour thermophysical properties can be used to probe selective absorption, the strongest indicators being high boiling point and high polarity. Toluene has a similar MW to DCM (92.1 to 84.9), but much higher b.p. (111 vs. 39.6 °C). A similar trend holds for toluene vs. TCM/chloroform (MW= 119.4, but b.p.= 61.2 °C).

6. Conclusions

Upstream pharmaceutical manufacturing inevitably relies on significant solvent volumes for many (especially reaction/separation) uses, with great environmental repercussions. Efficiently capturing VOCs from pharmaceutical effluent streams is a strategic priority for advanced manufacturing striving to embrace greener practices via limiting waste, and VOC pollutant adsorption on activated carbon beds is an established industrial practice. The power of LV (PCA) methods (MacGregor et al., 2005; Kourti, 2005) comes to aid. The present paper has validated and used a PCA model to study VOC adsorption onto an activated carbon bed which processes multicomponent effluents of salbutamol synthesis. We have confirmed that toluene clearly behaves differently from all other VOCs emitted, with strong evidence suggesting its selective adsorption; its higher boiling point vs. the other compounds is an indication of strong adsorbate-adsorbent intermolecular forces. Furthermore, Industrial Methylated Spirit (IMS) appears to play a key role in adsorption, as it appears that it has substantial interaction with the carbon bed and is potentially inhibiting the removal of more hazardous VOCs (e.g. DCM, TCM) from the raw feed. Though selective toluene adsorption has already been suspected from empirical evidence, the potentially harmful role of IMS presence is a new key finding of this research study. Ensuring inlet concentration data acquisition is essential for subsequent studies, as it is clearly important to distinguish sources and composition of streams fed to the carbon bed. Further investigation of toluene and IMS effects via data from various batches/locations can elucidate their potential for detrimental bed interactions and preferential adsorption, towards measures (e.g. effluent feed scheduling) ensuring cost-effective performance.

Acknowledgements

The authors gratefully acknowledge the School of Engineering of the University of Edinburgh for a Kenneth Denbigh Fellowship for MSc in Adv. Chem. Eng. awarded to D.K., as well as a Royal Society Short Industrial Fellowship (2020-21) awarded to D.I.G.

References

- Balasubramanian et al., 2012. Biotrickling filtration of VOC emissions from pharmaceutical industries, *Chemical Engineering Journal* **209**, 102–112.
- Constable et al., 2007. Perspective on solvent use in the pharmaceutical industry, *Organic Process Research & Development* **11**(1), 133–137.
- D. Das, V. Gaur and N. Verma, 2004. Removal of volatile organic compound by activated carbon fiber, *Carbon*, **42**(14), 2949–2962.
- J.F. MacGregor et al., 2005. Data-based latent variable methods for process analysis, monitoring and control, *Computers & Chemical Engineering* **29**(6), 1217–1223.
- T. Kourti, 2005, Application of latent variable methods to process control and multivariate statistical process control in industry, *International Journal of Adaptive Control and Signal Processing* (Special Issue: *Recent Advances in Process Fault Monitoring*), **19**(4), 213–246.
- S. Perez-Vega et al., 2013. A system view of solvent selection in the pharmaceutical industry: Towards a sustainable choice, *Environment, Development & Sustainability*, **15**(1), 1–21.
- T. Rajalahti and O.M. Kvalheim, 2011. Multivariate data analysis in pharmaceuticals: A tutorial review, *International Journal of Pharmaceutics*, **417**(1–2), 280–290.
- C. Yang et al., 2019. Abatement of various types of VOCs by adsorption/catalytic oxidation: A review, *Chemical Engineering Journal*, **370**, 1128–1153.
- V. Venkatasubramanian et al., 2013. A review of process fault detection and diagnosis, Part II: Qualitative models and search strategies, *Computers & Chemical Engineering* **27**(3), 313–326.

A Study of Spectral Envelope Method for Multi-Cause Diagnosis using Industrial Data

Gizem Kuşoğlu Kaya^{a,*}, Pelin Döloğlu^a, Çağla Odabaşı Özer^a, Ocan Şahin^a, Ahmet Palazoğlu^b, Murat Külahçı^{c,d},

^a *Turkish Petroleum Refinery, Körfez, Kocaeli, 41790, Turkey*

^b *University of California, Davis, CA, 95616 USA*

^c *Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*

^d *Luleå University of Technology, 97187 Luleå, Sweden*

gizem.kusoglu@tupras.com.tr

Abstract

Petroleum refineries are complex systems that consist of multiple integrated units. This situation makes it difficult to track down the root cause of abnormal situations that occur during production. It is noted that abnormal situations usually trigger plant-wide oscillations in a number of measured process variables. Therefore, root cause detection is often attempted to be carried out by examining these trends in process data. Observing multiple effects and underlying problems at the same time presents a challenge in determining the root cause by examining trends only. In this study, spectral envelope is used to detect oscillations by identifying the variables and categorizing them based on a statistical hypothesis test which produces Oscillation Contribution Index (OCI) in order to isolate potential root cause variables. Two distinct abnormal events in the hydrocracker unit that occur simultaneously were successfully isolated and the root causes could be assigned by using the spectral envelope analysis.

Keywords: Root-cause analysis, Multi-effects, Plant-wide oscillation, Spectral Envelope

1. Introduction

Plant-wide oscillation is a general term used to express the propagation of an oscillatory effect throughout the entire plant originating at one location in the plant. Plant-wide oscillations may cause deterioration in the quality of the product, energy losses, sudden problems/shut-downs in the plant. Poor tuning in control loops, valve problems such as valve stiction and fouling, interactions between different plants are common root-causes of oscillations.

In the literature, different techniques have been proposed for detecting similar characteristics of the variables and diagnosing root-causes. Duan et al. (2014) provided a review of these detection and diagnostic techniques (Choudhury MAAS, 2011). High-density plot is one of the off-line visualization technique to show spectral plots of related variables. Thornhill et al. (2003) have tried to detect the oscillations that have a common frequency based on the regularity of zero-crossings of the control error signal. Jiang et al. (2007) have proposed a new method called the spectral envelope to cluster

the variables, which have similar spectral behaviors. This paper shows the capability of spectral envelope method for diagnosing two different root-causes based on a real plant data from a refinery in Turkey. One of the major root cause is the temperature effect associated with a stream coming from another plant. The other root cause is related to the fouling of a control valve for a level control loop in a downstream process drum that acts to compensate for the effect of the temperature change of this stream.

2. Detection and Diagnosis with the Spectral Envelope Method

2.1. The Spectral Envelope Method

Finite number of data samples is used in a data matrix of m variables and n samples represented as $X = [x(0), x(1), \dots, x(n-1)] \in R^{m \times n}$. Here, x is a multivariate, vector-valued time series on R^m as in Eq. (1):

$$x = [x_1(t) \ x_2(t) \ \dots \ x_m(t)]^T \quad t = 0, \pm 1, \pm 2, \dots \quad (1)$$

To simplify the calculations and obtain more meaningful results, it is preferable to use the matrix created with the normalized form of each variable (Jiang et al., 2007). The periodogram of the X matrix is calculated for a large sample set based on the fast Fourier transform assumption as:

$$\hat{I}_n(\omega_k) = \frac{1}{n} \left[\sum_{t=0}^{n-1} x(t) \exp(-2\pi i t \omega_k) \right] \left[\sum_{t=0}^{n-1} x(t) \exp(-2\pi i t \omega_k) \right]^* \quad (2)$$

where $\omega_k = k/n$, for $k = 1, 2, \dots, n/2$ are the Fourier frequencies. This approach is not exactly consistent with the power spectral density (PSD). Therefore, a smoothed periodogram estimate by creating the symmetric moving average of $\hat{P}_X(\omega_k)$ can be used:

$$\hat{P}_X(\omega_k) = \sum_{j=-r}^r h_j \hat{I}_n(\omega_{k+j}) \quad (3)$$

where h_j are symmetric weights ($h_j = h_{-j}$) and $\sum_{j=-r}^r h_j = 1$. The calculated periodogram will be smoother with higher r value. In the refinery case study, h_j is assumed as: $h_j = \frac{(r-|j|+1)}{(r+1)^2}$, with $|j| = 1, 2, \dots, r$ where $r = 2$. Once the moving

average of a periodogram is calculated using Eq. (3), we have $v_n = \left(\sum_{j=-r}^r h_j^2 \right)^{1/2}$. The

eigenvalues of $\hat{P}_X(\omega)$ in the decreasing order and the corresponding eigenvectors are represented as $\{\lambda_1(\omega), \lambda_2(\omega), \dots, \lambda_m(\omega)\}$ and $\{\beta_1(\omega), \beta_2(\omega), \dots, \beta_m(\omega)\}$, respectively.

When the maximum value of the eigenvalues $\{\lambda_1(\omega), \lambda_2(\omega), \dots, \lambda_m(\omega)\}$, say $\lambda(\omega)$, is plotted against the frequency $\omega = 1, 2, \dots, n/2$ and if a peak appears at any frequency, it implies that a plant-wide oscillation may be possible at that frequency. For further detection of oscillating variables, the asymptotic covariance matrix of the optimal scaling vector $\hat{\beta}(\omega)$, $V_{\beta}(\omega)$ is estimated according to the calculated eigenvectors and eigenvalues:

$$V_{\beta}(\omega) = v_n^{-2} \lambda_1(\omega) \sum_{l=2}^m \lambda_l(\omega) [\lambda_1(\omega) - \lambda_l(\omega)]^{-2} \beta_l(\omega) \beta_l^*(\omega) \quad (4)$$

The following distribution

$$\frac{2|\hat{\beta}_{1,j}(\omega) - \beta_{1,j}(\omega)|^2}{\sigma_j(\omega)} \quad (5)$$

is a chi-square distribution with 2 degrees of freedom, where $\hat{\beta}_{1,j}(\omega)$, $j = 1, 2, \dots, m$, j^{th} element of $\hat{\beta}_1(\omega)$, and $\sigma_j(\omega)$, j^{th} diagonal element of the estimate of $V_{\beta}(\omega)$. For the null hypothesis ‘ $\beta_{1,j}(\omega) = 0$ ’, the test statistics will be $\frac{2|\hat{\beta}_{1,j}(\omega)|^2}{\sigma_j(\omega)}$ and the critical value is set to $\chi_{2,\alpha}^2$ which is the upper $\alpha\%$ of the $\chi_{2,\alpha}^2$ distribution with 2 degrees of freedom. If the test statistic exceeds the critical value, the null hypothesis is rejected and hence it is concluded that the time series has an oscillation at that frequency for the j^{th} variable (David et al., 2000).

2.2. Root Cause Diagnosis via the Spectral Envelope Method

The root-cause determination of unit-wide oscillations and the isolation of the root cause are challenging tasks. To isolate key variables after applying the chi-square test, Jiang et al. (2007) proposed a new index called as the Oscillation Contribution Index (OCI):

$$OCI_j(\omega) = \frac{\hat{\beta}_{1,j}(\omega)}{2\sigma_{\beta}(\omega)} \quad (6)$$

where $\sigma_{\beta}(\omega)$ is the standard deviation of the optimal scalings of variables which have oscillations based on the chi-square test. This index shows the contribution of variables to the oscillation at that frequency. The criterion for the most contribution to the oscillation is $OCI(\omega) > 1$ at frequency ω .

3. Refinery Case Study

3.1. Process Description

The Steam-Methane Reforming Unit, which is known as the Hydrogen Generation Unit, is a common process due to the increasing demand on producing environmentally friendly and light-end products by cracking and desulfurization in refineries and other petrochemical industries. In this paper, a process anomaly experienced in a specific part of this unit has been studied.

Since the catalyst of the reforming reactor is very sensitive to sulfur and halogens, the steam-hydrocarbon mixture undergoes a pretreating process before entering the reforming reactor. In the pretreating reactors, sulfur and halogens in the natural gas and naphtha should be converted to H_2S and HF/HCl in order not to deactivate the reforming catalyst. After the pretreatment section, a mixture of CO , CO_2 , CH_4 and H_2 , which is also known as syngas, is obtained and goes to the shift reactor for conversion of CO to CO_2 .

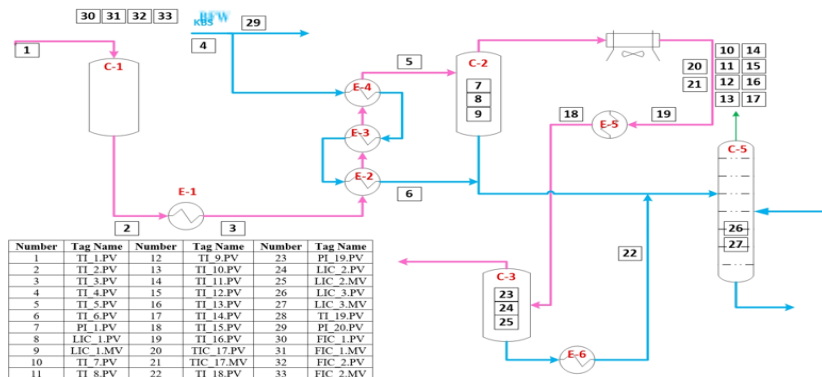


Figure 1. Process Flow Diagram of Part of the Hydrocracker Unit

In Figure 1, the process flow diagram of the section that includes the shift reactors in the unit is shown. The syngas coming out of the shift reactor (C-1 in Figure 1) is cooled first in the heat exchanger E-1 and then on the shell-side of E-2, E-3, E-4 heat exchangers with boiler feed water (BFW) on the other side. The purpose of the cooling of the stream after the reactor is the condensation of the steam and then, to remove the resulting water from the gas stream as liquid in the hot condenser drum (C-2 in Figure 1). Then, the water-gas mixture is separated into gas and water streams in C-2. The gas stream obtained from C-2 is cooled to a lower temperature in the E-5 and E-6 coolers and sent to the cold condenser drum (C-3 in Figure 1) for removing the water in the gas phase. The water stream obtained from C-2 is mixed with the output water stream of C-3 and fed into a stripper column (C-5 in Figure 1).

3.2. Case Study

As mentioned above, BFW is generally fed into the heat exchangers as a cold stream. However, BFW may sometimes be sent as a hot stream into the exchangers depending on the operations of the BFW generation unit. If the BFW stream coming from the unit

is hot, water present in the syngas cannot be condensed to the extent when BFW is charged as a cold stream. As a result, the upstream of C-2 contains much more water as steam and more steam condenses in E-5 and E-6 coolers. Condensed excess water causes an increase in the C-3 drum level and this situation is normally controlled by the valve connected to the level controller in the bottom of the drum. As time passes, however, the presence of water leads to fouling (often as scaling) inside the valve and so, in the case of a level increase, the valve stem opening cannot be effectively controlled, leading to a sticky valve situation. In this study, the simultaneous effect of temperature variation of the BFW stream and valve fouling were investigated.

3.3 Industrial Data Sets

Total of 33 tags from the hydrocracker unit in the Turkish Petroleum Refinery were studied. Four of these tags belong to the control loops and the remaining 29 tags to sensor transmitters. The points where the numbers in the general flow schema given in Figure 1. show the location of tags. The labels of both transmitter and controller information corresponding to the numbers are noted in the table in Figure 1. TI, PI, LIC, TIC and FIC represent temperature indicator, pressure indicator, level controller, temperature controller and flow controller, respectively. In this study, using historical data, the spectral envelope was estimated on a daily basis for the 64 days before the event occurred. The sampling interval was 60 s and so there were a total of 1440 data points in one day. The tags of daily data are placed in a matrix with columns as variables. The data matrix is standardized and then the spectral envelope on a daily basis is calculated using Eq. (2) to (6). Based on the estimated spectral envelope and determined oscillation frequencies, a statistical hypothesis test is performed to determine if the variables oscillate with these frequencies. Then, the OCI index is used to isolate key variables that cause (or associated with) the problem.

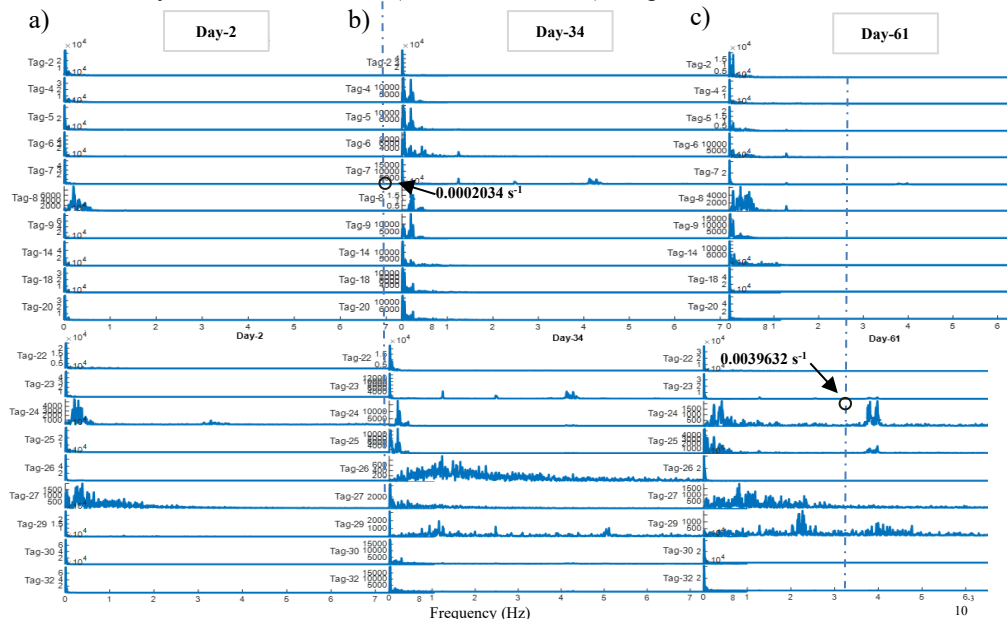


Figure 2. Power Spectrum Analysis for (a) 2nd, (b) 34th, (c) 61st days of operation

4. Results and Discussion

With the spectral envelope analysis performed for 64 days, the OCI index values show that about 20 variables could be root-cause variables contributing to the oscillation(s). Therefore, the rest of the article will examine these 20 variables known to have an effect, which are indicated on the y-axis in Figure-2. The common frequencies of the oscillations are determined by the spectral envelope and the results are verified by estimation of Power Spectra. Power spectra on the 2nd(without any problems), 34th(BFW temperature switch), 61st(with valve stiction) days are given as samples in Figure 2a, b and c, respectively.

It is expected that the effect of temperature change can be observed on days when the temperature of the BFW stream changes. If there is no temperature effect, the normal oscillation period of the level control loops will be observable. In Figure 2a, the effect of the level itself over oscillation is seen at Tag-8, Tag-24 and Tag-27. Temperature of BFW was supplied as slightly cooler to the plant on the 34th day. Therefore, the oscillations observed in all tags (Tag-4, 5, 8, 9, 14, 18, 20, 24, 25) could be attributed to the BFW temperature shift with a frequency of 0.0002034 s^{-1} as detected in Figure 2b. On the 64th day, the valve fouling reached a critical level and caused a shut-down in the unit. The fouling effect of the valve is also seen in the power spectrum of the data set from a few days before the event, i.e., 61 days, given in Figure 2c. Process and manipulated variables of the level controller (Tag-24 and Tag-25), on the C-3 drum are dominantly oscillating with a frequency of 0.0039632 s^{-1} . Power spectrum results are validated with spectral envelope as in Figure 3 and two distinct effects can be easily isolated. Variables contributing to the specified frequencies as OCI>1 were also noted on the below graph.

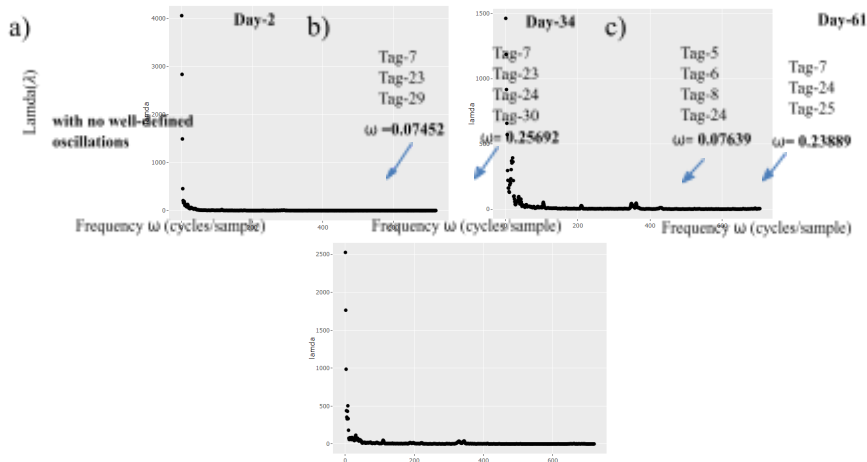


Figure 3. Spectral envelope of the 33 variables

In the early days, there was no frequency for valve contamination (as in Figure 2a). A new frequency (0.25692 s^{-1} or 0.23889 s^{-1}) with OCI>1 for both Tag-24 and Tag-25

appear over time (as in Figure 2b, 2c) and maintained its continuity close to the unit shut-down.

5. Conclusion

In this study, the simultaneous events affecting a process unit and how these effects can be isolated using a post-mortem analysis are studied. With the increasing fouling of the valve the unit has been experiencing a gradual build-up a failure mode. According to the results of spectral envelope method, the impact of the fouling can be detected up to a few days in advance without causing any serious problems in the unit, while also identifying a shift in the BFW stream temperature occurring during the same period.

References

- P. Duan, T. Chen, S.L. Shah, F. Yang, 2014, Methods for root cause diagnosis of plant-wide oscillations, *AIChE Journal*, 60, 2019-2034.
- M. Choudhury, Plantwide oscillations diagnosis-current state and future directions, *Asia Pac J Chem Eng.* 2011, 484-496.
- N.F. Thornhill, B. Huang, H. Zhang, 2003, Detection of multiple oscillations in control loops, *Journal of Process Control*, 13, 91–100.
- H. Jiang, M.A.A. Shoukat Choudhury, Sirish L. Shah, 2007, Detection and diagnosis of plant-wide oscillations from industrial data using the spectral envelope method. *Journal of Process Control*, 17, 143-155.
- D. S. Stoffer, D.E. Tyler, D.A. Wendt, 2000, The spectral envelope and its applications, *Statistical Science* 15, 224–253.

Dynamic Model Based Detection of Cyberattacks in Industrial Facilities

Kathrin Reibelt ^{a,*}, Hubert B. Keller ^a, Veit Hagenmeyer ^a, Jörg Matthes ^a

^a *Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany*

kathrin.reibelt@kit.edu

Abstract

The proceeding digitalisation of industrial facilities does not only provide meeting the requirements of flexibility and just-in-time production. It comes with increasing vulnerability against cyberattacks. This already manifests in high numbers of cyberattacks against industry 4.0 facilities (KasperskyLab, 2020). In these cyber-physical-systems, cyberattacks do not only affect the informational system, they can also cause damage in the physical part that is closely coupled to the informational part. Therefore, new methods for securing industry 4.0 facilities against cyberattacks are required. The method introduced in this paper continuously checks if the measurements and setpoints fulfil the systems physical relations utilizing data reconciliation and extends it by consideration of informational properties for detecting, localizing and selectively combatting cyberattacks.

Conventional protection methods like the comparison of each measurement with a specific range of safe service cover only basic sensor failures and trivial cyberattacks. A more sophisticated protection method, which analyses the consistence of the static system state with respect to a static system model of the process was introduced in (Reibelt et al. 2020) using static data reconciliation. The new method presented here is also applicable in dynamic system states. This paper shows the adaptation of the data reconciliation based approach for dynamic system relations and demonstrates its application.

Keywords: cyber-security, data reconciliation, industry 4.0, cyber-physical systems, dynamic states

1. Introduction

Cyberattacks targeting on physical damage are usually based on false data injection. Values of measurements (or actuator setpoints) are exchanged by different values. The manipulated values pretend a system state that is different from the actual one. For causing physical damage, the manipulated system variables can pretend to require setpoint values different from the ones the actual system state requires. This can cause control actions running the system into a critical state or prevent mandatory control interventions. Manipulated actuator setpoints can directly cause critical system states. Attacks can also aim at production losses, either by disturbance of product quality or by extending down times.

Many of these manipulations can be realized in a manner they are not detected by common protection systems on comparing values to ranges of safe service. As data reconciliation uses a more detailed description of the system, taking into account the

relations between variables, it is able to detect and localize false data injection in many more cases.

The manipulation can be implemented e.g. by modifying the measurement itself, by modifying the preprocessing of values in smart sensors, by modifying the communication interface, messages, or communication connections or by manipulating the values in the process control system. This variety of possible attack vectors as well as the existence of unknown ones cannot be satisfactorily faced with attacker models. For our detection method, prior knowledge that is more general is used to improve the detection, such as exposition to networks or informational commonalities effecting the vulnerability towards cyberattacks. This also prefers anomalies caused by manipulations against sensor failures that are also detected by the anomaly detection.

This paper will give a brief summary to gross error detection based on data reconciliation and our extension for the detection of cyberattacks in section 2. In section 3, the modelling of dynamic system relations and the adaptation for data reconciliation is described. Section 4 shows the evaluation of the detection performance. Therefore, a simulated example system with manipulations is used to generate receiver operating characteristic curves (ROC-Curves).

2. Extended Data Reconciliation Gross Error Detection

For the evaluation of system variable values, the physical relations, restrictions and conservation laws are described by mathematical equations. With the vector y containing all observed system variables, the equations are converted to one common system equation

$$\mathbf{A} \cdot y = 0 \quad (1)$$

Statistical errors of the variables as well as manipulations cause deviations from this equation that manifest in equation residuals $r = \mathbf{A} \cdot y$. Data reconciliation allows for calculating optimized variables \hat{y} that are the closest values to the noisy system values y , fulfilling the system equation (1). The gross error detection methods used in this paper, are based on the evaluation of the residuals r or on the measurement deviations $(y - \hat{y})$.

- For the *measurement test* the measurement deviation of every system variable is considered. The test statistics is calculated by weighting the deviation with the corresponding variance \mathbf{V}_y : $d = \mathbf{V}_y^{-1} \cdot (y - \hat{y})$ (Tahamhane, 2010)
- The *global test* evaluates, if the sum of the residuals $r = \mathbf{A} \cdot y$ is compatible to a system without manipulations, with respect to the variance in the residual space: $\gamma = r^T \cdot \mathbf{V}_r^{-1} \cdot r$. For localizing manipulations, the variables are removed one by one and the test is repeated. (Madron, 1985)
- For the *hypotheses test*, reference residuals f for norm manipulations are calculated. They form the column vectors of a matrix \mathbf{F} . These reference residuals are compared to actual residuals f by calculating the test statistics: $T = (\mathbf{F}^T \cdot \mathbf{V}_r^{-1} \cdot \mathbf{F})^{-1} \cdot (\mathbf{F}^T \cdot \mathbf{V}_r^{-1} \cdot r)^2$. (Narasimhan et al., 1987)

In all three test methods a threshold is applied to decide whether current residuals/deviations can be explained by measurement noise or indicate a cyberattack.

In order to facilitate the detection of manipulations of combinations (subsets) X_i of variables Y with common properties P_i : $X_i = \{Y|P_i\}$ (such as common subnet or operation system of the sensors, having a common type of vulnerability towards cyberattacks) in this paper an adaption of the threshold is proposed. Therefore, the general threshold is multiplied by an adaptation factor $(1 - S_{MT/HT} \cdot I_i)$ for measurement/

There are publications using an optimization method they call dynamic data reconciliation (e.g. Bai et al. 2006), but they use a tuning factor for adapting the predicted values $x(k+1)$ to the measured values similar to the Kalman-Filter. This adaption hides manipulations that do not show fast changes, therefore it is not usable for manipulation detection.

4. Evaluation

The dynamic system example was constructed on the standard example of data reconciliation shown in Figure 1 (Narasimhan et al., 2000). Between the measurement/actuator points dampings with time constants of $T_2 = 2$ s, $T_3 = 3$ s, $T_4 = 4$ s, $T_5 = 5$ s, $T_6 = 6$ s are introduced. As varying input signal a sine function of amplitude 1 and period time of 2π s was used. The total time considered is 10 s.

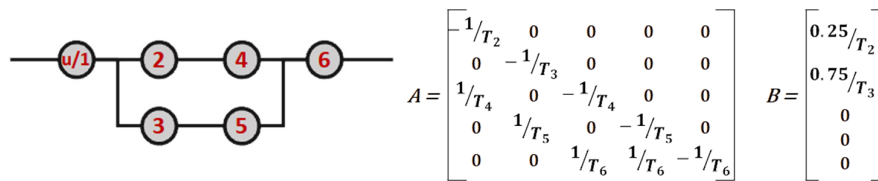


Figure 1: Standard example, cooling circuit with six connected devices, matrices for continuous dynamic system model.

From these matrices describing the dynamic system in time-continuous form, the time-discrete form is calculated for sample time 1 s (zero order hold) and finally the matrix \mathcal{A} is determined.

The variables are not coupled with the same strength, so an additional normalization of the test statistics is necessary. Therefore, for a normalized manipulation of all variables and combinations, the test statistics of the variables or combinations manipulated indeed are calculated without statistical errors. For demonstration in Table 1 the test statistics for manipulations in single variables are shown for a manipulation height of 1 and 2 for measurement test and hypotheses test. For the global test all test statistics would be 0, since the manipulated variable is directly eliminated for calculating the test statistics.

Table 1: Test statistics of single variable in case of manipulation of the respective variable. Manipulation amplitudes are 1 and 2.

Var	Measurement Test		Global Test		Hypotheses Test	
	manip 1	manip 2	manip 1	manip 2	manip 1	manip 2
1	0.9792	1.9584	0	0	0.0641	0.2566
2	0.7228	1.4456	0	0	0.3028	1.2115
3	0.5025	1.0050	0	0	0.1069	0.4276
4	0.4145	0.8290	0	0	0.0567	0.2267
5	0.3205	0.6410	0	0	0.0392	0.1569
6	0.1265	0.2531	0	0	0.0156	0.0626

In the case of linear systems for hypotheses test, the test statics for manipulations with amplitude of 2 are always 4 times the value of the ones for manipulations with amplitude of 1. For measurement test they are only double the value. These relations are the same for all combinations as well. Therefore, a normalization with the test statistics value of the normalized manipulations does make sense for measurement test and hypotheses test. The relations also allow for determining the threshold adaptations S for measurement test and hypotheses test. If e.g. an unexpected manipulation should be detected at double the

amplitude of an expected one, the adapted threshold can be half the general threshold for measurement test ($\Rightarrow S_{MT} = 0.5$), a quarter of the threshold for hypotheses test ($\Rightarrow S_{HT} = 0.75$). For global test, this kind of determining the threshold adaption is not feasible, as without statistical errors, the test statistics values always become zero (test statistics for GT in Table 1). For the evaluation example the threshold was chosen to be raised by up to three quarters ($\Rightarrow S_{GT} = 0.75$).

For evaluation of the new detection method, the statistic errors are simulated with a variance of 1. More than 23,000 manipulations are simulated with amplitudes of ± 3 . The combinations of variables X_i with common properties (common type of vulnerability towards cyberattacks, i.e. expected manipulations) are listed in Table 2.

Table 2: Example for prior knowledge, used for evaluation.

Property P_i	Factor I_i	Variables $X_i = \{Y P_i\}$
e.g. online configurable	1	1
e.g. value transmitted online	1	3
e.g. non-decrypted Wifi communication	1	6
e.g. common operating system	1	1, 5
e.g. shared hardware	1	4, 6
e.g. common firmware producer	1	2, 3
e.g. common communication protocol	1	1, 2, 3
e.g. common network layer with high exposition	1	3, 4, 5
e.g. common software producer	1	2, 4, 6
...	1	3, 5, 6

For creating comprehensible results, the vulnerability factor is set to maximum ($I_i = 1$) for each property P_i . When testing for expected manipulations X_i , the adapted threshold is considered, otherwise the general threshold is used.

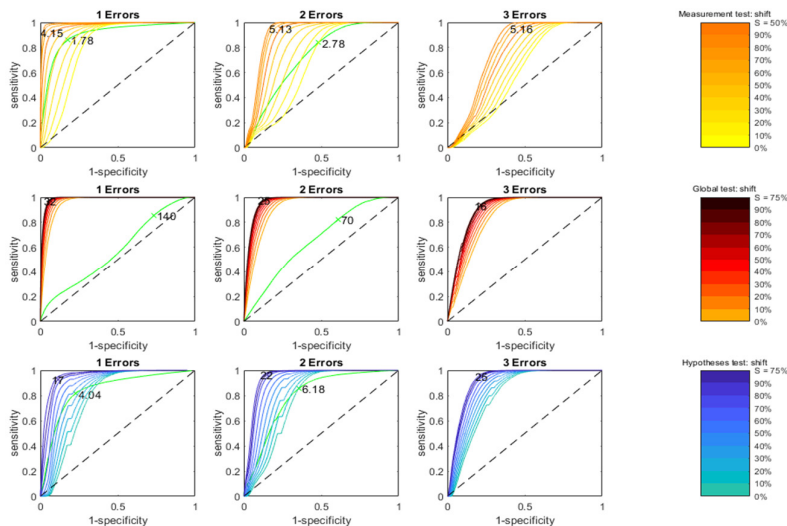


Figure 2: ROC-Curves for the three tests for different numbers of manipulated variables and different threshold adaptations.

Based on all simulated manipulations, for determining the ability of the tests and the optimal general thresholds, ROC-curves (Fawcett, 2005) are generated. ROC-curves are determined by plotting the true positive rate (sensitivity) against the false positive rate (1-

specificity) for different general thresholds (Figure 2). For optimal selectivity, the general threshold referring to the point with the highest distance to the diagonal is selected. This threshold is labelled in the plots. For all detection methods, the ROC-curves are determined for manipulations of single variables (1 Error), for conjoint manipulation of two variables (2 Errors) and for conjoint manipulation of three variables (3 Errors). To show the effect of the threshold adaption, the ROC-curves are calculated for different threshold adaptations S . The brightest coloured curve is the ROC-curve without threshold adaption ($S = 0$), and the darkest coloured curve the one for the adaption values given above. The green coloured curves refer to the test for selecting higher order combinations as candidates for the actual manipulated variables. For measurement test (yellow) and hypotheses test (blue) the threshold is used to check if single variables or certain combinations of variables are a subset of the actual manipulated variables. The area between the ROC-curves and the diagonal (random guesses) is a measure for the performance of the test. Already when using the general thresholds (adaption $S = 0$), the tests performs much better than random guesses. By introducing prior knowledge of commonalities, using adaptations $S > 0$, all three test methods show a high performance, as the majority of the area in the left upper half is below the ROC-curves.

5. Conclusion

In this paper, the adaption for dynamic systems of our new model-based detection method for cyberattacks is introduced. It is shown to come with a high performance in detection using classical data reconciliation extended by prior knowledge regarding informational properties of the system variables. The dynamic model based detection enables the application for industry 4.0 facilities during non-static system states, which are currently rarely covered by security measures. Our method is scalable, following the real system size, as most calculations can be prepared prior to its application. As the hypotheses test and the global test both evaluate resulting residuals, they can easily be adopted for any kind of system, including nonlinear systems. In case of the measurement test, an analytical reconciliation would be required. The method enables the selection of targeted countermeasures for reducing the risk of damages and losses.

References

- Kaspersky Lab ICS CERT, 2020, Threat landscape for industrial automation systems in H1 2020, Moscow, Russia, ics-cert.kaspersky.com
- Kathrin Reibelt, Jörg Matthes, Hubert B. Keller, Veit Hagenmeyer, 2020, Identification and Localization of Cyber-Attacks in Industrial Facilities, 30th European Symposium on Computer Aided Process Engineering Part A, pp.1747-1752
- Shuanghua Bai, Jules Thibault, David D.McLean, 2006, Dynamic data reconciliation: Alternative to Kalman filter. *Journal of Process Control* 5 (2006) 16, S. 485 – 498.
- Ajit C. Tahamhane, 2010, A Note on the Use of Residuals for Detecting an Outlier in Linear Regression, *Biometrika* 69(2), pp. 488–489
- Frantisek Madron, 1985, A New Approach to the Identification of Gross Errors in Chemical Engineering Measurements, *Chemical Engineering Science - CHEM ENG SCI* 40 (1985), pp. 1855–1860
- Tom Fawcett, 2005, An introduction to ROC analysis, *Pattern Recognition Letters* 27, pp. 861–874

Evaluation of control strategies in CO₂ capture unit

Pavan Kumar Veldandi ^a, Abhisek Roy Chowdhary, Anuradha Durvasula,
Rajkumar Krishnamoorthy

*Cognizant Technology Solutions, CDB-CCP- Industrial Operations, Hyderabad,
500032, India*

Pavan.veldandi@cognizant.com

Abstract

Post Combustion CO₂ Capture process (PCC) is a popular technology for regulating CO₂ emissions from Coal-based power generation plants. The process uses Mono Ethanol Amine (MEA) solvent to absorb CO₂ from Flue gas. The CO₂ rich solvent is regenerated in a stripping tower with kettle reboiler using process steam as heating media. The lean MEA solution is recirculated back to absorption column, and the recovered CO₂ is taken for further processing depending on the nature of demand.

The PCC unit control set-up has challenges in coping up with process upsets caused by variations in power demand resulting in lower operating rate for the power plant. The rapid reduction of flue gas flow rate to the absorber leads to a decreased load on the regeneration stripping tower, necessitating steep reduction of the steam flow to the reboiler. The problem is caused by the dynamic response rate of the regenerator-reboiler combine, which must match the fast-dynamics of the flue gas feed rate. The process upset in the absorber-stripper system takes time to settle down, leading to CO₂ slip from the PCC unit. The loss of CO₂ is not only an economic concern but an environmental issue as well. This study intends to observe this problem in a simulated environment and evaluate options to ensure faster stabilization of the MEA absorber-regenerator system; to minimize loss of CO₂ during changes in production rate of the power plant. The study involves developing a high-fidelity dynamic simulation model of the MEA absorber-stripper unit using 1st principles unit-op models and rigorous thermodynamics on a commercial dynamic simulator platform. Once the steady state conditions and disturbance caused in the system due to plant turn-down have been successfully replicated; the model is used to study the impact of process upsets on key operating parameters like solvent flow rate to absorber, regenerator bottom and top temperature, steam flow-rate and temperature of the reboiler, CO₂ slip from the absorber and the key parameter of overall CO₂ capture by the system. The next step is evaluation of various control schemes and operating strategies in their effectiveness of controlling the key process parameters viz. lean solvent flow rate to absorber, ratio of solvent flow to feed gas rate and regenerator bottom / reboiler temperature. In the final step the optimal control scheme and operating strategy to be followed by plant operators for the absorber-stripper system is identified. The selection is made based on the capability of the strategy to quickly react to the process upsets caused by reduction in power plant production rate, and to establish steady state operations in the least amount of time; reducing CO₂ loss during the process transients.

Keywords: Power plant, CO₂ capture, MEA solvent, Dynamic simulation, Control Strategy.

1. Introduction

Post combustion CO₂ capture unit is an important part of fossil fuel-based power plants and plays a key role in minimizing CO₂ emission. A typical challenge faced by the PCC unit is the disturbance in the flue gas feed rate to the PCC unit due to variation in demand for power production from the power plant. Timely controller action to minimize the effect of variation in feed gas rate and reestablishment of steady state conditions are required to ensure minimal impact on overall CO₂ capture quantity and environmental consequences. Several studies on control strategy of PCC unit have been done in recent times. Sharifzadeh [1], Rua [2], Mejdell [3], Chen [4] and Dutta [5] have studied several possible operating scenarios for an integrated PCC - Power plant complex and evaluated multiple control strategies for the PCC unit in detail. One specific scenario, which is quite challenging, involves rapid ramp down of power plant production rate. In this scenario, it is observed that the heating steam flowrate to the CO₂ regenerator reboiler fluctuates significantly and poses a tough challenge to quick reestablishment of steady state conditions. In our study, we have focused on this specific challenge and evaluated different configurations, controller schemes and transient state operating strategies to identify the best option to address this scenario. We have also considered an unconventional operating approach in the study, which involves rapid reduction of solvent flowrate to the CO₂ absorber; to match the reduction in incoming flue gas flow rate, while ensuring consistent CO₂ slip within acceptable operating range.

2. Process Description

A simplified process flow diagram of a typical MEA solvent based PCC unit is depicted in Figure 1. Flue gas from the power plant furnace is fed to the absorber, where CO₂ is absorbed by the circulating MEA solution. The rich MEA solvent is then regenerated in the CO₂ stripper, while the CO₂ is taken for further processing depending upon the nature of requirement. The regenerator tower is typically accompanied with a kettle type reboiler and uses process steam as heating utility. The reboiler tower is typically accompanied with a kettle type reboiler and uses process steam as heating utility.

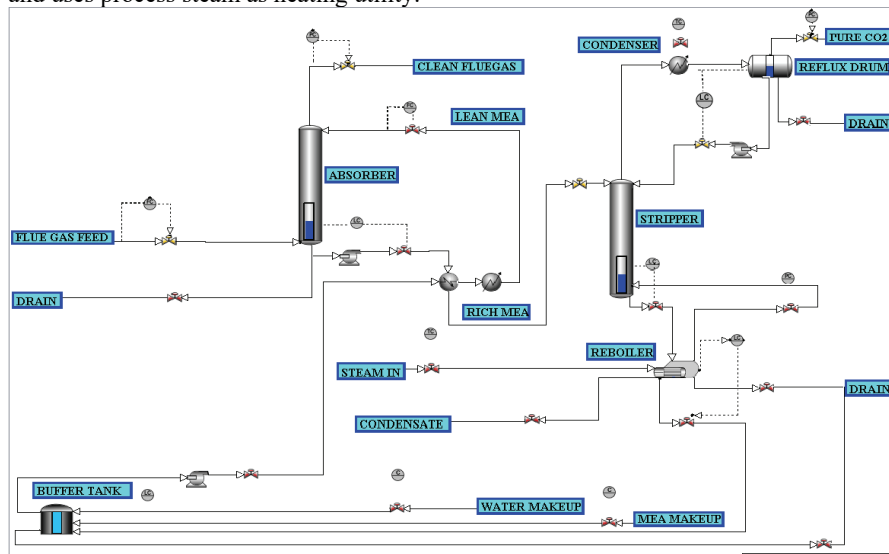


Figure 1 - Process flow diagram of a typical MEA solvent based PCC

In the process used for study, the solvent is 30 wt% aqueous Lean MEA solvent, which enters the absorber at 40 °C after being cooled by an air cooler. The solvent captures 85% of the CO₂ present in the feed. The rich solution exiting the absorber is heated by the stripper effluent stream and then fed to the stripping tower. The rich solution is regenerated in the stripping tower, and the kettle reboiler is maintained at 120 °C using steam as heating media. The CO₂ gas after condensation from stripper exit is also cooled in an air cooler, and the recovered condensate is recycled back to the stripper column as reflux. We have developed a rigorous dynamic simulation model of the above process, using a state of the art dynamic simulator, AVEVA™ Dynamic Simulation (formerly known as DYNASIM™), from AVEVA to carry out the operational study and validation of various control schemes / operating strategies.

3. Evaluation of Control Strategy

The study focusses on the operating scenario of 50% drop in power plant production, leading to a rapid drop of flue gas feed rate to the PCC unit. The details of the control strategy used for controlling the process are as per Table 1.

Table 1. Base case control strategy

Controlled Variable	Manipulated variable
Lean Amine Flow to Absorber	Inlet Valve Position
Reboiler Outlet Temperature	Steam Inlet Flow
Absorber Level	Absorber Outlet Valve Position
Stripper Level	Stripper Bottom Outlet Valve Position
Stripper Pressure	Stripper Top Outlet Valve Position
Absorber Pressure	Reflux Drum Outlet Valve Position
Reboiler Level	Reboiler Outlet Valve Position

In our study, the primary objective was to study the transient behaviour of key operating parameters, for the scenario where the feed gas flow ramps down to half of the normal value in about 30 mins and comes back to normal in about the same time. We have studied the effectivity of different control schemes in handling this process upset. The control schemes are evaluated by their capability to maintain expected operating condition in both the absorber and regenerator, i.e., avoid sharp change in gas / solvent flow rate, avoiding flooding or weeping in the towers, maintain sump levels and ultimately the CO₂ capture %.

The following three control schemes / strategies have been studied for their response to the above mention process scenario:

- a. Reboiler bottom outlet temperature control (kettle type reboiler)
- b. Stripper top outlet temperature control (kettle type reboiler)
- c. Reboiler bottom outlet temperature control (thermosyphon reboiler)

In all these scenarios, the flowrate of lean solvent to Absorber is controlled manually, i.e. the set point of the flow controller is reduced gradually by operator every 5 minutes, until the solvent flow corresponding to 50% absorber load is achieved.

4. Results and Discussion

The charts below depict all the key parameters recorded during the transient scenario, plotted against time (in minutes).

In **scenario (a)**, the reboiler outlet temperature is controlled to 120°C by varying the heating steam flow rate. Figures 2-9 show the response of the key process parameters.

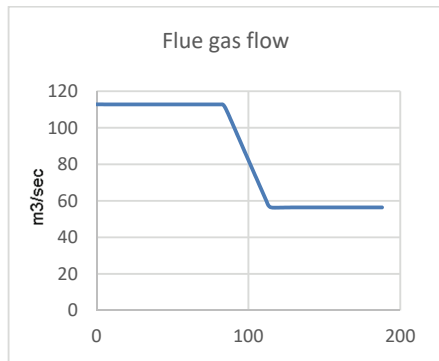


Figure 2. Flue gas flow ramp down

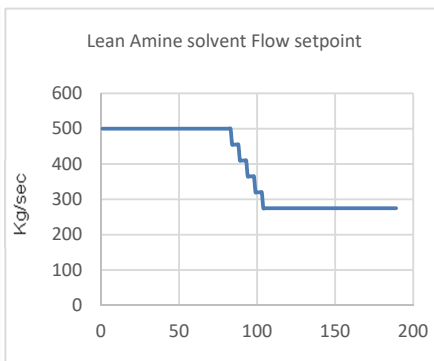


Figure 3. Solvent flow ramp down

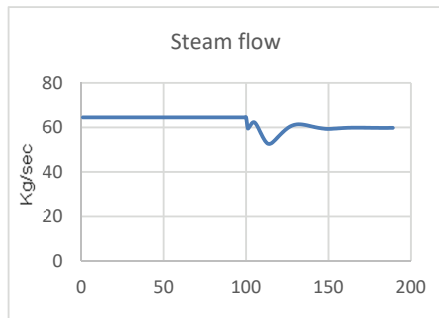


Figure 4 Steam Flow

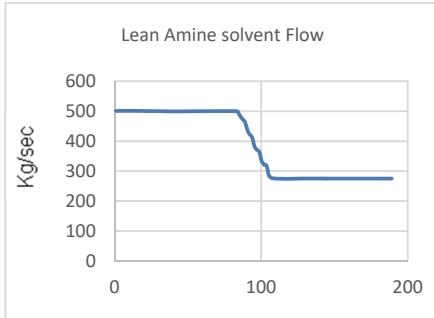


Figure 5. Lean Amine Solvent Flow

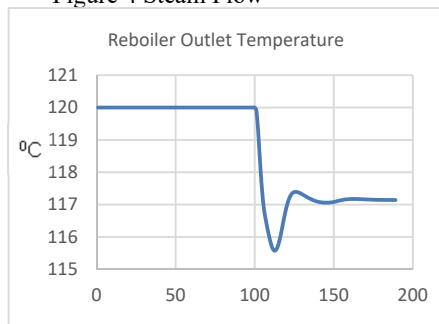


Figure 6. Reboiler Outlet Temperature

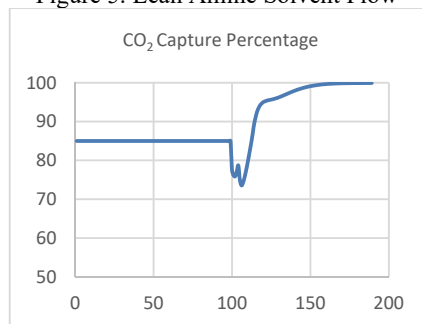


Figure 7. CO₂ Capture Percentage

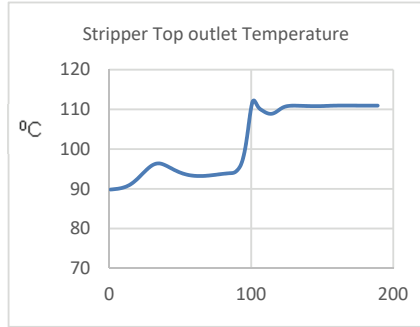


Figure 8. Stripper Top Outlet Temperature

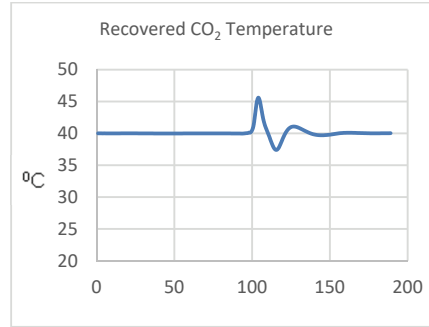


Figure 9. Recovered CO₂ Temperature

In **scenario (b)**, the stripper top outlet temperature is controlled, by varying the steam flow rate to the reboiler. The main reason for considering this scenario is, in scenario (a), it is observed that the Stripper top temperature (figure 8) increases significantly during the ramp down phase. Figures 10-13 give trends of key operating parameters during the transient. The trends indicate that the steam flow, reboiler outlet temperature and stripper top temperature do not stabilize any faster than scenario (a). Further, significant oscillations are introduced into the system by this control strategy.

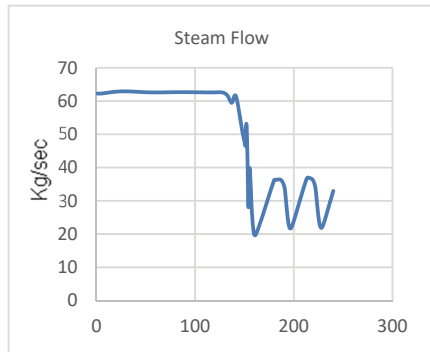


Figure 10. Steam flow to the Reboiler

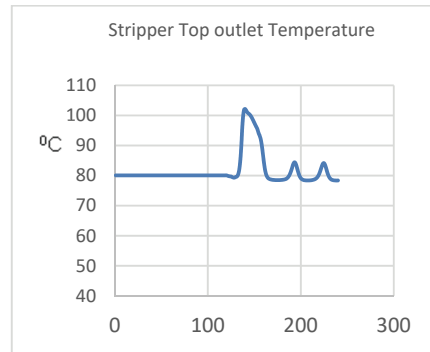


Figure 11. Stripper Top outlet Temperature

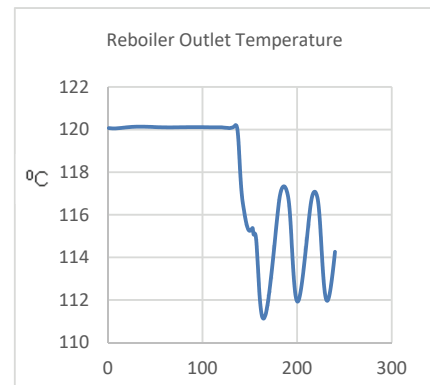


Figure 12. Reboiler Outlet Temperature

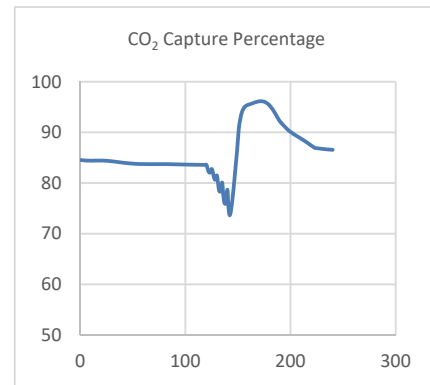


Figure 13. CO₂ Capture percentage

In **scenario (c)**, the kettle type reboiler is replaced with an equivalent thermosyphon reboiler, and the reboiler outlet temperature is controlled with steam flow as variable. The key parameters response for this configuration is depicted in figures 14 and 15. This configuration produced oscillations in the reboiler outlet temperature, steam flow and Percent CO₂ Capture.

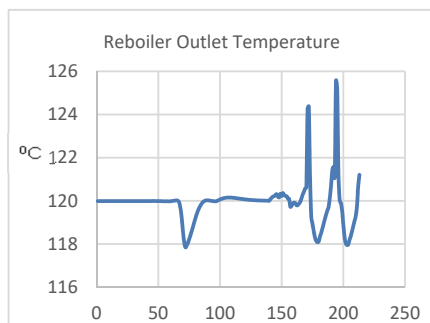


Figure 14. Reboiler outlet Temperature

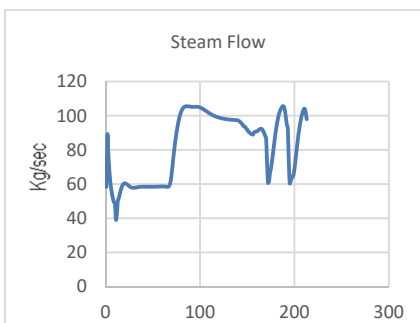


Figure 15. Steam flow rate

5. Conclusions

A state-of-the-art dynamic simulator was utilized to perform studies and identify an optimal control strategy to minimize the impact of variation in power demand on the effectiveness of the PCC unit. The results from the study show that the reboiler temperature control plays an important role in ensuring quick stabilization of the process and should be the deciding criteria for recommending the optimal control strategy. Among the three scenarios studied, scenario (a), where the kettle type reboiler temperature is controlled, turns out to be the most optimal strategy. In this strategy the ramp down scenario was most effectively handled, and the steady state operation was achieved in a shorter period without introducing any significant oscillations.

References

1. Mahdi Sharifzadeh, Nilay Shah, MEA-based CO₂ capture integrated with natural gas combined cycle or pulverized coal power plants: Operability and controllability through integrated design and control, *Journal of Cleaner Production*, Volume 207, 2019, Pages 271-283, ISSN 0959-6526, <https://doi.org/10.1016/j.jclepro.2018.09.115>
2. Jairo Rúa, Mai Bui, Lars O. Nord, Niall Mac Dowell, Does CCS reduce power generation flexibility? A dynamic study of combined cycles with post-combustion CO₂ capture, *International Journal of Greenhouse Gas Control*, Volume 95, 2020, 102984, ISSN 1750-5836, <https://doi.org/10.1016/j.ijggc.2020.102984>
3. Thor Mejdell, Geir Haugen, Alexander Rieder, Hanne M. Kvamsdal, Dynamic and Control of an Absorber - Desorber Plant at Heilbronn, *Energy Procedia*, Volume 114, 2017, Pages 1231-1244, ISSN 1876-6102, <https://doi.org/10.1016/j.egypro.2017.03.1378>
4. Chen, Y.-H.; Shen, M.-T.; Chang, H.; Ho, C.-D. Control of Solvent-Based Post-Combustion Carbon Capture Process with Optimal Operation Conditions. *Processes* 2019, 7, 366. <https://www.mdpi.com/2227-9717/7/6/366#cite>
5. Rohan Dutta, Lars O. Nord, Olav Bolland, Selection and design of post-combustion CO₂ capture process for 600 MW natural gas fueled thermal power plant based on operability, *Energy*, volume 121, 2017, Pages 643-656, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2017.01.053>

Modelling and Temperature Uncertainty Analysis of a Fermentation Bioreactor at an Industrial Brewery

Wilfredo Angulo^a, Dany De Cecchis^a, Santiago D. Salas^a, Yris Gozález^a and Carolina Massay^a

^aEscuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Naturales y Matemáticas, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador

Abstract

A mathematical model is adjusted to mimic the operational data of an industrial brewery during the fermentation process. The model is composed of a set of nonlinear ordinary differential equations. The parameters of the model were estimated using the least-squares method and the Levenberg-Marquardt algorithm. Once the model is adjusted, stochastic simulations run for different temperature profiles which are obtained using a Latin-hypercube sampling strategy. The results show the impact of the temperature uncertainty on the fermentation sub-products responsible of the final products' off-flavors.

Keywords: Mathematical modelling, Temperature uncertainty, Brewery, Off-flavor.

1. Introduction

Beer is one of the most widely consumed alcoholic beverages worldwide. Its market is estimated to be around 500 billion USD. Its competitiveness motivates breweries to encompass their production processes towards enhancing the operation to improve profitability and the ultimate success of the brewery. A key process in the industry corresponds to the fermentation stage, which not only demands high processing-times but it also guarantees key organoleptic characteristics of the beer. The industrial-scale fermentation process takes place inside a type-batch bioreactor, and the biochemical transformation of the wort or substrate into ethanol occurs through the action of yeast.

The dynamic understanding of the beer fermentation process has received considerable interest in recent years. Main reasons include the requirement of brewers to improve the process efficiency under different market conditions while complying with quality constraints. A number of mathematical models have been proposed for simulating the fermentation stage in breweries (Gee et al., 1994; Andres-Toro et al., 1998). In this context, a phenomenological-based methodology addressed the time evolution of the concentration of substrate, biomass, ethanol and off-flavors in a batch-type bioreactor (Gomez et al., 2008; Alvarez et al., 2009).

Despite these developments, the current operation in fermentation reactors can frequently be far from its desired conditions. In this sense, the incorporation and study of the uncertainty appears to be a tool of remarkable importance for model based frameworks (Salas et al., 2017, Gonzalez et al., 2020).

In this work, a mathematical model is fitted to the operational data of an industrial brewery. The model consists of a nonlinear system of ordinary differential equations, which solutions were approximated for certain conditions. The parameters of the model

were estimated using the least-squares method for nonlinear functions along with the Levenberg-Marquardt algorithm.

Once the model is adjusted, a total of 5000 stochastic simulations run for different temperature profiles, resembling common process disturbances. Finally, the results are statistically evaluated showing the effects of the temperature in the outputs of the system.

2. Mathematical Modelling

2.1. Background information

The proposed dynamic model is based in a material balance carried out on a batch-type fermenter. The model is reduced to a kinetic-type such as that proposed by (Andres-Toro et al., 1998) The state variables correspond to the concentrations of latent yeast X_{lag} ; active yeast X_{lact} ; dead yeast X_{dead} ; suspended yeast X_{sus} ; substrate C_S ; ethanol C_E ; diacetyl C_{DY} and ethyl acetate C_{EA} . Also, it is necessary a temperature profile given by a continuous function $T(t)$ with respect to time $t \geq 0$. These variables describe the two essential phases in the brewing process (see details in (Rodman et al., 2016)). During the first (lag) phase, the biomass mostly consists of latent yeast cells and, while they are activated, minimal fermentation takes place. The time interval which corresponds to this phase is $\Omega_1 = [0, t_{lag}]$, with t_{lag} the time around the lag phase ends. Once approximately half of the suspended cells are activated, the second (fermentation) phase begins. The active cell concentration is enough for inducing enzymatic effects, converting the sugar substrate into ethanol product. This phase time interval is denoted by $\Omega_2 = [t_{lag}, +\infty)$. The model considers the complete interval $\Omega = \Omega_1 \cup \Omega_2$. For defining the kinetics terms regarding the dynamics of active, dead and suspended yeast in each phase, we define the characteristic functions $\chi_i(t)$, with value of 1, if $t \in \Omega_i$, and 0 otherwise, for $i = 1, 2$. Therefore, the matrix function depending on the variables $C_S(t)$, $C_E(t)$ and $T(t)$ is defined as

$$A = \begin{pmatrix} -\mu_L & 0 & 0 & 0 \\ \mu_L(\chi_1 + \chi_2) & (\mu_X - \mu_{DT})\chi_2 & 0 & 0 \\ 0 & \mu_{DT}\chi_2 & -\mu_{SD}(\chi_1 + \chi_2) & 0 \\ 0 & \mu_X\chi_2 & \mu_{SD}(\chi_1 - \chi_2) & 0 \end{pmatrix}, \quad (1)$$

where $\mu_j = \mu_j(C_S(t), C_E(t); \theta_j)$ are given functions based on Michaelis-Menten-Arrhenius kinetics (Trelea et al., 2001) for $j \in \{E, S, X, SD\}$, and $\mu_k = \mu_k(T(t); \theta_k)$ are Arrhenius functions of T that represent specific rate for $k \in \{L, AB, DT, DY\}$. The parameters for each μ_j and μ_k are denoted as the vectors $\theta_j \in \mathbb{R}^{n_j}$ and $\theta_k \in \mathbb{R}^{n_k}$, respectively, with n_j and n_k representing their correspondent vector size. These parameters are estimated using data provided by an industrial brewery.

The model consists on determining, for all $t \in \Omega \setminus \{0\}$, the vector function, $\mathbf{X}(t) = [X_{lag}(t), X_{act}(t), X_{dead}(t), X_{sus}(t)]^T$, and the functions $C_S(t)$, $C_E(t)$, $C_{EA}(t)$ and $C_{DY}(t)$, such that,

$$\dot{\mathbf{X}}(t) = A\mathbf{X}(t), \quad (2)$$

$$\dot{C}_S(t) = -\mu_S X_{act}(t), \quad (3)$$

$$\dot{C}_E(t) = f\mu_E X_{act}(t), \quad (4)$$

$$\dot{C}_{EA}(t) = Y_{EA}\mu_X X_{act}(t), \quad (5)$$

$$\dot{C}_{DY}(t) = \mu_{DY}C_S(t)X_{act}(t) - \mu_{AB}C_{DY}(t)C_{EA}(t), \quad (6)$$

subject to initial conditions in $t = 0$. In Eq. (4), f is a function that depends on C_E (inhibition factor). In Eq. (5) the stoichiometric factor Y_{EA} is also an Arrhenius function of system temperature $T(t)$, both functions are taken from (Andres-Toro et al., 1998).

2.2. Parameter estimation

The vectors of parameters θ_j and θ_k , for $j \in \{E, S, X, SD\}$ and $k \in \{L, AB, DT, DY\}$, respectively, are estimated by minimizing in each $t_\ell \in \Omega$ the relative squared error. The operational data of the concentrations of substrate $C_{S_{exp}}(t_\ell)$ and ethanol $C_{E_{exp}}(t_\ell)$ are compared with respect to the concentrations obtained from the model depending on θ_j and θ_k , denoted by $C_S(t_\ell; \theta_j; \theta_k)$ and $C_E(t_\ell; \theta_j; \theta_k)$. Here, t_ℓ is the sampling time for $\ell = 1, \dots, N$, where $N \in \mathbb{N}^*$ is the number of fitted data points. Thereafter, the Levenberg-Marquardt algorithm estimates the parameters by solving a dynamic optimization problem (Bellavia et al., 2018) subject to the system of equations (2)-(6) and constraints. The feasible domain of variables θ_j^p , with $p = 1, \dots, n_j$, and θ_k^q , with $q = 1, \dots, n_k$, are defined by their lower and upper bounds l_j^p, l_k^q and u_j^p, u_k^q , respectively; this means $l_j^p \leq \theta_j^p \leq u_j^p$ and $l_k^q \leq \theta_k^q \leq u_k^q$.

The initial conditions of the state variables are set based on the operational data. The vectors of parameters θ_j and θ_k are constrained to guarantee the physical realizability of the parameters. The initial guests and the upper and lower bounds for the variables θ_j^p and θ_k^q are obtained from (Andres-Toro et al., 1998) and (Rodman et al., 2016).

To adjust the parameters, a set of 10 batches of a specific beer recipe were considered. A total of 19 parameters related to the kinetics of the fermentation process obtained residual norms of the difference of the experimental and simulated data between 0.002 and 0.057. The following study used the parameter's mean values for all the batches.

3. Stochastic dynamic modelling under uncertainty

To study the variability of the system dynamics when introducing temperature uncertainty, a series of stochastic simulations are carried out. In our practice a Python environment is utilized, in which the libraries NumPy, SciPy, pandas, pyDOE, matplotlib and seaborn facilitate the obtainment of meaningful results.

Initially, a total of 5000 temperature profiles are generated using the Latin-hypercube sampling method available in pyDOE. A standard deviation (σ_x) of 0.4 is considered in this case. The temperature profile changes at different pre-established time intervals. The objective of this experimental design is to represent process disturbances that affect the products of a brewery at the fermentation stage. The ODEs are solved using `solve_ivp`, available in the module `SciPy.Integrate`. The resultant state variable vectors, for each temperature profile, are stored in a data frame for further processing. This study focuses on the dynamic visualization of the process when temperature uncertainty is introduced. Thus, the library `seaborn` allows to observe its effects.

4. Results

The variation of the temperature from its recipe profile is captured in Figure 1. Notice that the bold line corresponds to the average, and the bold area the region between 25th and 75th percentiles. The temperature LHC varies as piece-wise function at the times (hour) $t = 0, 60, 72, 84, 96, 108, 120$, and 140. The blue dots correspond to the experimental values. Figure 2 illustrates the influence of the temperature uncertainty to the different yeast growth as well as the substrate (sugars) and ethanol production. It is important to notice that the active, dead, and suspended yeast show important relative variability from the input. On the other hand, active and suspended yeast only show important variability at the end of the fermentation process. The substrate plot shows the relationship between the variability and the experimental values. In this case, the model is not able to capture the variations at the beginning and at the end of the process. Although ethanol shows certain variability in the middle of the process; and after time $t = 100$, it tends to narrow the variability, showing that the result of producing ethanol is stable regarding the temperature variation.

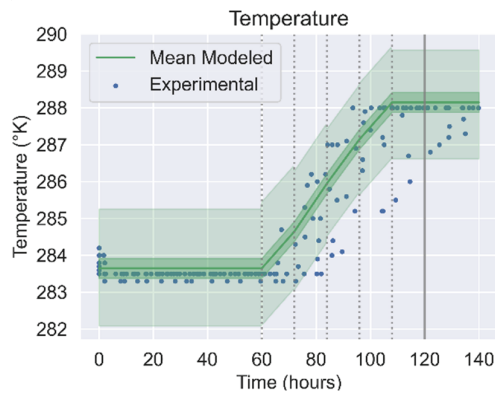


Figure 1: Temperature including uncertainty obtained using LHS. The vertical dotted lines correspond to the times (hour) $t = 60, 72, 84, 96, 108$. The solid line corresponds to theoretical final fermentation time at $t = 120$ hours. The blue dots are the experimental values of temperature.

Regarding the off-flavors, Figure 3 illustrates the influence of the temperature uncertainty to their dynamics, which are represented by the concentrations of ethyl-acetate and the diacetyl in parts per million (ppm). The upper panel shows the concentrations of both ethyl-acetate and diacetyl. The plot inset, included at the upper east side, allows to make the comparison of the concentration of both off-flavors after the time $t = 118$ hours. The lower panel shows the σ_x of the variability of both off-flavors, providing a dynamic quantitative evaluation of the actual uncertainty introduced to the outputs. Even though they are in the magnitude of ppm, unpleasant flavors can be perceived in the taste of beer, and after certain levels they could even become toxic. The diacetyl shows an important sensitivity to the temperature uncertainty at the middle of the process. However, after fermentation time $t = 100$ hours, this uncertainty reduces significantly. Indeed, a common practice in breweries is to measure off-flavors' concentration at certain time intervals to quantify whether the fermentation should last longer or not. These results show that a strict control of the temperature seems to be critical in this stage in order to reach diacetyl concentration below 0.15 ppm (Olaniran et al., 2017).

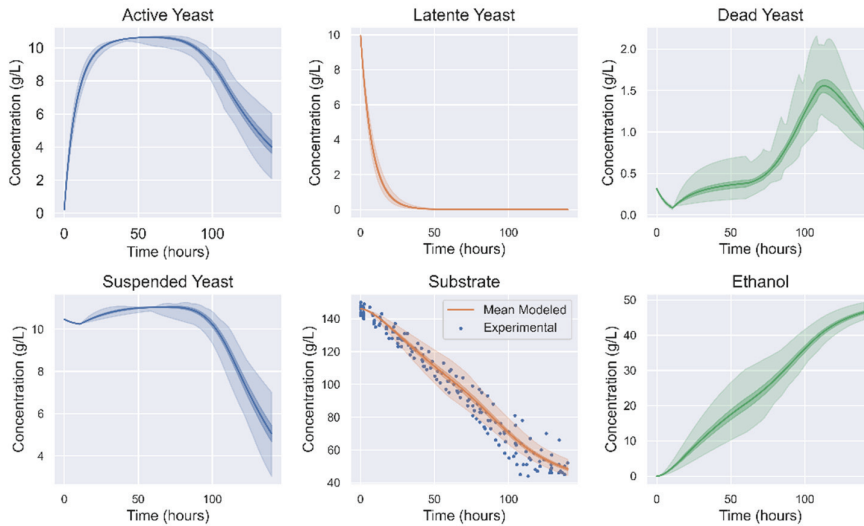


Figure 2: Concentration of active, latent, dead and suspended yeast, substrate and ethanol. The painted region corresponds to minimum and maximum values, the darker zone corresponds to the values between 25th and 75th percentiles. The blue dots in center-lower panel are the experimental values of substrate.

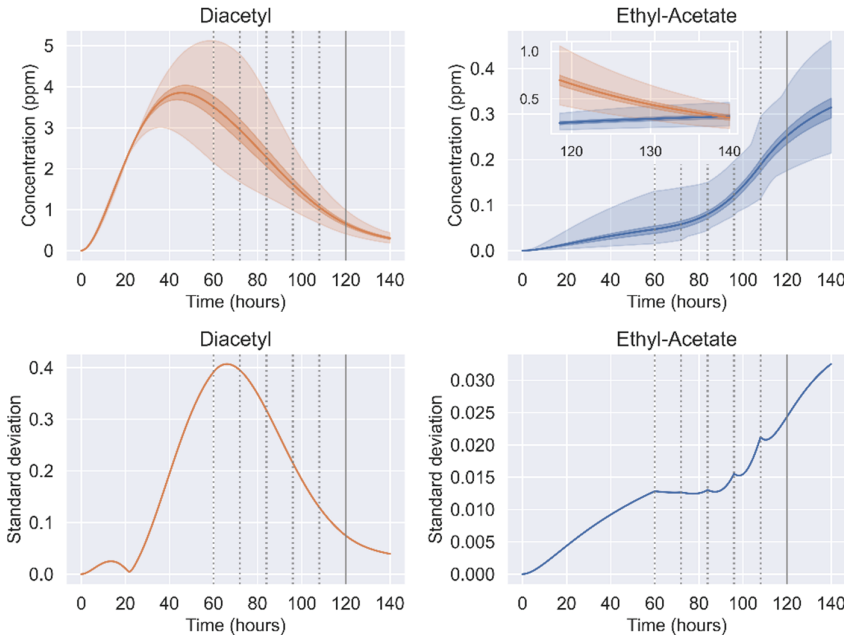


Figure 3: Ethyl-acetate and diacetyl concentration (upper panel). The painted region correspond to minimum and maximum values. The darker zone correspond to the values between 25th and 75th percentiles. Lower panel shows their respective σ_x . The vertical dotted lines correspond to the times (hour) $t = 60, 72, 84, 96, 108$. The solid line corresponds to theoretical final fermentation time at 120 hours.

5. Conclusions

In this work, the validated mathematical model of the fermentation process in an industrial brewery is studied by introducing uncertainty to the temperature. The model is adjusted using real experimental data from an industrial facility, the least-squares method, and the Levenberg-Marquardt algorithm. Stochastic simulations run for different temperature profiles which are generated using a Latin-hypercube sampling strategy setting a value of the standard deviation. The results show the impact of the temperature uncertainty to fermentation products. Among these outputs certain sub-products, responsible of the final products off-flavors, are analyzed. It is observed from the results that the off-flavors could be corrected by suitably extending the reaction and managing a well adjusted trade-off between them. Nevertheless, an extended operation could impact the economy of a brewery because it introduces an inefficient use of energy and other resources such as the tanks due to time delay in production. In this sense, these types of tools could benefit brewers for decision making as well as for adequate process scheduling.

References

- H. Alvarez, R. Lamanna, P. Vega, S. Revollar, 2009. Metodología para la obtención de modelos semifísicos de base fenomenológica aplicada a una sulfitoradora de jugo de caña de azúcar. *Revista Iberoamericana de Automática e Informática Industrial RIAI* 6 (3), 10 – 20. S.
- S. Bellavia, S. Gratton, E. Riccietti, 2018. A levenberg–marquardt method for large nonlinear least-squares problems with dynamic accuracy in functions and gradients. *Numerische Mathematik* 140 (3), 791–825.
- B. De Andrés-Toro, J. Girón-Sierra, J. López-Orozco, C. Fernández-Conde, J. Peinado, F. García-Ochoa, 1998. A kinetic model for beer production under industrial operational conditions. *Mathematics and Computers in Simulation* 48 (1), 65 – 74.
- D. A. Gee, W. F. Ramirez, 1994. A flavour model for beer fermentation. *Journal of the Institute of Brewing* 100 (5), 321–329. C. A. Gomez, Y. A. Calderon, H. Álvarez, 2008. Construcción de modelos semifísicos de base fenomenológica. caso proceso de fermentación. *Biotecnología en el Sector Agropecuario y Agroindustrial* 6 (2), 28–39.
- Y. Gonzalez, W. Angulo, D. De Cecchis, M. Lucena, S. D. Salas, 2020. Optimization of a cyclone reactor for biomass hydrolysis through global sensitivity analysis and stochastic optimization. In: *Computer Aided Chemical Engineering*. Vol. 48. Elsevier, pp. 667–672.
- A. O. Olaniran, L. Hiralal, M. P. Mokoena, B. Pillay, 2017. Flavour-active volatile compounds in beer: production, regulation and control. *Journal of the Institute of Brewing* 123 (1), 13–23. A.
- D. Rodman, D. I. Gerogiorgis, 2016. Multi-objective process optimisation of beer fermentation via dynamic simulation. *Food and Bioproducts Processing* 100, 255 – 274.
- S. Salas, A. Geraili, J. Romagnoli, 2017. Optimization of renewable energy businesses under operational level uncertainties through extensive sensitivity analysis and stochastic global optimization. *Industrial & Engineering Chemistry Research* 56 (12), 3360–3372.
- I. C. Trelea, M. Titica, S. Landaud, E. Latrille, G. Corrieu, A. Cheruy, 2001. Predictive modelling of brewing fermentation: from knowledge-based to black-box models. *Mathematics and Computers in Simulation* 56 (4), 405 – 424.

Symbolic transfer entropy for root cause analysis of process disturbances

Ching-Mei Wen, Yuan Yao *

*Department of Chemical Engineering, National Tsing Hua University, Hsinchu, Taiwan
300044, ROC*

yyao@mx.nthu.edu.tw

Abstract

Root cause analysis is an important step in process monitoring for the disclosure of causes of process disturbances. In the past research and practice, transfer entropy has been widely adopted for its capability to handle process nonlinearity. However, the conventional transfer entropy is not robust to noise and computational inefficient, which affects its application performance. These problems can be solved by incorporating the technique of symbolization. To the best of our knowledge, symbolic transfer entropy (STE) has seldom been implemented in the field of process monitoring. In this work, the concepts of statistical process control (SPC) and STE are integrated to achieve satisfactory diagnosis results. As known, SPC charts, e.g. Shewhart charts, exponential weighted moving average charts, and cumulative sum charts, are useful tools for process monitoring, where the sample locations on the control chart indicate certain process status. Therefore, it is reasonable to symbolize process measurements according to such information. After symbolization, STE is conducted to reveal the causality among process variables which is then visualized with a signed directed graph. The case study on the Tennessee Eastman process shows that the control chart-based STE possesses a capability to understand the causes of disturbances and offers the convenience of root cause analysis.

Keywords: root cause analysis, diagnosis, symbolic transfer entropy, causality analysis, process monitoring.

1. Introduction

Disclosure of the fault root causes is a critical step in process monitoring, which is important for controlling the risks associated with process failures and preventing problem recurrence. Based on the concept of statistical process control (SPC) (Woodall and Montgomery, 1999), a large number of process monitoring methods have been developed in recent decades (Ge et al., 2013). Most of these research works focus on timely detection of process faults or disturbances, as well as accurately isolation of critical process variables, while root cause diagnosis is relatively overlooked. Conventionally, process knowledge is a necessity to achieve reasonable diagnosis results (Iri et al., 1979). In recent years, statistical causality analysis techniques have been adopted to this field so as to relieve the dependency on the prior knowledge of the investigated process (Li et al., 2016), among which transfer entropy (TE) (Schreiber, 2000) is a representative method because of its capability to handle nonlinearity. A

number of improved versions of TE have been proposed in the past years. For example, permutation entropy (Bandt and Pompe, 2002) has been proved to provide robustness to noise, which estimates transfer entropy using symbolization. The concept of symbolic transfer entropy (STE) was then extended (Staniek and Lehnertz, 2008) by using a surrogate-based testing scheme. Applications of STE demonstrated its advantages of better computational efficiency and accuracy comparing to the conventional TE.

In this work, a new scheme is proposed for root cause analysis of process disturbances, which integrates the concepts of STE and SPC. The application results illustrate the feasibility of this method.

2. Methodology

2.1. Transfer entropy

As an information theoretic measure, TE was first introduced by Schreiber (2000), the basics of which are as following. Supposing X and Y represent l^{th} and k^{th} order Markov processes, respectively, the transfer entropy from X to Y is defined as:

$$T_{X \rightarrow Y} = \sum_{y_{i+1}, y_i^{(k)}, x_i^{(l)}} p(y_{i+1}, y_i^{(k)}, x_i^{(l)}) \log_2 \frac{p(y_{i+1} | y_i^{(k)}, x_i^{(l)})}{p(y_{i+1} | y_i^{(k)})} \quad (1)$$

where $x_i^{(l)}$ and $y_i^{(k)}$ are the past l and k -step states of process, i.e. $x_i^{(l)} = x_i, \Lambda, x_{i-l+1}$ and $y_i^{(k)} = y_i, \Lambda, y_{i-k+1}$. The transfer entropy from Y to X can be defined in a similar way. As an asymmetric measure, TE allows the quantification of the directional coupling between two processes. Ideally, a TE value greater than zero and indicates the existence of causality. In Eq. (1), the complete or conditional probability density functions can be estimated by kernel methods (Silverman, 1986).

2.2. Symbolic Transfer entropy

In 2008, Staniek and Lehnertz (2008) proposed an STE as a method of calculating information transfer between time series processes with discrete states. The STE for two discrete-time random processes \tilde{X} and \tilde{Y} can be expressed as Eq. (2):

$$T_{\tilde{X} \rightarrow \tilde{Y}} = \sum_{\tilde{y}_{i+1}, \tilde{y}_i^{(k)}, \tilde{x}_i^{(l)}} p(\tilde{y}_{i+1}, \tilde{y}_i^{(k)}, \tilde{x}_i^{(l)}) \log_2 \frac{p(\tilde{y}_{i+1} | \tilde{y}_i^{(k)}, \tilde{x}_i^{(l)})}{p(\tilde{y}_{i+1} | \tilde{y}_i^{(k)})} \quad (2)$$

Here, the symbols have similar definitions to those in Eq. (1), except that the data used for estimate STE are discretized. The discretization can be achieved by partitioning the data into a finite number of bins. The size of each bin can be determined based on the user-specified upper and lower bounds or by choosing specific quantiles of the empirical distribution of the data. In detail, if the bounds of n bins are denoted as q_1, q_2, \dots, q_n , where $q_1 < q_2 < \dots < q_n$, the symbolized data can be recorded in the following

way. For $x_i \leq q_1$, $\tilde{x}_i = 1$; for $q_1 < x_i \leq q_2$, $\tilde{x}_i = 2$; and so on and so forth. In practice, the selection of bin size requires balancing information conservation and computational burden. In general, the smaller the selected bin size, the more information retained. However, the reduction of noise and computational burden becomes insignificant in such cases.

In this work, we propose to divide the bins by adopting the concept of SPC. In detail, a modified exponential weighted moving average (EWMA) chart (Hunter, 1986) is utilized. The calculation of the EWMA statistic for monitoring the time series X is as Eq. (3)

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1} \quad (3)$$

Conventionally, the control limits of an EWMA chart is calculated as:

$$\text{UCL, LCL} = z_0 \pm ks_{ewma} \quad (4)$$

where

$$s_{ewma}^2 = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}] s^2 \quad (5)$$

z_0 and s are the process target, i.e. the mean of the historical data, and the estimated standard deviation, respectively, and k is user-specified parameter controlling the type I error rate. Therefore, totally seven bins can be set up according to the values of k . For example, for $z_i \leq z_0 - 3s_{ewma}$, $\tilde{x}_i = 1$; for $z_0 - 3s_{ewma} < z_i \leq z_0 - 2s_{ewma}$, $\tilde{x}_i = 2$; ...; for $z_0 - s_{ewma} < z_i \leq z_0 + s_{ewma}$, $\tilde{x}_i = 0$; ...; and for $z_i \geq z_0 + 3s_{ewma}$, $\tilde{x}_i = 7$. Note that the control limits approach to constant values (

$z_0 \pm k\sqrt{\frac{\lambda}{2 - \lambda}}s$) with the increased number of samples. Therefore, when the process status is out of statistical control, i.e. when it is necessary to conduct root cause diagnosis, there is a large chance that the symbolized process measurements have an unchanged value, e.g. 1 or 7, and cannot reflect the process dynamics. To solve this problem, in the proposed scheme, the estimate of standard deviation s is updated online using a moving window technique. In doing this, the EWMA control chart may have a lower alarm rate. However, note that the purpose of using the control chart here is not to detect whether the process is in statistical control. Instead, the symbolization needs to reflect the process dynamic characteristics and the information transfer between variables.

2.3. Estimation of Significance Level

No matter the conventional TE or STE is used, the estimates are seldom exactly zero even if there is no information transfer between two time series. Practically, subjective selected thresholds are usually needed for judge whether the statistical causality exists. Herein, in order to make objective judgements, it is desired to use a more statistical method. Therefore, the p-values are calculated based on bootstrap samples of the transfer entropy estimates (Behrendt et al., 2019). After conducting STE, the results are used to construct a signed directed graph (SDG), in which each vertex corresponds to a process variable, and the arrows on the edges denote the direction of information transfer, i.e. the inferred causality. In addition, a weight is assigned to each edge, whose value equals to the p-value calculated in the previous step. Only the edges with p-values smaller than 0.05 are connected.

3. Case study

3.1. Tennessee Eastman process and variable selection

Herein, the benchmark Tennessee Eastman process, developed by Downs and Vogel (1993), is used as an illustrative case. The process consists of five main units, namely reactor, product condenser, gas-liquid separation tower, product stripper, and recycle compressor. The gaseous reactants include A, C, D, and E, while the products are denoted as G and H. In addition, there is an inert component B and a by-product F. The disturbance IDV1 is used as a case study to illustrate the feasibility of the proposed method, where the composition of A in the feed stream 4 changes from 48.5 mol% to 45.5 mol% and at the same time the composition of C in the same stream changes from 51 mol% to 54 mol%. According to the previous research (Liu et al, 2014), nine process variables as listed in Table 1 were chosen as the candidates for root cause diagnosis.

Table 1. variables selection of IDV1 case study

V1	A feed (stream 1)	V20	Compressor work
V4	A and C feed (stream 4)	V31	Component C (stream 9)
V8	Reactor level	V38	Component E (stream 11)
V16	Stripper pressure	V50	Stripper steam valve
V18	Stripper temperature		

3.2. Parameter setting and symbolization

A few parameters were set before conducting symbolization. First, the average values used in the control charts were calculated based on the normal operating data. Second, the size of the moving window used for standard deviation estimation was chosen to be 5, 10, and 20, respectively, for comparison. In addition, the values of k and l used in STE were chosen to be 1 according to the literature (Yu and Yang, 2015).

Two symbolization methods were adopted for the STE calculation. In the first method, the Shewhart individual control charts are utilized, while the second method utilizes the EWMA control charts combined with the moving window strategy proposed in this paper.

3.3. Results and discussions

Three indices were used to evaluate the performance of different root cause diagnosis methods, as shown in Table 2. Sensitivity measures the correctness in identified causal correlations, specificity calculates the correctness of the identification of the non-causal relations, and edge density indicates the sparsity of the resulting SDGs. The first two indices have values between 0 to 1, where larger values are preferred. For edge density, a smaller value corresponds to a sparser graph.

Table 2. evaluation indexes of STE results

	Window size	Sensitivity	Specificity	Edge density
Shewhart	20	0.724	0.837	0.26
EWMA	5	0.86	0.58	0.43
EWMA	10	0.89	0.76	0.38
EWMA	20	0.931	0.860	0.37

As shown in Table 2, the EWMA chart-based STE methods outperformed the Shewhart chart-based method in general no matter what size of the moving window was selected. In addition, the modified EWMA chart with a moving window sized 20 performed best in this case study, because it led to the largest sensitivity and specificity values. It is noted that the edge density of the results of this model is not the lowest. However, this index is not critical to the accuracy of root cause diagnosis. In addition, the generally good performance of all three EWMA-based STE methods show that the proposed scheme is not very sensitive to the parameter setting. The SDG achieved by the EWMA chart-based STE method with the moving window sized 20 is plotted in Figure 1.

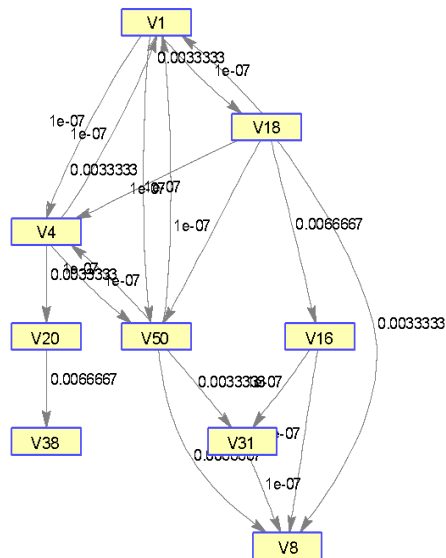


Figure 1. SDG resulted from EWMA-based STE

4. Conclusion

Symbolic transfer entropy has been used for the estimation of functional connectivity in networks. However, its applications to root cause diagnosis of process disturbances are rarely reported. In this work, STE based on a modified EWMA control chart is proposed, whose feasibility is illustrated using the benchmark Tennessee Eastman process. In the future work, the SDG obtained by using STE will be simplified to further highlight the root cause variables. As shown in Figure 1, the reciprocal causations between several variables may lead to the difficulty in identify the variable closest to the root of the disturbance. This may be partially caused by the improper selection of the time slot to be analysed, because the fault propagation may distort the information of the disturbance root cause. Therefore, how to determine the data segment for analysis is also an important topic for future research.

References

- C. Bandt, and B. Pompe, 2002, Permutation entropy: a natural complexity measure for time series, *Physical review letters*, 88, 17, 174102.
- S. Behrendt, T. Dimpfl, F. Peter, D. Zimmermann, 2019, RTransferEntropy — Quantifying information flow between different time series using effective transfer entropy, *SoftwareX*, 10, 100265.
- J. Downs, and E. Vogel, 1993. A plant-wide industrial process control problem, *Computers & Chemical Engineering*, 17, 3, 245-255.
- Z. Ge, Z. Song, and F. Gao, 2013, Review of recent research on data-based process monitoring, *Industrial & Engineering Chemistry Research*, 52, 10, 3543-3562.
- J. Hunter, 1986, The exponentially weighted moving average, *Journal of quality technology*, 18, 4, 203-210.
- M. Iri, K. Aoki, E. O'Shima, and H. Matsuyama, 1979, An algorithm for diagnosis of system failures in the chemical process, *Computers & Chemical Engineering*, 3, 1-4, 489-493.
- G. Li, S. J. Qin, and T. Yuan, 2016, Data-driven root cause diagnosis of faults in process industries, *Chemometrics and Intelligent Laboratory Systems*, 159, 1-11.
- J. Liu, D. Wong, and D.-S. Chen, (2014). Bayesian filtering of the smearing effect: Fault isolation in chemical process monitoring, *Journal of Process Control*, 24, 3, 1-21.
- T. Schreiber, 2000, Measuring information transfer, *Physical review letters*, 85, 2, 461-464.
- B. Silverman, 1996, Density estimation for statistics and data analysis, *Monographs on Statistics and Applied Probability*, London: Chapman and Hall.
- M. Staniek, and K. Lehnertz, 2008, Symbolic transfer entropy, *Physical review letters*, 100, 158101.
- W. Woodall, and D. Montgomery, 1999, Research issues and ideas in statistical process control, *Journal of Quality Technology*, 31, 4, 76-386.
- W. Yu, and F. Yang, 2015, Detection of causality between process variables based on industrial alarm data using transfer entropy, *Entropy*, 17, 8, 5868-5887.

Quality-by-Control of continuous drug substance isolation: study on a novel unit for integrated filtration-drying

Francesco Destro,^a Vivian Wang^b, Mesfin Abdi^b, Xin Feng^b, Erin Wood^b, Simon Coleman,^c Paul Firth,^c Alastair Barton,^c Massimiliano Barolo,^a Zoltan K. Nagy^{d,*}

^a*CAPE-Lab, University of Padova, 35131 Padova, PD, Italy*

^b*Office of Pharmaceutical Quality, Center for Drug Evaluation and Research, Food & Drug Administration, Silver Spring, MD, USA*

^c*Alconbury Weston Ltd, Stoke-on-Trent, UK*

^d*Purdue University, West Lafayette, IN 47906, USA*

zknagy@purdue.edu

Abstract

In the last decade, pharmaceutical manufacturing has been undergoing a modernization trend, promoted by the US Food and Drug Administration (FDA) with the Quality-by-Design (QbD) initiative. However, in most of the cases, the product quality is still controlled in open-loop, and pharmaceutical operation is carried out in the traditional batch mode, rather than continuously. In this work, we present a framework for closed-loop quality control of a novel continuous integrated filtration-drying unit. The unit addresses the current lack of technology for continuous drug substance isolation, which is the main bottleneck in end-to-end continuous pharmaceutical manufacturing. With the proposed closed-loop control framework, the product conformity is improved, showing how a Quality-by-Control (QbC) strategy can outperform earlier QbD approaches based on open-loop control.

Keywords: Quality-by-Design, Quality-by-Control, Process control, Continuous pharmaceutical manufacturing, Continuous filtration, Fluid-solid separation

1. Introduction

The QbD initiative of the US FDA (FDA, 2004) has generated a modernization momentum in the pharmaceutical industry within the last decade. In the QbD framework, the manufacturing process is developed with a risk-based approach relying on process knowledge, to inherently build the quality into the product. The critical process parameters (CPPs), the feed properties (critical material attributes, CMAs) and the product critical quality attributes (CQAs) are first identified. Then, the design space (DS) is determined, as that region of the multivariable space of CPPs and CMAs, where the CQAs meet the target values. QbD represents a major step forward with respect to the previous Quality-by-Testing (QbT) approach, that consisted in directly testing the

quality of the end-product. In the industrial practice and in the vast majority of the academic works, the DS is used in combination with a control system at open loop with respect to the CQAs. Although active process control is used for controlling the CPPs, their set-points are moved at open-loop within the DS in response to measured changes in the CMAs. As acknowledged by regulators, practitioners and academics (Yu et al., 2014), closed-loop control of the product quality can lead to a higher level of quality assurance and can enable real time release. In a control strategy with active quality control, the CQAs are monitored in real time, and the CPPs set-points are automatically adjusted in response to registered changes in the CMAs and/or CQAs. Recent literature publications demonstrated closed-loop control implementation and its advantages for a number of pharmaceutical processes, including plant-wide applications (Lakerveld et al., 2013). These results are leading to the development of a novel QbC framework (Su et al., 2019). QbC is not an independent system with respect to QbD, but it rather is an evolution, in which active process control is the core of the control strategy.

Another recent modernization of the pharma industry brought by the QbD initiative involves the shift to a more continuous production mode, compared to the traditional batch processing (Lee et al., 2015). When feasible, continuous manufacturing (CM) is the preferred choice in the process industry for many reasons, an important one being cost saving. For pharmaceutical manufacturing, the main motivation to the transition to CM is the inherently greater product consistency that can be achieved, thanks to the greater process controllability. Generally speaking, closed-loop control for continuous processes is typically easier and has been explored more than for batch processes. Overall, a tight interconnection emerges between the two main directions of evolution of QbD: the transition to CM requires the QbC framework for systematic control system design, while QbC implementation is gaining much momentum under the increased interest in continuous processing. So far, CM and QbC publications (Su et al., 2019) mainly focused on reactive systems, on crystallization and on drug product manufacturing (e.g.: tableting lines). The intermediate steps of filtration, washing and drying of crystallization slurries, and their integration, have not been studied yet in a CM/QbC perspective. Nonetheless, they are of pivotal importance to obtain the pure drug substance for downstream drug product manufacturing.

In this study, we make a step forward towards end-to-end automated CM by developing a QbC framework for a novel continuous carousel for integrated filtration, washing and drying of drug substances. The carousel filter-dryer system, manufactured by Alconbury Weston Ltd (UK), is a breakthrough technology in the field, as it is one of the few solutions available in the market for continuous upstream filtration, washing and drying. The rest of the manuscript is organized as follows. In Section 2, the carousel technology is described, while in Section 3 the conceived QbC system is discussed and the response of the control system to a disturbance is presented. The concluding section follows.

2. The continuous carousel for integrated filtration, washing and drying

Alconbury Weston Ltd designs and manufactures different models of carousels for solid/liquid separation. In this study, we refer to a prototype installed in the Crystallization Systems Engineering laboratory at Purdue University, used for isolating active pharmaceutical ingredients (API) from crystallization slurries. The P&ID of the process is showed in Figure 1. The controllers reported in the “Level 0” box (with the exception of the pressure controllers) are all implemented in the in-built programmable

logic control (PLC) of the unit. The carousel presents a main rotating cylindrical body, with five processing stations (Figure 2a). However, in the P&ID, each processing station is sketched as an independent tank (V104-108), to better illustrate the functioning of the control system. During carousel operation, every fixed cycle duration Δt_{cycle} the unit rotates, moving the content of each station to the following one. In the P&ID, the transfer of material upon carousel rotation is symbolized through the material streams situated among tanks V104-108 (controller FC102 opens the stream valves at each cycle switch). Processing stations V104-107 present filter meshes at the bottom (F101-104), while V108 is open for discharge. Starting from the inlet of the process, the slurry from the crystallizer (V101) is transported into an intermediate storage tank (V102) with a peristaltic pump (P101). From there, every Δt_{cycle} , an amount of slurry, controlled by FC101 based on the set-point $V_{slurry,sp}$, is drawn into the charge cell V103 through the action of the vacuum pump P102. P102 is connected to V103 only during the charging phase at the beginning of every cycle, through the three-way valve. During the rest of carousel operation, the vacuum pump is connected to the filtrate receiver V110 and, hence, to the bottom of the processing stations V104-107, providing the driving force for filtration, washing and drying. After the charging phase, the slurry enters V104, and filtration starts. Meanwhile, in V105-107 the material loaded during the previous cycles is processed. In V105, cake washing occurs, to reduce the amount of impurities in the cake (V109 is the wash solvent tank), while cake deliquoring (i.e. mechanical drying) is carried out in V106. A final drying step with hot air is delivered in V107, while the product (dry API crystals cake) is discharged from V108 through the action of a piston. The fan P103 can be used for increasing the pressure drops in each station (driving force for the processing steps), or the top portion of V104-107 can be directly connected to the atmosphere (i.e., eliminating F103 from the P&ID). PC101 and PC102 are, respectively, the in-built pressure controllers of P103 and P102. The heater H101 is responsible for heating the drying air, and its jacket temperature is manipulated by TC101 to control the temperature of the air entering V107. If the jacket temperature reaches 150°C, for safety reasons it is not heated up anymore, thanks to the action of TA101 and to the low selector. Additional sensors have been installed for monitoring and control purposes. PI101 measures the pressure at the bottom of the processing stations, FI101 measures the gas flowrate entering the carousel, FI102 (a scale placed below V110) measures the flowrate of filtrate coming from the ports, and TI101-103 are, respectively, the thermocouples for the jacket temperature, for the gas temperature at the dryer inlet, and for the gas temperature at the dryer outlet. The automated filter meshes cleaning system is not reported in the P&ID for conciseness. The control systems in the “Level 1” (Section 3) box have been implemented in this study for controlling the CQAs of the process, namely the residual solvents and impurities in the discharged cake. Overall, a risk analysis (Destro et al., 2021) showed that $\Delta t_{cycle,sp}$ (set-point of Δt_{cycle}) and $V_{slurry,sp}$ are the CPPs of the process.

3. A QbC framework for the continuous carousel

We develop the control system for the carousel following a recently proposed hierarchical three-level approach for control system design (Su *et al.*, 2019), based on the ISA-95 Enterprise-Control System Integration Standard. The Level 0 of the control system consists in the built-in PLCs of the carousel and of P102-103, already described in Section 2, that allow controlling CPPs and control variables. The designed Level 1 and Level 2 are instead outlined in this Section.

The Level 1 of the control system (Figure 1) uses advanced PID loops for controlling the CQAs. Since the solvents and impurities content cannot be measured online, the CQAs are controlled by inference through AC101, directly controlling the outlet temperature profile measured by TI103. Actually, the temperature profile of the outlet drying gas is strictly correlated to the residual solvents and impurities content in the cake. At the beginning of drying, the temperature drops, due to the latent heat of vaporization. While drying progresses, the temperature starts rising, since the residual solvents and impurities, and the energy consumed by their volatilization, are always lower. A reference temperature profile $T_{drying,sp}^{out}(t)$ is obtained in normal operating conditions, and is then tracked by AC101 to control the CQAs during operation. If the cake entering the dryer

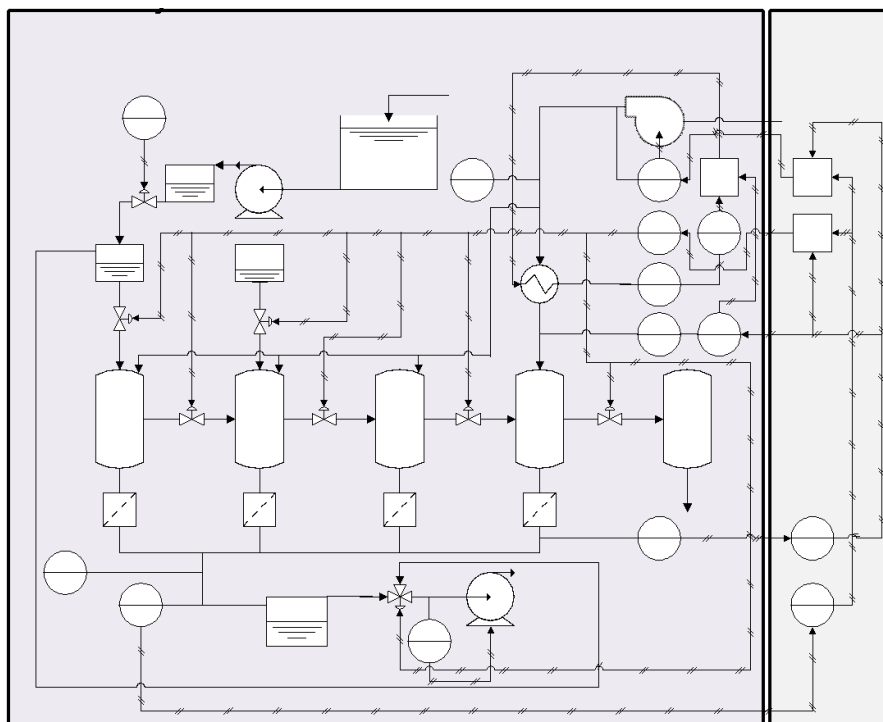


Figure 1. P&ID of the continuous carousel for integrated filtration, washing and drying.

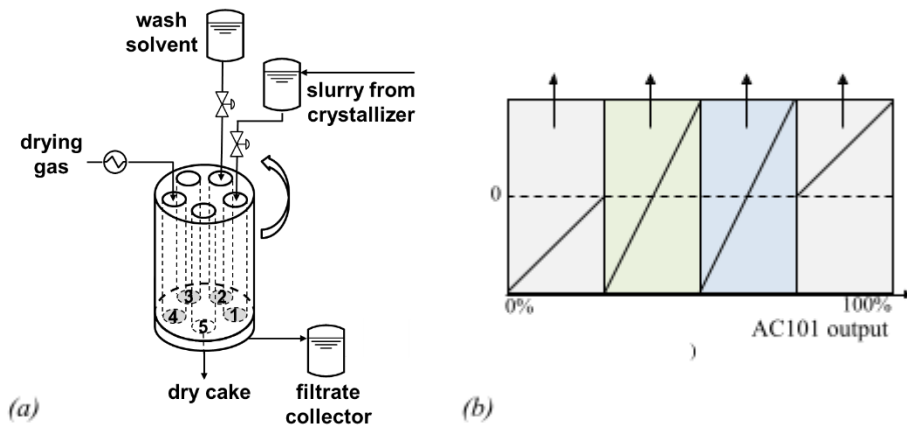


Figure 2. (a) Carousel schematic diagram. Stations 1-4 present a filter mesh at the bottom, connected to the collector. (b) Split range controller AC101: manipulated variable (deviation form) per range of controller output (proportional to $T_{drying,sp}^{out}(t_k) - T_{drying}^{out}(t_k)$).

contains more solvents than usually, due to issues in the previous section of the process, or if drying is progressing more slowly than usually, for example due to F104 fouling, TI103 will measure an abnormally low temperature. AC101, a split range controller (Figure 2b), will react increasing the set-points of one or more of the following: the pressure drop in the processing stations ΔP_{sp} (PC101), the drying gas inlet temperature $T_{drying,sp}^{out}$ (TC101) and/or the cycle duration $\Delta t_{cycle,sp}$ (FC102). The choice of increasing $\Delta t_{cycle,sp}$ only at last (Figure 2b), after that the maximum values of ΔP_{sp} and of $T_{drying,sp}^{in}$ are reached, is meant to maximize the process throughput. Instead, if the cake entering the dryer is purer than usually, the controller will react in the opposite direction, eventually reducing $\Delta t_{cycle,sp}$ and increasing the process throughput. The flowrate controller FC103 is also designed to help controlling the CQAs. FC103 (split-range controller) controls the filtrate flowrate by increasing/decreasing ΔP_{sp} and/or of $\Delta t_{cycle,sp}$ if the filtrate flowrate is low/high. This controller compensates for the disturbances that increase the filtration duration, such as fouling of F101-104, increase of crystals concentration in V101 or increase of cake resistance. Since both AC101 and FC103 act on ΔP_{sp} and $\Delta t_{cycle,sp}$, two high selectors are needed (Figure 1).

The response of Level 0 and Level 1 control system to a same filter mesh fouling disturbance is reported in Figure 3 for the isolation of aspirin from an aqueous slurry, as simulated by means of a detailed digital twin of the carousel, described elsewhere (Destro *et al.*, 2021). With Level 0 control, the batches produced in Cycles 7-71 have to be rejected, as they do not meet the desired quality (Figure 3a), which is reached again only after the meshes are cleaned (from Cycle 72 on). Instead, Level 1 controllers react very fast to the abnormal CQAs, by increasing ΔP_{sp} and $T_{drying,sp}^{in}$. After about 15 min (~ 7 cycles) from the start of the process, ΔP_{sp} and of $T_{drying,sp}^{in}$ saturate, and Level 1 controllers start increasing $\Delta t_{cycle,sp}$ (Figure 3b). Due to the increased $\Delta t_{cycle,sp}$, with Level 1 control only 68 cycles are completed in the same amount of time in which Level 0 control completes 115 cycles (Figure 3a). However, with Level 1 control the CQAs

always respect the quality threshold. When the filter meshes are cleaned (150 min from the start of the process), a higher $\Delta t_{cycle,sp}$ is not needed anymore, and Level 1 controllers start decreasing $\Delta t_{cycle,sp}$ (Figure 3b), effectively increasing the throughput again. Level 2 improves carousel operation through model-based control, exploiting the in-house developed model of the carousel (Destro *et al.*, 2021). Due to space limitation, Level 2 is not described in detail in this manuscript, but it includes a state estimator and a real time optimization/model predictive control component (RTO/MPC).

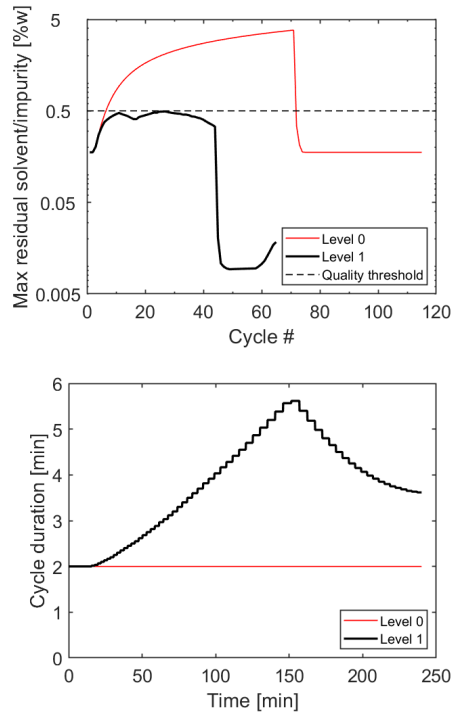


Figure 3. Control system response to a fouling disturbance: Level 0 vs Level 1. The solvent or impurity presenting the highest residual content in the discharged cake (CQAs) is reported on the left, while the profiles of $\Delta t_{cycle,sp}$ (CPP) are on the right.

The RTO/MPC updates the control vector $\mathbf{u} = [V_{slurry,sp}, \Delta t_{cycle,sp}, \Delta P_{sp}, T_{drying,sp}^{in}]$ by solving the following optimization problem:

$$W_1 \frac{V_{slurry,sp}}{\Delta t_{cycle,sp}} + (\mathbf{u} - \mathbf{u}_{ref})^T W_2 (\mathbf{u} - \mathbf{u}_{ref})$$

subject to: $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}), \mathbf{u}_{min} < \mathbf{u} < \mathbf{u}_{max}, \mathbf{x}_{min} < \mathbf{x} < \mathbf{x}_{max}$,

(1)

where \mathbf{x} is the states vector, \mathbf{W}_1 and \mathbf{W}_2 are suitable weighing matrices, f is the carousel model, \mathbf{u}_{ref} is the reference control vector and \mathbf{u}_{min} , \mathbf{u}_{max} , \mathbf{x}_{min} and \mathbf{x}_{max} are suitable bounds for \mathbf{u} and \mathbf{x} . The first term of the cost function represents the overall carousel throughput, which is maximized by the RTO/MPC respecting the quality constraints

(included in the constraints relation for \mathbf{x}). This represents a step forward with respect to Level 1, in which some sort of throughput maximization was only be achieved through split range control.

4. Conclusions

We presented a QbC three-level control system for a continuous carousel for integrated filtration, washing and drying of crystallization slurries. The control system successfully rejected a disturbance related to filter mesh fouling. This study represents a novel contribution to CM and QbC literature, as this type of process is scarcely studied. Future work will focus on Level 2 and on implementing the control system in the carousel.

5. Acknowledgements

Funding for this publication was made possible, in part, by the Food and Drug Administration through grant (U01FD006738). Views expressed in written materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government. F.D. gratefully acknowledges the “CARIPARO Foundation” for his PhD scholarship and “Fondazione Ing. Aldo Gini” for the financial support.

References

- Destro, F., Coleman, S., Firth, P., Barton, A., Barolo, M., Nagy, Z.K., 2021. Continuous integrated filtration, washing and drying of aspirin: digital design of a novel intensified unit. Proc. of the 11th IFAC Symposium on Advanced Control of Chemical Processes, Venice (Italy), June 13-16 2021. *In press*.
- FDA, 2004. Guidance for industry, PAT-A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance. <http://www.fda.gov/cder/guidance/published.html>. Accessed on 11/26/2020.
- Lakerveld, R., Benyahia, B., Braatz, R.D., Barton, P.I., 2013. Model-based design of a plant-wide control strategy for a continuous pharmaceutical plant. *AIChE J.* 59, 3671–3685.
- Lee, S.L., O’Connor, T.F., Yang, X., Cruz, C.N., Chatterjee, S., Madurawe, R.D., Moore, C.M.V., Yu, L.X., Woodcock, J., 2015. Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production. *J. Pharm. Innov.* 10, 191.
- Su, Q., Ganesh, S., Moreno, M., Bommireddy, Y., Gonzalez, M., Reklaitis, G. V., Nagy, Z.K., 2019. A perspective on Quality-by-Control (QbC) in pharmaceutical continuous manufacturing. *Comput. Chem. Eng.* 125, 216–231.
- Yu, L.X., Amidon, G., Khan, M.A., Hoag, S.W., Polli, J., Raju, G.K., Woodcock, J., 2014. Understanding pharmaceutical quality by design. *AAPS J.* 16, 771–783.

Control of Batch and Continuous Crystallization Processes using Reinforcement Learning

Brahim Benyahia,^{a,*} Paul Danny Anandan,^a Chris Rielly^a

^aLoughborough University, Department of Chemical Engineering, Epinal Way, Loughborough, Leicestershire, LE11 TU, UK

B.Benyahia@lboro.ac.uk

Abstract

In crystallization processes, the control of particle size distribution, shape and purity are crucial to achieve the targeted critical quality attributes of the final drug product and meet the pharmaceutical regulatory requirements. This work presents novel optimal trajectory tracking control strategies for batch and continuous cooling crystallization processes using reinforcement learning (RL). The cooling crystallization of paracetamol in water was used as a case study. A model-based reinforcement learning technique is implemented to achieve large crystal size by reducing the deviation from targeted reference trajectories namely process temperature, supersaturation and particle size. This multioutput tracking control strategy was developed to address quality and performance challenges commonly encountered in batch and continuous crystallization processes. Various training strategies and reward functions were investigated to enhance the learning capabilities and robustness of the reinforcement-learning-based control. Despite the computational costs inherent to reinforcement learning, the later demonstrated robust control capabilities compared the benchmark control strategies such as model predictive control.

Keywords: Reinforcement learning; Trajectory tracking control; Batch crystallization; Continuous crystallization; Supersaturation control; Crystal size distribution.

1. Introduction and Background

The pharmaceutical industry is subject to stringent regulatory requirements. The control of quality is critical as small deviations from the quality target profiles may have a significant impact on drug safety and efficacy, which may lead to failure of the clinical trials, at early development stages, or drug recalls causing dramatic economic consequences. The development of robust, flexible and agile control strategies is the cornerstone of the current Quality-by-Design paradigms. The development and successful integration of new process analytical technologies (PAT) allowed real time measurements of the critical quality attributes which has changed dramatically the pharmaceutical manufacturing landscape and paved the way for more advanced and reliable real-time optimization and control strategies (Lakerveld et al. 2015).

The emergence of continuous manufacturing in pharma opened new opportunities for more cost-effective and flexible technologies (Benyahia 2018; Mascia et al. 2013). However, the successful development of continuous processes requires more reliable and advanced control strategies to cope with the different sources of variation and

perturbation as well the complexity associated with the large number of process outputs, interactive control loops and intrinsic process dynamics, particularly during start-up, shut-down and ramp-up.

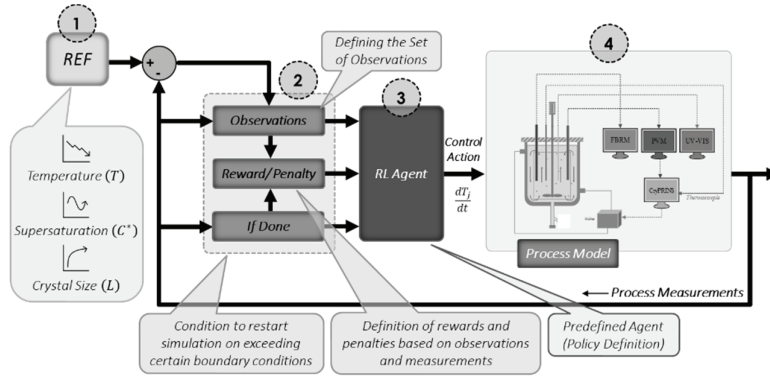
Crystallization is used extensively in the pharmaceutical industry to purify reaction intermediates and final active pharmaceutical ingredients. As such, several crystallization steps may be used at different stages of the manufacturing process (Benyahia et al., 2012). A successful crystallization requires tight control over crystal size distribution, purity, and polymorphism. These properties impact dramatically the critical quality attributes of the final pharmaceutical product, such as solubility and safety, besides the downstream processability of the crystals, such as flowability and filterability. Various control and optimization techniques have been developed and implemented over the last 2 decades ranging from model-free to model-based techniques (Lakerveld and Benyahia, 2020). Despite the significant progress in crystallization process control, there is still an increasing demand for advanced, versatile, robust, and cost-effective real-time optimization and control strategies, driven by a broader adoption of Quality-by-Design and the emergence of pharma 4. Mathematical models and digital twins are increasingly used in all aspects of process design, optimization, control, and decision making.

Over the last few years, artificial intelligence has witnessed a resurgence of interest in very broad research and industrial areas. Reinforcement Learning (RL) is particularly interesting and has the ability address dynamic problems encountered, for instance, in robotics and automotive industries (Roveda et al., 2018). These advantages have motivated recent implementation of RL as a new control strategy for bioprocesses (Petsagkourakis et al., 2020). This paper outlines the first application of RL in the field of crystallization. The approach consists in designing and training a RL agent to perform a tracking control over multiple reference trajectories, namely process temperature, supersaturation and particle size in batch and continuous crystallization processes. The performance of this model based RL control, which aims at minimizing the training costs, was compared against well-established and benchmark control strategies such as MPC.

2. Problem Formulation

The primary objective of this work is to develop a controller able to follow a set of dynamic performance trajectories effectively and reliably. This trajectory tracking control will help achieve shorter batch times as well optimal start-up, shut down and transitions between productions rates (production ramp-up). The approach will also help maintain the product critical quality attributes, such as crystal size distribution, within the acceptable and safe margins. The reference trajectories can be computed using model-based open-loop dynamic optimization for both batch and continuous crystallization processes. As a first attempt, an optimal linear cooling profile is suggested.

A model-based RL controller was developed in MATLAB Simulink. The idea is to train an agent in closed loop to achieve the best performance (highest reward) over a series of training episodes. The agent takes dynamic actions by changing the jacket temperature or cooling rate (manipulated variable) and receives rewards and penalties (negative rewards) according to its dynamic performance against the reference trajectories and process constraints. The setup of the closed loop training strategy is shown in **Figure 1**. The first block contains the reference trajectories, which are three in this case (i.e. temperature, supersaturation, and particle size). The second part receives the observations, quantifies the rewards, and decides whether the training episode should be skipped or restarted if



certain criteria are met. The third block in this sequence is the RL Agent that constitutes the policy and the training algorithm. The fourth block is the process model which is represented by a system of differential algebraic equations.

The dynamic mathematical model of the continuous crystallization of paracetamol in water was derived from the population, mass, and energy balances. The kinetic parameters of the growth and nucleation were obtained from the literature (Nagy et al., 2008). The standard method of moments is used to capture the population dynamics. The primary nucleation and size independent growth were considered whereas the secondary nucleation, agglomeration and breakage were negligible. The moment equations, mass and energy balances are described by the set of nonlinear ordinary differential equations described below.

$$\frac{d\mu_0}{dt} = \frac{F_{seed}}{V} \mu_{0,seed} - \frac{F_{out}}{V} \mu_0 + B \quad (1)$$

$$\frac{d\mu_i}{dt} = \frac{F_{seed}}{V} \mu_{i,seed} - \frac{F_{out}}{V} \mu_i + iG\mu_{i-1} \quad (2)$$

$$\frac{dC}{dt} = \frac{F_{in}}{V} C_{in} - \frac{F_{out}}{V} C - \rho_c k_v \frac{d\mu_3}{dt} \quad (3)$$

$$\frac{dT}{dt} = T_{in} \frac{F_{in}}{V} - T \frac{F_{out}}{V} - 3 \frac{\Delta H k_v \rho_c}{\rho_t C_{pt}} G \mu_2 + \frac{UA(T_j - T)}{V \rho_t C_{pt}} \quad (4)$$

where F_{in} , F_{out} and F_{seed} are the inlet, outlet and seed flowrates respectively, μ_i is the i th moment, V is the volume of the crystallizer, C is the solution concentration, ρ_c is the density of the crystals, k_v is the volumetric shape factor, T_{in} , T and T_j are the inlet, process and jacket temperatures respectively, ΔH is the enthalpy of crystallization, U is the overall heat transfer coefficient and A the jacket surface area. The nucleation and growth rates are described below.

$$B = k_b \Delta C^b \quad (5)$$

$$G = k_g \Delta C^g \quad (6)$$

$$R_t = \underbrace{(10(|T_e| \leq 1) - 1(|T_e| \geq 1))}_{\text{Reward and penalty for maintaining the error within a acceptable range}} - \underbrace{\left(0.02 \left(\int |T_e|\right)\right)}_{\text{Penalty to minimise the effect of the accumulated error}} + \underbrace{\left(25 \left(\frac{T_s}{T_f}\right)\right)}_{\text{Reward for every incremental time step of the simulation period}} - \underbrace{(100(280 \geq T \geq 315))}_{\text{Penalty if the simulation is stopped in the middle by exceeding the temperature bounds}}$$

where $\Delta C = C - C^*$ is the absolute supersaturation. The mathematical model of the batch process can be obtained by eliminating the inlet and outlet terms ($F_{in} = F_{out} = 0$).

The proposed RL strategy relies on a Recurrent Neural Network (RNN) to parameterize the policy. RNN uses both process measurements and control actions as an input to deliver a probability distribution as an output. The efficiency of the policy strongly depends on the properties of the neural network. In this work we used the Deep Deterministic Policy Gradient (DDPG) method.

The defining of an appropriate reward function is critical in RL owing its evident impact on the quality and performance of the training. The decomposing of the rewards is the best way to teach the agent with small incentives for every little positive state observed and help achieve a good compromise between exploration and exploitation. Several reward functions have been used with different levels of complexity. **Figure 2** provides an example of the reward functions used in this work. Where R_t - is the reward given to the RL agent at time 't' of the simulation; T_e - the error between the process temperature and the reference profile; T_s - the time interval during which the RL agent takes a single control action; T_f - the final simulation time.

3. Results and Discussion

The RL results were obtained using the following settings: The actor network comprises 3 hidden layers and the critic networks has 25 hidden layers for the actions and observations. Each layer comprises 100 neurons, and each training episode consists of 300 steps. In addition to this several training parameters were optimized to achieve the best training performance which will not be provided for the sake of the brevity.

The RL approach was implemented to two different case studies. The first case focused on the trajectory tracking control of a batch crystallization, aiming at minimizing the deviation from a cooling profile while achieving optimal supersaturation and mean crystal size trajectories. This RL-based control strategy can be designed to minimize the batch time while maximizing quality. The second case considers the same set of process and product quality attributes but uses different reference trajectories. The approach can be designed to minimise start-up time and achieve high quality steady state performance.

In each case, the agent was trained using several training sets based on bespoke reward functions to achieve optimal training and control performances. The trained agent was validated against the reference trajectories in both cases and compared to a standard MPC, as shown in **Figure 3**. Overall, both MPC and RL perform well in terms of trajectory tracking. The RL seems to achieve better in the case of continuous crystallization.

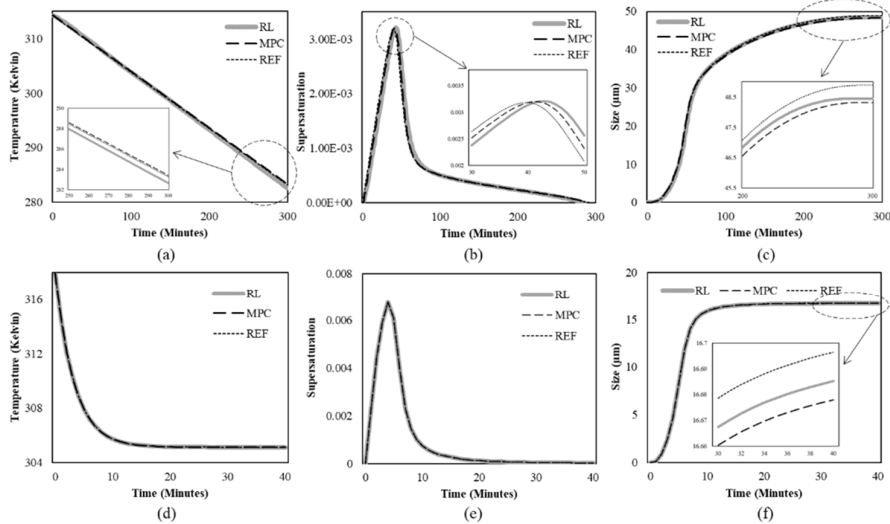


Figure SEQ Figure * ARABIC 3: Performance comparison of RL against standard MPC. (a), (b) and (c) for a batch system and (d), (e) and (f) for a continuous system.

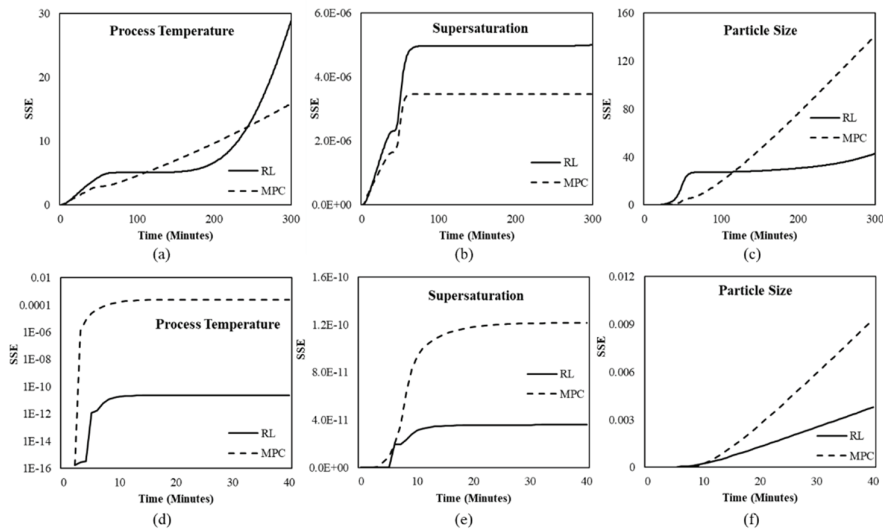


Figure SEQ Figure * ARABIC 4: Performance comparison between RL and MPC using the sum of the squared errors.

To analyze more effectively the performance of the RL-controller against the MPC, the sum of the squared errors (SSE) were computed systematically for each trajectory in both cases, batch and continuous systems, as shown in **Figure 4**. This quantitative approach makes it possible to gain more accurate insights into the control performance in each case.

4. Conclusion

A model based RL controller was designed and trained to address tracking control problems for batch and continuous crystallization processes. Three critical reference trajectories were considered namely: process temperature, supersaturation, and mean

crystal size. The jacket temperature was used as a control variable. The RL agent was trained using a range of reward functions to achieve the highest reward score, the fastest training and the best compromises between exploration and exploitation. The resulting RL-based controller was validated and compared against benchmark control strategies such as MPC. It was shown that the RL can achieve excellent performance, based on both qualitative and quantitative key performance indicators, despite the challenges inherent to the simultaneous tracking of three reference trajectories. As such, the RL can be a promising advanced process control candidate, where dynamic trajectory tracking, and constraint handling are required. Despite the computational burden, the development of model based RL, using reliable and predictable models, has a real potential to cut the high costs commonly associated with the training of the agent in real world processes. However, it is critical to assess the robustness of the agent in presence of model uncertainties, measurement noise and process upsets. This is precisely the current focus of the research group.

Acknowledgements: Funded by the EPSRC (EP/R032858/1) ARTICULAR project.

References

- Benyahia, B., Lakerveld, R., Barton, P.I., 2012. A plant-wide dynamic model of a continuous pharmaceutical process. *Industrial and Engineering Chemistry Research*. <https://doi.org/10.1021/ie3006319>
- Benyahia, B. 2018. Applications of a plant-wide dynamic model of an integrated continuous pharmaceutical plant: Design of the recycle in the case of multiple impurities. In *Computer Aided Chemical Engineering* (1st ed., Vol. 41). Elsevier B.V.
- Lakerveld, R., Benyahia, B., 2020. CHAPTER 4: Process Control, in: Yazdanpanah, N., K Nagy, Z. (Eds.), *The Handbook of Continuous Crystallization*. London Royal Society of Chemistry, pp. 172–218.
- Lakerveld, R., Benyahia, B., Heider, P.L., Zhang, H., Wolfe, A., Testa, C.J., Ogden, S., Hersey, D.R., Mascia, S., Evans, J.M.B., Braatz, R.D., Barton, P.I., 2015. The Application of an Automated Control Strategy for an Integrated Continuous Pharmaceutical Pilot Plant. *Organic Process Research and Development*. <https://doi.org/10.1021/op500104d>
- Mascia, S., Heider, P.L., Zhang, H., Lakerveld, R., Benyahia, B., Barton, P.I., Braatz, R.D., Cooney, C.L., Evans, J.M.B., Jamison, T.F., Jensen, K.F., Myerson, A.S., Trout, B.L., 2013. End-to-end continuous manufacturing of pharmaceuticals: Integrated synthesis, purification, and final dosage formation. *Angewandte Chemie - International Edition*. <https://doi.org/10.1002/anie.201305429>
- Nagy, Z.K., Chew, J.W., Fujiwara, M., Braatz, R.D., 2008. Comparative performance of concentration and temperature controlled batch crystallizations. *Journal of Process Control*. <https://doi.org/10.1016/j.jprocont.2007.10.006>
- Petsagkourakis, P., Sandoval, I.O., Bradford, E., Zhang, D., del Rio-Chanona, E.A., 2020. Reinforcement learning for batch bioprocess optimization. *Computers and Chemical Engineering* 133, 106649. <https://doi.org/10.1016/j.compchemeng.2019.106649>
- Roveda, L., Pallucca, G., Pedrocchi, N., Braghin, F., Tosatti, L.M., 2018. Iterative Learning Procedure with Reinforcement for High-Accuracy Force Tracking in Robotized Tasks. *IEEE Transactions on Industrial Informatics*. <https://doi.org/10.1109/TII.2017.2748236>
- Fysikopoulos, D., Benyahia, B., Borsos, A., Nagy, Z. K., & Rielly, C. D. 2019. A framework for model reliability and estimability analysis of crystallization processes with multi-impurity multi-dimensional population balance models. *Comp. and Chem. Eng.*, 122, 275–292.

31ST EUROPEAN SYMPOSIUM ON
COMPUTER AIDED PROCESS
ENGINEERING

PART C

31ST EUROPEAN SYMPOSIUM ON
COMPUTER AIDED PROCESS
ENGINEERING

PART C

Edited by

Metin Türkay
Koç University

Rafiqul Gani
PSE for SPEED



ELSEVIER

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

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

Copyright © 2021 Elsevier B.V. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

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

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

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

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN (Part C): 978-0-323-98327-9

ISBN (Set) : 978-0-323-88506-5

ISSN: 1570-7946

For information on all Elsevier publications visit our website at <https://www.elsevier.com/>



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

Typeset by STRAIVE

Contents

212. Process Monitoring under Uncertainty: An Opportunity for Bayesian Multilevel Modelling	
<i>Andrew J. Radcliffe, Gintaras V. Reklaitis</i>	1377
213. Modelling and Health Monitoring of a Pressure Regulating System	
<i>Shivom Sharma, Matthew B Leo, Shamsuzzaman Farooq, Iftekhar A Karimi</i>	1383
214. Unsupervised Monitoring of Flocculation Processes based on Recurrence Theory	
<i>Hooman Ziaei-Halimejani, Nima Nazemzdeh, Reza Zarghami, Krist V. Gernaey, Martin Peter Andersson, Seyed Soheil Mansouri, Navid Mostoufi</i>	1389
ENERGY/WATER/FOOD NEXUS AND SUSTAINABILITY	
215. Optimal Supply Chain for Renewable Furfural Production Involving Economic, Environmental and Social Criteria	
<i>Gabriel Contreras-Zarazúa, Mariano Martin-Martin, José María Ponce-Ortega, Juan Gabriel Segovia- Hernandez</i>	1395
216. Levels of Approximation for the Optimal Design of Distributed Energy Systems	
<i>Ishanki A. De Mel, Oleksiy V. Klymenko, Michael Short</i>	1403
217. A Coordinated Framework for the Optimization of Municipal Solid Waste Management	
<i>Aurora del Carmen Munguía-López, Victor M. Zavala, José Ezequiel Santibañez-Aguilar, José María Ponce-Ortega</i>	1409
218. A Probabilistic Scenario Generation Framework for Optimal Decision Making in Turkish Renewable Energy Market	
<i>Hasan Sildir, Handan Akulker, Erdal Aydin</i>	1415
219. Dynamic optimization modelling for a national-scale energy system in transition	
<i>Xiang Li, Dario Müller, Subhash Kumar, François Maréchal</i>	1421
220. Carbon Policy Assessment in Process Integration	
<i>Mohammad Lamah, Dhabia Al-Mohannadi, Patrick Linke</i>	1427
221. Back-End Design and Development of an Energy Systems Analysis Tool	
<i>Maryam Arbabzadeh, Sapna kumari, Ragini Sreenath, Emre Gençer</i>	1433
222. Compression system power requirements for various CO₂ sources and transportation options	
<i>Mathew Dennis Wilkes, Sanjay Mukherjee, Solomon Brown</i>	1439

223.	Life Cycle Optimization of Hybrid Energy Systems towards Carbon Neutrality <i>Xueyu Tian, Fengqi You</i>	1445
224.	Value of aggregator-led community energy storage systems with degradation constraints <i>Jude O. Ejeh, Flora A. V. Biggins, Solomon F. Brown</i>	1453
225.	Clean Electricity Generation from Sewage: a novel Ultra-Supercritical Water Oxidation Technology with integrated CCS <i>Sanjay Mukherjee, Solomon Brown</i>	1459
226.	Design of Food-Energy-Water-Waste Nexus Systems in New York State under COVID-19 Pandemic <i>Ning Zhao, Fengqi You</i>	1465
227.	Assessment strategies of different molten salts for heat transfer in tri-generation systems <i>Rachid Klaimi, Sabla Y. Alnouri, Mirko Stijepovic</i>	1473
228.	Techno-economic Assessment of Load Following Operation for Super-critical Power Plants Equipped with Carbon Capture Feature <i>Calin-Cristian Cormos, Ana-Maria Cormos, Cristian Dinca</i>	1479
229.	Utilisation of CO₂ in transnational LNG supply chain for the enhancement of jet fuel production in globally decentralised GTL industries <i>Ali Attiq Al-Yaeeshi, Tareq Al-Ansari</i>	1485
230.	River Water Quality Prediction in Malaysia Based on Extra Tree Regression Model Coupled with Linear Discriminant Analysis (LDA) <i>Danny Hartanto Djarum, Zainal Ahmad, Jie Zhang</i>	1491
231.	Synthesis and Design of Sustainable Integrated Process, Water Treatment, Energy Supply Networks and Carbon Utilization Networks Under Uncertainty <i>Yue Li, Fengqi You, Zhihong Yuan, Rafiqul Gani</i>	1497
232.	Optimising the Sustainability Performance of an Industrial Park: an Energy-Water-Food Nexus <i>Jamileh Fouladi, Ahmed AlNouss, Tareq Al-Ansari</i>	1505
233.	A Study to Target Energy Consumption in Wastewater Treatment Plant using Machine Learning Algorithms <i>Akash Das, Piyush Kumar Kumawat, Nitin Dutt Chaturvedi</i>	1511
234.	Development of sustainable integrated biorefinery networks in pulp and paper industries <i>Ghochapon Mongkhonsiri, Amata Anantpinijwatna, Pongtorn Charoensuppanimit, Amornchai Arpornwichanop, Rafiqul Gani, Suttichai Assabumrungrat</i>	1517

235. **Optimization of biogas autothermal reforming integrated with PEM based on renewable energy**
Jeongdong Kim, Meng Qi, Il Moon 1523
236. **Integration of CO₂ capture to power plants: the effect of fuel and gas composition in process configuration**
Ana Gabriela Romero-García, Nelly Ramírez-Corona, Eduardo Sánchez-Ramírez, Heriberto Alcocer-García, Juan Gabriel Segovia-Hernandez 1529
237. **Optimal Integration of Renewables, Flexible Carbon Capture, and Energy Storage for Reducing CO₂ emissions from Fossil Power Plants**
Manali S. Zantye, Mengdi Li, M. M. Faruque Hasan 1535
238. **Circular Economy Systems Engineering: A case study on the Coffee Supply Chain**
Stefanos G. Baratsas, Efstratios N. Pistikopoulos, Styliani Avraamidou 1541
239. **Towards a Circular Economy Calculator for Measuring the “Circularity” of Companies**
Stefanos G. Baratsas, Naseem Masoud, Valentini A. Pappa, Efstratios N. Pistikopoulos, Styliani Avraamidou 1547
240. **Economic Optimization of Qatar’s Hydrocarbon-based Fuels for Sustainable Maritime Applications**
Ahad Al-Enazi, Eric C. Okonkwo, Yusuf Bicer, Tareq Al-Ansari 1553
241. **Unmanned aerial vehicles in precision agriculture towards circular economy: a process system engineering (PSE) assessment**
Mohammed Yaqot, Brenno Menezes, Tareq Al-Ansari 1559
242. **Utilisation of Carbon Dioxide and Gasified Biomass for the Generation of Value Added Products**
Ahmed AlNouss, Gordon Mckay, Tareq Al-Ansari 1567
243. **Synthesis of the EU Supply Networks for the Gradual Transition to 100 % Renewable Energy Production**
Sanja Potrč, Lidija Čuček, Zdravko Kravanja 1573
244. **Simultaneous optimization and heat integration of a macroalgae-based biorefinery**
Fernando D. Ramos, Andrés I. Casoni, Vanina Estrada, M. Soledad Diaz 1581
245. **Integrated optimization of Multi-Energy System operation and thermal comfort management for buildings**
Lavinia Marina Paola Ghilardi, Alessandro Francesco Castelli, Luca Moretti, Mirko Morini, Emanuele Martelli 1587

246. **Towards a Rational, Quantum-Chemistry-Based Selection and Screening of Green Solvents for Liquid-Liquid Phase Transfer Catalysis**
Abhimanyu Pudi, Adam P. Karcz, Rofice Dickson, Suojiang Zhang, Seyed Soheil Mansouri, Martin P. Andersson 1593
247. **A Model-Centric Multivariate Analysis for Reducing H₂S Emissions in Wastewater Treatment Plants**
Wilfredo Angulo, Yris Gonzalez, Santiago D. Salas, Dany De Cecchis, Alexander Espinoza 1599
248. **Gas-to-liquid strategy for liquid fuel production from industrial residue gases: Process development and technoeconomic analysis**
Thai Ngan Do, Chanhee You, Jiyong Kim 1605
249. **Multi-Objective Optimization of Lignocellulosic Ethanol Production Based on Regional Land and Water Availability**
Rashi Dhanraj, Varun Punnathanam, Yogendra Shastri 1611
250. **Deep Learning and AutoML for Dynamic Modeling of LNG Regasification Process Using Seawater**
Yongbeom Shin, Dongil Shin 1617
251. **Carbon or Nitrogen-based e-fuels? A comparative techno-economic and full environmental assessment**
Diego Freire Ordóñez, Thorsteinn Halfdanarson, Caroline Ganzer, Gonzalo Guillén-Gosálbez, Niall Mac Dowell, Nilay Shah 1623
252. **Techno-economic assessment of a renewable energy-based electric vehicle fast-charging station in Qatar**
Abdulla Al Wahedi, Yusuf Bicer 1629
253. **Simultaneous Synthesis and Optimization of Refrigeration Cycles and Heat Exchangers Networks**
Matteo Martinelli, Cristina Elsidio, Ignacio E. Grossmann, Emanuele Martelli 1635
254. **Waste- to-DME: Process Synthesis and Sustainability Evaluation**
Lefu A. Lithoko, Bilal Patel 1641
255. **Long-distance hydrogen delivery using a toluene-based liquid organic hydrogen carrier system**
Mariana C. S. Carvalho, João Marques, Henrique A. Matos, José F.O. Granjo 1647
256. **Assessment of energy efficiency improvement in ceramic industry through waste heat recovery modelling**
Miguel Castro Oliveira, Muriel Iten, Henrique A. Matos 1653

257. Targeting and Optimisation of Industrial and Urban Symbiosis for Circular Economy	
<i>Limei Gai, Petar Sabev Varbanov, Hon Huin Chin, Jiří Jaromír Klemeš, Sandro Nižetić</i>	1659
258. Negative CO₂ emissions in biomass gasification process with hybrid amine-deep eutectic solvents	
<i>Cristian Dinca, Nela Slavu, Calin-Cristian Cormos, Eliza-Gabriela Mihaila</i>	1665
259. Development and optimization of an energy efficient and self-sustained integrated process for nitrogen removal for baseload LNG plant	
<i>Ajinkya Pal, Easa I. Al-musleh, Iftekhar A Karimi</i>	1671
260. Integration of Green Power in a Gas to Liquid Process	
<i>Mohammad Ostadi, Emre Gençer and Magne Hillestad</i>	1677
261. Comparative Study of Different Hybrid Life Cycle Assessment Methodologies Applied to Biomass-based Chemical Production	
<i>Yuqing Luo, Robert O'Dea, Thomas Epps, Marianthi Ierapetritou</i>	1683
262. Optimising the energy, water and food nexus node to support decision making for sustainable food security in Risky Environments	
<i>Maryam Haji, Rajesh Govindan, Tareq Al-Ansari</i>	1689
263. Synthesis of Sunlight, Seawater and CO₂ based Industrial Parks	
<i>Elizabeth J. Abraham, Farah O. Ramadan, Dhabia M. Al-Mohannadi, Patrick Linke</i>	1695
OPERATIONS AND SUPPLY CHAINS	
264. Planning production of solar grade silicon to yield solar panels involving behaviour of population	
<i>Esbeydi Villicaña-García, César Ramírez-Márquez, Juan Gabriel Segovia-Hernández, José María Ponce-Ortega</i>	1701
265. Bi-level Mixed-Integer Data-Driven Optimization of Integrated Planning and Scheduling Problems	
<i>Burcu Beykal, Styliani Avraamidou, Efstratios N. Pistikopoulos</i>	1707
266. A Prospective Approach in the Design of Lignin-based Products Supply Chains	
<i>Brunelle Marche, Javier A. Arrieta-Escobar, Vincent Boly, Juan C. Solarte-Toro, Carlos A. Cardona Alzate</i>	1715
267. Flexibility Analysis in Supply Chain Management: Application to the Traveling Salesman Problem	
<i>Alessandro Di Pretoro, Stephane Negny, Ludovic Montastruc</i>	1721

- 268. Production scheduling of continuous make-and-pack processes with byproducts recycling**
Apostolos P. Elekidis, Georgios P. Georgiadis, Michael C. Georgiadis 1727
- 269. Robust Optimization for Scheduling and Lot-sizing of a Single Machine with Sequence-dependent Changeovers**
Hossein Mostafaei, Fabricio Oliveira 1733
- 270. Simultaneous minimization of minimum resource and storage requirements in batch process**
Nitin Dutt Chaturvedi, Rakesh Kumar Sinha 1741
- 271. A Hybrid Solution Approach for Large-scale Batch Scheduling with Quantum Computing**
Akshay Ajagekar, Fengqi You 1747
- 272. A Digital Twin Framework for Business Transactional Processes in Supply Chains**
Hector D. Perez, Satyajith Amaran, Esra Erisen, John M. Wassick, Ignacio E. Grossmann 1755
- 273. A Data-Driven Approach to Plan Electricity Production from Diesel Engines with Constrained Parameters**
Piyush Kumar Kumawat, Nitin Dutt Chaturvedi 1761
- 274. An Efficient Computational Strategy for Process Scheduling: a Case Study of the Omsk Lubricant Plant**
Andrei Kostin, Yulia Mashchenkova, Vladimir Reflov, Alexander Tkachev 1769
- 275. Centralised vs. decentralised production and storage: optimal design of a decarbonised hydrogen supply chain with multiple end uses**
Renato Luise, Annabelle Brisse, Catherine Azzaro-Pantel 1775
- 276. Optimal Design of the Reverse Logistics Network of Empty Agrochemical Containers**
Glenda N. Yossen, Gabriela P. Henning 1781
- 277. Designing a Rectification Strategy for Managing Disruptions in LNG Supply Chain**
S.V. Ramanan, Rajagopalan Srinivasan 1787
- 278. Statistical Decision-Theoretic Risk Management for Planning Renewable Energy Pathways**
Houd Al-Obaidli, Rajesh Govindan, Tareq Al-Ansari 1795
- 279. Supply Chain Planning with Vehicle Allocation for Gas Industry**
Yena Lee, Sivaraman Ramaswamy, Jose M. Pintob, Lazaros G. Papageorgiou 1803

280.	Optimal Day-Ahead Load Curtailment for Central Chiller Plants <i>Gustavo Campos, Yu Liu, Nael H. El-Farra, Ahmet Palazoglu</i>	1809
281.	Interplaying of food supply chain resilience, industry 4.0 and sustainability in the poultry market <i>Ahmed H. A. Mohamed, Brenno C. Menezes, Tareq Al-Ansari</i>	1815
282.	Integrated Spatially Explicit Landscape and Biofuel Supply Chain Network Design <i>Eric G. O'Neill, Christos T. Maravelias</i>	1821
283.	Techno-economic and environmental assessment of Gasoline produced from GTL and MTG processes <i>Saad A. Al-Sobhi, Ahmed AlNouss, Mohammad Alhamad</i>	1827
284.	Simplified targeting models for Sustainable Supply Chains retrofitting in process industries <i>Shabnam Morakabatchiankar, Fernando D. Mele, Moisés Graells, Antonio Espuña</i>	1833
285.	Flexibility-expansion planning for enhanced balancing-power market participation of decentralized energy systems <i>Niklas Nolzen, Ludger Leenders, André Bardow</i>	1841
286.	A fair profit allocation model for the distribution plan optimisation of refined products supply chains <i>Bohong Wang, Jiří Jaromír Klemeš, Taicheng Zheng, Yongtu Liang</i>	1847
287.	Markov Decision Process to Optimise Long-term Asset Maintenance and Technologies Investment in Chemical Industry <i>Hon Huin Chin, Bohong Wang, Petar Sabev Varbanov, Jiří Jaromír Klemeš</i>	1853
288.	Cost, environmental and exergy optimization of hydrogen and methane supply chains: application to Occitania region, France <i>Eduardo Carrera, Catherine Azzaro-Pantel</i>	1859
289.	Planning of non-conventional gas field development with parametric uncertainties <i>Jui-Yuan Lee, Pei-Fan Bai, Raymond E. H. Ooi, Dominic C. Y. Foo, Raymond R. Tan</i>	1865
BIOLOGICAL SYSTEMS AND PROCESSES		
290.	Effect of the lignin extraction process on the economics of a woody-based biorefinery <i>Juan Camilo Solarte-Toro, Javier A. Arrieta-Escobar, Brunelle Marche, Carlos A. Cardona Alzate</i>	1871

291.	Economic and environmental assessment of bio-hydrogenated diesel production process from waste cooking oil integrated with reforming process <i>Suksun Amornraksa, Lida Simasatitkul, Parinya Khongprom, Akekapob Sektaweelarp, Suttichai Assabumrungrat</i>	1877
292.	Modelling and simulation of a multiple feedstock integrated biorefinery for the production of aviation biofuel and other biofuels <i>Araceli Guadalupe Romero-Izquierdo, Fernando Israel Gómez-Castro, Claudia Gutiérrez-Antonio, Salvador Hernández, Massimiliano Errico</i>	1885
293.	Coordinated Management of Agricultural Waste and High-tech Value Bioproducts <i>Heriberto Alcocer-García, Juan G. Segovia-Hernandez, Eduardo Sanchez-Ramirez, Ana Gabriela Romero-García, Yicheng Hu, Victor M. Zavala</i>	1891
294.	A design and optimization framework for (bio-)chemical processes based on exergo-economic and environmental aspects <i>Jaime David Ponce-Rocha, Martín Picón-Núñez, Ana Carvalho, Andreia Santos, Fernando I. Gómez-Castro, Ricardo Morales-Rodríguez</i>	1897
295.	Integrating Process Simulation and Life Cycle Assessment to Evaluate the Economic and Environmental Performance of Algae Biodiesel <i>S. Zapata-Boada, M. Gonzalez-Miquel, M. Jobson, R. Cuellar-Franca</i>	1903
296.	Optimization of biogas production from cattle manure by anaerobic digestion using a gradient-based algorithm <i>Tina Kegl, Anita Kovač Kralj</i>	1909
297.	Photosynthetic Bioplastics Production with Cyanobacteria by Coupled Growth-Production Mutants <i>Romina Lasry Testa, Vanina Estrada, Claudio Delpino, M. Soledad Diaz</i>	1917
298.	Parametric analysis of mammalian cell (GS-NS0) culture performance for advanced mAb biopharmaceutical manufacturing <i>Wil Jones, Dimitrios I. Gerogiorgis</i>	1923
299.	A technoeconomic assessment of an on-site biocrude production from sewage sludge in Qatar's wastewater treatment plants <i>Mohammad Alherbawi, Tareq Al-Ansari, Hamish R. Mackey, Gordon McKay</i>	1929
300.	Production costs of advanced biofuels using a multi-component learning curve model <i>Paraskevi Karka, Filip Johnsson, Stavros Papadokostantakis</i>	1937
301.	Simulation of the Biofuel Production Process from Organic Fraction of Municipal Solid Waste (OFMSW) <i>José Rocha-Rios, Arturo Pérez-Roman, Teresa Lopez-Arenas, Mauricio Sales-Cruz</i>	1943

- 302. Comparison of different processing routes for the valorisation of olive tree pruning wastes**
Luis David Servián Rivas, Ismael Diaz, Manuel Rodriguez, María González-Miquel, Emilio J. González 1949
- 303. The Use of GVL for Holistic Utilization of Biomass**
Antreas Pateromichelakis, Melina Psycha, K. Pyrgakis, A. Kokossis 1955
- 304. Valorization of sugarcane bagasse to lactic acid: Life cycle assessment and Techno-economic evaluation in Indian scenario**
Munagala Meghana, Yogendra Shastri, Kakasaheb Konde, Sanjay Patil 1963
- 305. Hybrid Cybernetic Modeling of the Microbial Production of Polyhydroxyalkanoates Using Two Carbon Sources**
Stefanie Duvigneau, Robert Dürr, Lena Kranert, Annette Wilisch-Neumann, Lisa Carius, Rolf Findeisen, Achim Kienle 1969
- 306. Microbial Production of Polyhydroxyalkanoates – Modeling of Chain Length Distribution**
Robert Dürr, Stefanie Duvigneau, Achim Kienle 1975
- 307. Identifiability of metabolic flux ratios on carbon labeling experiments**
Rafael D. de Oliveira, Caroline S. M. Nakama, Vânia Novello, José G. C. Gomez, Galo A.C. Le Roux 1983
- 308. Metabolic Network Reconstruction from Time-Series of Concentration Data: Evaluation of the Numerical Matrices Methods**
Severo Balasbas III, Ivan Ivanov, Kai Sundmacher 1991
- 309. Assessment of intermediate storage and distribution nodes in personalised medicine**
Andrea Bernardi, Maria Papathanasiou, Matthew Lakelin and Nilay Shah 1997
- 310. Thermal treatment in dairy processes: Validation of protein deposition models**
Wilfred Kwabena Darko, Federico Lozano Santamaria, Laurent Bouvier, Guillaume Delaplace, Sandro Macchietto 2003
- 311. Optimal efficiency of biomass conversion from bio-based byproducts to biofuel production in the Ethiopian sugar industry: A case study in Wonji-Shoa sugar factory, Ethiopia**
Shumet Sharew, Ludovic Montastruc, Abubeker Yimam, Stephane Negny, Jean- Henry Ferrasse 2009
- 312. Data-driven prediction of antiviral peptides based on periodicities of amino acid properties**
Chris A. Kieslich, Fatemeh Alimirzaei, Hyeju Song, Matthew Do, Paige Hall 2019

313. Allosteric Regulation of CRYs in Mammalian Circadian Clock <i>Onur Ozcan, Seref Gul, Ibrahim Halil Kavakli</i>	2025
EDUCATION AND TRAINING	
314. Reflections on the development of scenario and problem-based chemical engineering projects <i>A. Tsatse, E. Sorensen</i>	2033
315. Programming skills across the (bio)engineering curriculum – a students’ perspective <i>Simoneta Caño de las Heras, Carina L. Gargalo, Krist V. Gernaey, Ulrich Krühne</i>	2039
316. Engineering Social Responsibility: Rising to the Challenge of Sustainable Development <i>Iain Duncan Stalker, Rinkal K. Desai</i>	2045
317. Virtualization of Pressure Control Demonstration Unit – A Step Towards Virtual Laboratory <i>Miloš Bogataj, Nejc Arh, Zdravko Kravanja</i>	2051
318. Complementarities of part-time and full-time postgraduates in process systems engineering (PSE): when industry and academia merges <i>Brenno C. Menezes, Robert E. Franzoi, Mohammed Yaqot, Mohamed E. H. Sawaly, Adnan A. M. A. Al-Banna, Salman Ashkanani, Jeffrey D. Kelly</i>	2057
319. Building Knowledge Capacity for Quantum Computing in Engineering Education <i>Deborah Carberry, Amirhossein Nourbakhsh, Jay Karon, Mark N. Jones, Mojgan Jadidi, Kyarash Shahriari, Christian Beenfeldt, Martin Peter Andersson, Seyed Soheil Mansouri</i>	2065
320. Teaching PSE Mastery During, and After, the COVID-19 Pandemic <i>Daniel R. Lewin</i>	2071
321. Real-Time Chemical Process Monitoring with UMAP <i>Zachary Webb, José Romagnoli</i>	2077
322. Efficient Energy Integration and Design of Distillation Separation Sequences <i>José A. Caballero, Juan Javaloyes</i>	2083
323. Enhancing operability during early stage of process synthesis considering flexibility and inherent safety simultaneously <i>Andreja Nemet, Klavdija Zirngast, Zdravko Kravanja, Zorka Novak Pintarič</i>	2089

Process Monitoring under Uncertainty: An Opportunity for Bayesian Multilevel Modelling

Andrew J. Radcliffe,* Gintaras V. Reklaitis

*Purdue University, 480 Stadium Mall Drive, West Lafayette, IN 47907 United States
andrewjradcliffe@gmail.com*

Abstract

Bayesian modelling has recognized utility for system design, process monitoring and control strategy wherein decisions must be made under uncertainty. Applications to operation and quality assurance are particularly promising as the Bayesian paradigm enables unique multivariate probabilistic control strategies through the use of multi-level regression with post-stratification. This work demonstrates application to real-time quality control in a dropwise additive manufacturing system for pharmaceuticals. A variety of multilevel hierarchical models are developed, and rigorous model selection is performed using information criteria. Then, the probabilistic quality control strategy is presented in general form using transformations of the posterior predictive distribution, with results assessed on out-of-sample data.

Keywords: hierarchical Bayesian, multilevel modelling, machine learning, pharmaceuticals, additive manufacturing

1. Introduction

When a mechanistic model of a system cannot be fully defined, data-driven modelling provides an adaptable means of relating inputs to outputs, but often without allowing ready interpretation of the relationships obtained. The Bayesian paradigm enables the specification of known mechanistic structure, which can then be supplemented by the relationships implicit in the data, thus providing improved interpretability, while the use of probability as uncertainty measure provides a ready framework for decision analysis.

Bayesian approaches have attracted considerable interest for predictive purposes in process monitoring, control strategy, and systems design – where decisions must be made under uncertainty (Stamatis et al., 2018; Mockus et al., 2020). The ability to sensibly set up complex models with many parameters and multi-layered structure enabled by hierarchical Bayesian methods provides unique opportunities for addressing challenging problems in which a measurement model is used to predict quantities of interest from real-time sensor data. A simple example in which the data-level model is seemingly sufficient is the thermocouple, in which the measured voltage is related to temperature through its characteristic function under the assumption that the function is invariant under process conditions. In more complex sensors, the process conditions themselves may interact with the measured quantity in ways that cannot be accounted for by the measurement model. Such a scenario arises as the thermocouple ages, the thermoelectric properties change, and with them, the characteristic function. Hence the basic problem under consideration is how to utilize relevant information that arises from the process and device conditions, but which is external to the physics-based measurement model.

For inherently multi-level problems such as this, the hierarchical Bayesian approach enables creation of a model structure which reflects the relationship between phenomena, and therefore enables modelling of the parametric inter-/relations within levels, in addition to their relationship to the overarching physical variable spaces, thereby making full use of the extra information to improve predictive accuracy. As the posterior predictive distribution (PPD) is the method by which to rigorously evaluate event probabilities for new/out-of-sample data, this has direct use in control strategies based on model predictions.

In this work we demonstrate these ideas with an application to an image-based process monitoring and control strategy enacted on a dropwise additive manufacturing (DAMP) system for pharmaceutical dosage units. In previous work, a hierarchical Bayesian approach was used which involved categorical partitioning to construct hierarchical groupings across a database of 41 batches of drug products, spanning a wide range of process conditions and ink formulations (Radcliffe and Reklaitis, 2020). We expand these into multilevel models via nested hierarchical regressions on the coefficients of the image-derived explanatory variables using the numerous group-level explanatory variables which arise from the interaction of process conditions, fluid properties, and particle properties.

2. Modelling approach

The DAMP system produces dosage units through the sequential dropwise deposition of a suspension onto a substrate. The control strategy involves multiple aspects – fluid composition/rheology, temperature, substrate position; the process output – drops – is monitored in real time using an online image acquisition system. With the content of each tablet/capsule determined by the volume of material added to it, the online image data provides the means to enact quality control strategies at several levels of complexity.

2.1. Construction of the Explanatory Variables for Multi-level Regression

In our approach the historical process data is used to inform a multi-level model which takes as inputs at the data-level the features extracted from the online images, and at the upper-level the explanatory variables constructed from the process conditions. Physical reasoning plays a critical role in building the explanatory variables at both levels to achieve a hybrid model whose structure reflects the physics. Such hybrid approaches provide several advantages: enhanced interpretability, improved prediction at out-of-sample conditions, and lower requirements on volume of training data.

For use in the data-level model, each image is reduced to 8 explanatory variables: 1) primary drop volume, 2) summed satellite volume(s), 3) cylindrical approximation for partially observed satellites, 4) spherical approximation for partially observed satellites, 5) approximation for partially-observed primary drop at bottom border, 6) approximation for partially-observed primary drop at top border, 7) Euclidean distance from centroid of primary drop to image centre, and 8) mean of Euclidean distances from centroids of satellite drops to image centre. Partial observations necessitate the addition of parametric approximations for the missing volumes based on the geometry presented by each image and category of defect. The image processing, methodology and supporting logic for the image-derived explanatory variables are described in (Radcliffe and Reklaitis, 2020).

The operating conditions and suspension properties have substantial effects on drop formation. Though such variables do not belong in the data-level model, the existence of

interaction between the drops observed in the images and the underlying physics is undeniable. Including such effects requires a multi-level model in which the data-level parameters are themselves modelled using a regression on a set of upper-level explanatory variables derived from the process conditions.

In the DAMP operating with suspensions, the process conditions include ejection velocity, shear rate, nozzle radius, number of pump volume strokes, fluid properties (viscosity, density, surface tension), particle properties (size, shape, density) and particle loading. Dimensional analysis can be used to combine the numerous physical properties and process conditions, and thereby re-express them using the relevant dimensionless groups for drop formation and suspension rheology: Weber number, Ohnesorge number, suspension viscosity scaling with particle loading $f(\phi/\phi_{max})$, Stokes number and particle Reynolds number; the last two can be modified by the particle aspect ratio to produce shape-scaled variants.

2.2. Hierarchical Structure

In building a model, the extent to which data sets are similar is of fundamental importance in modelling as this determines the pooling choices and hence the model structure. Traditional methods offer two approaches: no pooling, or complete pooling; hierarchical modelling offers a third path: partial pooling, the extent of which is determined by the data. As the historical data consists of clearly portioned sets of dosage units (“a batch”) made distinct by the dosage properties themselves (e.g. target content, number of drops per unit) and their respective set of dimensionless variables, a hierarchical model can provide the necessary flexibility while avoiding overfitting.

In the context of the multi-level regression, each batch of doses (a matrix of image-derived explanatory variables) is allocated its own regression coefficient vector at the data level; the hierarchy is constructed based on the exchangeability of these vectors and is modelled through a multivariate prior. At the upper-level regression, there exists a vector of explanatory variables for each batch (dimensionless parameters), while the upper-level regression coefficients and covariance matrix are shared. This is shown in in Eq.(1)-(5).

$$y_n \sim normal\left(X_n B_{j[n]}, \sigma_{j[n]}\right) \quad n = 1, \dots, N \quad (1)$$

$$\beta_j \sim multinormal\left(U_j G, \Sigma_\beta\right) \quad j = 1, \dots, J \quad (2)$$

$$\Sigma_\beta = diag(\tau)\Omega diag(\tau); \quad \Omega \sim LKJcorr(2), \quad \tau_k \sim Cauchy(0, 2.5) \quad k = 1, \dots, K \quad (3)$$

$$G_{l,k} \sim normal(0, 5) \quad l = 1, \dots, L, \quad k = 1, \dots, K \quad (4)$$

$$\sigma_j \sim half-normal(a + bw_j, \psi) \quad j = 1, \dots, J \quad (5)$$

Support for the data-level model is based on the physics: the outcome (deposited volume, y_n) should be the sum of the contributions from the primary drop and satellite drop(s) weighted by the respective coefficients $\beta_{j,k}$. Given the high reproducibility of the process and apparent linearity of the deposited volume to image relation, a normal likelihood (Eq.(1)) is assigned to the outcome. The prior distribution on the $j=1, \dots, J$ vectors of regression coefficients is assigned a multivariate normal distribution (Eq.(2)). For each β_j there is a vector of upper-level explanatory variables, U_j ; the components of the matrix of regression coefficients, G , are assigned weakly informative priors (Eq.(4)). The covariance matrix of the multivariate normal is decomposed into a vector of scale parameters, τ , and a correlation matrix, Ω (Eq.(3)); the components of τ are assigned

weakly informative priors with the potential for heavy tails using the half-Cauchy distribution (Eq.(3)). The correlation matrix is assigned a Lewandowski-Kurowicka-Joe correlation (LKJcorr) matrix distribution with some amount of concentration toward the identity using $\eta = 2$ (Eq.(3)). The scale parameter of the data-level model is modelled parametrically (Eq.(5)) using the half-normal distribution with location parameter $a+bw_j$ and scale ψ , given the relevant explanatory variable, $w_j = 1/n_{drops}$; the slope and intercept of the regression are shared. Hyperpriors on a, b are the standard normal and $\psi \sim \text{half-normal}(0,1)$.

2.3. Model Comparison and Selection

With the form of the model established, there remain many possible choices for the data-level explanatory variables, X , and upper-level explanatory variables, U . Physical reasoning provides general guidelines, but it may not be readily apparent which to include/exclude. This problem can be approached systematically by evaluation of multiple models followed by comparison utilizing information criteria which account for the ability of a large/ more complex model to fit data better. One such fully Bayesian measure is the WAIC (widely applicable information criterion) (Vehtari et al. 2017). This measure can be used to screen hundreds of models to obtain a handful which are then examined further using graphical checks, posterior predictive performance on out-of-sample data, etc.

For the problem at hand, there are 6 possible combinations of X 's and 24 possible combinations of U 's, yielding 144 models. The possible combinations of X 's can be reasonably well-constrained by physical reasoning, e.g., $X_{1,2,3,4,5,6}$ should always be present as these are direct contributions to drop volume. However, it is not immediately clear whether the Euclidean distance measures ($X_{8,9}$) should be included; furthermore, there is the option of adding a constant intercept ($X_7 = 1$). For the upper-level explanatory variables, it is reasonable to forcibly include dimensionless groups related to drop formation – We (U_{j1}), Oh (U_{j2}) – and some measure of suspension behavior through $f(\phi/\phi_{max})$ (U_{j4}); alternately, $Oh:f(\phi/\phi_{max})$ can be used (U_{j3}). However, for other variables for which it is difficult to justify inclusion/exclusion *a priori*, terms are sequentially added to/ removed from the upper-level variable vector to construct 24 possible choices; these are denoted in the results as $U_j_l_*$, where the wildcard represents the member indices of included variables. Aspect-ratio (AR) scaling for the Stokes (U_{j5}) and particle Reynolds (U_{j6}) numbers is used for both or none. Pump volume strokes is also considered (U_{j7}).

3. Computational Methodology

Hamiltonian Monte Carlo with the No-U-Turn-Sampler, implemented in Stan (Carpenter et al., 2017), is used to draw samples from the joint posterior distribution of the model parameters given the training data. The historical dataset was partitioned into two equal halves: for each batch of doses, the odd dose indices ($i=1,3,5,\dots$) are allocated to the training data and the even indices ($i=2,4,6,\dots$) to the out-of-sample data. For each model, 4 Markov chains are simulated with warm-up period of 20,000 followed by sampling period of 20,000 iterations; this yielded 80,000 samples upon mixing. Convergence is assessed using the rank normalized \hat{R} , which must below 1.01 for a simulation to be considered for use; additionally, the effective sample size must be ≥ 400 to keep Monte Carlo error low.

Scalability of such computations can present unique challenges related to dynamic memory requirements, particularly during predictive uses, as the number of stored values

scales with proportionality $n_{\text{data points}} \times n_{\text{MCMC samples}}$. The Markov chain simulations themselves can be parallelized to fully utilize available compute resources: multi-process (1 chain per process), and multi-threaded within each process; the latter enables efficiency improvements via map-reduce, which is very helpful for log-likelihood computations.

4. Results and Discussion

4.1. Using the posterior predictive distribution

The major uses of the joint posterior distribution, $p(\theta|y,X)$, involve the PPD, $p(\tilde{y}|\tilde{X},y,X)=\int p(\tilde{y}|\tilde{X},\theta)p(\theta|y,X)d\theta$, in which θ denotes the model parameters. Predictive accuracy scoring utilizes the training data to compute scores based on the log predictive density; in this vein, WAIC is used to score predictive performance for each of the 144 models and identify superior variants, as shown in Figure 2. Detailed checks of the fit of the model to the training data, X, U, W , utilize the PPD to predict the distribution of the alternate possible realizations of the outcomes, y^{rep} ; then, test variables can be set up for this model and used to assess model performance (Radcliffe and Reklaitis, 2020).

To make predictions with regression models given out-of-sample data, $\tilde{X}, \tilde{U}, \tilde{W}$, the joint posterior draws are used to sample from the PPD of \tilde{y} . The PPD of \tilde{y} can be transformed by any function of interest, with full inference obtained on the transformed quantity since probability is conserved through transformations. Consequently, event probabilities can be evaluated in a straightforward manner on both \tilde{y} and transformed quantities; this enables probabilistic control strategies at multiple levels.

For the problem at hand, quality control can be enacted at the individual level (each dose) or for entire batches. For each new batch, there is a target value for content, T and a target tolerance, z , such that the quality specification requires each dose be within $[T(1-z), T(1+z)]$. The probability that a dose (index i) is within specification can be computed by transforming the samples from \tilde{y} using Eq.(6), then computing the expectation of the transformed variable, $E[f(\tilde{y}_i, T, z)]$. The distribution of the fraction of the batch within specification can be obtained by transforming the joint distribution of the transformed \tilde{y}_i 's using Eq.(7), in which I is the index set of doses in the batch. The samples from Eq.(7) can then be used to compute the probability that a desired fraction of the batch, denoted D , is within specification by transforming using Eq.(8) and taking the expectation of the result. In addition to event probabilities, transformations such as Eq.(7) can also be used to assess the predictive performance on new/out-of-sample data, as shown in Figure 1.

$$f(\tilde{y}_i, T, z) = \begin{cases} 1 & T(1-z) \leq \tilde{y}_i \leq T(1+z) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$g(f(\tilde{y}, T, z), I) = \sum_{i \in I} f(\tilde{y}_i, T, z) / |I| \quad (7)$$

$$h(g(f(\tilde{y}, T, z), I), D) = \begin{cases} 1 & g(f(\tilde{y}, T, z), I) \geq D \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

The ability to compute event probabilities at multiple levels of abstraction provides a powerful method for multi-level control schemes. For example, suppose that $\Pr(g(f(\tilde{y}, T, z), I) \geq D)$ is low, and it is proposed that some doses from the batch be discarded in order to increase the probability. The decision of identifying the best candidates to exclude can be posed as an optimization problem in which the optimal solution is the index (sub)set which maximizes $\Pr(g(f(\tilde{y}, T, z), I) \geq D)$, subject to the constraint that the number of doses in the reduced batch be greater than some constant. Alternately, this can

be inverted: maximize the number of accepted doses, subject to the constraint that $\Pr(g(f(\tilde{y}, T, z), I) \geq D)$ must exceed some threshold.

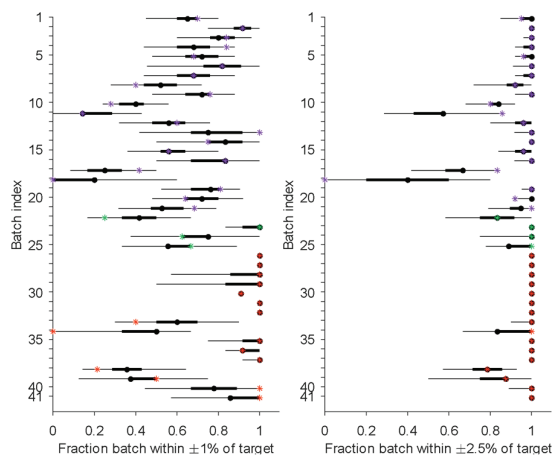


Figure 1. Percentiles (2.5, 25, 50, 75, 97.5) of $g(f(\tilde{y}, T, z), I)$ respective to each batch at $z = 0.01, 0.025$. Actual values displayed as

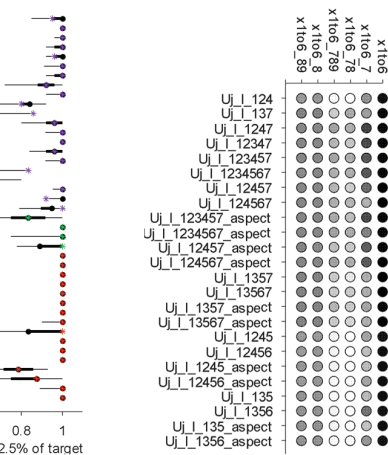


Figure 2. WAI score for 144 models. Lighter color indicates better predictive accuracy.

5. Conclusions

In this work a multilevel hierarchical Bayesian model was developed for probabilistic quality control in a DAMP system. The hybrid approach utilized physical understanding in building the models, thereby increasing interpretability of the joint posterior and improved confidence in prediction at new conditions. Working in the Bayesian paradigm enabled rigorous evaluation of event probabilities through transformations of the PPD, which provides unique capabilities for optimization and quantitative decision analysis. Future work will expand on use in batch/lot optimization, in addition to considering more abstract problems wherein a sensible physical model is not available, but the data exhibits innate multi-level structure which can be leveraged.

References

- A.J. Radcliffe, G.V. Reklaitis, 2020, Bayesian Hierarchical Modeling for Online Process Monitoring and Quality Control, Computers and Chemical Engineering (in review)
- L. Mockus, G.V. Reklaitis, K. Morris, D. LeBlond, 2020, Risk-based Approach to Lot Release, Journal of Pharmaceutical Sciences, 109, 2, 1035-1042
- S.D. Stamatis, L. Mockus, L.E. Kirsch, G.V. Reklaitis, 2018, Bayesian Hierarchical modeling of Gabapentin absorption and disposition with application to dosing regimen assessment, Computer Aided Chemical Engineering Vol 42, pp. 111-137
- Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing, 27(5), 1413-1432.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A. 2017. Stan: A probabilistic programming language. Journal of statistical software, 76(1).

Modelling and Health Monitoring of a Pressure Regulating System

Shivom Sharma,^a Matthew B Leo,^a Shamsuzzaman Farooq,^{a,*} Iftekhar A Karimi^{a,*}

^aDepartment of Chemical & Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore

chesf@nus.edu.sg; cheiak@nus.edu.sg

Abstract

A pressure regulating station (PRS) is used to step down the pressure of gas distribution pipelines suitable for domestic or industrial users, irrespective of gas demand. A gas distribution network can have hundreds of PRSs. Hence, they are key assets, and have to be continuously monitored for condition deterioration and/or sudden faults. A PRS can be modelled using first principles, but it is computationally expensive. In this study, we developed a computationally cheaper machine learning based surrogate for a PRS. Subsequently, weekly cumulative flow is calculated for a PRS to monitor the filter choking and to predict the remaining useful life of the filter. Finally, a dashboard has been developed for monitoring the health of PRSs in a gas distribution network that can be integrated with online operational data for their condition monitoring and predictive maintenance. This can increase the reliability of a gas distribution network and reduce the maintenance cost.

Keywords: Gas Distribution Network, Pressure Regulating System, First Principles Modelling, Health Monitoring, Prescriptive Maintenance.

1. Introduction

Natural or town gas distribution requires a complex network of pipelines to supply gas from a supplier to the residential, commercial, and industrial end-users. Suitably placed pressure regulating stations (PRSs) are used to step down the supply pressure and maintain it within a specified range for the end-users, irrespective of the gas demand. Any issue with the PRS system should be identified and rectified before the supply of gas is adversely affected. As PRSs are critical assets of the gas distribution networks, they are carefully monitored for any health deterioration and potential faults (Leo et al., 2020).

Some vital potential health degradation and faults that may occur in a PRS are filter choking, spring decay, valve seat damage, diaphragm malfunction, and overloading of network. The health of PRS should be monitored for gradual decrease in the performance and/or sudden faults. Real-time operational data is critical for online monitoring and automated prescriptive maintenance of a PRS.

Depending upon the nature of operational data available for a PRS, both model-based and data-based techniques may be applied for its health monitoring. In this work, we applied a rigorous first principles model in for a PRS (Nabi and Dayan, 2000), which is validated using technical data provided by the vendor (RMG by Honeywell). However, a first principles based health monitoring may be computationally heavy for developing a fast

PRS health monitoring methodology for a large network. As an alternate, we have developed a computationally cheaper but equally effective machine learning (neural networks: NN) based surrogate to replace the first principles based model.

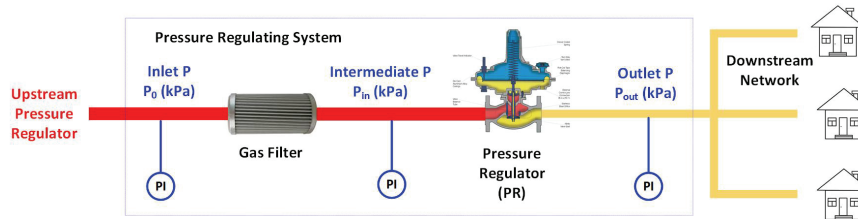


Figure 1: Overview of PRS with downstream network

Leo et al. (2019) have applied the PRS health monitoring methodology on a simulated case study. In this study, actual operational data from a gas distribution company is used. The schematic of a representative PRS is shown in Figure 1. Three potential pressure reading locations are shown in the figure. Our developed methodology for PRS health monitoring should be able to identify fault in filter, with high reliability and accuracy. In some cases, the intermediate pressure readings (P_{in}) are not available. Our methodology allows monitoring the filter chocking for both cases. For easy reference, let us call them Case 1 and Case 2. In case 1, available data set comprise all three pressure (P_0 , P_{in} , P_{out}). In Case 2, data for only two pressures (P_0 and P_{out}) are available.

Estimating Remaining Useful Life (RUL) is critical for the prescriptive maintenance of an equipment. The PRS health monitoring procedure includes prediction of RUL for its parts. Finally, a dashboard (a user-friendly GUI) has been developed for monitoring the health of all the PRSs in the gas distribution network. This dashboard can be further integrated with the real-time PRS operational data to extend health by advising on the prescriptive maintenance. In addition to supervising their gradual decline, the PRS health monitoring dashboard may also monitor swift abnormality in any PRS.

2. First Principles Model and NN Surrogate for PRS

Leo et al. (2019) have developed a first principles model for HON 280 pressure regulator. They have considered mass balance and energy balances across the flow chamber, and force balance on the assembly of valve seat and diaphragm. A pressure regulator reaches steady state almost instantaneously due to its faster dynamics compared to the downstream distribution network. Hence, solving the pressure regulator model assuming a steady state operation was found adequate.

The developed first principles model for a pressure regulator accurately predicts the flow rate using intermediate and downstream pressures (P_{in} and P_{out}). This first principles model is expected to be computationally expensive for online monitoring and fault detection of hundreds of pressure regulators in a very large network. Hence, we have developed a simpler and computationally cheaper surrogate. The neural network-based surrogate was developed using the available operational data (pressures) for a gas distribution network.

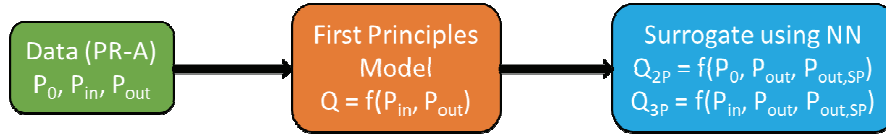


Figure 2: Development of neural network-based surrogates for 2- and 3-pressures cases

The NN-based surrogates (with 2 hidden layers) were developed using operational data (P_0 , P_{in} , and P_{out}) for the pressure regulator A (PR-A). The performance of the first principles model and surrogates were compared on the operational data of four pressure regulators (B with P_0 and P_{out} ; C, D, E with P_{in} and P_{out}). In other words, we developed surrogates for regulator A, and used them to predict the flows for other pressure regulators. Note that P_{in} (intermediate pressure in Figure 1) is not available for pressure regulator B, and $P_{in} = P_0$ (no pressure drop across the filter) is assumed for calculating flow rate using the first principles model. Figure 3 presents hourly flow rate calculated using the first principles model and NN-based surrogates. It can be seen that surrogates were able to predict the flow rate with high accuracy (~0.6 to 2.9% deviation from the first principles model). Table 1 compares the performance of the first principles model and surrogates for four pressure regulator data set. The NN-based surrogates require negligible computation time compared to the first principles model.

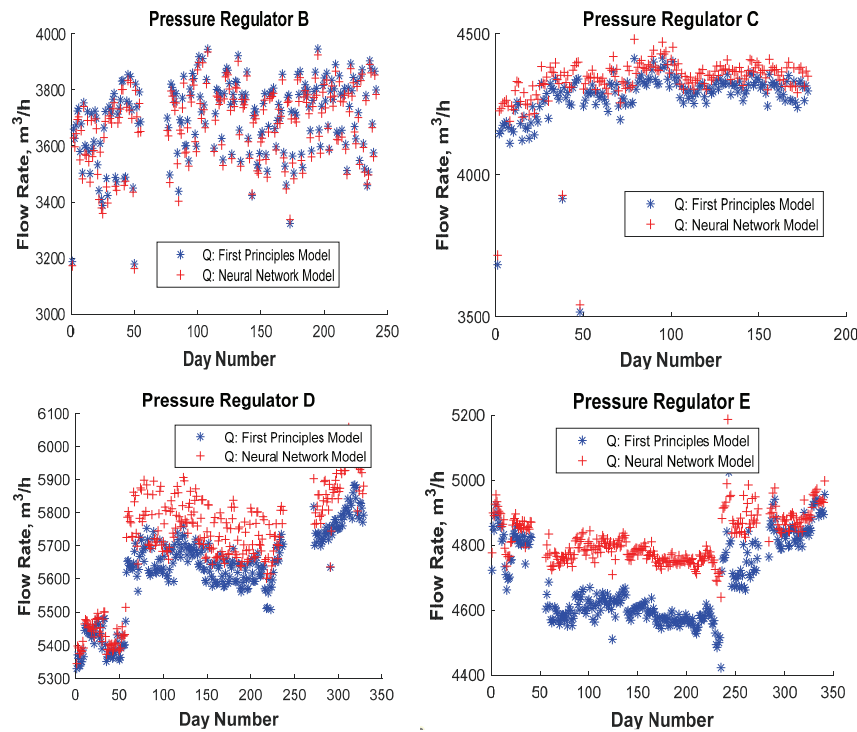


Figure 3: Hourly flow rate comparison using first principles model and neural network-based surrogates, for pressure regulators B, C, D and E

Table 1: Performance comparison between first principles (FP) model and surrogates (computer processor: i5-8265u @ 1.60 GHz & 1.80 GHz, RAM: 8 GB)

PR Name	Number of Samples	Computation Time (s)		Flow Difference b/w FP Model and Surrogates	
		FP Model	Surrogate	%	Absolute, m ³ /h
B (2P)	242691	2132	0.24	0.60	22.16
C (3P)	249753	3869	0.29	-1.28	-54.90
D (3P)	98057	859	0.20	-1.61	-91.19
E (3P)	105954	832	0.21	-2.90	-134.21

3. Monitoring Filter Choking on Gas Distribution Network Data

PRs are vital components in a gas distribution network. Hence, PR health monitoring is essential for a reliable gas supply to the customers. Leo et al. (2019) studied four PR faults on a simulated case study. Filter choking and spring decay cause an under pressure in downstream distribution network, whereas valve seat damage causes an over pressure of downstream distribution network. A PR does not work properly, if there is any damage in diaphragm. Leo et al. (2019) studied following three scenarios:

Scenario 1: P_{out} data at short intervals and cumulative flow at long intervals

Scenario 2: P_{out} and flow data at same intervals

Scenario 3: P_{in} and P_{out} data at same intervals

Mostly gas distribution companies do not measure flow rate for different pressure regulators. The gas flow is usually measured at the gas consumption sites (household, commercial or industry). Further, only inlet (P_0) and outlet (P_{out}) pressures are measured for many pressure regulators. Thus, we present a methodology for the two cases defined earlier, in the next section.

3.1. Weekly Cumulative Flow

In Case 1, monitoring of filter choking is easier as discussed by Leo et al. (2019). As stated above, both filter choking and spring decay cause under pressure downstream distribution network. Filter choking occurs much faster than the spring decay. Thus, spring constant can be assumed constant for monitoring filter choking over a relatively short time period.

An ideal pressure regulator will supply gas at P_{out} (= set point) regardless of the flow. For a real pressure regulator, if P_{in} is fixed, P_{out} will decrease with increase in flow (droop curve, Emerson, 2019). For a constant value of the inlet pressure (P_0), the intermediate pressure (P_{in}) decreases with increase in filter choking. In order to compensate for the decrease in P_{in} , P_{out} also decreases that increases cumulative flow calculated by the PR model. In this study, we have estimated weekly cumulative flow through the pressure regulators using the surrogates. The RUL of the filter can be estimated by fitting a line through the weekly cumulative flows. Maintenance should be carried out when weekly cumulative flow exceeds some threshold value. Figure 4 shows variations in the weekly cumulative flow for five pressure regulators. In order to predict the RUL, we used a

threshold value of 10^6 m^3 for weekly cumulative flow. The RUL values for different pressure regulators are also given on plots in Figure 4.

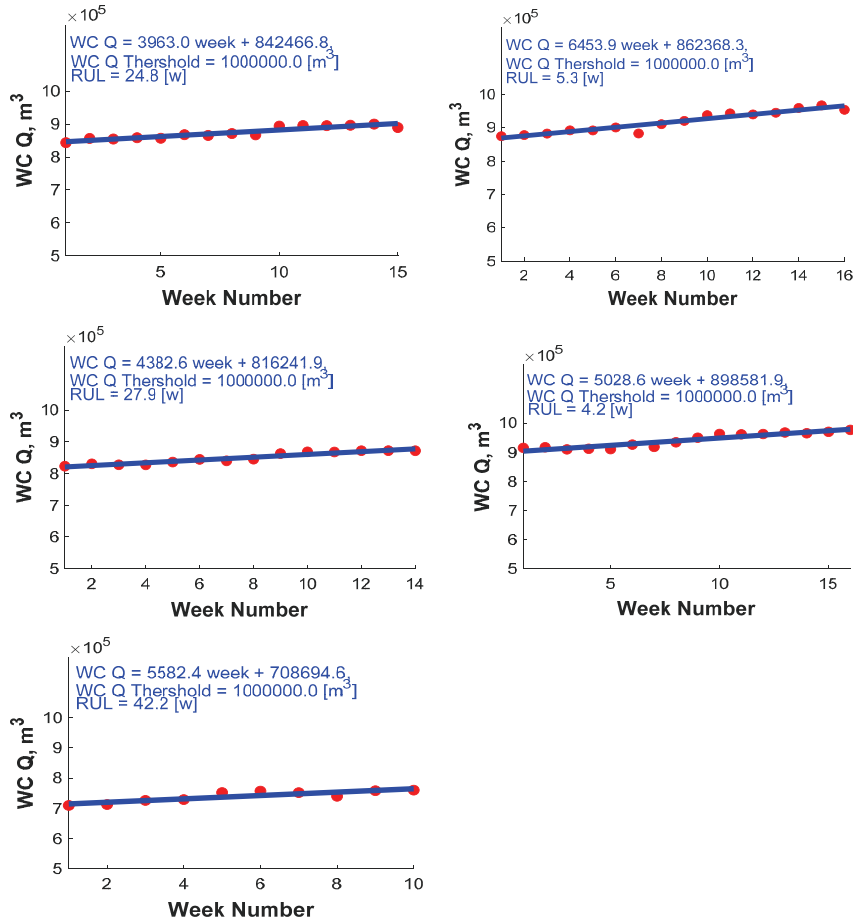


Figure 4: Weekly cumulative flow-based RUL predictions for five pressure regulators

3.2. Pressure Regulator Health Monitoring Dashboard

We have developed a dashboard for the health monitoring of the PRSs in a gas network. The pressure regulator dashboard has two parts: the ‘PR Group Analysis’ is used to analyse the performance of the pressure regulators in a group or region, and the ‘PR Performance Comparison’ is used to compare the performance of selected pressure regulators from the same or different groups or regions (see Figure 5). Some important flexible features of the dashboard are automatic update of data files, possibility of rearranging pressure regulators in different groups, selecting some or all pressure regulators for performance analysis, selecting start and end dates of analysis, and interactive display of results/plots.

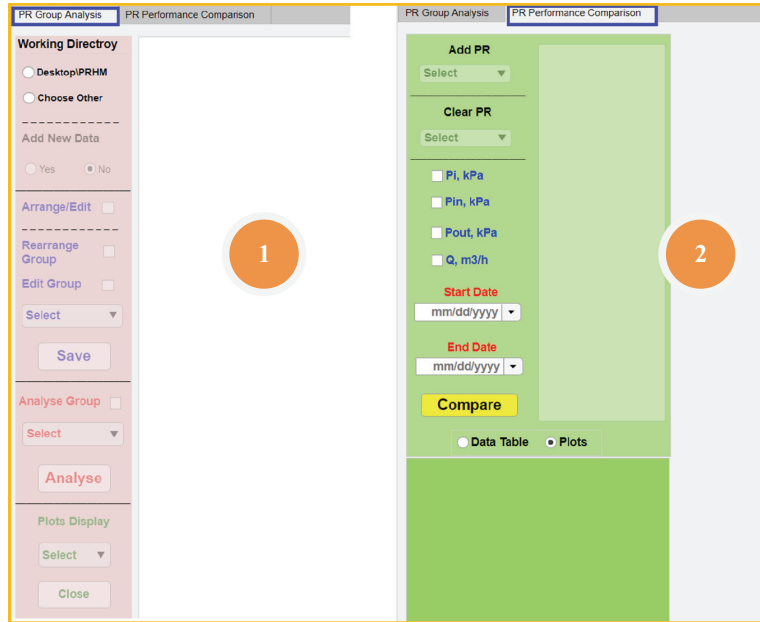


Figure 5: User interface of PRS health monitoring dashboard

4. Conclusions

The PRSs are key assets of a gas distribution network. This study compares the performance of first principles model and machine learning-based surrogate for some representative PRSs. It is demonstrated that weekly cumulative flow through the pressure regulators can be used to monitor the filter chocking and predict the RUL of the filter even if P_{in} data is not available. It is further shown that the PRS health monitoring dashboard under construction, if integrated with real time operational data of gas network, can monitor the filter health, and prescribe a predictive maintenance schedule. Other faults under investigation will be gradually included in the dashboard.

Acknowledgements

This research work is financially supported by the Energy Market Authority, Singapore (Project no - NRF2017EWT-EP003-020).

References

- M. B. Leo, A. Dutta, S. Farooq, I. A. Karimi, 2020, Simulation and health monitoring of a pressure regulating station, *Computers and Chemical Engineering*, 139.
- A. Nabi, J. Dayan, 2000, Dynamic model for a dome-loaded pressure regulator. *J. Dyn. Syst. Meas. Control. Trans. ASME* 122, 290–297.
- RMG by Honeywell. Gas Pressure Regulator Technical Datasheet, https://www.honeywellprocess.com/library/marketing/brochures/HON_280-284%20Brochure.pdf (accessed Sep 20, 2020).
- Emerson. The Industry Standard for Pressure Regulators <https://www.emerson.com/documents/automation/catalog-regulators-mini-catalog-fisher-en-125484.pdf> (accessed Oct 15, 2019).

Unsupervised Monitoring of Flocculation Processes based on Recurrence Theory

Hooman Ziaei-Halimejani^a, Nima Nazemzdeh^b, Reza Zarghami^{a,*}, Krist V. Gernaey^b, Martin Peter Andersson^b, Seyed Soheil Mansouri^b, Navid Mostoufi^a

^a*Multiphase Systems Research Lab, School of Chemical Engineering, College of Engineering, University of Tehran, Tehran 11155/4563, Iran*

^b*Department of Chemical and Biochemical Engineering, Søtofts Plads, Building 228A, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
rzarghami@ut.ac.ir

Abstract

Continuous monitoring of abnormal conditions during operation is an important requirement to increase the quality, and efficiency of chemical processes, and to optimize operating costs. In this study, fault diagnosis of abnormal conditions is considered for flocculation processes, which due to the complexity of these processes requires more attention. To this end, an unsupervised learning method is developed to diagnose the faults in chemical processes based on recurrence analysis. This method consists of two stages of pre-processing and clustering. The pre-processing stage is carried out by transferring the time series from time space to state space and converting the data into a two-dimensional recurrence plot. Quantitative parameters of recurrence analysis can be extracted from this plot. Then, in the clustering stage, the density-based spatial clustering of applications with noise (DBSCAN) method was used for clustering different operating conditions and diagnosing faults. By comparing the results with conventional methods, such as independent component analysis (ICA) and Kernel ICA (KICA) it was found that the developed method is more powerful and shows the best performance. Application of this method was illustrated throughout a laboratory scale flocculation of silica particles in water. An on-line non-invasive sampling method was used for monitoring the size distribution of particles with a dynamic image analysis sensor.

Keywords: Unsupervised learning, Recurrence plot, DBSCAN, Flocculation process, Fault diagnosis.

1. Introduction

Advances in bioprocessing equipment and instrumentation have increased the likelihood for diagnose of abnormal conditions while improving the process productivity and usability. Therefore, development of more powerful methods for monitoring the operating conditions of these processes is of particular importance. Process monitoring includes several stages of fault detection, diagnosis, and causal analysis. Fault diagnosis can be a significant help for operators and process engineers as it identifies the type of fault. Data-driven methods are widely used for this purpose since they do not need basic principles knowledge of the process (Tidriri et al., 2016). Among these methods, due to the frequent lack of information about the operating conditions, development of a method based on unsupervised learning is of greater importance. In general, any unsupervised fault

diagnosis method involves two general stages of pre-processing and clustering. In addition to the importance of clustering methods, the development and proper application of pre-processing methods is very important (Zheng and Zhao, 2020). The main research conducted in this field is based on the principal component analysis (PCA), independent component analysis (ICA) and other extensions such as Kernel PCA (KPCA), dynamic PCA (DPCA), and many more methods provided by (Liang et al., 2019, Bounoua and Bakdi, 2020). Sebzalli and Wang (2001) developed an approach based on PCA and fuzzy c-means clustering to classify the operational spaces of the RFCC reaction mechanism and grouped 303 cases in four clusters or operating zones. Also, Singhal and Seborg (2005) modified the *k*-means clustering algorithm with two similarity factors based on PCA and Mahalanobis distance to cluster time-series multivariate data from both batch and continuous chemical systems. However, despite the capability of time series analysis methods, only a few studies have been carried out in this field. Recurrence plots (Marwan et al., 2007) are among the powerful methods in the field of analysis and extraction of information from nonlinear, non-stationary and short time series (Ziaei-Halimejani et al., 2017). In this regard, Ziaei-Halimejani et al. (2021) conducted an unsupervised method using recurrence plot for fault diagnosis of Tennessee Eastman and four water tanks processes. They found that recurrence plot is a powerful method for isolating and capturing information of process time series. Accordingly, in this research recurrence plots were adopted for process monitoring of silica flocculation process. As a result, the main contributions of this research are: (i) development of a hybrid method of recurrence plot and density-based spatial clustering of applications with noise (DBSCAN) without the need for previous information about the operating conditions (unsupervised learning), and (ii) fault diagnosis of the silica flocculation process using the image processing technique

2. Experiment

Flocculation is a two-step aggregation phenomenon in which a large number of small particles form a small number of large flocs. In this study, a silica particle flocculation laboratory setup was employed to monitor the evolution of particle size distribution over time through data analysis (Nielsen et al., 2020). The particle size distribution was determined by the image segmentation algorithm developed by ParticleTech ApS (Farum, Denmark) (Nazemzadeh et al., 2020). As a result, time series of particle size distribution with particle area, minimum Feret size, maximum Feret size, mean Feret size, Feret ratio and other particle properties were obtained. The batch experiments were carried out for a range of process settings. Variation of the process settings affects particle interaction, which leads to variations in the flocculation rate. Accurate adjustment of the process settings improves the process efficiency, otherwise, abnormal conditions in the process may be encountered. The adjusted process settings are stirrer rotational speed and pH of the suspension. The stirrer rotational speed was adjusted by varying the voltage of the power supply of the stirrer and pH was varied by adding HCl (1 M) to move towards acidic conditions and NaOH (1 M) to move towards alkaline conditions. The particles used in this process were silica (acquired from Sigma-Aldrich, CAS no.: 7631-86-9) in demineralized water (0.015 wt% concentration). The suspension was then transferred to a stirring tank of 1.5 L. Also, a dynamic optical scanning device (oCelloscope by BioSense Solutions ApS, Farum, Denmark) was used to record images to monitor particle size distribution. The images were processed by ParticleTech Analyzer software, developed by ParticleTech ApS. In order to correctly diagnose different faults, six

different faults were artificially introduced into the process, which included step change, pulse and variation in the pH and stirrer rotational speed.

3. Theory

3.1. Recurrence Analysis

The Recurrence plot (RP) is a method based on the analysis of the recurrence/return behavior of time series (Eckmann, 1987). In this method, the time series is transferred from the time domain to the state space. By comparing all points of the time series trajectories in the state space and considering a radius threshold, recurrence plots can be constructed. The recurrence plot is obtained based on the following definition:

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - \|x_i^p - x_j^p\|) \quad i, j = 1, \dots, N \quad (1)$$

where N is the number of measured points x_i , ε is the radius threshold, Θ is the Heaviside function and $\|\cdot\|$ is the norm for measuring the distance between trajectories. If the difference between two trajectories is greater than ε , the Heaviside function is, $R_{i,j}(\varepsilon) = 0$. On the contrary, if the distance is less than ε , the Heaviside function becomes one, $R_{i,j}(\varepsilon) = 1$. For plotting this matrix, a black dot is shown in place of 0 and a white dot in place of 1. According to its definition, the RP is a symmetric plot and its main diagonal line is black. According to this definition, each point shows the recurrences of dynamical system states.

Complex structures in the recurrence plots include many information about the dynamic behavior of the time series. For example, diagonal lines indicate uniform and predictable behavior, and single points indicate noisy and random behavior. In addition to qualitative analysis, quantitative analysis of these patterns is required in order to accurately extract information from the time series. In this regard, the determinism (*DET*) is defined as the ratio of diagonal structures recurrence points to all recurrence points (Ziaei-Halimejani et al., 2018).

3.2. Clustering

Data clustering is a type of unsupervised classification in which a set of data is automatically placed in different clusters without any labelling. There are various methods in this field, most of which are statistical data analysis and data mining. Clustering methods are generally divided into five categories: hierarchical, partitional, distribution-based, density-based, and grid-based methods (Xu and Tian, 2015). In this study, the density-based DBSCAN method was used.

This algorithm, like other clustering methods, requires finding the proximity of data. Euclidean distance can be used to measure distance (similarity). To describe the algorithm, it is necessary to determine distance threshold (d) and minimum points of each cluster (*MinPts*). Each point in the dataset is different from other points. Any point for which the distance from a given point is less than d is considered a neighbor of that point. Also, any assumed point that has neighboring *MinPts* is a central point. In general, in the DBSCAN algorithm, a point is optionally selected that has not been visited before. The neighborhood of this point is checked with distance threshold (d). If it has the required minimum number of neighborhood points, a cluster is created; otherwise, it is labelled as a noise point. Note that after visiting all points, this point may be in the neighborhood of other points and become a part of another cluster (Saxena et al., 2017). In order to evaluate the performance of the proposed method, the adjusted rand index (*ARI*) was used, calculation of which is described by Hubert and Arabie (1985).

3.3. Method Framework

Fault diagnosis of the silica flocculation process using particle size distribution properties was carried out in the steps of the flowchart shown in Figure 1.

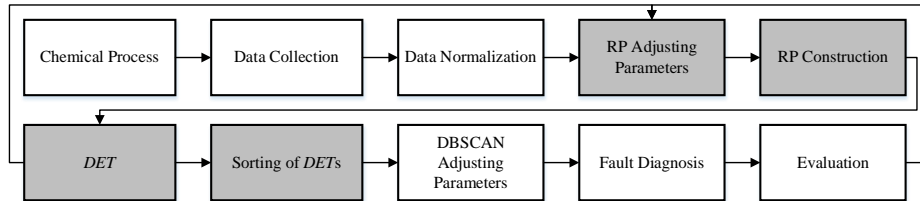


Figure 1. Overall workflow of the developed method for process fault diagnosis

4. Application on silica flocculation process

In order to diagnose faults associated with silica flocculation process, the particle size distribution time series was extracted from the reactor at normal operating conditions and used as input to proposed algorithm. In the pre-processing stage, after normalizing the data, the input parameters of the recurrence plot, including time delay (τ), embedding dimension (m), sub-series length (S_i) and radius threshold (ε) were determined. In this study, these values were calculated as 1, 2, 35 and 0.06, respectively. In this step, recurrence plots of time series were obtained. An example of recurrence plot for the particle size in the flocculation process during the normal operating condition is shown in Figure 2. The diagonal lines of recurrence plot correspond to the recurrence behavior of this process, at which its dynamics can be predicted by recurrence plots. In addition, the absence of the same patterns and the presence of single points in the plot indicate chaotic and nonstationary behavior. In order to extract this information in the form of quantitative index without any uncertainty, the determinism parameter was used.

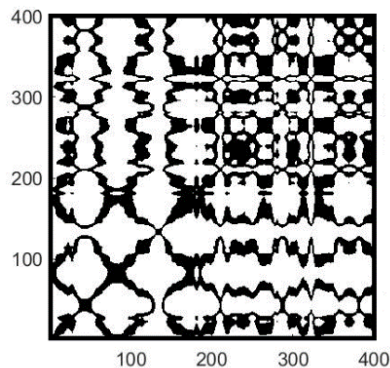


Figure 2. Recurrence plot of silica particle diameters during normal operating condition

After this step, by calculating the determinism of the RP and comparing it with normal operating condition values, three variables with the highest rate of change were selected and a three-dimensional graph was obtained. Figure 3 shows the geometric location of the three variables with the highest rate of changes for various faults of the silica flocculation process. As shown in this figure, various faults are isolated by values of their determinism in different geometric locations. It can be seen that the determinism has properly isolated various faults of the process in different geometric locations. However,

in order to properly understand the extent of this quality, the DBSCAN method was used for clustering of the obtained three-dimensional graph.

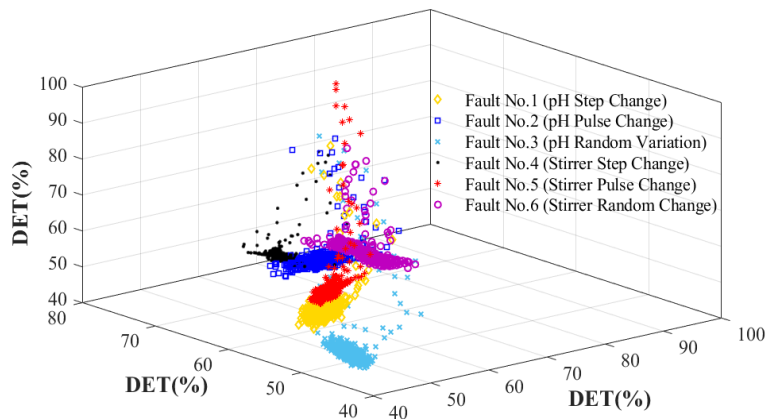


Figure 3. Geometric location of various kinds of faults in silica flocculation process

Also, the results of the clustering were evaluated with the previously available methods, which were displayed in the *ARI* criterion. The results for the *ARI* criterion were given in Table 1. In order to compare the results of the developed method with other available methods, the results of the ICA and KICA method were also given in Table 1. As shown in this table, the RP-DBSCAN method is a powerful method in fault diagnosis of the silica flocculation process.

Table 1. Evaluation criteria result of different fault diagnosis scheme for silica flocculation process

Method of Diagnosis	Evaluation Criteria (<i>ARI</i>)
ICA-DBSCAN	0.68
KICA-DBSCAN	0.73
RP-DBSCAN	0.85

5. Conclusions

Fault diagnosis of abnormal condition in chemical processes is one of the main phases of on-line monitoring, which requires more attention due to the very high complexity of these processes. Therefore, in this study, flocculation of silica particles was considered for fault diagnosis. To monitor the particle size distribution of particles with an image analysis device, an on-line sampling method was used. In the proposed method, by transferring time series from time domain to state space and converting them into a two-dimensional image of a recurrence plot, the pre-processing stage was accomplished. At the end, the time series was converted into quantitative recurrence parameters. Then, the DBSCAN method was used in the clustering stage to cluster various operating conditions and diagnose faults. It was found that the quality of the developed method stood at a high level and showed the best performance by assessing and comparing the results with conventional methods.

Acknowledgement

We would like to thank the Department of Chemical and Biochemical Engineering at the Technical University of Denmark, Novozymes A/S and Greater Copenhagen Food Innovation program (CPH-FOOD) for co-financing this research and also for their support during this project. National Elite Foundation of Iran is acknowledged for providing financial support for H.Z-H for research stay at the Technical University of Denmark.

References

- W. Bounoua and A. Bakdi, 2020, Fault Detection and Diagnosis of Nonlinear Dynamical Processes through Correlation Dimension and Fractal Analysis based Dynamic Kernel PCA, *Chemical Engineering Science*, 229, 116099.
- J. Eckmann, 1987, Recurrence Plots of Dynamical Systems, *Europhys Lett*, 4, 973-977.
- L. Hubert and P. Arabie, 1985, Comparing Partitions, *Journal of Classification*, 2, 1, 193-218.
- X. Liang, F. Duan, D. Mba and B. Ian, 2019, Centrifugal Compressor Diagnosis using Kernel PCA and Fuzzy Clustering, *Asset Intelligence through Integration and Interoperability and Contemporary Vibration Engineering Technologies*, 373-381.
- N. Marwan, M. C. Romano, M. Thiel and J. Kurths, 2007, Recurrence Plots for the Analysis of Complex Systems, *Physics Reports*, 438, 5-6, 237-329.
- N. Nazemzadeh, L. W. Sillesen, R. F. Nielsen, M. N. Jones, K. V. Gernaey, M. P. Andersson and S. S. Mansouri, 2020, Integration of Computational Chemistry and Artificial Intelligence for Multi-scale Modeling of Bioprocesses, *Computer Aided Chemical Engineering*, 48, 295-300.
- R. F. Nielsen, N. Nazemzadeh, L. W. Sillesen, M. P. Andersson, K. V. Gernaey and S. S. Mansouri, 2020, Hybrid machine learning assisted modelling framework for particle processes, *Computers & Chemical Engineering*, <https://doi.org/10.1016/j.compchemeng.2020.106916>.
- A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding and C.-T. Lin, 2017, A Review of Clustering Techniques and Developments, *Neurocomputing*, 267, 664-681.
- Y. Sebzalli and X. Wang, 2001, Knowledge Discovery from Process Operational Data using PCA and Fuzzy Clustering, *Engineering Applications of Artificial Intelligence*, 14, 5, 607-616.
- A. Singhal and D. E. Seborg, 2005, Clustering Multivariate Time-Series Data, *Journal of Chemometrics: A Journal of the Chemometrics Society*, 19, 8, 427-438.
- K. Tidriri, N. Chatti, S. Verron and T. Tiplica, 2016, Bridging Data-Driven and Model-based Approaches for Process Fault Diagnosis and Health Monitoring: A Review of Researches and Future Challenges, *Annual Reviews in Control*, 42, 63-81.
- D. Xu and Y. Tian, 2015, A Comprehensive Survey of Clustering Algorithms, *Annals of Data Science*, 2, 2, 165-193.
- S. Zheng and J. Zhao, 2020, A New Unsupervised Data Mining Method Based on the Stacked Autoencoder for Chemical Process Fault Diagnosis, *Computers & Chemical Engineering*, 135, 106755.
- H. Ziaei-Halimejani, R. Zarghami, S. S. Mansouri and N. Mostoufi, 2021, Data-Driven Fault Diagnosis of Chemical Processes Based on Recurrence Plots, *Industrial & Engineering Chemistry Research*, <https://doi.org/10.1021/acs.iecr.0c06307>.
- H. Ziaei-Halimejani, R. Zarghami and N. Mostoufi, 2017, Investigation of Hydrodynamics of Gas-Solid Fluidized Beds using Cross Recurrence Quantification Analysis, *Advanced Powder Technology*, 28, 4, 1237-1248.
- H. Ziaei-Halimejani, R. Zarghami and N. Mostoufi, 2018, Recognition of Particle Size Changes in Fluidized Beds by Recurrence and Cross Recurrence Quantification Analyses, *Industrial & Engineering Chemistry Research*, 57, 34, 11778-11784.

Optimal Supply Chain for Renewable Furfural Production Involving Economic, Environmental and Social Criteria

Gabriel Contreras-Zarazúa^a, Mariano Martin-Martin^b, José María
Ponce-Ortega^c, Juan Gabriel Segovia- Hernandez^{a*}

^a *Department of Chemical Engineering University of Guanajuato, Noria Alta
S/N,36000, Guanajuato, Gto., Mexico.*

^b *Department of Chemical Engineering, University of Salamanca, Plz. Caidos 1-5,
37008, Salamanca, Spain.*

^c *Department of Chemical Engineering, Universidad Michoacana de San Nicolás de
Hidalgo, Francisco J. Mújica SN, Ciudad Universitaria, Morelia, Mich, 58060, Mexico*

**Corresponding author. Email: gsegovia@ugto.mx*

Abstract

Biochemicals produced from lignocellulosic residues appear to be a feasible alternative to replace current chemicals derived from oil. Nevertheless, some important challenges need to be overcome in order to achieve a feasible transition towards biochemicals. These challenges are the biomass seasonality and availability, selection of suitable feedstocks, potential biorefineries allocation, as well as numerous economic environmental, social restrictions. Those challenges can be addressed through the design of a suitable supply chain. In this work a multi-period supply chain to produce furfural from agricultural wastes of Mexico is proposed, this model was written and optimized in GAMS. The demand of furfural was selected in order to replace the raw materials used to produce the terephthalic acid, which is imported to Mexico. Economic, environmental, and social criteria were chosen to evaluate the supply chain performance. Finally, the results obtained show that it is feasible the replacement of the current raw materials for producing terephthalic acid by furfural produced from agricultural wastes. The supply chain solution with the best tradeoff among the different metrics consists of a net profit of 1000 million USD/year, 19000 jobs generated and 370 million of eco-points /year.

Keywords: Furfural Production, Supply chain, Agricultural wastes, Social Impact, Biorefinery.

1. Introduction

Nowadays, the several problems associated with the climate change, pollution and the depletion of fossil resources have forced humanity to seek renewable resources and develop new technologies for using them. In this sense, the lignocellulosic residues

generated by agricultural activities are considered as promissory alternative to fossil resources due to their abundant availability, low costs and because they are not in competition with food, preventing ethical dilemmas. In the case of Mexico, it has 20 million of hectares destined for agricultural purposes, according with data provided by the Ministry of Agriculture and Fishing of Mexico (SAGARPA). This land generates about $5.86 \cdot 10^8$ metric tons of lignocellulosic residues each year, however, only the 5% of those residues are reused or leveraged. On the other hand, data from Ministry of Ecology and Climate Change (INECC) indicate that a significant amount of these agricultural wastes are burned in the harvest site generating around the 5% of the total CO₂ emissions of Mexico (INECC,2017; SAGARPA,2019).

Based on the aforementioned, the reutilization of these agricultural wastes for manufacturing new generation of bio-chemicals, bio-materials and fuels provides important and benefits since economic, environmental and social point of view. In this sense, the U.S. National Renewable Energy Laboratory (NREL) has catalogued the 30 most important chemicals produced from biomass and with the capability to compete with chemicals derived from fossil resources. In this list, the furfural stands out over other bio-chemicals because its several industrial applications for producing a wide range of commodities. Currently, the furfural is employed to produce lubricant oils, diols, solvents, furan resins and fuels (Cai et al., 2013). However, in recent years, novel furfural applications have been proposed, in this way, the manufacture of polymers and plastics to such as polyethylene terephthalate (PET), polyester or Nylon 6-6 has attracted special attention. Despite, the immense potential of bio-chemicals such as furfurals, some obstacles like biomass seasonality and availability, farmers planting decisions must be overcome in order to achieve a feasible transition towards the eco-friendly. These obstacles generate a not continuous supply of agricultural wastes and therefore an inconsistent production of bio-chemicals, as consequence several difficulties to satisfy the market demand are presented. These challenges can be overcome by an appropriate supply chain (SC) design where all those aspects are considered. A properly designed SC offers adaptability for providing feedstocks to a biorefinery and at the same time, it provides the ability to generate an optimal operational schedule and planning in order to satisfy the production required.

This work proposes the use of a supply chain model as a tool to determine the feasibility of producing furfural from agricultural wastes in Mexico at industrial scale. The demand for furfural is assumed to be the one that replaces the importation of chemicals required to produce terephthalic acid (TA). The TA is high added value chemical used in the manufacturing of polyethylene terephthalate (PET), polyester fibers, molding resins, adhesives. This study allows to determine the feasibility of replacing the current chemicals by renewable ones in the production of important current commodities. Based on sustainability criteria suggested by Garcia and You, (2015) the Economic, environmental, and social aspects were considered as the most important indicators to evaluate the supply chain solution. The economic objective is maximization of net profit, whereas the environmental impact and social objectives are the minimization of eco-indicator 99 and the maximization of the jobs generated, respectively.

2. Methodology

This work proposes the use of the four most abundant lignocellulosic residues produced each year in Mexico. The choice of raw materials was carried out according with data

provided by the Ministry of Agriculture and Fishing of Mexico (SAGARPA,2019). Based on the data provided by SAGARPA the most abundant lignocellulosic wastes produced each year in Mexico are corn stover, wheat straw, sorghum bagasse and sugarcane bagasse. In order to take into account, the biomass seasonality in each harvest site, a multiperiod inventory planning with one-year of time horizon is proposed. This horizon has been divided into the four different time periods ($t \in T$), one period for each season of the year. Figure 1 illustrates the superstructure considered for the supply chain. With the aim of providing a more realistic distribution of biomass, the Mexican territory was discretized into 59 different regions. This discretization was performed based on population data, where the zones with less population have larger lignocellulosic wastes availability, the biomass availability of each zone was fixed according with data provided by SAGARPA.

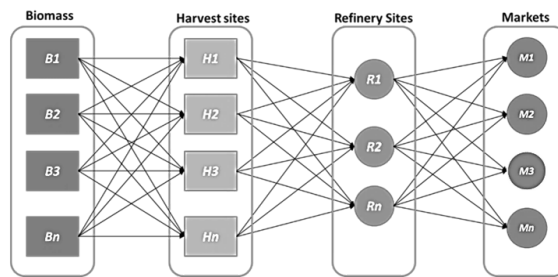


Figure 1. Supply chain superstructure.

In order to solve the problem in a simpler way some assumptions are considered. The centroid of each zone is the point where all the agricultural residues are located. This consideration could be easily relaxed only increasing the discretization zones to generate smaller sub-regions. Additionally, it is considered that the biomass is stored at the harvesting sites, which prevents the monetary losses provoked by the transportation of biomass that will not be used due to its degradation in the biorefineries. Finally, only one biorefinery per region is allowed, the biorefineries are characterized by significant use of water, this assumption avoids the excessive use of water and at the same time, it avoids the depletion of the total biomass located per region, which can be used for other purposes as cattle feed. The transportation distances and centroids locations are calculated using the Rhumb line method, which is a common method used in cartography to estimate distances. Furthermore, all the parameters required by this method can be easily obtained from Google Maps.

3. Mathematical Model

This section provides the mathematical formulation for the supply chain for furfural production and the corresponding objective functions used to evaluate the supply chain performance. The mathematical model was formulated as multiperiod Mixed Integer Lineal Programing (MILP).

The mass balances at the harvest sites can be modelled using the follow equation:

$$A_{i,j,t}^{RM} = A_{i,j,t-1}^{RM} (1 - \alpha_{j,t}) + b_{i,j,t}^{RM} - \sum_k S_{i,j,k,t}^{RM} \quad i, j \in J^{RM}, t \quad (1)$$

Where $A_{i,j,t}^{RM}$ and $A_{i,j,t-1}^{RM}$ are the inventory levels at a harvest site for a specific time and previous period, respectively. $\sigma_{i,t}$ represents the loss factor coefficient by biomass degradation, $S_{i,j,t}^{RM}$ is the shipments of biomass j from a harvest site i to a furfural plant k . Finally, $b_{i,j,t}^{RM}$ represents the lignocellulosic wastes required at a harvest site, it is calculated as follows:

$$b_{i,j,t}^{RM} \leq \varphi_{i,j,t} \quad i, j \in J^{RM}, t \tag{2}$$

Where $\varphi_{i,j,t}$ is the availability of lignocellulosic waste j at a harvest site i at a period t .

The shipments of lignocellulosic wastes from harvest to the furfural plants can be constrained as follows:

$$\sum_i b_{i,j,t}^L \cdot y_{i,j,k} \leq \sum_i S_{i,j,k,t}^{RM} \leq \sum_i y_{i,j,k} b_{i,j,t}^U \quad i, j \in J^{RM}, k \tag{3}$$

Where $b_{i,j,t}^L$ / $b_{i,j,t}^U$ represents the lower/upper bounds for the wastes required in a harvesting site, $y_{i,j,k}$ is the binary variable for selecting a specific shipment arc $i \rightarrow k$ (shipments from harvest to furfural plant).

In order to guarantee a continuous biomass supply to furfural plants, a cyclic inventory was used, which can be implemented assuming that the biomass storage in the last period ($t=T$) is equal to the storage at beginning period ($t=0$), according with the follow equation:

$$A_{i,j,t=T}^{RM} = A_{i,j,t=0}^{RM} \quad i, j \in J^{RM}, t \tag{4}$$

The mass balances for the furfural plants (biorefineries) were modeled in similarly. Additionally, three different criteria have been considered to evaluate the performance of the supply chain. These criteria are the net profit, environmental impact, which is calculated using the Eco indicator 99 and finally the jobs generated in the different activities of supply chain, the jobs generated are calculated using the Jobs and Economic Development Impact methodology (JEDI), more information about these objectives is provided by Santibanez-Aguilar et al., (2014). Mathematically, the three objectives function can be expressed as follows:

$$\min \quad Z = \{-\text{Profit}; EI99, -\text{Jobs}\} \tag{5}$$

Subject to:

$$\gamma^L \leq \sum_k S_{j,k,m,t}^P \leq \gamma^U \quad j \in J^P, m, t \tag{6}$$

The Eq.6 indicates as the shipments ($S_{j,k,m,t}$) are constrained between upper (γ^U) and lower bound (γ^L) respectively. Finally, this multi-objective optimization problem was solved using the ϵ - constrained method, in order to find a tradeoff among different objectives, more information about this method is provided by Diwekear (2008)

4. Results

The model proposed consists of 194,207 equations, 16,284 binary variables and 282,316 continuous variables. This model was solved using a computer with AMD Ryzen 5-1600 @3.2GHz, and 16GB of RAM @2400MHz. Each Pareto point was solved in an average time of 600s. The solver used was CPLEX with a relative gap of 1%. Figure 3 shows the Pareto fronts for all combinations of objectives. The form of these graphics can be easily explained based on Eq.6. Please note, that this equation restricts the solution between a lower and upper bound for the demand. The lower bound represents the demand required and the upper bound is the minimum demand plus an excess of 10%. Therefore, the extreme lines of the Pareto fronts correspond to different solutions operating between lower upper bounds. The average difference between the lower and upper bounded solutions is about 11% which is expected owing to the difference between bounds is 10%.

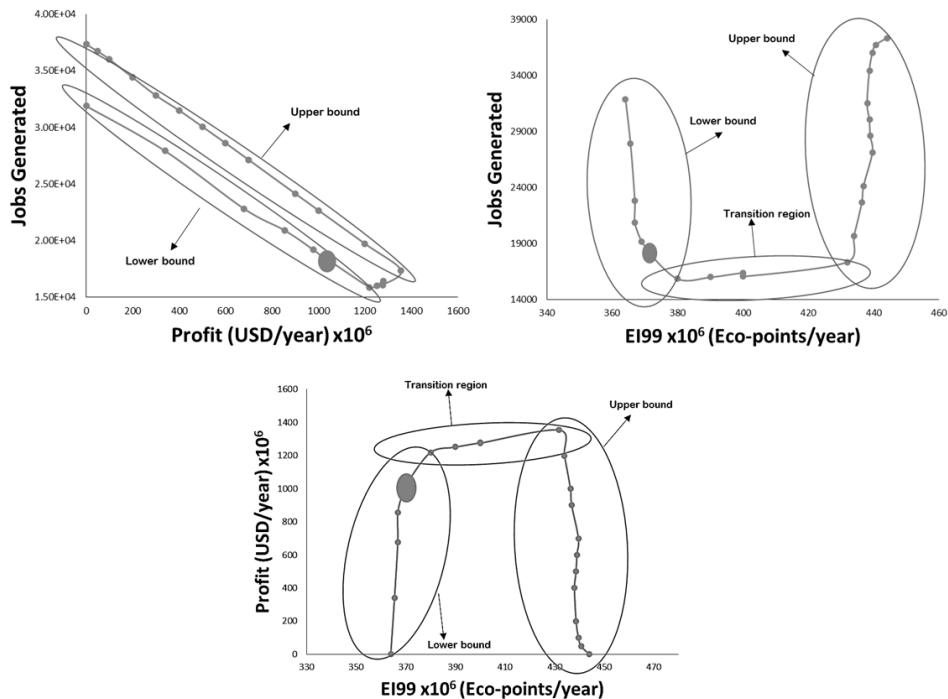


Figure 2. Pareto Fronts for furfural supply chain.

Please note, the large dots in Figure 2 correspond to the solution considered with most appropriate trade off among different metrics, thus this point was selected as the best SC solution. In this point the SC has around 75% of the maximum profitability, also the EI99 is close to its lower limit and the jobs generated are 19500 jobs, which signifies almost the 50% of the maximum possible jobs. It is important to highlight that this number of jobs is associated with different activities of the supply chain, such as recollection and distribution of raw materials and products, manufacturing in different furfural plants etc. In addition the profit was performed considering current funeral sale price of 2500USD ton. The biomass inventory in this point is given in Figure 3a. The results show that only corn stover and wheat straw are required to produce furfural due to the lower manufacturing cost than sorghum bagasse and sugarcane bagasse. Moreover, based on Figure 3a, it results evident that the production of furfural is based on corn stover principally, because the availability of wheat straw is less, especially during the period $t=1$ and $t=2$ (spring and summer). For this reason, significant amounts of corn stover must be stored in order to keep a constant furfural production.

Figure 3b illustrates the shipments of lignocellulosic wastes, during each time period. The results show a clear relation between the biomass availability and the shipments. During spring and summer ($t=1$, $t=2$), when the wheat straw availability is low the shipments are mainly based in corn stover. However, during autumn and winter seasons ($t=3$, $t=4$) the wastes shipments are based on wheat straw. These results also indicate that most furfural plants that use corn stover as raw material will be active during seasons in which corn stover is abundant, the same behavior occurs with wheat straw.

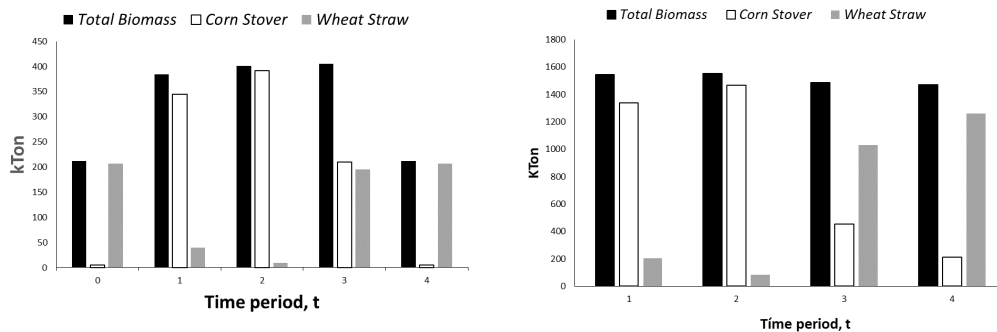


Figure 3. Biomass inventory and shipments levels for one-year time horizon.

5. Conclusions

In this work a supply chain model for furfural production is proposed. This mathematical model is used as a tool to determine the feasibility of producing furfural at industrial scale using as raw materials agricultural residues of Mexico. Economic, environmental and social aspects, where considered as indexes to evaluate the supply chain solution and to improve the sustainability of supply chain. Based on the results, the furfural production using agricultural residues to replace chemicals derived from petroleum is feasible. The results showed important profits up to 1250 Million USD/year. The optimal SC solution consist in a profitability of 1056 million USD/year, 19300 jobs generated and 370 million of eco-points /year.

Acknowledgements

Authors acknowledge JCYL SA026G19 and CONACYT

References

Ministry of ecology and climate change (IECC) (2020)

<https://www.gob.mx/inecc/prensa/presentacion-de-los-resultados-del-inventario-nacional-de-emisiones-de-gases-y-compuestos-de-efecto-invernadero?idiom=es> (Accessed 10 March,2020).

Ministry of Agriculture (SAGARPA), (2019)
<http://infosiap.siap.gob.mx/gobmx/datosAbiertos.php> (Accessed 10 July,2019)

Cai, C. M., Zhang, T., Kumar, R., & Wyman, C. E. (2013). Integrated furfural production as a renewable fuel and chemical platform from lignocellulosic biomass. *Journal of Chemical Technology & Biotechnology*, 89(1), 2-10.

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

Santibañez-Aguilar, J. E., González-Campos, J. B., Ponce-Ortega, J. M., Serna-González, M., & El-Halwagi, M. M. (2014). Optimal planning and site selection for distributed multiproduct biorefineries involving economic, environmental and social objectives. *Journal of cleaner production*, 65, 270-294

Diwekar, U. (2008). Introduction to applied optimization (Vol. 22). Springer Science & Business Media.

Levels of Approximation for the Optimal Design of Distributed Energy Systems

Ishanki A. De Mel^a, Oleksiy V. Klymenko^a, Michael Short^{a*}

^aDepartment of Chemical and Process Engineering, University of Surrey, Guildford, United Kingdom GU2 7XH

**m.short@surrey.ac.uk*

Abstract

Optimisation models for the design of distributed energy systems (DES) often exclude inherent nonlinearities and constraints associated with alternating current (AC) power flow and the underlying distribution network. This study aims to assess this gap by comparing the performance of linear and nonlinear formulations of DES design models, connected to and trading with an AC grid. The inclusion of the optimal power flow (OPF) constraints within the DES design framework is demonstrated in the methodology. A residential case study is used to test both models and compare the designs obtained from the two formulations. The results highlight that DES designs obtained are different when constraints related to the underlying distribution network are added, particularly when electricity storage is not considered. Overall, this study highlights the need for future modelling efforts to include OPF within DES optimisation frameworks to obtain practically feasible designs, rather than considering them as standalone problems.

Keywords: distributed energy system, optimal power flow, mixed-integer nonlinear programming

1. Introduction

The implementation of DES has become increasingly attractive as the power industry aims to integrate more renewable energy resources into existing networks while avoiding costly grid upgrades. DES consist of local, often privately-owned, small-scale energy resources, many of which are renewable. Optimisation models for design and operation of DES are commonly used to determine the suitability and location of the energy resources, and their performance over the system lifetime (Mavromatidis et al., 2019). However, these models often exclude inherent nonlinearities associated with AC power flow, despite most DES generating AC power and interacting with the main electricity grid. Power flow constraints, which are nonlinear and nonconvex, are included in OPF models which determine capacities of distributed generators and calculating active and reactive powers within the network, while maintaining voltage limits (Frank & Rebennack, 2016). OPF has often been treated as a standalone problem and not adequately accounted for in DES design formulations. Modellers often choose to exclude these nonlinear and nonconvex constraints in DES design models to obtain more computationally tractable models, and achieve an accuracy-complexity balance (De Mel et al., 2020). Mixed-integer linear programming (MILP) models are the most used formulations for DES design, with direct current (DC) approximations that exclude reactive power flows. However, the consequences of doing so have not been adequately analysed, with preliminary studies suggesting that either excluding or linearising certain key constraints can result in practically infeasible designs (Mashayekh et al., 2017;

Morvaj et al., 2016). This study addresses the increasing need to compare these different levels of approximation and establish which level is most suitable to achieving an accuracy-complexity balance and practically feasible designs.

2. Methodology

Four model formulations for DES design are employed in this study: 1) an MILP model with a general active power balance, based on works by Mehleri et al. (2012) for PV constraints and general power balances, and Mariaud et al. (2017) for battery representations and parameters, 2) a nonlinear programming model (labelled NLP-1) serving as a post-optimisation check with fixed binary decisions and capacities obtained from the MILP, and detailed AC power flow constraints, 3) an NLP model similar to model 2 (labelled NLP-2), however with capacities unfixed and treated as variables, and 4) a mixed-integer nonlinear programming (MINLP) model with detailed AC power flow constraints. All models minimise total annualised cost, which includes technology investment costs, operation and maintenance costs, and income from relevant tariffs. Each of the buildings comprising the DES can install one or more of the following distributed energy resources (DERs): solar photovoltaics (PVs), lithium-ion batteries, and boilers.

An OPF formulation based on the bus injection model in polar form (Frank & Rebennack, 2016) for a single-phase, balanced, low voltage network is utilised. It includes nonlinear and nonconvex constraints for net active (Eq. (1)) and reactive power (Eq. (2)), considering node $n \in \mathbf{N}$ and branch $(n, m) \in \mathbf{L}$:

$$P_{n,t} = V_{n,t} \sum_{m=1}^N V_{m,t} ((G_{nm} \cos(\theta_{n,t} - \theta_{m,t}) + (B_{nm} \sin(\theta_{n,t} - \theta_{m,t}))) \quad \forall n \in \mathbf{N} \quad (1)$$

$$Q_{n,t} = V_{n,t} \sum_{m=1}^N V_{m,t} ((G_{nm} \sin(\theta_{n,t} - \theta_{m,t}) - (B_{nm} \cos(\theta_{n,t} - \theta_{m,t}))) \quad \forall n \in \mathbf{N} \quad (2)$$

where the variables voltage magnitude and angle are denoted by $V_{n,t}$ and $\theta_{n,t}$, respectively. Note that parameters line conductance (G_{nm}) and line susceptance (B_{nm}) are obtained via bus admittance calculations (Frank & Rebennack, 2016). The OPF formulation also includes inequality constraints to ensure voltage magnitude and angles remain within specified lower and upper bounds. The layout of the network and connections are specified using a bus connectivity matrix for the set of all branches \mathbf{L} . Note that a node which injects power has positive power flow, while a node absorbing power has negative power flow. Current limitations are introduced in Eq. (3) where current magnitude in each branch, expressed as a function of voltage, cannot exceed the network-specified maximum current I^{max} (Frank & Rebennack, 2016):

$$\begin{aligned} & (V_{n,t} \cos \theta_{n,t} - V_{m,t} \cos \theta_{m,t})^2 + (V_{n,t} \sin \theta_{n,t} - V_{m,t} \sin \theta_{m,t})^2 \\ & \leq \frac{(I^{max})^2}{y_{nm}^2} \quad \forall (n, m) \in \mathbf{L} \end{aligned} \quad (3)$$

where y_{nm} refers to the magnitude of series branch admittance (which is usually calculated in complex form using the parameters series resistance (R_{nm}) and reactance (X_{nm}) for each branch). The OPF and MILP are linked by the following equality constraints for each house i connected to node n , forming the overall MINLP formulation for DES design (Frank & Rebennack, 2016):

$$P_{n,t} = P_{i,t}^{Gen} - P_{i,t}^{Load} \quad (4)$$

$$Q_{n,t} = Q_{i,t}^{Gen} - Q_{i,t}^{Load} \quad (5)$$

These determine whether the respective node is an injection or load node by calculating the net active and reactive powers, respectively. Note that the same OPF formulation is also used in the NLP models, with linear DES constraints and fixed binary decisions.

3. Results

3.1. Case study

The models are tested using the residential DES network available in Morvaj et al. (2016) in a UK context, using seasonal averaged values for electricity demand and heating demand as inputs (note that cooling is not considered in this study). Solar irradiance data obtained from Met Office (2020) are also fed as inputs to the model. Demands, available area for PV installation, and available volumes for battery installation at each location are summarised in Table 1. Note that the model considers 24 hours (one representative day) for each season, discretised into hourly timesteps. Other input parameters include DER efficiencies, capacity and area of a solar panel, and battery parameters such as maximum state of charge/discharge, volumetric energy density, etc. The DES is assumed to operate under the Feed-In-Tariff scheme (Ofgem, 2019) over 20 years, which provided prosumers with tariffs for renewable energy generation and export. Separate day and night prices are used for buying electricity from the grid to encourage battery installation.

Table 1. Average daily demands for electricity and heating per day

Building	Peak Electricity (kW)	Peak Heat (kW)	Area available (m ²)	Volume available (m ³)
A	3.8	10.0	150	5
B	18.4	47.6	700	5
C	14.1	27.1	600	5
D	3.8	6.4	150	5
E	12.0	31.3	550	5
<i>Total</i>	<i>52.2</i>	<i>122.3</i>	<i>2150</i>	<i>25</i>

The NLPs and MINLP are fed with network-related parameters, such as nominal voltage, resistance (R) and reactance (X), as presented in Figure 1. Note that the power factor (PF) for solar power is set to unity.

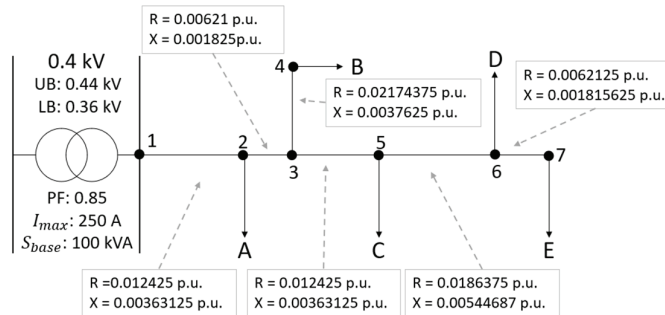


Figure 1. The distribution network and parameters, where S_{base} is the base apparent power. Two scenarios are tested 1) PVs and boilers only, and 2) PVs, batteries and boilers.

3.2. Results and discussion

The MILP, NLP and MINLP models are solved using CPLEX (IBM, 2019), CONOPT (Drud, 1985), and SBB (Bussieck and Drud, 2001), respectively, all on an Intel® Core™ i7-10510U processor. It is not possible to guarantee global optimality in the MINLP and NLP solutions, as the formulation is nonconvex and we experienced difficulties using global solvers. The statistics of the models (with all the technologies) are summarised in Table 2. The MINLP and NLPs are initialised using the MILP solution.

Table 2. Number of variables and constraints in the complete models employed in this study.

Model	Continuous variables (#)	Binary variables (#)	Constraints (#)
MILP	4,473	980	9,317
NLP	8,217	-	16,113
MINLP	8,217	980	16,133

The most notable differences in results were observed in Scenario 1 (PVs and boilers only), which are summarised in Table 3. While the MILP provides the best objective value and chooses the highest total PV capacity of 307 kWp, the post-optimisation check conducted via NLP-1 suggests that this solution is unattainable when grid constraints are considered. It is evident that curtailment of solar power is carried out in NLP-1 at nodes 1 and 2 in summer due to current violations in the network, which is also reflected by the lower income and thus an objective value approximately 6% greater than that of the MILP. This is further confirmed in Figure 2, which shows the current in branch (1, 2) (which carries the highest current out of all branches) for all four models. Both NLP-2 and MINLP opt to install lower PV capacities of 294 kWp and 287 kWp, respectively, to achieve lower objective values than NLP-1, while maintaining grid constraints. The capacities of boilers installed across all four models are the same (122 kWth), which is reflected in the boiler investment and operation costs.

Table 3. Solutions obtained from all four models for the PVs and boilers only scenario.

Breakdown	MILP	NLP-1	NLP-2	MINLP
Time taken (s)	5.29	5.57	5.84	6.86
Objective value (£)	43,793	46,380	46,158	46,102
Relative optimality gap	0	-	-	0
PV investment (£)	54231	54231	51825	50693
Boiler investment (£)	480	480	480	480
Battery investment (£)	0	0	0	0
Grid electricity (£)	37944	37944	38294	38448
PV operation (£)	5446	5382	5160	5055
Boiler operation (£)	5240	5240	5240	5240
Generation income (£)	50976	48938	47310	46498
Export income (£)	8572	7959	7532	7316

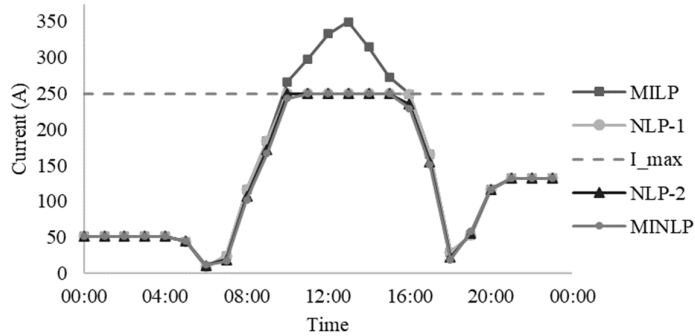


Figure 2. Current in branch (1, 2) calculated for all four models. Note that the MILP branch current was calculated via the NLP with current constraints deactivated.

Interestingly, in Scenario 2 where batteries are included in the model formulation, all four models produce the same objective value of £37,566. The total capacities for PVs (307 kWp), batteries (322 kWh), and boilers (122 kWth) across the models are the same. These results suggest that batteries allow the network to compensate for any potential violations by storing power rather than selling all excess power to the grid. Despite the capacities being equivalent, the operational strategies proposed in the models vary, particularly, the quantities of excess electricity and times it is sold as shown in Figure 3, and/or the batteries are charged. Note that the integrals of these variables across all timesteps and nodes are the same across all four models. In Figure 3, although it appears as if the excess power sold in the nonlinear models are bounded at ~100 kW, this is a result of including the current constraint (Eq. (3)) and the value varies between 102.8 kW and 103.1 kW in the models. Note that it is necessary to use an OPF formulation to determine the maximum amount of power that can be sold without violating this constraint for each case study, as it is sensitive to the design and parameters used. The MILP operational strategy is tested using a modified version of NLP-1, with fixed variables for selling electricity and charging the battery. This renders the model infeasible, confirming that the MILP strategy is unattainable without violating current constraints.

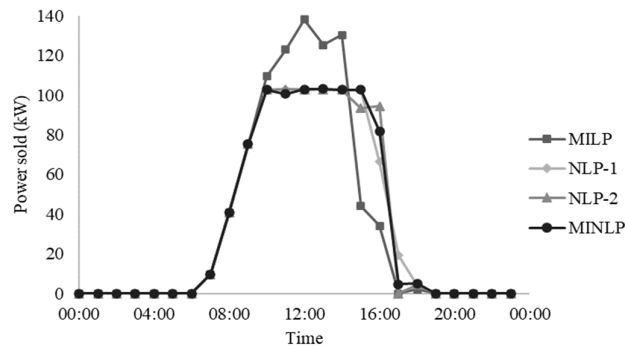


Figure 3. Total excess power sold at each time point.

Despite including nonconvex constraints, the models solved within seconds as highlighted in Table 3 (this included initialisations). However, computational expense may increase significantly if the network is enlarged to include more nodes, particularly if solved as an MINLP. In that case, using a model such as NLP-2, where an MILP is

solved first to obtain and fix binary decisions, may enable the modeller to achieve an accuracy-complexity balance without excluding network constraints.

4. Conclusions

This study focuses on consolidating two previously standalone problems, design of distributed energy systems and optimal power flow, and comparing results from linear and nonlinear model formulations for DES design. The methodology demonstrates the inclusion of nonlinear and nonconvex active and reactive power balances when formulating optimisation models for DES design, as well as other network-related constraints. Results highlight that inclusion of nonconvex constraints guarantees that the design is practically feasible and allows the inclusion of key network constraints that rely on both active and reactive power flow calculations. While the MILP model produces a better objective value for the presented case study when electricity storage is not considered, this is misleading as the inclusion of OPF constraints shows that this would not be realised in practice. Including these constraints also significantly impacts the operation of the DES, regardless of whether electricity storage is available or not, and therefore should be investigated further using DES operational models. Future work includes the extension of the nonlinear model to consider a larger, three-phase unbalanced network where voltage violations may also occur. Other network-related constraints such as line thermal limits and transformer limits will also be considered. Using sensitivity analysis to quantify the effect of uncertainties present in the system and therefore analyse the intermittence of renewable generation may further shed light on the significance of including network constraints.

References

- M.R. Bussieck, A. Drud, A., 2001, SBB: A new solver for Mixed Integer Nonlinear Programming.
- I.A. De Mel, O.V. Klymenko, M. Short, M., 2020, Balancing Accuracy and Complexity in Optimisation Models of Distributed Energy Systems and Microgrids: A Review, arXiv.
- A. Drud, 1985, CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems, *Math. Program.*, 31, 153–191.
- S. Frank, S. Rebennack, 2016, An introduction to optimal power flow: Theory, formulation, and examples, *IIE Trans.*, Institute Ind. Eng., 48, 1172–1197.
- IBM, 2019. IBM ILOG CPLEX Optimization Studio V12.9.
- S. Mashayekh, M. Stadler, G. Cardoso, M. Heleno, 2017, A mixed integer linear programming approach for optimal DER portfolio, sizing, and placement in multi-energy microgrids, *Appl. Energy*, 187, 154–168.
- G. Mavromatidis, K. Orehounig, L. A. Bollinger, M. Hohmann, J.F. Marquant, S. Miglani, B. Morvaj, P. Murray, C. Waibel, D. Wang, J. Carmeliet, 2019, Ten questions concerning modeling of distributed multi-energy systems, *Build. Environ.*, 165.
- E.D. Mehleri, H. Sarimveis, N.C. Markatos, L.G. Papageorgiou, 2012, A mathematical programming approach for optimal design of distributed energy systems at the neighbourhood level, *Energy*, 44, 96–104.
- B. Morvaj, R. Evins, J. Carmeliet, 2016, Optimization framework for distributed energy systems with integrated electrical grid constraints, *Appl. Energy*, 171, 296–313.
- A. Mariaud, S. Acha, N. Ekins-Daukes, N. Shah, C.N. Markides, 2017, Integrated optimisation of photovoltaic and battery storage systems for UK commercial buildings, *Appl. Energy*, 199, 466–78.
- Ofgem, 2019, About the FIT scheme [WWW Document], URL <https://www.ofgem.gov.uk/environmental-programmes/fit/about-fit-scheme> (accessed 3.18.19).

A Coordinated Framework for the Optimization of Municipal Solid Waste Management

Aurora del Carmen Munguía-López,^a Victor M. Zavala,^b José Ezequiel Santibañez-Aguilar,^c José María Ponce-Ortega^{a,*}

^a*Chemical Engineering Department, Universidad Michoacana de San Nicolás de Hidalgo, Francisco J. Mujica S/N, Ciudad Universitaria, Morelia 58060, Mexico*

^b*Department of Chemical and Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Dr., Madison 53706, USA*

^c*School of Engineering and Science, Tecnológico de Monterrey, Av. Eugenio Garza Sada 2501, Monterrey 64849, Mexico*
jmponce@umich.mx

Abstract

The lack of collection and processing infrastructure for waste management in developing countries is an overwhelming concern. Moreover, the growing generation of waste makes this problem even more complex. Specifically, in Mexico, the waste generation per day is approximately 0.13 MT. Landfill space is becoming increasingly constrained in some large urban centers where there are no other processing technologies, thus the use of open dump systems might be needed. The Mexican environmental protection agency (SEMARNAT) has reported that in 2012 from all the waste generated in Mexico, 72% was disposed of at sanitary landfills and regulated sites, while 23 % was disposed of at open dumps and the rest 5% was recycled. Considering these issues, we propose to optimize municipal solid waste management systems using a coordinated framework. The framework accommodates distinct stakeholders involved in the system, such as suppliers of waste, consumers of waste and derived products, and providers of transportation and processing services. The proposed optimization formulation seeks to maximize the collective profit and to balance supply and demand for waste and derived products. To achieve this, in the proposed framework, the stakeholders submit bids to a market coordinator that solves the optimization problem to find the allocations and clearing prices that guarantee that no stakeholder loses money. The framework also allows the monetization of environmental impacts, such as open dump disposal. Therefore, a taxation scheme to account for this environmental impact is evaluated. To illustrate the applicability of the presented approach, a municipal solid waste (MSW) system in Mexico was analyzed as a case study. Results show how taxation incentivizes the provision of services for all stakeholders. Furthermore, the minimum tax required to avoid diverting waste to open dump systems was identified.

Keywords: waste management, stakeholders, coordinated markets, taxation.

1. Introduction

MSW management is a complex problem that has been addressed through different approaches. Such as a taxation framework to incentivize recycling (Ko et al., 2020), a waste management cycle to guide policy regulations (Jiang et al., 2020), and mathematical models for the optimization of the MSW supply chain (See for instance:

Santibañez-Aguilar et al., 2013). Particularly in developing countries, the lack of infrastructure has led to great environmental issues including open dump disposal.

In the presented work, we propose to address the optimization of MSW management systems using a coordinated framework. We consider distinct options for the generated waste including recycling, sanitary landfills, and open dumps. Within the coordinated framework, the following stakeholders are identified: suppliers of waste, consumers of waste, consumers of products, providers of transportation, and providers of processing technologies. Specifically, in the MSW system, these stakeholders are urban centers that generate waste, sanitary landfills that consume waste, processing facilities that consume waste, urban centers that consume derived products, transportation providers that move waste or products, and transformation providers that treat different types of waste. We consider that the waste that is not allocated to consumers (sanitary landfills or processing facilities) ends up in open dumps. This problem is common in developing countries. Currently, there are not specific economic regulations associated with this practice. Therefore, to monetize this environmental impact, we include a taxation scheme. In this approach, we show how the coordinated framework allows identifying suitable tax structures and maximizing the collective profit of all stakeholders.

2. System Description

The stakeholders of the system and the proposed superstructure are presented in Figure 1. Here, we can identify the consumers, suppliers, and providers (transportation and transformation) that participate in the MSW system. Note that each of these stakeholders manages different types of waste or products at a particular geographical location (urban centers). The players are categorized by the type of waste they handle plastic, metal, organic, glass, and non-recyclables. Furthermore, subtypes are involved in some waste. For instance, for glass, we consider clear, green, and brown glass. We consider that each urban center involves a specific generation rate of waste and has available sanitary landfills, open dumps, and processing facilities. Furthermore, the transformation providers offer different types of treatment and technologies for each type of waste. The transportation providers can move waste to sanitary landfills and processing facilities. These providers can also move products to final consumers. Regarding open dump disposal, we observe that the waste that is not sent to sanitary landfills or processing facilities is sent there. It is assumed that this type of disposal does not involve any economic cost. However, the environmental cost is considered through the proposed taxation scheme. To avoid open dump disposal, we include a tax for the waste sent to such systems. Within the coordinated framework, this tax is considered as a service that the environment provides to the waste suppliers.

Into the coordination system, the suppliers, consumers, and service providers submit bids to an independent system operator (ISO). This coordinator uses the bidding information to clear the market by identifying the optimal allocations and clearing prices for all stakeholders. The next section describes these considerations.

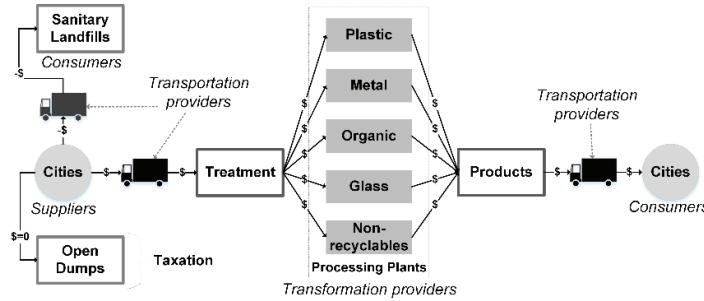


Figure 1. Superstructure for the coordinated system.

3. Coordinated Framework

The proposed framework guarantees that no cleared stakeholder loses money (payments collected are equal to payments made). This is possible because the solutions provided by the ISO satisfy a set of economic properties (Sampat et al., 2019). The solution procedure is as follows. First, the bids are submitted to the ISO (they can be positive or negative). In Figure 1, we can observe that almost all stakeholders offer positive bids (\$), however, the sanitary landfills offer negative bids (-\$). This negative bid refers to a payment that landfill suppliers are willing to give the market for taking away their waste. Furthermore, the negative bid of landfill consumers involves that the landfill will take the waste only if it is paid for this action (such as a disposal fee). Once this information is provided, the ISO solves an optimization problem to clear the market. Here, the collective profit of all stakeholders is maximized. As part of the clearing process, the allocations, prices, and profits that balance supply and demand are found. It is considered that, when a player is not cleared, no product is allocated, and this stakeholder does not participate in the market. The cleared stakeholders are paid based on their allocations and clearing prices. Specifically, transportation providers are paid considering the differences in prices at the source and destination locations. Similarly, the transformation providers are paid considering the prices of their input and output products.

4. Model Formulation

The formulation of the coordination problem involves geographical locations N , products P , consumers D , suppliers S , transportation providers K , and transformation providers M . The bidding information $(\alpha_d, \beta_s, \gamma_k, \delta_m)$ and the maximum capacities for the demand, supply, and service providers are given parameters. As shown in Eq. (1), the objective function seeks to maximize the collective profit. This profit is the difference between the demand served and the costs of supply, transportation, and transformation. The solution to the problem includes finding the optimal allocations (c_d, g_s, q_k, f_m) .

$$\max \sum_{d \in D} \alpha_d c_d - \sum_{s \in S} \beta_s g_s - \sum_{k \in K} \gamma_k q_k - \sum_{m \in M} \delta_m f_m \quad (1)$$

These allocations satisfy the physical conservation laws and capacity constraints of the following equations. The clearing prices are also part of the solution and they are estimated through the dual variables $(\pi_{n,p})$. These variables act as market clearing prices because they set values for products P at each geographical locations N . Here, $\zeta_{m,p}$ is the conversion factor for each technology and product.

$$\text{s.t } \sum_{s \in S_{n,p}} g_s - \sum_{d \in D_{n,p}} c_d + \sum_{k \in K_{n,p}^{\text{in}}} q_k - \sum_{k \in K_{n,p}^{\text{out}}} q_k + \sum_{m \in M_n} \zeta_{m,p} f_m = 0, \quad (2)$$

$$(n, p) \in N \times P, (\pi_{n,p})$$

$$0 \leq c_d \leq c_d^*, \quad d \in D \quad (3)$$

$$0 \leq g_s \leq g_s^*, \quad s \in S \quad (4)$$

$$0 \leq q_k \leq q_k^*, \quad k \in K \quad (5)$$

$$0 \leq f_m \leq f_m^*, \quad m \in M \quad (6)$$

These allocations and prices are used to charge consumers and remunerate providers. For the clearing prices of each stakeholder, we use the notation $\pi_d, \pi_s, \pi_k, \pi_m$. These prices along with the bids and the allocations are used to compute the profits of stakeholders as follows. Eq. (7) refers to the profit for consumers that is the difference between the allocated demand ($\alpha_d c_d$) and the payment made to the market ($\pi_d c_d$). For suppliers, their profit is estimated as shown in Eq. (8). Here, $\pi_s g_s$ refers to their revenue and $\beta_s g_s$ represents their operating cost. Similarly, the profits for the transportation and transformation providers are computed by Eq. (9) and Eq. (10).

$$\phi_d^D = (\alpha_d - \pi_d) c_d, \quad d \in D \quad (7)$$

$$\phi_s^S = (\pi_s - \beta_s) g_s, \quad s \in S \quad (8)$$

$$\phi_k^K = (\pi_k - \gamma_k) q_k, \quad k \in K \quad (9)$$

$$\phi_m^M = (\pi_m - \delta_m) f_m, \quad m \in M \quad (10)$$

The resulting model is linear and was solved through the GAMS modeling environment using the solver CPLEX.

5. Results and Discussion

To illustrate the applicability of the approach, we evaluate a case study of an MSW system in the central-west region of Mexico. Here, five urban centers (Morelia, Celaya, Apatzingan, Lazaro Cardenas, and Leon) are considered. These centers act as suppliers and consumers. We use the notation $l-5$ to refer to where each stakeholder is situated. Each stakeholder has a specific flow, product type, capacity, location, and bidding cost. Also, we assume that 1) the urban centers have equal technologies available for treatment, 2) the transportation bids exist to move the waste and products between locations, and 3) landfill suppliers are willing to pay for the service of taking away their waste. For the

taxation scheme, the minimum tax required to avoid open dump disposal was identified and evaluated for the case study. To compare the impact of the taxation, two scenarios were analyzed. We consider the Scenario I): a base case without taxation in which the impact of open dumps is ignored, and the Scenario II): a tax of 5.1 USD/t is applied to the waste disposed of at open dumps. We found that the collective profit of the system is 871,744 USD and 784,061 USD for Scenarios I and II, respectively. Note that when the environmental impact of open dump disposal is penalized the profit decreases. For the plastic, organic, glass, and non-recyclable waste, we found that no transformation providers participate (for both scenarios). Therefore, the suppliers (*S1-S5*), consumers (*D1-D5*), and transportation providers (*K1-K5*) refer to the waste sent to the landfill. For Scenario I, there is no waste sent to the landfills. On the other hand, for scenario II, the waste is sent to sanitary landfills and open dump disposal is avoided because of the taxation. Only for metal waste, the processing providers are cleared. Thus, part of this waste is sent to treatment. In the following, we present some of the obtained results. Figure 3 shows the profits for glass waste. The types of glass are denoted by *G1*, *G2*, and *G3*. Here, we observe that the transportation providers attain the smallest profits. On the contrary, the landfill suppliers make the highest profits. We can see that stakeholders *S1* and *K1* attain the highest benefits of the suppliers and providers, respectively. For the consumers, the stakeholder *D5* makes the highest profit. Note that all profits are positive. Regarding the types of glass, *G3* (corresponding to brown glass) represents most of the total profit. These results are similar in behavior to the solutions for plastic, organic, and non-recyclable waste since the same stakeholders are cleared. The specific profits vary due to the involved flows, bids, and prices. However, through the presented results we intend to illustrate the type of solutions that can be obtained using the coordinated framework. For metal waste, there is an important variation in the solutions since the transformation providers participate here (see Figure 4). Here, we observe the profits for involved stakeholders. As expected, the profits of the plant supply are the highest. We can see that only one plant consumer is cleared (*D1*). For the transportation providers, the profits are greater for Scenario I because no metal is sent to the landfill. For the processing providers, only the stakeholder *M1* is cleared. The profit of this stakeholder is 1,070,035 USD for both scenarios. Note that the profits for the transformation providers do not change with the taxation scenario since the tax is not involved on the recycled waste.

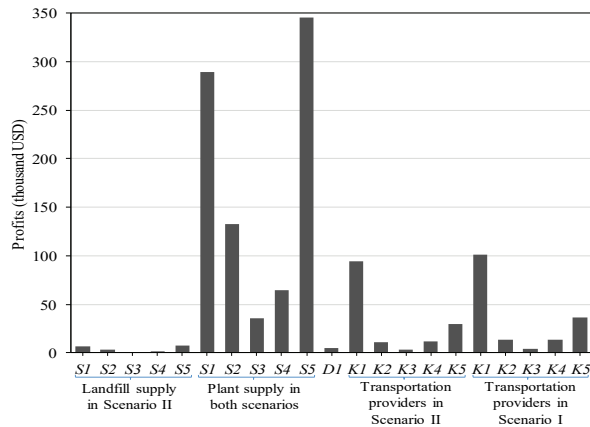


Figure 2. Profits for supply, demand, and transportation providers of metal for Scenarios I and II.

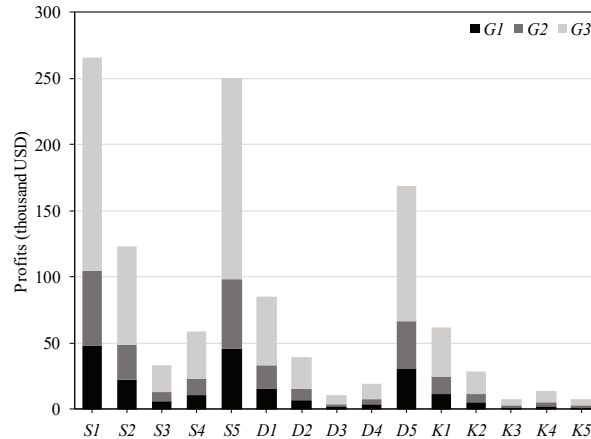


Figure 3. Profits for landfill supply, demand, and transportation providers by types of glass in Scenario II.

6. Conclusions

This work presented a coordinated framework for multiple stakeholders in MSW systems. We considered the waste sent to treatment, sanitary landfills, and open dumps. A taxation scheme was included to account for this environmental impact. We analyze two different scenarios (with and without taxation). The results show how the tax incentivizes stakeholders to avoid open dump disposal. Furthermore, the optimal prices and allocations for the suppliers, consumers, and services providers of all types of waste were found. The collective profit was maximized, and the individual profits were identified (all of them non-negative). We evaluated a case study to show the applicability of the formulation, however, the model can be applied to any case study.

References

- P. Jiang, Y. Van Fan, J. Zhou, M. Zheng, X. Liu, J.J. Klemeš, 2020, Data-driven analytical framework for waste-dumping behaviour analysis to facilitate policy regulations, *Waste Management*, 103:285-95.
- S. Ko, W. Kim, S.C. Shin, J. Shin, 2020, The economic value of sustainable recycling and waste management policies: The case of a waste management crisis in South Korea, *Waste Management*, 104:220-7.
- A.M. Sampat, Y. Hu, M. Sharara, H. Aguirre-Villegas, G. Ruiz-Mercado, R.A. Larson, V.M. Zavala, 2019, Coordinated management of organic waste and derived products, *Computers & Chemical Engineering*, 128:352-63
- J.E. Santibañez-Aguilar, J.M. Ponce-Ortega, J.B. González-Campos, M. Serna-González, M.M. El-Halwagi, 2013, Optimal planning for the sustainable utilization of municipal solid waste, *Waste management*. 33(12):2607-22.

A Probabilistic Scenario Generation Framework for Optimal Decision Making in Turkish Renewable Energy Market

Hasan Sildir^a, Handan Akulker^{b,c} and Erdal Aydin^{b*}

^a*Department of Chemical Engineering, Gebze Technical University, Kocaeli 41400, Turkey*

^b*Department of Chemical Engineering, Bogazici University, Bebek, Istanbul 34342, Turkey*

^c*Department of Chemical Engineering, Ondokuz Mayıs University, Samsun 55139, Turkey*
eaydin@ku.edu.tr

Abstract

Turkey is one of the richest countries in terms of renewable energy resources. At the same time, the largest portion of the account deficit of Turkey is due to energy import. Optimization studies for design, integration and management of renewable energy is therefore crucial in terms of increasing overall energy efficiency. In addition, energy sources and demands in Turkey have significant uncertainty and meeting the market conditions in a profitable manner is a challenging task. Stochastic programming is an efficient approach to deal with the aforementioned challenge. It requires introducing representative and comprehensive scenarios for the optimal design and scheduling. In this study, the aim is to propose a systematic and generic scenario generation method which is compatible with historical data, dependable for forecasts, and easily tunable for the scenario likelihood. Main contributions of this work are as follows: The uncertainty in the model parameters are propagated to the forecasts to obtain prediction intervals under a desired confidence, which provides probabilities of each scenario over the whole time horizon with predetermined likelihood the generated scenario set (e.g., likely, rare or statistically very low probability). Thus, scenario reduction is not needed. We implemented our method for Yalova, a developing city in Turkey, for the scenario generation of wind speed, population, air temperature, electricity consumption and solar irradiance, with a prediction horizon of 20 years. We also developed a preliminary mixed-integer linear programming (MILP) decision making model which computes both the optimal equipment investments and the optimal sub-hourly scheduling sequences of the equipment based on these scenarios and economic considerations.

Keywords: energy systems integration, renewable energy, scenario generation, mixed-integer linear programming, stochastic programming.

1. Introduction

Optimal design and management of energy systems must take uncertainties and volatilities into account. Deterministic models represent only the most likely behavior in contrast to stochastic models where the uncertainties due to varying demands, market conditions and other variables are considered additionally. In other words, random distributions with problem specific tunable parameters are incorporated to the

optimization in order to introduce the probabilistic nature of energy systems (Zakaria et al. (2020)).

In stochastic optimization, scenario tree construction is a major approach in which a particular distribution function is discretized. In this approach, different probabilistic scenarios are generated with predetermined and limited branches in order to deal with computational load. Monte-Carlo sampling, optimal quantization and moment matching are prominent methods based on statistics for scenario tree construction (Löhndorf (2016)). In Monte-Carlo sampling, a uniform probability random sampling is done over the region (Ekblom and Blomvall (2020)). On the other hand, in optimal quantization, a distribution function is estimated by optimization whilst in moment matching method, optimization is performed by considering the higher order moments because first few moments do not represent the whole distribution (Löhndorf (2016); Høyland et al. (2003)).

In this study, empirical time series models with tunable parameters are formulated for the air temperature, solar irradiance, wind speed, population and energy demand. For the prediction intervals of associated variables, the covariance matrix of the model parameters are calculated. Those prediction intervals having a particular probability distribution function are used for the generation of scenarios based on their likelihood. The highlights of the study are summarized as: Empirical models for the instantaneous air temperature, electricity demand, solar irradiance and population prediction are obtained through training actual historical data. Using parameter uncertainty propagation, covariance matrices and prediction intervals of prediction variables are calculated. Sampling for scenario generation is performed over these predefined intervals eliminating the need for scenario reduction to avoid additional computational effort. In addition, parameter uncertainty propagation builds a wider domain for scenario generation that is expected to increase the optimal design's robustness by taking account of both the most likely representing usual operating conditions and the most unlikely cases representing extreme conditions such as sharp changes in the load demand or weather conditions. The proposed method is implemented on the optimal design of a hybrid renewable energy micro grid to be placed in Yalova, which is a developing city in Turkey. This is a pioneer and sole case study in which real historical Yalova data are used to generate scenarios based on uncertainty propagation (Yilmaz and Dincer (2017); Turkdogan (2020); Talebi et al. (2016)).

2. Theoretical Background

2.1. Formulation of Empirical Models

All wind speed, solar irradiation and ambient temperature data are obtained from real data using the Darksky API (API (2020)) and electricity consumption data were taken from UEDAS, the official electricity distributor of Yalova. The empirical parameters related to wind speed, solar irradiance, electricity demand etc, with index j , are calculated by using nonlinear optimization as the following:

$$\min_{p_j} \sum_{i=1}^N \| f_j(t_i, u_i, p_j) - y_i \| \quad (1)$$

where N is the number of measurement samples; p_j is the vector of parameters in equation j ; t_i is the time instant of measurement; u_i is the input to the equation; y_i is the measurement; f_j is the formed empirical equation for the prediction of y_i . Corresponding empirical time series models are obtained by parameter estimation using historical data. Formulated empirical models for daily and instantaneous air temperature and solar irradiance show

sinusoidal trend while wind speed model exhibits Weibull distribution. For brevity, only electricity demand and population models are given explicitly in the results part. The daily electricity demand is a function of population and time:

$$E_{daily} = p_{e,daily,1} \sin\left(\frac{2\pi}{360}(t + p_{e,daily,2})\right) + p_{e,daily,3} + p_{e,daily,4} P_t \quad (2)$$

where P_t is the population of the city at time t . The daily electricity demand demonstrates both seasonal and hourly differences by the presence of sine function. P_t is determined through:

$$P_t = P_0 + P_{p,1}t + P_{p,2}t^2 \quad (3)$$

where P_0 is the reference population. The hourly electricity demand is calculated from:

$$E_h = (p_{E,h,1}t_h^4 + p_{E,h,2}t_h^3 + p_{E,h,3}t_h^2 + p_{E,h,4}t_h)E_{daily} \quad (4)$$

2.2. Uncertainty Propagation

Parameter covariance matrix is a measure of variations from the mean parameter values. Different parameter samples which result in different prediction profiles are obtained after parameter combinations are generated with Cholesky decomposition based methods. By this way, ultimate uncertainty is calculated considering parameters' probable variations (Chakraborty (2006)). Taylor series expansion using only first order term is utilized for uncertainty propagation in this study (Tellinghuisen (2001)):

$$cov_y = Jcov_pJ^T \quad (5)$$

cov_y is the output covariance matrix; cov_p is the parameter covariance matrix; J is the Jacobian matrix evaluated at the desired prediction regime. cov_y is a measure of prediction uncertainty (Borenstein et al. (2011)) to be considered in scenario generation by participating in upper and lower bound of prediction interval as follows (Lane and Dumouchel (1994); MathWorks (2020)):

$$f_j(t_i, u_i, p_j) - \lambda\sqrt{cov_y + MSE} \ll y_i \ll f_j(t_i, u_i, p_j) + \lambda\sqrt{cov_y + MSE} \quad (6)$$

λ , which is the coefficient of standard deviation of the mean value of i^{th} output variable, depends on the confidence level and is calculated through the inverse of Student's t cumulative distribution function. MSE is the mean squared error coming from the Eqn 1.

3. Scenario Generation

Inverse transform sampling is performed (Sugiyama (2015)) over a predefined interval as shown in the Eqn 6 to create the scenarios under their desired likelihood. This sampling provides to manage the probabilities of scenarios such as the most likely or the most unlikely. For example, a likely scenario which has one standard deviation interval is sampled more closely to the mean rather than the tails of the probability distribution function. On the other hand, a mid-likely and unlikely scenario which are sampled more closely to the tails than the mean have two and three times standard deviation interval, respectively.

4. Results

Despite its small area, Yalova is an important city with a large population density and open to development due to its geographical and industrial location. It is surrounded by big industrial cities such as Istanbul, Bursa, and Kocaeli and there are ferry lines with these three cities having major industrial and financial companies. The climate of the city is very suitable for growing agricultural products with high commercial value. Because of all these reasons, Yalova is the city that has increased its share in the gross national product the most in the last ten years.

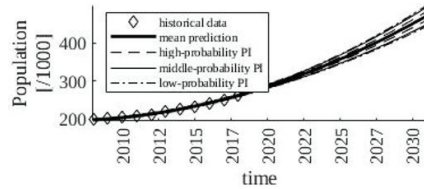


Figure 1a: Population PIs

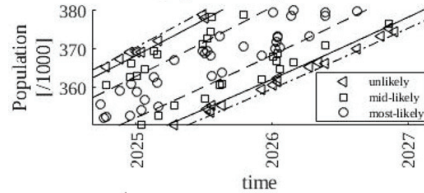


Figure 1b: Population scenarios

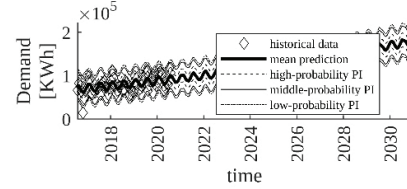


Figure 2a: Electricity demand PIs

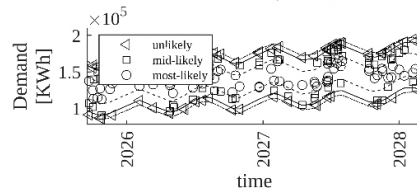


Figure 2b: Electricity demand scenarios

Figure: 1. Population prediction intervals and scenarios. 2. Electricity demand prediction intervals and scenarios.

As shown in Fig.1.a and 2.a, since the cumulative uncertainty propagates along with time, range of the prediction interval becomes wider while time horizon is extending. Fig. 1.b and 2.b show the generated scenarios over a constrained time interval because when the scenarios are drawn over the whole time period, the figures look unclear. In the figures, the scenarios sampled from the unlikely, mid-likely and most-likely prediction interval are indicated instead of the whole interval for the representation of the distribution.

5. Case Study

We developed a deterministic MILP decision making model whose objective is to maximize the profit of the design and management of an integrated renewable energy system grid for the city of Yalova, for the upcoming 20 years. This model computes the optimal equipment investment decisions together with their rated powers and capacities, and the optimal sub-hourly scheduling sequences of these equipment. Exact linearization is applied on the nonlinear unit commitment constraints. The model also takes into account spinning reserves, peak shaving, time value of money as correction factor, piecewise generation equation formulations for both solar and wind energy and strict emission limits for CO₂ (Mavromatidis et al. (2018)).

The objective function of the model is written as:

$$\max f^{\text{revenue}} - CRF f^{\text{initial}} - \sum_{k \in K} J_k^{\text{operational}} \quad (7)$$

where $f^{revenue}$ is the income from sales, CRF is capital recovery factor, $f^{initial}$ is the initial cost representing the fixed and variable costs for the investment, $f_k^{operational}$ is operational cost for year k , and K is investment period. The deterministic MILP model is formulated to supply all of the electricity demand of the city of Yalova and it is assumed that the selling price of the supplies of the system are 5 percent less than the usual price. All wind speed, solar irradiance, ambient temperature, population and electricity demand data were collected as mentioned in Section 2. Afterwards, these data are fed to the suggested scenario generation framework to obtain the required future predictions as a scenario. The decision making model is coded into the GAMS language and solved with the CPLEX solver. 32 different equipment, including both renewable and non-renewable, are characterized as candidates for the superstructure. As expected, optimal solution consists of only a subset of these equipment. Optimal equipment decisions are shown in Table 1 with the corresponding rated power values.

Table 1: Installed equipment in the optimal solution of model

Installed Units	Rated Power (kW)	Capacity (kWh)
Wind Turbine-1	200000 (upper limit)	
Wind Turbine-2	200000 (upper limit)	
Wind Turbine-3	29278.96	
Battery-1	69616.24	199019.49
Heat Pump	2006.9	
Reciprocating Internal Combustion Engine-3	9341	
Fuel Cell-1	1400	

Optimal profit is found as 13.01 billion \$ for 20 years plant life. It is observed that the wind turbines are chosen over PV units. Although they have very similar cost and capacity values, photovoltaic cells are not suggested to be installed. The reason is related to the operating power of PV units which depends on temperature and UV lights. Furthermore, the surplus electricity penetration of wind during night time brings about larger profit margins by storing the electricity in the battery. Accordingly, in daytime where the electricity prices are relatively expensive, decision making model prefers to choose the battery to supply electricity instead of buying it from the national grid. In addition to the suggested renewable energy equipment, combined heat and power generators, which produce both electricity and heat at certain ratio, are also included into the optimal configuration by the MILP model. The return on investment of such a project is around 8-9 years. On the other hand, we should mention that this study is mainly deterministic and more reliable results could be obtained using stochastic programming framework.

6. Conclusion

In conclusion, in this study, empirical mathematical models including the uncertainty of parameters are formulated and validated with historical data from Yalova, Turkey. One of the most significant feature of this study is the propagation of uncertainty with covariance matrix of parameters, which are then used to determine prediction intervals having particular distribution functions. Thus, the proposed method allows to generate scenarios with user-defined probabilities. In addition, defining predetermined intervals

removes the need for scenario reduction causing extra computational effort and time. Moreover, the model is able to provide independent scenarios for each time interval. On the other hand, there is a restriction with this method that high number of empirical models and parameters may deliver different results due to various possible future forecast behavior. Proposed method is open to adding extra system parameters such as economic conditions or unit efficiency parameters. This flexibility may generate a wider range of scenarios. Furthermore, the method may also result in broad parameter uncertainty regions to capture the wider prediction uncertainty. Handling the impact of uncertainty is a critical issue for energy system design. Finally, we also plan to integrate expert knowledge to the scenario generation method to forecast the future costs of the renewable energy systems reliably. We will employ the proposed scenario generation method for stochastic studies for Yalova and Turkey. Created scenarios with different probabilities will be fed into the stochastic version of the decision making model to assess the value and contribution of the stochastic programming to deal with uncertainty.

7. Acknowledgement

This publication has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C245). However, the entire responsibility of the publication belongs to the owner of the publication. We thank Uludag Electricity Distribution Company (UEDAS) for sharing the real electricity consumption data for Yalova. We also thank Su Meyra Tatar for her support.

References

- D. API, 2020. Darksy api. Last accessed 16 August 2020. URL <https://darksy.net/dev>
- M. Borenstein, L. V. Hedges, J. P. Higgins, H. R. Rothstein, 2011. Introduction to meta-analysis. John Wiley & Sons.
- A. Chakraborty, 2006. Generating multivariate correlated samples. *Computational Statistics* 21 (1), 103–119.
- J. Eklblom, J. Blomvall, 2020. Importance sampling in stochastic optimization: An application to intertemporal portfolio choice. *European Journal of Operational Research* 285 (1), 106–119.
- K. Høyland, M. Kaut, S. Wallace, 2003. *Comput. Optim. Appl* 24, 2–3.
- T. P. Lane, W. H. Dumouchel, 1994. Simultaneous confidence intervals in multiple regression. *The American Statistician* 48 (4), 315–321.
- N. Löhndorf, 2016. An empirical analysis of scenario generation methods for stochastic optimization. *European Journal of Operational Research* 255 (1), 121–132.
- MathWorks, 2020. Confidence and prediction bounds. Last accessed 16 August 2020. URL <https://www.mathworks.com/help/curvefit/confidence-and-prediction-bounds.html>
- G. Mavromatidis, K. Orehounig, J. Carmeliet, 2018. Design of distributed energy systems under uncertainty: A two-stage stochastic programming approach. *Applied energy* 222, 932–950.
- M. Sugiyama, 2015. Introduction to statistical machine learning. Morgan Kaufmann.
- S. Talebi, A. F. Ariza, T. V. Nguyen, 2016. High-level multi-objective model for microgrid design. In: 2016 Annual IEEE Systems Conference (SysCon). IEEE, pp. 1–8.
- J. Tellinghuisen, 2001. Statistical error propagation. *The Journal of Physical Chemistry A* 105 (15), 3917–3921.
- S. Turkdogan, 2020. Design and optimization of a solely renewable based hybrid energy system for residential electrical load and fuel cell electric vehicle. *Engineering Science and Technology, an International Journal*.
- S. Yilmaz, F. Dincer, 2017. Optimal design of hybrid pv-diesel-battery systems for isolated lands: A case study for kilis, turkey. *Renewable and Sustainable Energy Reviews* 77, 344–352.
- A. Zakaria, F. B. Ismail, M. H. Lipu, M. A. Hannan, 2020. Uncertainty models for stochastic optimization in renewable energy applications. *Renewable Energy* 145, 1543–1571.

Dynamic optimization modelling for a national-scale energy system in transition

Xiang Li^{a,*}, Dario Müller^a, Subhash Kumar^a and François Maréchal^a

^a*Ecole Polytechnique Fédérale de Lausanne, Rue de l'industrie 17, Sion 1951, Switzerland*

xiang.li@epfl.ch

Abstract

Despite numerous energy system optimization models for long-term planning on a national level, the plausibility assessment of suggested pathways is seldom involved taking into consideration: 1) the impact of existing capacity stocks on the future energy system; 2) self-corrective functionality if the pathway diverges from the projected ones *a posteriori*. In order to improve the reliability of energy planning models, a novel dynamic modeling methodology is proposed and applied to the Swiss energy systems covering power, heat, and mobility. It provides freedom for users to generate tailored pathways according to region-specific inputs and to define model horizons determined by the variable start year, end year, and time slice, taking into account dynamic minus-plus of existing stocks and new installation across 140 energy conversion technologies. Four typical emission-driven pathways representing different mitigation strategies for Switzerland are defined and analyzed. The results show that an exponential strategy of carbon mitigation seems cost-effective compared to the three other suggested pathways towards carbon neutrality.

Keywords: Energy transition pathways, dynamic optimization, carbon neutrality.

1. Introduction

In the context of transition from fossil-dominant to renewable-based society, energy system modeling plays an essential role in guiding long-term decision-making, particularly for achieving the 2°C temperature rise goal formulated by IPCC (2014). Many European countries have set net zero carbon emissions in 2050 as objective. In terms of Switzerland, the government published the Energy Strategy 2050 (SFOE, 2018) and recently declared its fulfillment to net zero emission. One of the most important decisions is to phase out nuclear power plants before 2035. How to fill the gap raised by the power production deficit, and at the same time ensuring low carbon emissions are becoming overarching in the system design. Concretely, the following questions need to be answered: 1) Is the net zero emission objective achievable? 2) Is there a risk of energy deficit in the future, especially in the period when nuclear is completely decommissioned? 3) What measures and which technologies should be prioritized for development, and how much should be deployed? What's the consequence of final energy consumption (FEC)? 4) What is the cost for achieving the energy transition?

In order to answer these questions, IPESE (Industrial Process and Energy Systems Engineering) lab and Energy center EPFL have developed Energyscope (Codina Gironès

et al., 2015), a snapshot bottom-up model based upon mixed-integer linear programming (MILP) for cost minimization subject to various constraints such as supply-demand balance, energy conversion efficiency, storage continuity, availability of resources and so on. Detailed mathematical formulations were reported in previous studies (Li et al., 2020). Energyscope assists in characterizing the theoretically optimal configuration of the energy system in a "steady" state. One of the limitations lies in the "snapshot" nature, which disables its application in depicting pathways for the energy transition. Therefore, a model capable of handling dynamic changes of the energy system is expected in practice in favor of policy-making given the *status quo* and the ultimate objectives. In this study, we propose a novel model - Energyscope Transition Pathway (ESTP) - adapted from Energyscope by applying a "glutton" algorithm for decision-making in each step. Although the results obtained from ESTP may not be strictly optimal compared to global optimization algorithms widely used over the whole time series, such as Panos et al. (2019), it is assumed to be of more realistic guidance for making plausible strategies by simulating the system's dynamic behavior in evolution. In parallel, the convergence speed of the optimization problem gets curtailed from hours to minutes depending on the granularity of time steps.

In the following sections, we firstly elucidate the modelling structure and key data (Section 2), followed by the definition and analyses of four typical scenarios representing different mitigation strategies (Section 3). Finally we summarize the key discoveries and project future research directions (Section 4).

2. Model description

2.1. Modeling structure

ESTP analyzed firstly the historical evolution of energy technologies from 1990-2018 with respect to annually newly installed capacities and energy production, leading to an estimation of the capacity deficits in future years given the lifespans of the considered technologies. Secondly, the model was applied to the reference year for calibrating energy and carbon balance. Thirdly, the optimization algorithm was executed in order to obtain the results of the current period based upon all the information already given as

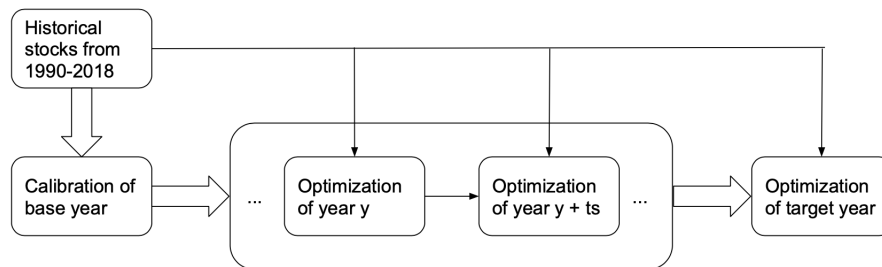


Figure 1: Model structure. Each optimization block is a MILP snapshot model. The historical stocks exert influence on all time steps from the base year to the target year. Each decision-making depends on previous optimization results.

inputs or available as the outputs from previous iterations, until the end year. Within each iteration, the total cost of the system was minimized taken into consideration the projected end energy demands (EUDs) driven by population, GDP, fuel price evolution and carbon emission objectives, as well as over 140 conversion technologies. Figure 1 shows the modeling structure. In this study, 2018 is set as the reference year with 4-year time steps towards 2050. Within each year, 12 periods represent the 12 months respectively.

The objective function for each year y within the considered horizon is expressed by Eq. 1, in which $\Phi(y' \leq y)$ represents the inventory of installations in all years y' starting from 1990 until a given year $y \in [2018, \dots, 2050]$. Different from the well-known Markov Chain where the probability of each event depends merely on the state attained in the previous event, the optimal total cost of the year y in ESTP is determined by the whole historical time series, including the previous state $y - 1$, as well as the specific parameters in the year y , e.g. the energy demands, interest rate etc. $\mathbf{F}(y)$ and $\mathbf{F}_t(y)$ represent the installed capacity and the used capacity in the month $t \in [1, \dots, 12]$ in the year y respectively.

$$\min_{\mathbf{F}(y), \mathbf{F}_t(y)} (TotalCost(y) | \Phi(y')) = (CAPEX(y) + OPEX(y) | \Phi(y')) \quad (1)$$

All the parameters used for calculating the annualized investment $CAPEX$ and operational cost $OPEX$ are year- and technology/resource-specific. Detailed mathematical formulations for one year were reported in Li et al. (2020) and the corresponding data are available in the same reference, as well as in Moret (2017) and Stadler et al. (2019).

2.2. Data: historical energy statistics

The historical data in Figure 2 on the installed capacity over 1990-2018 are based upon Swiss official reports OFEN (2020). Some missing data are calculated by historical production data and capacity factor assumptions according to Moret (2017). The calibration of the base year was conducted from two perspectives: FECs balance and EUDs balance.

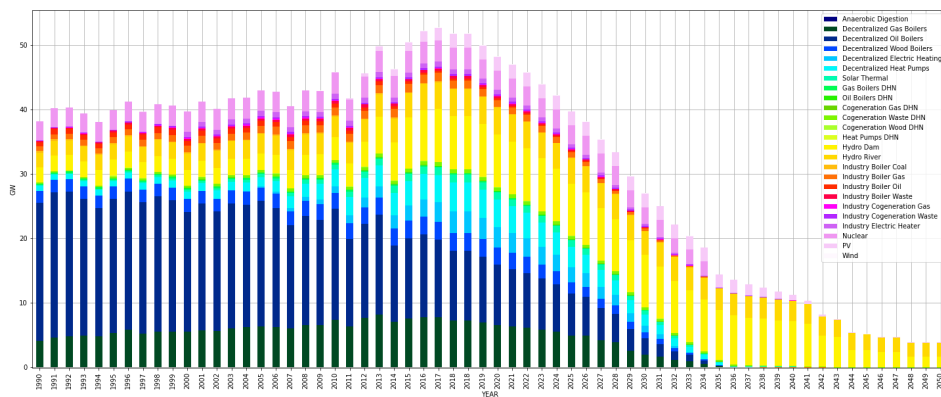


Figure 2: Historical accumulated installations from 1990-2018 and expected decommissioning from 2019-2050 in the Swiss energy system. In this figure: DHN district heating

First, all the resources obtained by the model were compared with the data reported in OFEN (2020); secondly, the EUDs deduced by all the technologies in terms of electricity, industrial process heat, low temperature heat (space heating and hot water), and mobility, were calibrated with the data from JASM (2020) database. Based upon the model, the total emissions of the Swiss energy system excluding land use, land-use change, and forestry sum up to 36038 kt in 2018, with approx. 3% relative error compared with Statista (2019).

Based upon the calibrated historical data, we estimated the decommissioning time during 2019-2020 of all historical stocks before 2018. For instance, the new capacity of PV in 2015 is supposed to be decommissioned in 2040 given a typical life span of 25 years. From Figure 2, the majority of stocks phase out gradually before 2040, except hydro dam and river plants which are of long life span and encouraged by domestic policy to carry on operation for those with positive results from re-evaluation of techno-ecological feasibility after concession (Barry et al., 2015).

3. Scenarios analyses

In this study, four scenarios are defined representing four typical mitigation strategies from 2018 towards the net zero emission objective in 2050:

- (a) Linear: mitigating the emission in a constant speed;
- (b) Logarithmic: advocating a rapid decrease in the short term then slows down gradually;
- (c) Exponential: limited decarbonization in the initial stage and prompt action in the long term, in contrast to (b);
- (d) Logistic (S): allowing a quick mitigation in the mid-term whilst keeping the two ends smooth, widely applied in simulation of population growth and disease propagation.

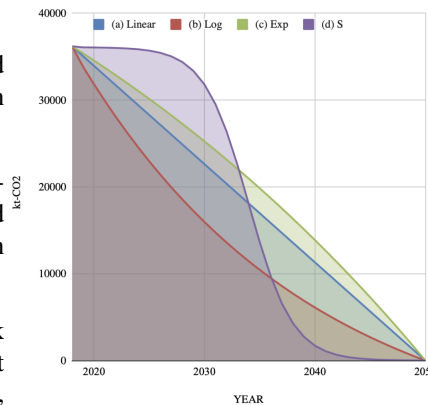


Figure 3: Carbon mitigation pathways

Without loss of generality, we take the results of scenario (a) for analysis, as shown in Figure 4. RES represents renewable resources and WASTE_BIO and WASTE_FOS represent biogenic and fossil wastes respectively. Two typical periods could be categorized:

- Pre-2035: the transition refers substantially to the shift from oil to gas, reflected by a considerable drop of light fuel oil (LFO) and motor fuel liquids, as well as the disappearance of coal utilization in industry. Instead, the utilization of natural gas increases significantly from 33 TWh in 2018 to 50-60 TWh in the 2030s.
- Post-2035: radical revolution is witnessed from fossils to renewables, where PV, geothermal (RES_GEO), wind and biomass keep rapid increasing, reaching their corresponding potentials in 2050.

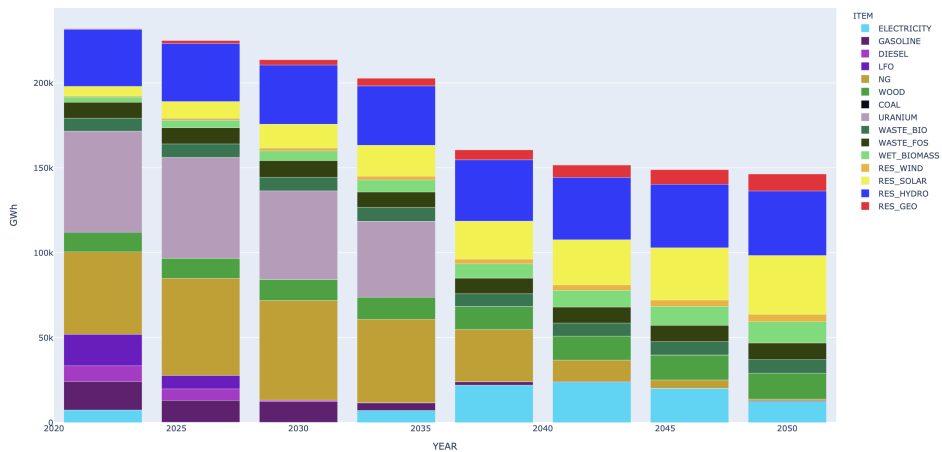


Figure 4: Evolution of resource utilization in Scenario a.

From an overall perspective, the total utilization of energy resources in the system decreases over the periods, as a result of the synergy of energy-efficient technologies, such as heat pumps, district heating, electric vehicles, and increasing reduction on the heat demand side from building and industry. Massive electrification, particularly renewable power, is witnessed in Figure 4 after 2035. Electricity import (in light blue) appears necessary in the proximity of 2035 where a drastic drop of domestic power production is foreseen due to the decommissioning of all the remaining nuclear power plants in Switzerland. Around 22 TWh net import amount is needed in that period in order to compensate the supply. Decreasing reliance on power import is expected afterwards, rendering finally an import amount of 10 TWh in 2050, approx. 1/3 of today’s value.

Associated to each mitigation scenario, Figure 5 shows the annual total system cost evolution from the first optimization year 2022 to the target year 2050. In general, cost declines in all scenarios over time predominantly due to the reduction of technology specific costs, and energy demand saving. An abnormality is observed in the scenario (d) where a steep increase is observed from 2034 to 2038. This results from a dramatic decreasing slope of the S-curve in between 2034-2038 that enforces the system to adopt more expensive choices in a short term, e.g. heat pumps, and compulsorily shutting down a part of carbon intensive gas plants that are still in operation as well to meet the climate demand. Among the four scenarios, scenario c with exponential mitigation strategy appears most cost-effective, which could be interpreted from two aspects: before

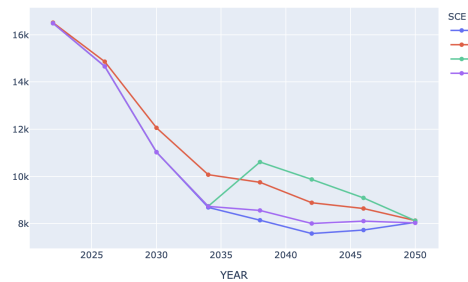


Figure 5: Evolution of energy system total cost [MCHF/year] by different mitigation strategies.

2035, less stringent emission limitation implies less investment; after 2035, emerging technologies today are supposed to become relatively mature, e.g. electrolysis, of which the specific investment is promising to be half of today's value, by consequence, the cost impact due to rapid decarbonization process could be dampened. By contrast, the most expensive pathway appears in scenario b with logarithmic mitigation strategy.

4. Conclusion

Our research is dedicated for policy-makers to define pathways and quantify their impacts towards net zero emission. The dynamic modeling approach by stepwise minus-plus representation of the energy system evolution allows flexible adjustment if the pre-defined pathway diverges from reality. The results consolidate our conclusion in previous studies that the carbon neutrality of Swiss energy system could be achieved with massive penetration of renewables (over 90 %) and wide deployment of CCUS technologies as well as sufficient infrastructure support, by showing specific pathways towards this objective taking into account the impact of decommissioning of all historical stocks and new capacities. It should be highlighted that the curvature of the carbon mitigation pathway would lead to different strategies. Therefore, exploring as exhaustively as possible the space of transitional carbon pathways, e.g. via parameterization in combination with uncertainty analysis, would be necessary in further research.

5. Acknowledgements

The authors express their gratitude to the financial support from the Swiss Competence Center for Energy Research - Joint Activity Scenario & Modeling (SCCER-JASM).

References

- M. Barry, P. Baur, L. Gaudard, G. Giuliani, W. Hediger, F. Romerio, M. Schillinger, R. Schumann, G. Voegeli, H. Weigt, 2015. The Future of Swiss Hydropower - A Review on Drivers and Uncertainties. SSRN Journal.
- V. Codina Gironès, S. Moret, F. Maréchal, D. Favrat, Oct. 2015. Strategic energy planning for large-scale energy systems: A modelling framework to aid decision-making. *Energy* 90, 173–186.
- IPCC, 2014. Climate Change 2014 Mitigation of Climate Change: Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.
- JASM, 2020. Database: The Joint Activity Scenarios and Modelling (JASM). <https://sccer-jasm.ch/>.
- X. Li, T. Damartzis, Z. Stadler, S. Moret, B. Meier, M. Friedl, F. Maréchal, 2020. Decarbonization in Complex Energy Systems: A Study on the Feasibility of Carbon Neutrality for Switzerland in 2050. *Frontiers in Energy Research* 8, 17.
- S. Moret, 2017. Strategic energy planning under uncertainty, Ph.D thesis, No. 796. Ph.D. thesis, EPFL.
- OFEN, 2020. Statistique Globale Suisse De L'énergie 2019, Office fédérale de l'énergie. Tech. rep.
- E. Panos, T. Kober, A. Wokaun, Oct. 2019. Long term evaluation of electric storage technologies vs alternative flexibility options for the Swiss energy system. *Applied Energy* 252, 113470.
- SFOE, 2018. Swiss Federal Office of Energy, Energy Strategy 2050 Once the New Energy Act Is in Force.
- Z. Stadler, B. Meier, M. Friedl, T. Damartzis, S. Moret, X. Li, M. Borasio, F. Marechal, 2019. Carbon Flows in the Energy Transition. Tech. rep., Swiss Federal Office of the Environment.
- Statista, 2019. Switzerland: Carbon dioxide (CO2) emissions 2019. <https://www.statista.com/statistics/449824/co2-emissions-switzerland/>.

Carbon Policy Assessment in Process Integration

Mohammad Lameh, Dhabia Al-Mohannadi, Patrick Linke*

*Department of Chemical Engineering, Texas A&M University at Qatar, Education City,
PO Box 23874, Doha, Qatar*

patrick.linke@qatar.tamu.edu

Abstract

The rising concerns about climate change have led to the emergence of several pathways for mitigating the increasing emissions of greenhouse gases, especially CO₂. Several technologies and actions constitute such pathways, whether through reducing the production of CO₂ streams by implementing renewable energy options for example or through processing them via CO₂ capture, utilization, and storage. For such projects to be implemented, they need to satisfy an economic favourability that serves as an incentive for the industrial stakeholders to invest. Process Integration tools have been developed recently to optimize the planning of CO₂ reduction through identifying the cheapest pathways. However, the interactions between the different stakeholders are not considered in such approaches. The existence of such interactions may lead to a non-optimal implementation of CO₂ reduction pathways under the applied policy framework. It is important for the enforced policies to guarantee the implementation of the lowest cost pathways to avoid recessive consequences associated with emissions mitigation. This work tries to identify the effect of such interactions on the feasibility of the optimal solution. CO₂ reduction policy can then be assessed based on the resulting feasible solution in comparison with the optimal solution. A conducted case study showed that a high tax on CO₂ emissions may result in deviations from the optimality. Subsidizing CO₂ capture and renewable energy would help in reducing the economic impact on the stakeholders, leading to a feasible optimal solution.

Keywords: carbon policy, process integration, minimum cost CO₂ reduction, CO₂ capture, renewable energy.

1. Introduction

The growing concern about the issue of global warming associated with the rising emissions of greenhouse gases has led to the emergence of national and international commitments to limit the pollution. CO₂ is a major greenhouse gas which is heavily produced from industrial and power plants. Different pathways exist for CO₂ reduction; however, the associated high costs prevent the required implementation (Tapia et al., 2018). Hence, it is important to identify optimal CO₂ reduction pathways to avoid unnecessary costs. In the field of process integration, various optimization methods have been developed as decision support platforms in CO₂ reduction planning (Manan et al., 2017). Such models consider different stages at which CO₂ reduction pathways can be implemented and provide detailed designs for cost-optimal implementation. Al-Mohannadi and Linke (2016) developed an optimization model to minimize the cost of implementation of CO₂ capture, utilization, and sequestration (CCUS). A multi-objective multi-period optimization was solved to determine optimal CO₂ reduction through

integration under conflicting environmental and economic objectives (Al-Mohannadi et al., 2019). Al-Mohannadi et al. (2020) considered a multiperiod optimization of CCUS and energy mix to assess CO₂ reduction planning. Since the economic motive is a major factor affecting the decisions taken, the implementation of CO₂ reducing pathways needs to achieve an economic favourability for the stakeholders to be implemented. The governments enforce different policies and regulations that act as incentives for the industries to implement pollution reduction options. As a result, symbiotic relations emerge between different interacting stakeholders that ensure the feasibility of such projects for all sides (Yazdanpanah et al., 2018). Different approaches have been developed to model such interactions. Tang et al. (2015) established an agent-based model to investigate the application of carbon emissions trading scheme on the Chinese case. Fraccascia et al. (2017) used a multi-agent model to study the effect of introducing landfill taxation and subsidy policies on the symbiotic relations within industrial parks. Aguilar et al. (2017) proposed a bi-level fuzzy optimization model to determine a proper incentive policy that results in the plants adopting proper heat and solid waste network to minimize CO₂ emissions. Such approaches implement complicated tools, and the solutions are obtained without assessing the behavior leading to optimal performance. Hence it is important to develop a simple method to determine the optimal cost for CO₂ reduction and to understand the system with the various options at a higher level. This work investigates the impact of CO₂ tax and subsidies on the implementation of CCUS and alternative energy options. The resulting feasible solution based on the interactions is determined and assessed in context of the minimum cost pathways determined from process integration perspective.

2. Methodology

The aim is to identify optimal CO₂ reduction pathways for a set of emissions sources, and to describe the interactions between the possible stakeholders in order to determine a feasible solution under the set policy framework. The CO₂ policy framework is assessed through comparing the costs of the feasible and optimal solutions. The CO₂ reduction pathways considered are CO₂ capture utilization and storage (CCUS), beside energy shifting. The system contains a set of existing emissions sources characterized by CO₂ capture and compression costs (C_{si}) and secondary emissions (γ_{si}), and a set of available sinks characterized by their profitability (CO₂ cost R_{dj}) and CO₂ reduction efficiency (η_{dj}). Power sources are characterized by their costs C_{Ek} and CO₂ intensity ε_k . Note that these techno-economic parameters are specified by the user, and their estimation is subjected to the user's assumptions.

The optimal pathways are determined through representing the possible options on a Marginal Abatement Cost (MAC) curve which shows the pathways arranged from cheapest to most expensive as represented by Lameh et al. (2020a). After that, the interactions between the stakeholders are modelled as a cooperative game where the allocation of CO₂ between a source and a sink is feasible only if both players can achieve an economic benefit from the symbiosis. The interaction can be characterised by the material and monetary flows between the different components of the system. The government plays a coordinating role through setting CO₂ tax (CT) and subsidy rates (SR) where the CO₂ emitting sources need to pay for the emissions they produce, while the government can cover a fraction of the total cost of implementation of renewable energy or CO₂ capture as an incentive to the industrial stakeholders.

The feasibility of the different options is determined by the economic favourability of the decisions. The decision of capturing and allocating a CO₂ stream from a source to a sink should have a lower cost (or higher profit) than other possible decisions. A source may emit its CO₂ to the atmosphere and pay the corresponding tax, or it can pay for the subsidized capture and allocate the emissions to the sink. The sink can afford to pay if the implemented utilization process is profitable. The stakeholders representing the sink may decide to invest in installing and operating the corresponding technology if it can reach a convincing level of profitability under the agreed-upon CO₂ price. Otherwise, the sink won't be available. Hence, under a set policy, the source (*s_i*) would have a minimum CO₂ price (*CP_{si}*) that the sink needs to pay, and the sink (*d_j*) would have a maximum price (*CP_{dj}*) which it can pay while meeting its profitability target. The CO₂ prices are determined as follows:

$$CP_{si} > C_{si} \times (1-SR) - CT \times (1-\gamma_{si}) \quad (1)$$

$$CP_{dj} < R_{dj} \quad (2)$$

Feasibility condition:

$$CP_{dj} > CP_{si} \quad (3)$$

In a system with multiple sources and sinks, the price is determined through the CO₂ market dynamics. It depends on the CO₂ availability, as well as *CP_{si}* and *CP_{dj}* for the available sources and sinks. The resulting model can be solved graphically on the CO₂ price curve (Figure 1) to determine the feasible CO₂ network scale and the participating players. The plot consists of a supply profile representing the sources by their *CP_{si}* and CO₂ availability *F_{si}*, and a demand profile representing the sinks by their *CP_{dj}* and the capacity *F_{dj}*. The sources follow the law of supply and their segments are arranged in the increasing order of *CP_{si}*. The sinks follow the law of demand, and their segments are represented in the decreasing order of *CP_{dj}*. The representation is similar to the graphical analysis presented by Lameh et al. (2020b). The CO₂ market equilibrium is defined as the point at which the profiles intersect. This point divides the graph into two sections: before the carbon market equilibrium is achieved, there is a feasible region for establishing the network where *CP_{si}* < *CP_{dj}*. Beyond that region, none of the sources or the sinks can afford to participate in the network at the set equilibrium price.

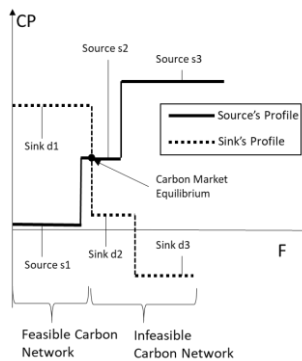


Figure 1 The graphical representation of the network with multiple sources and multiple sinks

The role of the government is played by the user through setting *CT* and *SR*. Note that the CO₂ prices for the sources can vary depending on the policy framework implemented. For example, if the tax (or the subsidy) is high, then the sources would be willing to sell

CO₂ at lower prices. Hence, the set CO₂ reduction policy can control the market equilibrium, and consequently the level of CO₂ reduction as well as the CO₂ price. For power generating sources that produce CO₂, beside the options of emitting the CO₂ streams or capturing them, they have the option of shifting to renewable energy sources. The renewable energy can be subsidized as well.

3. Case Study

Consider a system composed of different plants with different CO₂ sources. The different sources emit CO₂ streams with varying conditions which affect the total cost of CO₂ supply (capture and compression) and the resulting secondary emissions. Table 1 summarises the sources considered and the corresponding CO₂ capture cost and secondary emissions. The data used is based on case studies present in Al-Mohannadi et al. (2017a) and Lameh et al. (2020b).

Table 1 Source's data for the case study on carbon policy

Source	CO ₂ Produced (10 ⁶ tCO ₂ /y)	C _{si} (\$/tCO ₂)	γ _{si} (tCO ₂ -produced/tCO ₂ -captured)
GTL - Concentrated	4.12	2.5	0.03
Cement	1.38	32.5	0.27
GTL - Diluted	11.38	36.5	0.26
Methanol	0.64	36.5	0.26
Aluminium	3.30	36.5	0.29
Power Plant	3.8	40	0.13

The total emissions level of the considered sources is 24.62×10⁶ tCO₂/y. Amine absorption is considered as the capture process with 90% capture efficiency. The considered sinks are summarized in Table 2 (Al-Mohannadi et al., 2017b). Beside CCUS, solar energy can be used to generate up to 50% of the power plant's capacity to reduce CO₂ emissions at a marginal abatement cost of 31 \$/tCO₂-reduced. The determined cost is based on a 714 \$/kW capital cost for utility scale photovoltaic panels (IRENA, 2020), without storage and disregarding land cost, and assuming an emissions intensity of 0.54 tCO₂/MWh for the existing power plant.

Table 2 Sink's data for the case study on carbon policy

Sink (d _i)	F _{di} (10 ⁶ tCO ₂ /y)	R _{di} (\$/tCO ₂)	η _{di}
Enhanced Oil Recovery (EOR)	5	30	100%
Methanol B	2	20	99%
Storage	18	-10	100%

The MAC curve represents the optimal pathways to reduce CO₂ from the described sources, considering the CCUS and solar energy options (Figure 2). The analysis shows that the maximum profit of the CO₂ network is 113×10⁶ \$/y, achieved at 15% CO₂ reduction. The cost-neutral CO₂ reduction (X_{cost-neutral}) was found to be around 31%.

Three different scenarios for the implemented policy scheme were investigated (Table 3). A base case is considered where no policy is implemented, to examine the CO₂ reduction limit under the absence of authority interference. Two different policy scenarios were investigated to analyze the impact of CO₂ tax and subsidies on the system to achieve the 31% reduction (X_{cost-neutral}). Scenario 2 considered only the CO₂ tax as a policy, and Scenario 3 considers a mix between tax and subsidies.

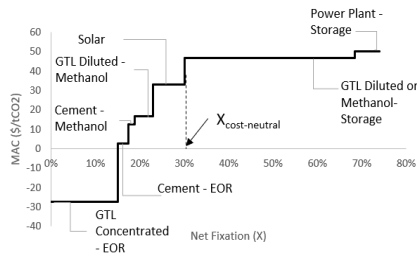


Figure 2 MAC curve for the case study on carbon policy

Table 3 CO₂ reduction policy scenarios

	CO ₂ Tax (\$/tCO ₂)	Subsidy-Capture (% of total capture cost)	Subsidy – Renewable Energy (% of the total renewable energy cost)
Scenario 1	0	0	0
Scenario 2	33	0	0
Scenario 3	5	40	85

The base case represented in the first scenario shows that there is only one feasible allocation: from the pure stream of the GTL to utilization in EOR. Such pathways would result in 15% reduction in the total CO₂ emissions, and would generate a total profit of around 113×10^6 \$/y, divided between the stakeholders of the GTL and the EOR, depending on the CO₂ price which is estimated to be within 20\$/tCO₂ to 30\$/tCO₂. This pathway abides by the optimum route for CO₂ reduction, as it is the only profitable option among the considered where a symbiotic opportunity that guarantees the economic profit for both the source and the sink exists.

After that, the CO₂ policy parameters were changed to achieve a CO₂ reduction target close to the cost-neutral CO₂ reduction suggested by the MAC curve analysis. Increasing the CO₂ tax rate would decrease the price at which the sources are willing to sell their CO₂ which pushes for more symbiotic opportunities and hence more CO₂ reduction. Since the sources would still have to pay carbon tax for the secondary emissions from the capture process, the impact of the tax differs among the sources depending on the corresponding secondary emissions factor γ_i . It was shown that the feasible CO₂ reduction pathways under the proposed carbon tax will be constituted of CCUS options between GTL-concentrated, cement, and power plant from the supply side, and from EOR and methanol on the demand side, with a CO₂ exchange flowrate of 7×10^6 tCO₂/y. Solar energy would be implemented at its maximum capacity. This showed a deviation from the MAC curve analysis where for a CO₂ reduction of 30%, power plant did not participate. Hence, the CO₂ tax resulted in a deviation from the optimal solution. The resulting feasible system from the CO₂ tax cost 5×10^6 \$/y in contrast to the zero cost of the optimal solution for the same level of CO₂ reduction.

Increasing the subsidy would have the same effect of increasing the tax in terms of providing an economic incentive that pushes down the required CO₂ price by the sources and increases the feasible CO₂ reduction limit. In the suggested policy of Scenario 3, polluting plants pay 5 \$/tCO₂, and the collected tax is used to cover 40% and 85% of the total cost of implemented CO₂ capture and solar power respectively. With the lower tax, the CO₂ price is affected by the cost of capture, and consequently, sources with lower capture cost participate in the CCUS network, beside the full implementation of solar power. This solution is coherent with the optimal pathway retrieved from the MAC curve analysis.

4. Conclusions

This work introduced a novel approach to assessing CO₂ reduction policy in context of process integration and optimization approaches. The methodology presented a simple tool that allows the identification of the symbiotic opportunities between different CO₂ sources and different CO₂ sinks based on the economics controlled by the set policy framework. A case study was then developed in which the methodology showed that different policy schemes can result in varying total costs to achieve the same CO₂ reduction limit. This approach helps in giving context to the solutions obtained from optimization models when it comes to the implementation under different stakeholder's interactions.

References

- Aguilar, K.D.T., Ubando, A.T., Culaba, A.B., Aviso, K.B., Tan, R.G.R., Chiu, A.S.F., 2017. Carbon dioxide reduction incentive for eco-industrial parks using bilevel fuzzy programming, *HNICEM. IEEE*, pp. 1-6.
- Al-Mohannadi, D.M., Abdulaziz, K., Alnouri, S.Y., Linke, P., 2017a. On the synthesis of carbon constrained natural gas monetization networks. *Journal of Cleaner Production* 168, 735-745.
- Al-Mohannadi, D.M., Hassiba, R.J., Abdulaziz, K., Linke, P., 2017b. A Natural Gas Monetization Approach with Carbon Dioxide and Excess Heat Integration in Industrial Parks, *Computer Aided Chemical Engineering* 44. Elsevier, pp. 1963-1968.
- Al-Mohannadi, D.M., Kwak, G., Linke, P., 2020. Identification of optimal transitions towards climate footprint reduction targets using a linear multi-period carbon integration approach. *Computers & Chemical Engineering* 140, 106907.
- Al-Mohannadi, D.M., Linke, P., 2016. On the systematic carbon integration of industrial parks for climate footprint reduction. *Journal of Cleaner Production* 112, 4053-4064.
- Al-Mohannadi, D.M., Linke, P., Shah, N., 2019. A multi-objective multi-period optimization of carbon integration networks in industrial parks, *Computer Aided Chemical Engineering* 46. pp. 487-492.
- Fraccascia, L., Giannoccaro, I., Albino, V., 2017. Efficacy of landfill tax and subsidy policies for the emergence of industrial symbiosis networks: An agent-based simulation study. *Sustainability* 9(4), 521.
- IRENA., 2020. Renewable power generation costs in 2019. International Renewable Energy Agency, Abu Dhabi.
- Lameh, M., Al-Mohannadi, D.M., Linke, P., 2020. Developing Minimum Cost Targets for Carbon Reduction in Different Geographical Regions. *Chemical Engineering Transactions* 81, 97-102.
- Lameh, M., Al-Mohannadi, D.M., Linke, P., 2020. Graphical analysis of CO₂ emissions reduction strategies. *Cleaner Engineering and Technology* 1, 100023.
- Manan, Z.A., Nawi, W.N.R.M, Alwi, S.R.W, Klemeš, J.J., 2017. Advances in Process Integration research for CO₂ emission reduction—A review. *Journal of Cleaner Production* 167, 1-13.
- Tang, L., Wu, J., Yu, L., Bao, Q., 2015. Carbon emissions trading scheme exploration in China: A multi-agent-based model. *Energy Policy* 81, 152-169.
- Tapia, J.F.D., Lee, J.-Y., Ooi, R.E., Foo, D.C., Tan, R.R., 2018. A review of optimization and decision-making models for the planning of CO₂ capture, utilization and storage (CCUS) systems. *Sustainable Production and Consumption* 13, 1-15.
- Yazdanpanah, V., Yazan, D.M., Zijm, H., 2018. Industrial symbiotic networks as coordinated games, *Autonomous Agents and MultiAgent Systems. International Foundation for Autonomous Agents and Multiagent Systems*, pp. 2145-2147.

Back-End Design and Development of an Energy Systems Analysis Tool

Maryam Arbabzadeh,^a Sapna kumari^a, Ragini Sreenath^a, Emre Gençer^a

^aMIT Energy Initiative, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA

Abstract

Low-carbon transformation of the energy system requires a combination of technology and policy options to ensure reliable, affordable, and clean energy. An assessment of plausible transition pathways can be guided with a set of tools that cover multi-sector dynamics of transitions and consider economy-wide and sectoral life-cycle analysis of numerous options. Sustainable Energy Systems Analysis Modeling Environment (SESAME) is a comprehensive system-level and pathway-level analysis model. Pathway-level analysis includes two main analysis types: life cycle assessment (LCA) and techno-economic assessment (TEA). The framework allows users to assess the sensitivity of key technological, operational, and geographical parameters associated with various energy pathways as measured by the emissions and costs. In this paper, we present the back-end design and implementation of the SESAME web application. The key contribution of web-based framework is that it supports easy integration of new LCA and/or TEA models and data sources in the backend so they can be made easily available to users for analysis via the front-end web interface. We achieve this by creating a Python-based library of classes and functions that model a higher-level code-based representation of LCA and TEA pathways. The tool is designed to capture the inherent modularity of SESAME, which makes it easier to add either new pathways or new data models and sources for existing pathways.

Keywords: energy systems modeling, computational tool, life cycle assessment, techno-economic assessment

1. Introduction

The global energy system is undergoing major transformations given the dual challenge of meeting increasing energy demand while reducing greenhouse gas (GHG) emissions. Such transformation requires a combination of technology selection and policy choices. Understanding the implications of these dynamics is challenging and requires a holistic approach to provide systems-level insights. Although there is considerable related research work emerging, there is a lack of readily available quantitative models and tools that consider a broad and robust life-cycle analysis approach for a range of plausible energy futures at regional and national levels. Such a tool is needed to help policy makers, industry, investors, and the financial sector to better understand and make decisions on energy choices and energy transitions.

One of MIT's modeling tools associated with energy choice evaluation is a comprehensive system-level and pathway-level LCA model called Sustainable Energy Systems Analysis Modeling Environment (SESAME). SESAME has a modular framework designed for supporting pathway-level and a system-level analysis (Gençer, 2020). Pathway refers to life-cycle stages of a particular product from raw material

extraction to production to its final end use. Pathway-level analysis is subdivided into main analysis types: LCA and TEA. LCA calculates GHG emissions associated with each of the various life-cycle stages whereas TEA calculates combined costs associated with the product or process through its life-cycle. Systems-level analysis models the emissions associated with large-scale energy sectors such as electric power system or a vehicle fleet. The goal is to allow users to explore variations in emissions and costs arising from different technological, operational, and geospatial parameters selected by the user. Given potential trade-offs that might arise between options, SESAME will include optimization models to guide users in selecting the optimal solution based on their preferred objectives and provide an extensive analysis. Future publications will include details of such optimization models.

The web-based framework of SESAME, developed in Python, is designed to capture the inherent modularity of the tool. The key contribution is the creation of a Python-based library of classes and functions to model a higher-level code-based representation of LCA and TEA pathways. Because LCA practitioners rely on a wide variety of models and data sources for analysing different pathways, the key aspect of our backend system design is to ensure its flexibility to incorporate new pathway models into the web framework using this library. Another contribution is defining a standard format for storage of model data for most effective ways to perform analytic calculations. A final contribution is the creation of useful HTTP end-points, abstracting away the analytic computations and programming-based back-end so that the front-end developer can easily use those endpoints to collect the required user inputs and display the results of analysis.

2. Methodology

2.1. Analysis Types

Our framework supports three main analysis types: LCA, TEA and LCA combined with TEA, with two separate metadata objects, one for LCA pathways and one for TEA pathways. Metadata object for LCA pathways is quite intricate since we don't simply design a separate model for every pathway; instead we design models for individual activities which are connected together to make up a pathway. This is because same activity may be shared across many different pathways and encoding the activity as part of a whole pathway instead of as a separate unit will lead to a lot of overhead in terms of repeated code and functions. For instance, upstream crude oil extraction activity has several different end-uses activities such as gasoline, diesel, and LPG, which make up many different energy pathways. The pathway metadata object for LCA analysis is a nested JSON data structure (Pezoa, 2016) that encapsulates all pathways available for LCA analysis in SESAME. The first layer comprises of life-cycle stages with second layer storing the activities in the stages and the third layer containing reference to the actual data source models within the activities that conduct LCA flow calculations. A new pathway for LCA analysis can easily be added to the SESAME framework by first implementing the new activity data source models and then adding all the relevant information about the implemented classes in the metadata object.

For every pathway available to LCA analysis, we build the corresponding TEA model as well, so the combined analysis is defined for all pathways. TEA metadata object is simply a one-layer JSON data structure that contains the reference to the TEA model class and the lookup table by the pathway name and id. Each TEA model class specifies the user inputs that will need to be collected for a given pathway and performs the cost calculations using both the collected user input values and data stored in various cost-specific data tables.

2.1. Pathway Representation via Python Classes

2.1.1. LCA Pathway Classes

Figure 1 shows the hierarchical relationship between the objects. Metadata is a list of stages where each stage is a list of activities. Each activity is subsequently a list of sources that contains the reference to actual activity-source class which implements the LCA flow calculations for that activity in that stage.

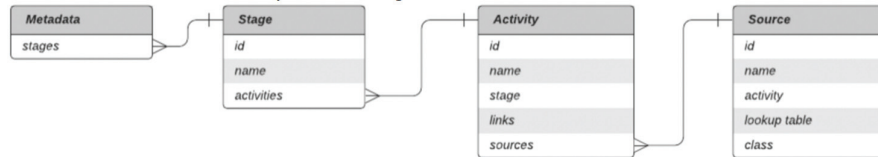


Figure 1 Pathway entity-relationships diagram

In implementation of these classes, we make use of the key concepts of object-oriented programming which include abstraction, encapsulation, and inheritance (D'Andrea and Gowda, 1990). Abstraction means that each object should only expose a high-level mechanism for using it, and hide the implementation details. While we describe the attributes and methods of all objects below, any user of this library only needs to know the high-level purpose of these methods and not the exact implementation detail. Inheritance means creating child classes from parent class where child classes can implement their own unique methods in addition to inheriting all the shared methods and common logic implemented by parent class. We make use of inheritance in creation of a parent Activity-Source class, which is described later.

Metadata: A Metadata object represents the set of all pathways defined in SESAME. The metadata class is initialized with an empty list of stages, and supports two main methods: register stage and get stage. The register stage method adds a new stage object, initialized with the name argument, to the stages list. The get stage method returns the stage object from the list of stages whose name matches the name passed as argument.

Stage: A Stage object represents all the activities that belong to a particular LCA stage. The stage class is initialized with attributes id, name, and an empty list of activities and supports two main methods: register activity and get activity. The register activity method adds a new activity object, initialized with the name passed as argument, to the activities list. The get activity method returns the activity object from the list of activities whose name matches the name passed as argument.

Activity: An Activity object represents a specific life-cycle activity. The Activity class is initialized with attributes id, name, stage, empty list of links and an empty list of sources, and supports three main methods: link, register source and get source. Links denote the set of activities in the next stage that this activity leads to. For instance, for transmission activity, the links would include all power production activity including NG, wind, and other electricity generation activities since transmission activity in gate to user stage implies the process activities must be power generation activities.

Source: Source objects denote the set of data source models that determine how the activity flow calculations will be performed. The Source object is initialized with attributes id, name, activity, lookup table, and class. The register source method takes as arguments a source name, the activity object, associated Activity-Source class, and lookup table and adds a new source object. The get source method returns the source object from the list of sources whose name matches the name passed as argument. The id attribute is initialized as null. The Activity-Source class is a reference to the specific class

that defines the set of user input parameters and implements the flow calculation methods. The lookup table is an optional table that may be read in the reference class and includes information for the input collection and stores associated emissions for selected pathways. When a source object is being registered to an activity object, it is also inserted as a record to a global sources database instance which stores all source records by their ids. The ids are created by concatenating the source name with activity identifier, thus ensuring their uniqueness. Sources database object is important because once the user interface collects all the user inputs for a particular source id, sources database is used to find the associated source object. The user inputs value collected are then passed to the corresponding Activity-Source class which is an attribute of this source object.

Activity-Source: Since most activity source objects share some common attributes and methods, we create a parent class called Activity-Source that implements the common methods. Creation of such a class helps avoid repeated functions and code redundancy through various activity source classes. Within an Activity-Source class, we can create child classes that can be used to implement method specific to a particular pathway. The set of relevant parameters and the exact method of calculation may vary not only between different activities but also between different source models for same activity. Hence, in our implementation, there is a separate class for every unique activity source model. For example, for NG power production activity, we have two separate source models, and thus, we have two classes which take different set of user inputs and calculate flows differently. Likewise, we have a separate class for NG transportation activity in midstream, and NG extraction activity in upstream. However, all activity source classes share the same model and they take as arguments the user inputs, an optional lookup table, and implement three methods which compute three flow types of input, output, and emission as shown in Figure 2.

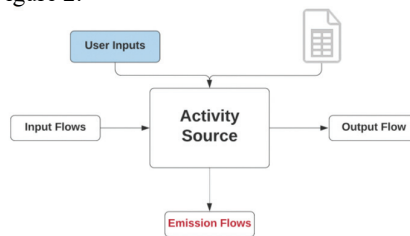


Figure 2 Activity source class model

2.1. User Input Classes

User inputs are a very important part of our analysis tool. In the user interface, we let the user specify a range of technological, temporal and geographical variables for each of the analysis type. Once the user inputs are all collected on the front-end, they are routed to the back-end where they are used in performing the necessary computations. The order of input collection matters because the activities between two different stages are linked by a shared input-output flow, and which flow type is passed from one activity to another differs. Regardless of the direction in which the user inputs are collected, whether it is from upstream to end use or from end-use to upstream, we choose to perform LCA calculations in one order on the back-end i.e. from end-use to upstream.

The set of inputs required of user before they can see the analysis results can be quite large especially in case of five-stage LCA analysis. Thus, to take away some burden of specifying all inputs from the user, we put in place a mechanism for specifying defaults for all input types. Another advantage of having defaults is that not every value might be known by the user and having pre-filled data can be helpful for non-experts.

2.1.1. Input Subclass

Based on the system needs, we defined three different sub-classes of input class in the back-end which include continuous, categorical and options. Continuous inputs represent input variables that take continuous numerical values such as turbine efficiency, plant lifetime or solar power intensity. Options inputs refer to any inputs with a fixed set of options that is passed as an argument to the class. Categorical inputs denote the inputs, which the options are read from a filtered data frame based on previous user input.

2.2. Web Framework Interface

For LCA analysis module, the front-end interface fetches the LCA pathway metadata object via an HTTP endpoint, and uses the object to dynamically collect the user inputs for all pathway stages starting from end use or upstream. The storage of activity links in the metadata object guides the appropriate user input collection for each subsequent stage based on the user activity selection in previous stage. Likewise, for TEA analysis module, the front-end fetches the TEA metadata object via a separate HTTP endpoint. For LCA-TEA module, both LCA and TEA metadata are wrapped in one object to be sent to front-end for input collection. The front-end posts the pathway analysis response object to back-end, which contains the user input values for a pathway for a particular analysis type. The backend then performs the analysis on either LCA, TEA model or both by calling relevant models specified in the metadata objects for the pathway specified by user. Back-end returns the analysis results object back with either emissions data or costs data or both depending on the analysis type. The front-end uses a graphing library to display the relevant emissions/costs plot sent in the analysis results object.

3. Case Examples

3.1. Pathway classes of power production

Stage objects corresponding to the five stages (upstream, midstream, process, gate to end-use, and end-use) in a pathway are created. Next, for each, their corresponding activity objects need to be created. For example, for the process stage object, we create an activity object called XPowerProduction. X refers to energy source i.e. NG, coal, wind, or solar. For this activity object, we register a source object Data1, which is the data source we use for this calculation. Now, for this source object, we create a child class XPowerData1 (with the parent class being an ActivitySource object) and link the file with the required data as the lookup table. Lastly, under the XPowerData1 class, we define a get_emissions method that calculates emissions based on the data in the lookup table and the specific inputs provided by the user. Figure 3 displays the GHG emissions results for multiple pathways across life cycle stages. As shown in this figure, the process stage is the main contributor to emissions in case of fossil fuels. The inputs default assumptions for each pathway is provided in Figure 4, which demonstrates the user's capabilities to test the results to technological and operational varieties.

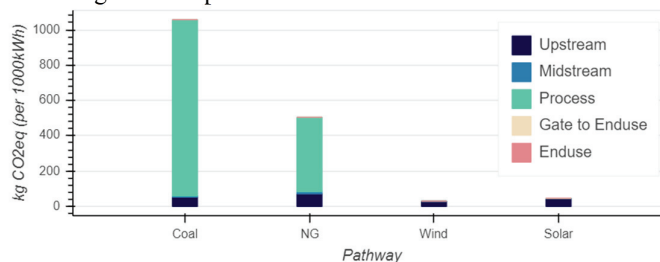


Figure 3 LCA GHG emissions of producing 1000 kWh electricity through multiple pathways

NG Power Production

Activity	Source	Generation Region	Generator Type
NGPowerProduct	GREET	US	Mix
Include Infrastructure Emissions?			
no			
Mode	Distance in miles	Loss Factor in %	
pipeline	750	0.12	
Natural Gas Type	Well Infrastructure	Leakage Parameter	
NG mix	No	EPA 2019	

Coal Power Production

Activity	Source	Coal Rank	Generation Region
CoalPowerProduct	GREETonly	Mix (50% bituminous)	US
Include Infrastructure Emissions?			
no			
Generator Type	Sub Feed	Mode	Distance in miles
Mix	mix	default mix	69991
Well Infrastructure		% Underground Mining Share	
No		31	
Loss Factor in %			
0			

Wind Power Production

Activity	Installation Type	Choose Capacity Factor or Wind Speed	Hub Height (m)
WindPowerProduct	onshore wind farm	Capacity Factor	80
Year Installed			
2014-2016			
Turbine Model		Lifetime (years)	
Typical Onshore mix		20	
Emissions in Manufacturing (gCO ₂ e/kWh)			
600			

Solar Power Production

Activity	Location of Installation	Cell Type	Installation Type
SolarPowerProduct	US NE (Boston)	multi crystal Si (mc)	utility, 1-axis tracker
Shading Losses (in %)			
2.5			
Lifetime (year)		Efficiency, STC Rated	Degradation Rate (per year)
20		20	0
Inverse Loading Ratio		Emissions in BOS Production (gCO ₂ e)	Shipping Distance from Panel Production to Installation (km)
1		660	20000

Legend: Process (green), Midstream (blue), Upstream (grey)

Figure 4 Input assumptions for 1000 kWh power production (end-use) through multiple pathways assuming 4.68% electricity loss in transmission (get-to-user). For wind and solar, the adjustment is made by computing all the flows in the process activity, but passing the computed flow values to the upstream activity for correct flow attributions.

3.2. System analysis - emissions impact of electric vehicles (EVs) charging pattern

The magnitude of the emissions reduction associated with EV deployment depends significantly on EV charging patterns as well as hourly power grid variations. In this case study, we use SESAME to estimate the emissions associated with EV deployment in 60 cases across the US using hourly grid data from 2018 and 2019 (alongside hourly charging, driving, and temperature data). The results show that the emissions impact of the charging pattern varies by region. For example, in California and New York, overnight EV charging produces ~70% more and ~20% fewer emissions than daytime charging, respectively (Miller, 2020).

References

- Gençer E, Torkamani S, Miller I, Wu TW, O'Sullivan F. Sustainable energy system analysis modeling environment: Analyzing life cycle emissions of the energy transition. *Applied Energy*. vol. 277. 2020. <https://doi.org/10.1016/j.apenergy.2020.115550>.
- Pezoa F, Reutter J. L, Suarez F, Ugarte M, and Vrgoč D. Foundations of JSON Schema. In *Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee*. 2016.
- D'Andrea R. J. and Gowda R. G. Object-oriented programming: concepts and languages. In *IEEE Conference on Aerospace and Electronics*. vol.2. 1990.
- Miller I, Arbazadeh M, Gençer E. Hourly Power Grid Variations, Electric Vehicle Charging Patterns, and Operating Emissions. *Environmental Science & Technology*. vol. 54. 2020. <https://dx.doi.org/10.1021/acs.est.0c02312>.

Compression system power requirements for various CO₂ sources and transportation options

Mathew Dennis Wilkes^a, Sanjay Mukherjee^a Solomon Brown^{a*}.

^a *Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom,*

s.f.brown@sheffield.ac.uk

Abstract

CO₂ compression and conditioning not only contributes to the overall energy penalty of Carbon Capture Utilisation and Storage (CCUS), it also influences the logistics surrounding the transportation network. The techno-economic feasibility of transporting CO₂ through pipelines or in storage vessels has been reported extensively. However, the conclusions in the existing studies are case specific, and focus on either the CO₂ source or the end-point. Whereas, in large CCUS clusters/networks, a range of sources will need to be interlinked and be able to supply conditioned CO₂ to various transportation options. With the prospect of multi-national CCUS transportation infrastructures, the link between CO₂ source and transport needs to be clearly characterised.

Herein, a process model for a CO₂ compression train is constructed in gPROMS gCCS, based off the IEAGHG base case B0 for post-combustion capture and compression. The model is capable of conditioning the CO₂ stream to the required outlet pressure, temperature and composition. The results compare various CO₂ streams from capture plants attached to a range of CO₂ sources including power generation and industry. The findings highlight the power requirement for conditioning to various end point characteristics, specifically the level of moisture control required for different pipeline options. This analysis can also be used alongside economic evaluations to enable the cost-effective deployment of CCUS transport networks, and aid in forming effective CCUS clusters.

Keywords: CCUS, CO₂ Compression, CO₂ Transport, CO₂ Sources

1. Introduction

One of the new set of challenges facing the Carbon Capture Utilisation and Storage (CCUS) industry is the formation of cluster networks, interlinking various CO₂ sources with different stream characteristics (pressure, temperature, and composition). The sources are mixed and delivered to potentially different storage options, posing technical and operational challenges for CO₂ transportation (Moe, et al., 2020). Alongside power industry focused CO₂ capture technologies (post-, pre-, and oxy-combustion), a growing interest is industrial sources such as the production of hydrogen, fertilisers, cement, iron and steel (Bui, et al., 2018).

An early study by Aspelund and Jordal (2007) investigated the interface between CO₂ capture and transportation, stating the energy requirement for CO₂ conditioning is between 90 and 120 kWh/tCO₂. However, this depends on the inlet stream characteristics (i.e. the source) and the conditioning method. Centrifugal compressors are the conventional choice in the power generation industry, but due to the low-pressure ratio (between 1.7-2:1) a multistage system is required (Martynov, et al., 2016).

Several studies have looked at optimising and comparing conditioning strategies, mainly focused on multistage compression and liquefaction processes. Witkowski & Majkut (2012) found power savings of up to 21% can be achieved through integrally geared centrifugal compressors, and over 45% using refrigerated subcritical liquefaction. IEAGHG (2011) also reported a reduced energy consumption when incorporating early liquefaction then liquid pumps, although it was only a 0.2 MW decrease, approximately 0.35% lower than conventional compression. This saving is potentially offset by an increased cooling water demand. Therefore, this study focuses on the conventional multistage compression process.

1.1. Aims and Objectives

Several parametric and techno-economic studies have analysed different CO₂ transportation options, highlighting the effects of impurities, operating phase, and system design. However, the conclusions in the existing studies are case specific, and focus on either the CO₂ source or the transportation option. The prospect of multi-national CCUS transportation infrastructures, means the link between CO₂ source and transport needs to be clearly characterised.

This study considers the impact different CO₂ sources and end point specifications have on the power demand for the compression and conditioning train. Figure 1 highlights the CO₂ capture technologies, the conditioning system, and the transportation options investigated in this study, focussing on dense phase pipelines for the transportation option (Harkin, et al., 2017). For each technology, the conventional multistage compression system will elevate the CO₂ stream pressure to 111 bar and 70°C as specified in IEAGHG (2011). Including intermediate storage and shipping in the transportation options is an unfair comparison, as the pressure requirements are much lower than for pipeline transportation. Therefore, less energy is required for conditioning. For more information on shipping transport, see MEP (2016). Three moisture control levels are used to compare different end-point composition guidelines:

- 20ppm – Beverage industry CO₂ specification as stated in EIGA (2016)
- 300ppm – Optimal pipeline moisture content (250-350ppm) from Brunsvold et al. (2016)
- 600ppm – Kinder Morgan pipeline specification stated in Jensen et al. (2014)

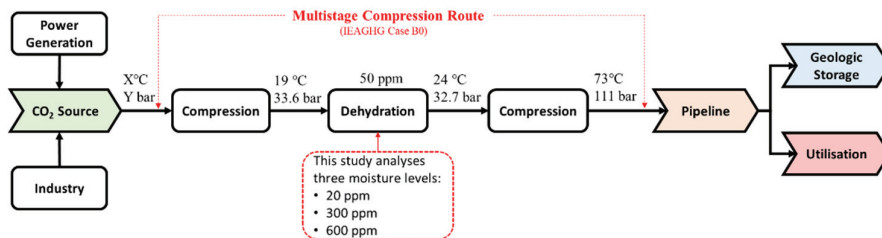


Figure : Study overview

2. CO₂ Sources

Table 1 shows the capture CO₂ stream characteristics used in this study. Impurities affect pipeline operation, therefore, H₂O (Brunsvold, et al., 2016) and N₂ (Peletiri, et al., 2019) are included alongside CO₂ in the composition of each stream. All other impurities are lumped as N₂.

In terms of technology readiness level (TRL) post-combustion capture using amines is TRL9 (commercial), so too is pre-combustion capture for natural gas processing using the Selexol processes (Bui, et al., 2018). In oxy-fuel combustion the purification process involves compression integrated with additional separation technologies (Pipitone & Bolland, 2009). Therefore, it is not included in this study. For post-combustion capture, the benchmark is 30wt.% MEA. Aqueous Ammonia (NH₃) was chosen as an alternative post-combustion solvent, due to the lower stripping duty 2.0-2.9 GJ/tCO₂ compared to 3.6-4.0 GJ/tCO₂ for MEA (Bui, et al., 2018). Methyl diethanolamine (MDEA) was chosen as an alternative pre-combustion capture technology to Selexol, due to the high capture rate (95%) and low stripping duty 0.99-1.34 GJ/tCO₂ (Romano, et al., 2010). Adsorption systems for post-combustion capture have advanced over past three decades, at TRL7 they are at demonstration level (Bui, et al., 2018). For smaller scale operations such as the simultaneous production of H₂ and CO₂, the Gemini PSA process can recover 94% CO₂ at 99.4% purity from steam methane reforming off-gas (Sircar & Golden, 2000). Membrane based CO₂ separation has been investigated for hydrogen production (TRL5), power generation (TRL6), and natural gas reforming (TRL7) (Bui, et al., 2018). Chung et al. (2018) highlighted the potential for using hollow-fiber Polaris™ membranes to capture high purity CO₂ from blast furnace gas in iron and steel plants.

Table 1: Captured CO₂ characteristics

Source	Temperature (°C)	Pressure (bar)	Composition (%)			Source
			CO ₂	H ₂ O	N ₂	
Post-MEA	38.00	1.6	95.88	4.11	0.01	(IEAGHG, 2011)
Post-NH3	20.00	6.00	99.00	0.40	0.00	(Yu, et al., 2011)
Pre-MDEA	30.00	1.10	96.02	3.92	0.02	(Romano, et al., 2010)
Pre-Selexol	-5.00	1.20	99.77	0.17	0.00	(IEAGHG, 2011)
	1.00	4.80	97.30	0.07	0.03	
SMR-PSA	21.00	1.00	99.40	0.00	0.00	(Sircar & Golden, 2000)
Steel-MEM	40.00	1.00	98.97	0.00	0.01	(Chung, et al., 2018)

3. Process Simulation

The base case B0 for post-combustion CO₂ compression from IEAGHG (2011) is used as the basis for the compression train model. The layout of the system is illustrated in Figure 2. The model is developed in gPROMS® gCCS 1.1.0, utilising the model library to construct a flowsheet of the compression system, including the compressor sections, inter-stage cooling, knock-out drums, surge valves and a dehydration unit.

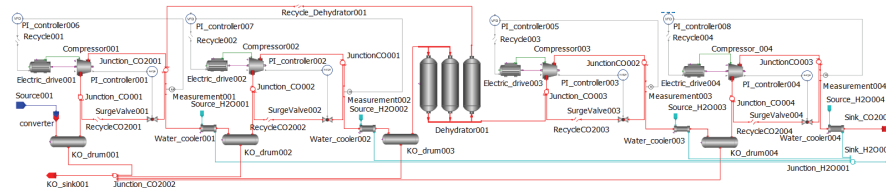


Figure : Model topology for multistage CO₂ compression, developed in gPROMS® gCCS. Each centrifugal compressor section consists of multiple stages, modelled via polytropic efficiency, with negligible hold-up and inertia of gas. The polytropic efficiency for the 1st, 2nd, 3rd and 4th compression stages are 79.38, 86.66, 82.92 and 80.82 %, respectively, within the range (74.50-87.45 %) highlighted in IEAGHG (2011). The thermo-physical properties and phase equilibrium of the fluid is determined through gSAFT. The foreign

object CompPerformFO determines the compressor maps and number of stages for a given set of input parameters (PSE, 2016).

Specific outlet conditions for the individual water coolers are not given in IEAGHG (2011), therefore, the two inter-stage water cooling steps are lumped into one and modelled via a heat exchanger unit. The cooling water supply is a set input parameter and each of the interstage cooling heat exchangers uses water at 6 bar and 12°C. The pressure drop over the inter-stage cooling and knock-out drums is 0.2 bar each. Ensuring the inlet conditions to the compression stages are identical to IEAGHG (2011). The dehydration unit removes H₂O from the CO₂ stream to a specified moisture content, and consists of three molecular sieve beds, two used for drying and one in regeneration mode operating simultaneously. The regeneration fraction (10%) and temperature (250°C) are specified parameters, and the model calculates the heat requirement and power consumption (PSE, 2016). Table 2 shows the input parameters used to simulate the process. The inlet flow contained 95.88 vol.% CO₂, 4.11 vol.% H₂O, and 0.01 vol.% N₂. The final CO₂ product contained 50 ppm moisture at 111 bar, ready for pipeline transportation.

Table 2: Input parameters and model validation

Parameters	Value	Simulation	Deviation (%)
Inlet Flowrate (kg/s)	154.57	154.57	0.00
Inlet Temperature (°C)	38.00	38.00	0.00
Inlet Pressure (bar)	1.60	1.60	0.00
Outlet Flowrate (kg/s)	151.90	148.17	2.46
Outlet Temperature (°C)	73.00	73.00	0.00
Outlet Pressure (bar)	111.00	110.76	0.22
1 st Compressor Power (MWe)	21.70	22.63	-4.29
2 nd Compressor Power (MWe)	24.10	20.60	14.52
3 rd Compressor Power (MWe)	8.00	6.97	12.88
4 th Compressor Power (MWe)	3.70	3.44	7.03
Total Power Demand (MWe)	57.50	53.68	6.64
Energy Demand (GJ/tCO ₂)	0.39	0.36	6.64
Energy Demand (kWh/tCO ₂)	107.77	100.61	6.64

For continuity across the CCUS chain the power demand has been converted into GJ/tCO₂ and kWh/tCO₂. Based off the data for the IEAGHG (2011) base case B0 the energy demand for compression is 0.39 GJ/tCO₂ or 107.77 kWh/tCO₂, within the bounds detailed in Aspelund & Jordal (2007). The simulation results (0.36 GJ/tCO₂) are 6.71% lower than expected. The simulation calculated the dehydration unit requires 4.04 MWe, to heat the regeneration stream to 250°C. The IEAGHG report does not give the power requirements for the dehydration unit. Therefore, if this additional power is included the simulations overall power demand is only 0.31% higher.

4. Results and Discussion

The flowrate basis is 1 kg/s for all technologies, in reality large CCUS clusters will have a range of flowrates attached with some having transient production, this is another challenge highlighted in Moe et al. (2020). Figure 3 highlights the power requirement for the compressors and dehydration unit, also included is the overall energy demand. All the technologies investigated have a low enough initial H₂O content that the knock-out drums remove moisture to <600 ppm. Removing the need and the small power demand associated with the dehydration unit, subsequently decreasing the inlet flowrate and power demand for the 2nd compressor. For the post-combustion MEA case this reduces

the overall power demand by 9.35 %, when going from 20 to 600 ppm. Interestingly, the energy demand for the high moisture MEA case (162.43 kWh/tCO₂) is lower than membrane separation for blast furnace gas treatment (174 kWh/tCO₂). Whereas the low moisture MEA case is 2.82% higher, due to a higher 2nd compressor power demand and the inclusion of the dehydration unit, although this only accounts for 7.76 kWh/tCO₂. Membrane separation has no H₂O in the capture CO₂ stream and requires no dehydration unit; however, the potential savings are offset by the the initial high temperature (40°C) and low pressure (1 bar).

Starting at a higher pressure reduces the number of compressions stages and overall pressure ratio, therefore, reducing the energy demand. For NH₃ this significantly reduces the power required for the first compressor which elevates the stream to 7 bar. Resulting in NH₃ requiring 19.23% less power overall than MEA, for the 20 ppm scenario. The Selexol process produces two CO₂ streams, it is assumed a 50/50 split between them, therefore, a pre-compression stage is required to get both streams to the same pressure. Slightly offsetting the savings of a higher initial stream pressure. Overall, there is no significant conditioning difference between the pre- and post-combustion capture technologies.

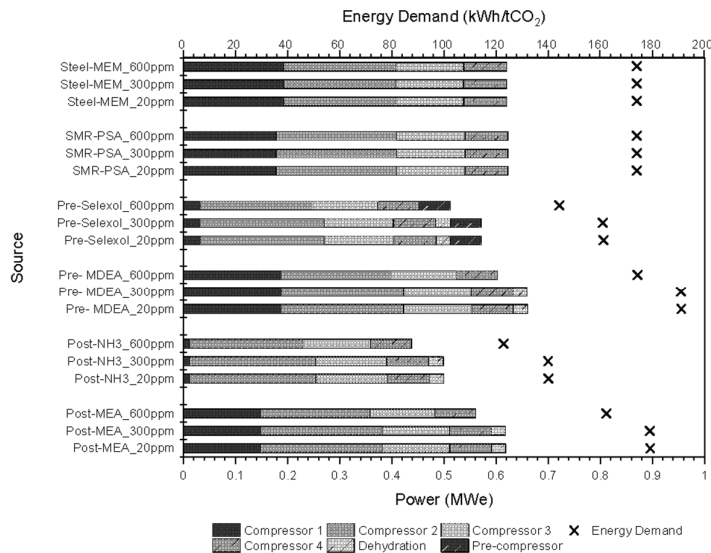


Figure : Individual unit power consumption and overall conditioning energy demand

Both of the industrial sources require no dehydration unit, and produce identical energy demands of 174 kWh/tCO₂. Comparable to the 20 ppm and 300 ppm MEA cases. The results show the end moisture level does have an effect on process performance, but not as significant as the initial stream pressure.

Energy demand in terms of kWh/tCO₂ in almost all cases is higher than expected, compared to Aspelund & Jordal (2007) and the validation work. A result of the small CO₂ flowrate basis, indicating economies of scale could play a major role in determining successful cluster partnerships. For example, a small-scale H₂/CO₂ PSA system, could be connected to much larger capture streams. Whereas a small-scale pre-combustion capture plant may be economically unfavourable.

5. Conclusions

The link between CO₂ source and transport needs to be clearly quantified to aid in the design and decision making of CCUS clusters. Within this study, process simulation has shown energy requirements for compressing and conditioning CO₂ streams from various power generation and industrial sources. This is purely a technical analysis and does not consider the economic impact of altering the inlet stream properties. End point moisture level has an effect on overall power consumption; however, the greater impact comes from the initial stream pressure. Higher starting pressures result in lower overall pressure ratios, and reduced number of compressor stages. The results indicate small-scale low-pressure capture sources may suffer economies of scale during compression and conditioning. Future studies should investigate the effects of different flowrates attached in the same CCUS cluster.

6. Acknowledgements

This work is funded by the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training in Carbon Capture and Storage and Cleaner Fossil Energy (reference: EP/L016362/1), and Drax Group Plc. The author would like to acknowledge the support of the Royal Academy of Engineering through the Industrial Fellowship (reference: IF\192046).

7. References

- Aspelund, A. & Jordal, K., 2007. Gas conditioning—The interface between CO₂ capture and transport. *International Journal of Greenhouse Gas Control*, 1(3), pp. 343-354.
- Brunsvold, A. et al., 2016. Key findings and recommendations from the IMPACTS project. *International Journal of Greenhouse Gas Control*, 54(2), pp. 588-598.
- Bui, M. et al., 2018. Carbon capture and storage (CCS): the way forward. *Energy & Environmental Science*, 11(5), pp. 1062-1176.
- Chung, W., Roh, K. & Lee, J. H., 2018. Design and evaluation of CO₂ capture plants for the steelmaking industry by means of amine scrubbing and membrane separation. *International Journal of Greenhouse Gas Control*, Volume 74, pp. 259-270.
- EIGA, 2016. *Carbon Dioxide Food and Beverages Grade, Source Qualification, Quality Standards and Verification*, Brussels: European Industrial Gases Association AISBL.
- Harkin, T. et al., 2017. Development of a CO₂ specification for a CCS hub network. *Energy Procedia*, Volume 114, pp. 6708-6720.
- IEAGHG, 2011. *Rotating equipment for carbon dioxide capture and storage*, s.l.: s.n.
- Martynov, S. B. et al., 2016. Impact of stream impurities on compressor power requirements for CO₂ pipeline transportation. *International Journal of Greenhouse Gas Control*, Volume 54, pp. 652-661.
- McMillan, G. K., 2010. *Centrifugal and Axial Compressor Control*. 10th ed. New York: Momentum Press.
- MEP, 2016. *Feasibility study for full-scale CCS in Norway*, s.l.: Ministry of Petroleum and Energy.
- Moe, A. M. et al., 2020. *A Trans-European CO₂ Transportation Infrastructure for CCUS: Opportunities & Challenges*, Brussels: Zero Emissions Platform.
- Peletiri, S. P., Mujtaba, I. M. & Rahmanian, N., 2019. Process simulation of impurity impacts on CO₂ fluids flowing in pipelines. *Journal of Cleaner Production*, Volume 240, p. 118145.
- Pipitone, G. & Bolland, O., 2009. Power generation with CO₂ capture: Technology for CO₂ purification. *International Journal of Greenhouse Gas Control*, 3(5), pp. 528-534.
- PSE, 2016. *gCCS Documentation*, London: Process Systems Enterprise Limited.
- Romano, M. C., Chiesa, P. & Lozza, G., 2010. Pre-combustion CO₂ capture from natural gas power plants, with ATR and MDEA processes. *International Journal of Greenhouse Gas Control*, 4(5), pp. 785-797.
- Roussanaly, S., Bureau-Cauchois, G. & Husebye, J., 2013. Costs benchmark of CO₂ transport technologies for a group of various size industries. *International Journal of Greenhouse Gas Control*, Volume 12, pp. 341-350.
- Sircar, S. & Golden, T. C., 2000. Purification of Hydrogen by Pressure Swing. *Separation Science and Technology*, 35(5), pp. 667-687.
- Witkowski, A. S. & Majkut, M., 2012. The impact of CO₂ compression systems on the compressor power required for a pulverized coal-fired power plant in post-combustion carbon dioxide sequestration. *The Journal of Committee on Machine Building of Polish Academy of Sciences*, 59(3), pp. 343-360.
- Yu, H. et al., 2011. Results from trialling aqueous NH₃ based post-combustion capture in a pilot plant at Munmorah power station: Absorption. *Chemical Engineering Research and Design*, 89(8), pp. 1204-1215.

Life Cycle Optimization of Hybrid Energy Systems towards Carbon Neutrality

Xueyu Tian, Fengqi You

Cornell University, Ithaca, New York, 14853, USA

xt93@cornell.edu

Abstract

In this work, we propose a novel superstructure of the hybrid energy systems with geothermal-natural gas-hydroelectricity, biomass and earth source heating, and deep water source cooling (DWSC). Based on the superstructure, a multi-period mixed-integer nonlinear fractional programming (MINFP) model is formulated to address the optimal design of the proposed hybrid energy systems to determine the optimal energy system configuration and operational conditions corresponding to the lowest leveled cost of energy, and simultaneously assess the life cycle carbon footprint. To tackle the computational challenges stemming from the combinatorial nature and nonconvexity of the resulting MINFP problem, a tailored global optimization algorithm integrating two state-of-the-art algorithms, namely the parametric algorithm and the branch-and-refine algorithm, is employed. The applicability of the proposed optimization framework for the hybrid energy systems is illustrated through the application based on Cornell campus hybrid energy system for accommodating the seasonal demand of cooling, heating, and power across the campus.

Keywords: superstructure optimization, life cycle assessment, hybrid energy systems.

1. Introduction

Approximately 78%-80% of the commercial energy around the globe comes from fossil fuels, namely petroleum, coal, and natural gas (Güney and Kaygusuz, 2010). Considering the rapid depletion and environmental concerns with respect to fossil resource utilization, it is imperative to reduce the dependence on conventional fossil resources and shift to carbon-neutral energy (Lawan and Abidin, 2020). With increasing research interest, one promising platform to achieve such carbon-neutral energy transition is hybridization of multiple energy sources. Hybrid energy systems couples two or more energy sources, potentially including fossil fuel, biomass, geothermal energy, or nuclear power, as well as other intermittent energy supplies, such as wind turbine, photovoltaic system, and hydropower (Davis et al., 2018). Hybrid energy systems can be used for, but are not restricted to, electric power generation. Other products from hybrid energy systems include heat, cooling, and hydrogen, etc (Ning and You, 2020). On the one hand, hybrid energy systems can mitigate the problem of intermittency and over-sizing of the overall systems based on single energy source (Beckers et al., 2015). On the other hand, hybrid energy systems that couple various forms of generation together and/or with storage technology may be able to provide more elastic supply to conform with less elastic demand (Dykes et al., 2020). The

engineering reliability, environmental implications, and economic feasibility are the central aspects that must be systematically considered when designing the hybrid energy systems. In this work, we propose a novel superstructure that incorporates geothermal-natural gas-hydro electricity, biomass and earth source heating, and deep water source cooling (DWSC) (Gong and You, 2015). A mixed-integer nonlinear fractional programming (MINFP) model is formulated to minimize the levelized cost of energy and simultaneously evaluate the carbon footprint of the optimal design and operations of the proposed hybrid energy system. The model explicitly considers the seasonality of heat and cooling demand, while the electric power demand is assumed to be relatively constant. To illustrate the applicability of the proposed modelling framework, a case study of Cornell's renewable campus energy systems is considered.

2. Superstructure Generation

In this work, we propose a novel superstructure of hybrid energy systems with geothermal-natural gas-hydro electricity, biomass and earth source heating, and DWSC. The geothermal energy and DWSC are regarded as the primary heating/cooling source. Occasionally, these major energy sources cannot completely accommodate the heating and cooling demand, or their availabilities are limited due to the geographical conditions. Therefore, conventional cooling technologies and heating supplement based on renewable biomass are considered as the auxiliaries to supply the peak heating and cooling demand (Yue et al., 2014). The overall superstructure of the proposed hybrid energy system is demonstrated in Figure 1. The electric power demand is satisfied by a set of generation technologies, which substantially vary regionally and thus are not completely reflected in the general superstructure.

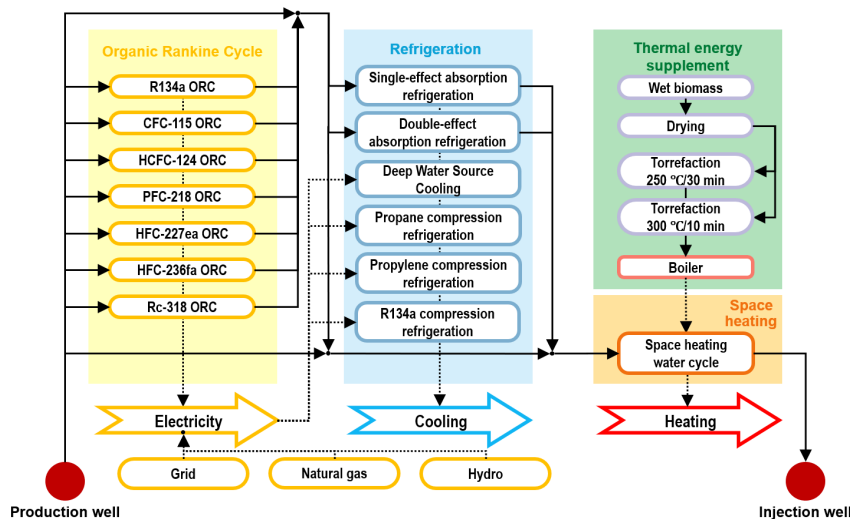


Figure 1. Overview of the proposed superstructure of hybrid energy system with geothermal-natural gas-hydro electricity, biomass and earth source heating, and DWSC.

3. Model Formulation

The general superstructure optimization model is subjected to five groups of constraints, namely, mass balance and configuration constraints, energy balance constraints,

temperature constraints, logic constraints, techno-economic evaluation constraints. The selection of technologies is represented by the integer decision variables. The number of well sets are integer decision variables. Other important decision variables such as the mass flow rates, energy flows, and technology capacities are continuous variables. The objective function, $LCOH$, includes integer variables such as the numbers of production wells and injection wells, and thus is a mixed-integer fractional term. The other nonlinear terms mainly come from the separable concave terms induced by economies of scale, as well as the bilinear terms. Therefore, the resulting problem is a MINFP problem. The general form of this MINFP problem is summarized as follows.

$$\min LCOH = \frac{\sum_{t=1}^{pl} (AIC + AOC - CREDIT)_t}{\sum_{t=1}^{pl} \frac{HEAT_t}{(1+r)^t}} \quad (1)$$

s.t. mass balance and configuration constraints; (2)

energy balance constraints; (3)

temperature constraints; (4)

logic constraints; (5)

techno-economic evaluation constraints; (6)

where $LCOH$ denotes the levelized cost of heat. AIC_t , AOC_t , and $CREDIT_t$ refer to the annualized investment costs, annual operating cost, and power and cooling credits in year t , respectively. $HEAT_t$ represents the heat produced in year t . r and pl denote the discount rate and lifetime of the plant.

Global Optimization Algorithm

```

1:  $q \leftarrow 0, obj \leftarrow +\infty, iter_{out} \leftarrow 0;$ 
2: while  $obj \geq tol_{out}$  do
3:    $iter_{out} \leftarrow iter_{out} + 1;$ 
4:    $lb \leftarrow 0, ub \leftarrow +\infty, gap \leftarrow +\infty, iter_{in} \leftarrow 0;$ 
5:   Initialize the parametric problem  $F(q)$  with one-piece approximation functions;
6:   while  $gap \geq tol_{in}$  do
7:      $iter_{in} \leftarrow iter_{in} + 1;$ 
8:     Solve the parametric problem for the optimal solution  $m^*$  and objective function value  $OBJ^*$ ;
9:     Estimate the objective function value  $OBJ'$  of the original problem based on  $m^*$  and concave functions;
10:     $lb \leftarrow \max\{lb, OBJ'\}, ub \leftarrow \min\{ub, OBJ^*\}, gap \leftarrow (ub - lb) / ub;$ 
11:    Update the parametric problem by adding a new partition point  $m^*$ ;
12:  end while
13:   $obj \leftarrow ub, q \leftarrow \left( \frac{\sum_{t=1}^{pl} (AIC + AOC - CREDIT)_t}{\sum_{t=1}^{pl} (1+r)^t} \right) \cdot \left( \frac{\sum_{t=1}^{pl} HEAT_t}{\sum_{t=1}^{pl} (1+r)^t} \right)^*$ ;
14: end while
15: return  $q$ 

```

Figure 2. The pseudocode of the tailored global optimization algorithm integrating the parametric algorithm and the branch-and-refine algorithm.

LCA is also conducted based on the optimization results. The goal of the LCA study is to quantitatively assess the life cycle environmental impacts of the optimal designs. Heat is widely used for residential, commercial, and manufacturing purposes, with approximately 50% of the total heat demand coming from the residential sector (McCabe et al., 2016). Therefore, it is worth looking into the environmental performance of heat generation, which arouses special research interests in the

literature. The functional unit of this LCA is defined as 1 kWh of heat generated. The impact category to evaluate the life cycle environmental performance of the hybrid energy system is dedicated to the global warming potential (GWP) to examine the extent of carbon neutrality. GWP is calculated using the IPCC 2013 method with 100-year time frame (Stocker et al., 2013).

The resulting MINFP problem for the optimal design of the proposed hybrid energy systems is computationally challenging due to its combinatorial nature and nonconvexity. In order to tackle the computational challenge, a tailored global optimization algorithm integrating two state-of-the-art algorithms, namely the parametric algorithm (Zhong and You, 2014) and the branch-and-refine algorithm (You and Grossmann, 2011), is applied. The pseudocode of the tailored global optimization algorithm is given in Figure 2, where a dual-loop structure can be identified.

4. Application to Cornell Campus Energy System

The proposed superstructure optimization model is applied to tackle the optimal design of an envisioned hybrid energy system with lake source cooling (LSC), hybrid earth source and biomass heating, as well as geothermal binary, natural gas and hydro-based power for meeting the seasonal demand of cooling, heating and power of the facilities (Tian et al., 2019). Historical demand data for Cornell's campus in Ithaca, New York are used for developing the case study. We address the problem over 12 time periods, with each corresponding to one month. The geothermal fluid is extracted from at most five production wells and the deep lake water is pumped from the Cayuga Lake. Willow is chosen as the feedstock for torrefaction, because it is one of the main commercial woody bioenergy crops in U.S. All computational experiments are conducted on a PC with Intel (R) Core (TM) i7-6700 CPU @ 3.40GHz and 32GB RAM. The superstructure optimization model and solution procedures are coded in GAMS 33, with CPLEX 12.10 as the MILP solver and BARON (version 20.10.16), SCIP (version 7.0) being the mixed-integer nonlinear programming (MINLP) solvers. The relative optimality tolerances are all set to 10^{-6} .

The case study considers at most five production and five reinjection wells, because approximately five well sets are sufficient to meet base-load heating demand for Cornell's Ithaca campus (Beyers, 2018). The mass and energy balances for ORCs and compression refrigeration cycles are extracted from the HYSYS simulation results. The performance and cost data for the absorption refrigeration are retrieved from the literature. Mass and energy balance information for the biomass torrefaction plant is reported in existing publications. The standard levelized cost model is adopted with a constant discount rate of 5% (Lee et al., 2019). According to (Beckers et al., 2015), the geothermal gradient in Ithaca is assumed to be 25 °C/km. The well depth is assumed to be 5 km, and the average geofluid production temperature is around 125 °C (Tian et al., 2020). For the DWSC system, deep lake water (250 feet deep) at around 3.89 °C is pumped to a heat exchanger at the shore, which returns gradually to the lake at a temperature no higher than 12.8 °C (Facilities and Campus Services, 2005). The space heating supply/temperature is set to 82 °C/72 °C and 71 °C/65 °C at high and low heat demand, respectively.

For a better comprehension of the optimal design in terms of *LCOH*, the breakdown details of the *AOC* and the *AIC* are shown in Figure 3(a). *AIC* is calculated with the same project lifetime (20 years) and discount rate (5%) as the *LCOH* model. The sum of

AOC and *AIC* is the total annualized cost (*TAC*). The *TAC* for the optimal design is \$44.44 MM/yr, with *AOC* accounting for 55.5%. The major contribution from the natural gas turbine (61%) and geothermal well drilling (19%) can be easily observed in terms of capital investment. Moreover, the breakdown of the annual GWP of the optimal solutions is shown in Figure 3(b). Emissions from the gas turbine (up to 118.12 kton CO₂-eq/yr) dominates the annual GWP, while there are no direct emissions from hydro power. We note that the selection of technologies varies from month to month, as well as the operating level. We focus on three months, namely January, May, and August to represent the cold winter days, average weather conditions, and hot summer days, respectively. The DWSC is able to accommodate all the cooling demand in most months other than August, in which compression refrigeration cycle using R134a as working fluid is selected as the auxiliary cooling supplier. In January, torrefied wood is combusted to provide extra hot water for space heating.

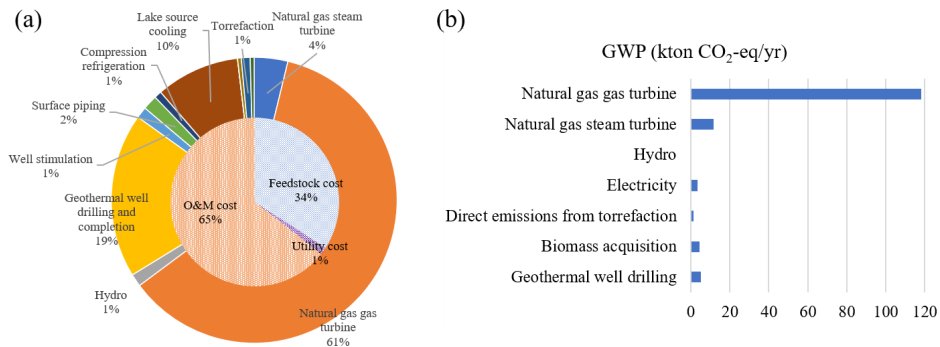


Figure 3. (a) Breakdowns of the annual operating cost (inner pie chart) and the annualized investment cost (outer donut chart); (b) Annual GWP.

Table 1. Computational performance for the case study using BARON (version 20.10.16), SCIP (version 7.0) and the tailored global optimization algorithm.

Solver	Discrete Variables	Continuous Variables	Constraints	Objective Value (\$/MMBTU)	Solution Time (CPUs)
BARON (version 20.10.16)	26,277	3,856	81,746	16.12	835.16
SCIP (version 7.0)	26,277	3,856	81,746	N/A ^a	-
Global Optimization Algorithm	26,294	3,874	81,806	16.12	24.13

^aNo upper/low bound is returned when reaching the time limit of two hours.

To demonstrate the computational performance, the proposed superstructure optimization problem is solved by general-purpose MINLP solvers BARON and SCIP, as well as the tailored global optimization algorithm. Table 1 summarizes the computational results for the case study. BARON and the tailored global optimization algorithm return the identical optimal solution. BARON returns the optimal solution within 835.16 CPU seconds, while the global optimization algorithm only requires 24.13 CPU seconds. We note that there is no upper and low bound information returned by SCIP solver before the program reaches the computation time limit of two hours. Overall, the tailored optimization algorithm is more efficient and is less dependent on initial values than general-purpose solvers. Since there is uncertainty embedded in the hybrid energy systems design problem including the seasonal utility demands in coming

years, the deterministic optimization problem proposed in this work that overlooks uncertainty may result in suboptimal or even infeasible solutions in certain uncertainty realization. It is crucial to hedge against such uncertainty using systematic approach in the future work after relevant data become available.

5. Conclusion

A novel superstructure optimization model was developed to assess the economic feasibility and environmental implications of the proposed hybrid energy systems with geothermal-natural gas-hydro electricity, biomass and earth source heating, and DWSC. The applicability of the proposed optimization framework was illustrated through a case study based on the Cornell campus energy systems. The resulting levelized cost is as low as \$16.12/MMBTU and the proposed tailored optimization algorithm showed high performance in solving the superstructure optimization model that outperformed the general-purpose solvers.

References

- K. F. Beckers, M. Z. Lukawski, G. A. Aguirre, S. D. Hillson, J. W. Tester, 2015, Hybrid Low-Grade Geothermal-Biomass Systems for Direct-Use and Co-Generation, 40th Workshop on Geothermal Reservoir Engineering Proceedings, Stanford University, Stanford, CA.
- S. Beyers, 2018, Water Thermal Storage at Cornell University, Personal communication, Dec 11.
- S. J. Davis, N. S. Lewis, M. Shaner, S. et al., 2018. Net-zero emissions energy systems. *Science*, 360, 6396.
- K. Dykes et al., 2020, Opportunities for Research and Development of Hybrid Power Plants, NREL Technical Report, Unpublished (forthcoming 2019).
- Facilities and Campus Services, 2005, How Lake Source Cooling Works. Available: <https://energyandsustainability.fs.cornell.edu/util/cooling/production/lsc/works.cfm>.
- J. Gong, F. You, 2015, Sustainable design and synthesis of energy systems. *Current Opinion in Chemical Engineering* 10, 77-86.
- M. Güney, K. Kaygusuz, 2010, Hydrokinetic energy conversion systems: A technology status review. *Renewable and Sustainable Energy Reviews* 14, 2996-3004.
- S. M. Lawan, W. A. W. Z. Abidin, 2020, A Review of Hybrid Renewable Energy Systems Based on Wind and Solar Energy: Modeling, Design and Optimization. In *Wind Solar Hybrid Renewable Energy System*, IntechOpen.
- I. Lee, J. W. Tester, F. You, 2019, Systems analysis, design, and optimization of geothermal energy systems for power production and polygeneration: State-of-the-art and future challenges. *Renewable and Sustainable Energy Reviews*, 109, 551-577.
- K. McCabe, M. Gleason, T. Reber, and K. R. Young, 2016, "Characterizing US Heat Demand for potential application of geothermal direct use," *GRC Transactions*, vol. 40.
- T. Stocker et al., 2013, IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I, Cambridge Univ. Press, Cambridge, UK.
- X. Tian, F. You, 2019, Carbon-neutral hybrid energy systems with deep water source cooling, biomass heating, and geothermal heat and power. *Applied Energy*, 250, 413-432.
- X. Tian, T. Meyer, H. Lee, F. You, 2020, Sustainable design of geothermal energy systems for electric power generation using life cycle optimization. *AIChE Journal*, 66, e16898.
- F. You, I. Grossmann, 2011, Stochastic inventory management for tactical process planning under uncertainties: MINLP models and algorithms, *AIChE Journal*, 57, 1250-1277.
- D. Yue, M. A. Kim, F. You, 2013, Design of Sustainable Product Systems and Supply Chains with Life Cycle Optimization Based on Functional Unit. *ACS Sustainable Chemistry & Engineering*, 1, 1003-1014.

- D. Yue, M. Slivinsky, J. Sumpter, et al., 2014, Sustainable Design and Operation of Cellulosic Bioelectricity Supply Chain Networks with Life Cycle Economic, Environmental, and Social Optimization. *Industrial & Engineering Chemistry Research*, 53, 4008-4029.
- N. Zhao, F. You, 2020, Can renewable generation, energy storage and energy efficient technologies enable carbon neutral energy transition? *Applied Energy*, 279.
- Z. Zhong, F. You, 2014, Globally convergent exact and inexact parametric algorithms for solving large-scale mixed-integer fractional programs and applications in process systems engineering, *Computers & Chemical Engineering*, 61, 90-101.

Value of aggregator-led community energy storage systems with degradation constraints

Jude O. Ejeh^a, Flora A. V. Biggins^a and Solomon F. Brown^{a,*}

^a*Department of Chemical & Biological Engineering, The University of Sheffield, Mappin Street, Sheffield, S1 3JD, United Kingdom
s.f.brown@sheffield.ac.uk*

Abstract

In this work, we present a mixed integer linear programming (MILP) model to assess the economic value an aggregator-led community electrical energy storage system provides, with the inclusion of online degradation constraints. Studies have shown that community energy storage devices present additional benefits to end users, such as an increased renewable generation self-consumption, peak shaving and reduced electricity costs, when compared to individually-owned devices. Further benefits may also be realised by giving a load aggregator partial access to the device to participate in the wholesale energy market, however previous studies have mostly assessed such benefits solely from an aggregator or community point of view. Also, degradation considerations, when considered, were implemented offline via a rolling horizon approach. The proposed model in this work includes online degradation considerations over the entire planning horizon to obtain the optimal schedule of the storage device. The community consists of a number of individual participants each having local renewable energy generation and batteries installed behind-the-meter, with the ability to export energy back to the grid. The LA has partial access to the community energy storage devices to participate in the UK day-ahead energy market. The model assesses the economic value a LA can realise under different storage sharing quotas with an overall goal to maximise profits without infringing on the interests of its participants. Results show that both the community and aggregator stand to gain with certain storage device and profit sharing quotas.

Keywords: Community energy storage, Energy arbitrage, Battery degradation, Mixed integer optimisation

1. Background

Electrical energy storage (EES) technologies have been shown to improve power system reliability, flexibility and security (Ma et al., 2018). This is more important in the present day as there is a growing reliance on intermittent renewable energy sources (RES) - solar and wind energy - owing to increasing commitments to the decarbonisation and decentralisation of the electric grid. EES devices have thus been identified as the “missing piece” in the integration of such RES (Jankowiak et al., 2020) towards reduced greenhouse gas emissions.

EES devices also have the advantage of being easily deployed both in front-of-meter and behind-the-meter (BTM) applications. Once deployed, they can be controlled in a centralized – by a large energy participant e.g. a system operator, load aggregator (LA) –

or a decentralized – by small, privately owned participants – manner (Castagneto Gisse et al., 2019). Decentralized devices may also be pooled and centrally controlled in what may be referred to as community energy storage (CES) (Dong et al., 2020). Studies have shown that sharing EES devices increase RES self-consumption, participant self-sufficiency, peak demand shaving and lowers costs for individual participants (Dong et al., 2020; Roberts et al., 2019). Such CES devices may also be connected to a LA. LAs give additional advantages of access to energy markets and ancillary services markets, increasing revenue potential and system reliability for a fee (Zhou et al., 2019).

A number of authors have proposed techno-economic models for scheduling CES devices in order to determine the benefits to end-users/participants. Parra et al. (2015) proposed a simulation-based model to assess current and projected performance and benefits of Lithium ion (Li-ion) and Lead-acid (PbA) batteries, showing that a community approach reduced the projected levelised cost by as much as 37%. Schram et al. (2020) considered a multi-objective optimisation-based (MOO) approach for the community where cost and CO₂ emission objectives were explored assuming participation in the day-ahead (DA) market. Terlouw et al. (2019) also proposed a MOO approach using different EES technologies but from the perspective of a LA, exploring scenarios of energy arbitrage with/without peak shaving, with Li-ion battery technology showing the best economic and environmental performance. Cycle-induced battery degradation was considered by Pimm et al. (2019) using a rolling horizon approach in a linear optimisation model for CES devices. Degradation considerations are important as the performance of EES devices has been shown to deplete with time, rate and frequency of charging/discharging and operating temperature (Ciez and Whitacre, 2016). Hence, including such considerations presents a more accurate estimation of the benefits to end-users.

In each of the aforementioned studies, and others, benefits - cost and/or CO₂ emission reduction - were almost always solely assessed from the LA's or community's perspective. In reality, shared ownership also exists. Communities may offer partial access of their CES device to a LA with access to additional energy services in exchange for a percentage of the profit. Hence, it becomes important for considerations to be made from both the community's and aggregator's perspective in order to ascertain sharing quotas that profits both parties. To this end, an optimisation model that minimises the total cost of a CES device is proposed. The CES device serves to increase RES consumption within the community, reduce local energy consumption as well as participate in the DA market through a LA. Using different device access and profit sharing quotas, the model seeks to quantify the economic benefits to both the community members and LA. 'Online' battery degradation is also considered in a linear formulation, as opposed to post-processing such values in a rolling horizon approach.

In the rest of the paper, the optimisation problem is outlined in section 2 and the proposed model described in section 3. Results obtained from applying the model to a case study (section 4) are discussed in section 5 with final remarks given in section 6.

2. Problem description

For the system illustrated in Figure 1, the optimisation problem is described as follows.

Given a community with a known number of domestic households (i) each with its energy demand (P_{it}^l), renewable energy generation (G_{it}), partial ($x\%$) access to a CES device with capacity, \hat{C}^{max} , and power, \hat{P}^{max} , and able to purchase power from a supplier at price, C_t^p ,

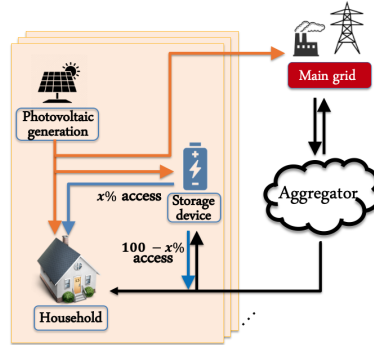


Figure 1: System description

to satisfy local demands as well as feed-in excess generation back to the grid at a price, C_t^s ; a LA (j) with partial access ($100 - x\%$) to the pooled CES device participating in the DA market;

Determine the optimal CES device schedule - charging (P_{it}^{b+}, P_{jt}^{b+}) and discharging (P_{it}^{b-}, P_{jt}^{b-}) power per time - so as to minimise the total cost of electricity purchased for the community and traded by the LA.

It is assumed that the allotted CES device quota to the community and LA are operated independently, and DA market prices (C_t^{DA}) are perfectly forecasted. The latter assumption is justified as DA prices tend to follow similar daily patterns and can be predicted with a high degree of accuracy.

3. Model description

The models for both the community and LA seek to minimise the total net cost of electricity purchased, and are described as follows. For the community, the objective function is given by eq. (1), which is the sum of the cost of power imported (P_{it}^{imp}) less the revenue generated from the amount fed into the grid (P_{it}^{exp}). The amount of power imported/exported is determined using eq. (2) with an additional restriction that the battery is solely charged using excess energy from the RES. The state of charge (SOC) of the battery (SOC_{it}) at each time period is calculated as the net power from charging and discharging the battery, where η^c and η^d are the charge and discharge efficiencies of the battery. Bounds on the battery power and SOC are enforced using eqs. (4) and (5).

$$\min \sum_{it} C_t^b \cdot P_{it}^{imp} - C_t^s \cdot P_{it}^{exp} \quad (1)$$

$$P_{it}^{imp} - P_{it}^{exp} = P_{it}^l + P_{it}^{b+} - P_{it}^{b-} - G_{it} \quad \forall i, t \quad (2)$$

$$SOC_{it} = SOC_{i,t-1} + \eta^c \cdot P_{it}^{b+} - \frac{P_{it}^{b-}}{\eta^d} \quad \forall t \quad (3)$$

$$P_{it}^{b+}, P_{it}^{b-} \leq x \hat{P}^{max} \quad \forall i, t \quad (4)$$

$$x SOC^{min} \bar{C}^{max} \leq SOC_{it} \leq x SOC^{max} \bar{C}^{max} \quad \forall i, t \quad (5)$$

The objective function for the LA is given by eq. (6) - to minimise the total cost obtained from trading in the DA market using the remaining portion ($100 - x\%$) of the CES device subject to eqs. (3) - (5) for the LA (j). It is assumed that the total quantity of electricity

traded by the LA has negligible effect on market prices, hence a 'price taker' strategy.

$$\min \sum_{jt} C_t^{DA} \cdot (P_{jt}^{b+} - P_{jt}^{b-}) \quad (6)$$

Battery degradation is modelled according to the power law relationship given by Ciez and Whitacre (2016). These expressions are given in eq. (7) for Li-ion and PbA batteries, where \hat{C}_t represents the fraction of the total capacity degraded at time t by reason of the SOC swing (SOC_t^{sw}) from battery actions in the previous time periods. The SOC swing is determined as the actual depth of discharge over the past 24 hours of battery operation, and the degradation models are implemented via a piecewise linear function.

$$\hat{C}_t = \begin{cases} \left(\frac{SOC_t^{sw}}{1307.4} \right)^{0.95}, & \text{Li_ion} \\ \left(\frac{SOC_t^{sw}}{12.838} \right)^{1.838}, & \text{PbA} \end{cases} \quad \forall t \quad (7)$$

Additional constraints are also included for members of the community and the LA to ensure that total energy discharged from the battery at each time period is always less than or equal to the net amount contributed to charging the battery. Hence, each participant may only use stored energy up to the net amount contributed in previous time periods.

4. Case study

For a case study, we consider a community with 1000 households each having a 3kWp PV cell and 1kW/2kWh Li-ion battery installed. Load profiles for groups of households for the month of January, 2019 were randomly generated using data from Elexon 10-year average load profile for domestic consumers in the United Kingdom (UK). The solar energy generated by each household is used to reduce household demands, the excess of which can either be exported to the grid at a fixed rate of 5.5p/kWh or used to charge the battery. Each of these batteries are pooled to form a CES device, a fraction ($x\%$) of which can be offered up to a LA in exchange for a percentage ($y\%$) of the profits realised from trading in the UK DA market. Households also purchase electricity to satisfy local loads using a time of day tariff. For this study, the EKO tariff (Greenenergy) is adopted.

These considerations are implemented using Pyomo 5.6.8 and solved with Gurobi 9.0 solver and an optimality gap of 1%. The piecewise linearization was implemented using Pyomo's "Piecewise function" with 11 equidistant SOC swing values ranging from 0 - 1 inclusive. In order to ascertain the value to both the community and LA, storage quotas ranging from 0 - 100%, with a step size of 20, were used to obtain results on the minimum costs to each party with/without degradation considerations.

5. Results & Discussion

Figure 2 shows the total household savings under differing CES device quotas and LA profit sharing percentages. The model was run for each scenario obtaining globally optimal solutions in less than 40s (20s for cases without degradation considerations). The total savings, with values indicated, are the sum of the average household savings and a percentage of the LA's profit. A CES device quota value of $x = 0\%$ corresponds to a case where the LA has full access, and the community members may only import power to satisfy residual load demand, or export excess solar energy at the feed-in price. This is also the base household cost value, below which the total household savings for the

community is calculated. A CES quota of 100% implies sole usage of the CES device by the community, and is the point above which additional economic benefit is realised by the community owing to LA participation.

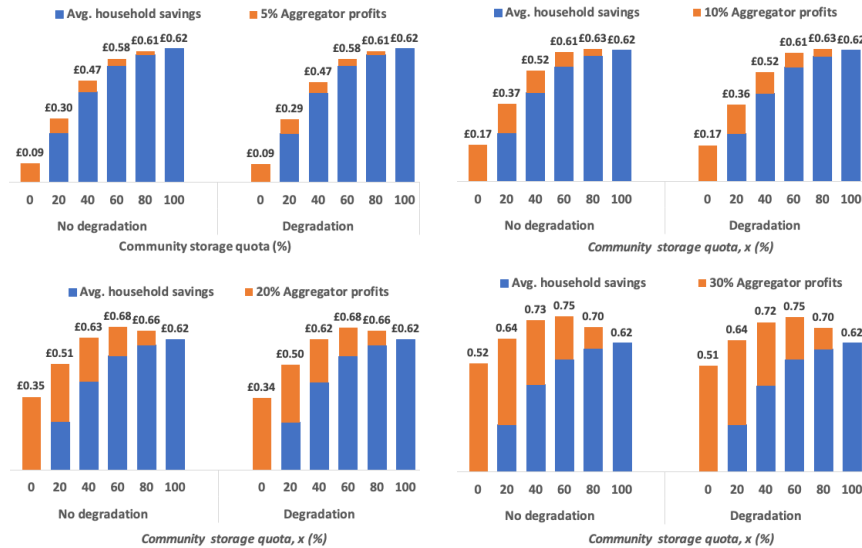
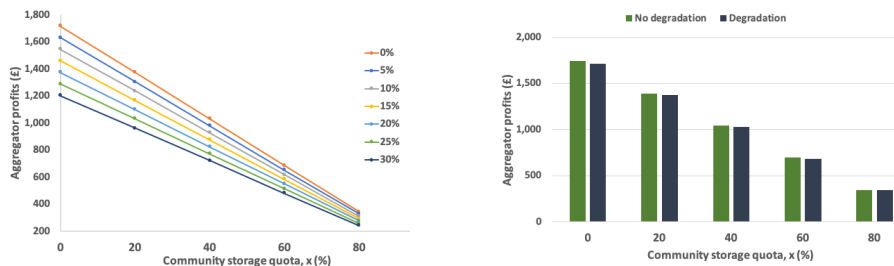


Figure 2: Total household savings



(a) Profits per revenue sharing quota (y%)

(b) Profits with/without degradation

Figure 3: Load aggregator profits

It can be first observed that for a 5% profit payment, the community stands to lose in giving the LA any form of access to the CES device. With increasing incentives payment percentages, the quota the community can let go of for a profit also increases up to 80% (community storage quota value of 20%) where 30% of the LA profits are remitted. Figure 3a also shows that the LA stands to gain from an increased profit sharing quota. This is as the community is more likely to give greater access to the CES device which leads to an increased overall profit for the aggregator.

The impact degradation has on the total household savings is barely seen as it is spread across all households for the period of consideration. Degradation considerations are better observed in the LA profits shown in Figure 3b where with higher access by the LA, and therefore greater use, the overall profits realised reduce. This is attributed to a reduction in available capacity owing to degradation over the period of consideration, and

represents a more realistic estimation of actual earnings.

6. Conclusion

In this work, the economic benefits an aggregator-led community EES device could realise were assessed from both the community's and LA's point of view. To achieve this a mixed integer linear programming (MILP) model was proposed to obtain the optimal schedule for a CES device part accessed by the community and LA. The community is assumed to consist of a set of households with energy demands, local generation and the ability to feed-in to the grid. Online degradation constraints were also included to accurately capture the final value to both parties based on use of the CES device.

Using differing CES device access quotas and LA profit sharing percentages, the economic value to the community - in terms of total household cost savings - and LA - in terms of total profits - were investigated. Results showed that both parties stand to gain at a LA profit sharing value above 5% under differing storage quota ranges. Conditions in which the community ran at a loss were also identified, and the impact of degradation, especially on the LA's profit, was highlighted.

Acknowledgement

This work was supported by the Energy Open Piazza - Power Forward Challenge project. The authors are grateful to the Department of Business, Energy & Industrial Strategy (BEIS), UK and Drax Power Ltd for their financial support. The final author would like to acknowledge the support of the Royal Academy of Engineering through the Industrial Fellowship (Reference: IF\192046).

References

- Castagneto Gissey, G., Subkhankulova, D., Dodds, P. E., Barrett, M., 2019. Value of energy storage aggregation to the electricity system. *Energy Policy* 128, 685–696.
- Ciez, R. E., Whitacre, J. F., 2016. Comparative techno-economic analysis of hybrid micro-grid systems utilizing different battery types. *Energy Conversion and Management* 112, 435–444.
- Dong, S., Kremers, E., Brucoli, M., Rothman, R., Brown, S., 2020. Techno-enviro-economic assessment of household and community energy storage in the UK. *Energy Conversion and Management* 205, 112330.
- Jankowiak, C., Zacharopoulos, A., Brandoni, C., Keatley, P., MacArtain, P., Hewitt, N., 2020. Assessing the benefits of decentralised residential batteries for load peak shaving. *Journal of Energy Storage* 32, 101779.
- Ma, T., Shen, L., Li, M., 2018. Electrical Energy Storage for Buildings. In: *Handbook of Energy Systems in Green Buildings*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1079–1107.
- Parra, D., Gillott, M., Norman, S. A., Walker, G. S., 2015. Optimum community energy storage system for PV energy time-shift. *Applied Energy* 137, 576–587.
- Pimm, A. J., Palczewski, J., Morris, R., Cockerill, T. T., Taylor, P. G., 2019. Community energy storage: A case study in the UK using a linear programming method. *Energy Conversion and Management* 205.
- Roberts, M. B., Bruce, A., MacGill, I., 2019. Impact of shared battery energy storage systems on photovoltaic self-consumption and electricity bills in apartment buildings. *Applied Energy* 245, 78–95.
- Schram, W. L., AlSkaif, T., Lampropoulos, I., Henein, S., van Sark, W. G., 2020. On the Trade-Off Between Environmental and Economic Objectives in Community Energy Storage Operational Optimization. *IEEE Transactions on Sustainable Energy* 11, 2653–2661.
- Terlouw, T., AlSkaif, T., Bauer, C., van Sark, W., 2019. Multi-objective optimization of energy arbitrage in community energy storage systems using different battery technologies. *Applied Energy* 239, 356–372.
- Zhou, X., Shi, J., Li, S., 2019. Optimal control strategy of load aggregators with demand response. *The Journal of Engineering* 2019, 1033–1036.

Clean Electricity Generation from Sewage: a novel Ultra-Supercritical Water Oxidation Technology with integrated CCS

Sanjay Mukherjee,^a Solomon Brown^{a,*}

^a*Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom*

s.f.brown@sheffield.ac.uk

Abstract

Waste management and clean electricity generation are amongst the most critical challenges and major barriers towards achieving sustainability. Ultra-Supercritical water oxidation (USWO) is an emerging technology that can possibly provide a solution for mitigating both these issues. It inherently allows the use of low-grade fuels, such as waste, for electricity generation with high efficiency. However, the true potential of USWO technology has not been extensively reported in the existing literature. This study focuses on modelling and optimisation of a novel USWO process working on the principles of Brayton cycle for electricity generation using sewage as fuel. The USWO process is integrated with carbon capture and storage (CCS) technology to reduce CO₂ emissions. The process design also incorporates water recovery facility from the sewage used as the fuel stream. An advanced system level process model is developed in Aspen Plus simulator using experimentally derived inputs together with FORTRAN programs. A sensitivity analysis is conducted to determine the impact of reactor temperature on the net electrical efficiency. The results indicate that the USWO technology can generate electricity from sewage with net electrical efficiency of 31.5%. It is also observed that integration of CCS can reduce the CO₂ emissions to zero with an efficiency penalty of 3.5% points. This efficiency penalty is comparable to the best performing CCS technologies for conventional gas- and coal-fired power plants. The sensitivity results show the electricity output is proportionate to reactor temperature and is primarily restricted by the material properties of the USWO reactor. The results also indicate nearly complete recovery of water from the sewage without any additional processing or energy requirement. Furthermore, the study also provides a systematic methodology to test other low-grade fuels (organic wastes) for the proposed USWO technology.

Keywords: Sewage Treatment, electricity generation, CO₂ capture, process modelling.

1. Introduction

With the rising world population, waste production has also soared substantially. Landfilling of waste can produce undesirable greenhouse gases, odour and leachate, which can contaminate the environment (Liu et al. 2017). Therefore, in the waste management hierarchy, landfilling is unquestionably the most unfavourable choice (Lukumon et al. 2013). Existing biochemical waste conversion processes are slow and have low conversion rates (Tyagi, 2013). Waste incineration, pyrolysis and gasification are attractive technologies that can generate waste-derived products such as electricity,

heat and synthetic fuels (Nakakubo et al. 2017, Lu et al. 2017, Gabriele and Siglinda, 2020). However, high moisture content in waste, due to its type (sewage and food waste) or lack of segregation at the source, means a considerable amount of energy intensive pre-processing required to make the waste suitable for use in the incinerator, gasifier or pyrolyser. This significantly impacts the overall efficiency of the process and can make the technology financial unviable in most cases (Syed-Hassan et al., 2017).

Supercritical water oxidation (SWO) is an emerging technology which has the potential to convert wastes with high moisture and low energy content to electricity, heat and synthetic fuels, more efficiently (Bermejo and Cocero, 2006). In a SWO technology, the oxidation of organic waste takes place in a supercritical water environment, which is an excellent non-polar solvent that dissolves the organic substance present in the sludge, thus, supporting the decomposition and reaction. Additionally, the heavy metals in the solid residues gets stabilised after the SWO (Hantoko et al., 2019). Sulphur and nitrogen in the waste gets deposited as inorganic salts in the supercritical water which means no NO_x or SO_x are released in the reactor exhaust (Guo et al., 2015). As the moisture in the waste becomes supercritical water in the reactor and helps in the conversion process, there is no requirement for waste pre-processing for moisture reduction.

Several experimentally derived reactor designs and process models for Supercritical Water Gasification or Oxidation has been extensively covered in the literature (Azadi and Farnood, 2011; Okolie et al., 2019). However, very limited studies on the USWO of waste for electricity generation is available. Moreover, almost all the published literature on SWO uses Rankine cycle for electricity generation (Bermejo and Corcero, 2006; Donatini et al., 2009, Tang et al., 2020). This work conducts a detailed performance evaluation of a novel USWO process for electricity generation using Brayton cycle, with integrated oxy-combustion CO_2 capture technology, based on electrical output, net electrical efficiency and CO_2 emissions. The work also estimates the energy penalties occurred due to carbon capture through modelling and simulation using common modelling assumptions.

2. Methodology

This section describes the assumptions and input parameters used to develop the model of the proposed USWO process.

2.1 Feedstock type

Sewage sludge contains high organic matter and therefore, can be potentially used as an energy source for electricity generation. This study uses sludge from wastewater treatment plants as feedstock in the USWO process. The characteristics of the sewage sludge used in the model are given in Table 1.

Table 1: Proximate and Ultimate analysis (wt % dry basis) of sewage (Hantoko et al., 2019).

Items	Proximate Analysis	Items	Ultimate Analysis
Fixed carbon	2.29	Carbon	18.94
Volatile matter	35.14	Hydrogen	2.21
Ash	62.57	Nitrogen	2.89
		Sulphur	0.6
Lower heating value	5.89 MJ/kg	Oxygen	12.79

2.2 Process description

A process model for the USWO of sewage sludge to produce electricity along with carbon capture is developed in Aspen Plus V10 by Aspen Technology Inc. A detailed schematic of the process is shown in Figure 1. The overall process primarily consists of five key stages, (i) Air Separation Unit (ASU), (ii) Feed Stock preparation, (iii) Reactor unit, (iv) electricity generation and (v) CO₂ capture.

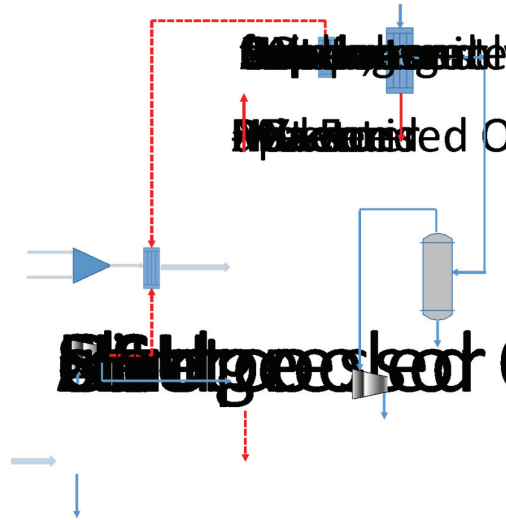


Figure 1: Schematic of the USWO process of sludge with carbon capture.

In the reactor of the USWO process, the feedstock is fully oxidised by pure oxygen (O₂) in an exothermic reaction under a supercritical water environment at 300 bar and 700 °C. The O₂ required is produced in an ASU with 95% purity (by volume) at 1.0135 bar. This O₂ is then compressed to the reactor pressure in a series of intercooled compressors. The heat released from O₂ compression is used to preheat the already pressurised reactor feedstock consisting of sewage sludge mixed with make-up water. For a given flowrate of sludge, the reactor temperature can be regulated by varying the flowrate of make-up water in the sludge-water mixture. The high-pressure and hot reactor exhaust, comprising of supercritical water and products of oxidation reactions, is expanded to atmospheric pressure in an expander-generator to produce electricity. The heat in the expander exhaust is used to partially preheat the reactor feed (Sludge + water) to up to 87 °C. The expander exhaust is then sent to a water-cooled condenser unit followed by a flash separator, where most of the condensed water is extracted from the gases. The outlet gases from the flash separator consist of over 90% CO₂ (by weight). The purity of CO₂ in the stream is increased by some pre-processing, followed by CO₂ compression for transport and storage.

2.3 Aspen Plus Model

A MIXCINC stream class is selected to represent the type of components flowing through the flowsheet streams. Components such as CO₂, O₂, methane, hydrogen, H₂O and other gases are entered as ‘Conventional’ in the components list. Sewage Sludge and ash comes in a variety of compositions and therefore, are the only components entered as ‘Non-

conventional' in the list. The Peng-Robinson-Boston-Mathias (PR-BM) is used as the property method for the conventional components whereas, enthalpy models HCOALGEN and DCOALIGT are used for the non-conventional components.

Table 2: Design assumptions and input parameters used for developing the process flowsheet model in Aspen Plus.

Unit operations	Parameters
ASU and O ₂ compression	O ₂ purity: 95% (vol); Energy consumption: 250 kWh/ton of O ₂ ; Compressor efficiency: 83 %; Pressure ratio: 2.97
Air compressor	Pressure ratio: 3.12
Reactor	Pressure: 300 bar; Temperature: 700 °C; Zero pressure drop; Conversion rate: 99.9 %.
CO ₂ compression	Delivery condition: 110 bar and 30 °C; 4 stage compression; Compressor efficiency: 83 %
Expander	Isentropic efficiency: 92 %; Outlet pressure: 0.046 bar for USWO with CCS and 1 bar for USWO without CCS

The Gibbs equilibrium reactor model RGIBBS is used for modelling the reactor. The key inputs needed for RGIBBS reactor model are pressure and heat duty. A multistage MCOMPR model is used to design the compressor units with inputs such as outlet pressure, temperature and efficiency. A 'COMPR' model is used to design the expander. The pressure, isentropic efficiency and mechanical efficiency are the key input parameters to the 'COMPR' model. A 'HeatX' and 'MHeatX' type heat exchanger models are used for modelling the heat exchanger units. A 'PUMP' model is used to pressurise the make-up and cooling water in the process.

3. Results and Discussions

The performance of the USWO of sewage sludge for electricity generation with CO₂ capture is estimated through modelling using Aspen Plus V10. The key performance indicators obtained from simulations are listed in Table 3. The USWO process without CCS generates 1,871 kWh of net electricity with 31.7 % efficiency. In comparison, the USWO process with CCS generates 207 kWh less net electricity with an efficiency of 28.2 % and captures almost all the CO₂ produced in the feedstock conversion process. The lower efficiency of USWO with CCS is mainly due to the significant drop in expander energy output resulting from the reduced mass flowrate of expanding gases caused by the separation of N₂ in its ASU. The expander in USWO with CCS is modelled to partially subdue the energy losses by expanding the gases to a much lower pressure of 0.046 bars. This is only possible since the gases (mainly CO₂) left after condenser and liquid-gas separator are compressed to 110 bars for transportation and storage, instead of venting. Whereas, in USWO without CCS, the expander outlet pressure is set to slightly above atmospheric pressure to overcome any resistance occurred while venting the gases to the environment. It is worth noting that even after reducing the expander outlet pressure, its energy output still experiences a loss of 17.7 %.

Table 3. Plant performance indicators for USWO process with and without CO₂ capture.

Plant Data	Units	With Capture	Without capture
Fuel input energy, LHV (A)	kWh _{th}	5,890	5,890
Expander energy output (B)	KWh	3,126	3,798
Energy consumed in ASU (C ₁)	KWh	504	-
Energy intake in O ₂ compression (C ₂)	KWh	390	-
Energy consumed in Air compression (C ₃)	KWh	-	1,823
Energy consumed in CO ₂ compression (C ₄)	KWh	440	-
Energy consumed by pumps (C ₅)	KWh	128	104
Total auxiliary energy (C=C ₁ +.....+C ₅)	KWh	1,462	1,927
Net electrical output (D = B - C)	KWh	1,664	1,871
Net electrical efficiency (E = D/A x 100)	%	28.2	31.7
Gross electrical efficiency (F = B/A x 100)	%	53.0	64.4
CO ₂ capture efficiency	%	100	-

The design of the USWO process with CCS inherently delivers some additional energy savings achieved by compressing only pure O₂ to 300 bars instead of air, as in the case of USWO process without CCS. Although, the USWO with CCS needs auxiliary energy for the ASU and CO₂ compression unit, this additional energy requirement is much less than the energy saved by avoiding air compression to 300 bars. Table 3 indicates that the inclusion of ASU and CO₂ capture in USWO process does not impose any additional energy penalties than that already imposed by the air compressor, instead, it saves 24 % in auxiliary energy requirement. It is observed that even after one fourth reduction in auxiliary energy, the net electrical efficiency drops by 3.5 % points on adding CCS, which is mainly because of a greater loss (672 kWh) in the expander energy generation than the auxiliary energy savings (465 kWh).

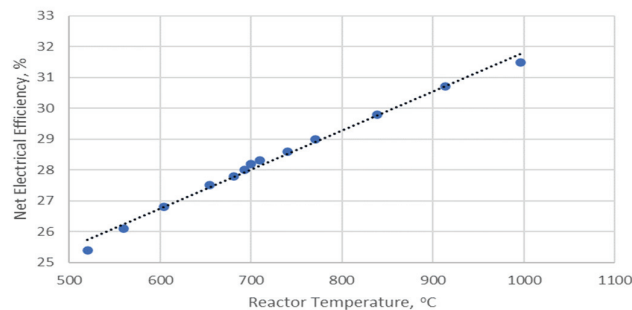


Figure 2: Relation between reactor temperature and net electrical efficiency.

Figure 2 showing the outcome of sensitivity analysis conducted to observe the relation between reactor temperature and net electrical efficiency indicates that a higher reactor temperature will work in favour of the electrical output. Although conventional Brayton cycles are known to operate at temperatures above 1500 °C, the suitability of existing metals and alloys needs to be tested at ultra-supercritical pressures and high temperatures. The outcomes from this sensitivity analysis will be a useful input for suitable reactor

material development and will also help to understand the correlation between additional electricity generation at higher temperatures and associated material cost.

4. Conclusions

This study investigates performance of a novel USWO process that uses sewage as fuel to generate electricity through modelling and simulation. The findings indicate a net electrical efficiency of 28.2 % with 100 % CO₂ capture, which is comparable to some existing coal power plants. The sewage feedstock is free, which makes the technology a useful tool for waste management along with clean and affordable electricity generation.

Acknowledgements

The final author would like to acknowledge the support of the Royal Academy of Engineering through the Industrial Fellowship (reference: IF\192046).

References

- D. Hantoko, Antoni, E Kanchanatip, M. Yan, Z. Weng, Z. Gao, Y Zhong, 2019, Assessment of sewage sludge gasification in supercritical water for H₂-rich syngas production, *Process Safety and Environmental Protection*, 131, pp. 63-72.
- F. Donatini, G. Gigliucci, J. Riccardi, M. Schiavetti, 2009, Supercritical water oxidation of coal in power plants with low CO₂ emissions, *Energy*, 34, pp. 2144-2150.
- G. Centi and S. Perathoner, 2020, Chemistry and energy beyond fossil fuels. A perspective view on the role of syngas from waste sources. *Catalysis Today*, 342, pp. 4-12.
- J. Lu, S. Zhang, J. Hai, M. Lei, 2017, Status and perspectives of municipal solid waste incineration in China: A comparison with developed regions. *Waste Management*, 69, pp. 170-186.
- J. Okolie, R. Rana, S. Nanda, A. Dalai, J. Kozinski, 2019, Supercritical water gasification of biomass: a state-of-the-art review of process parameters, reaction mechanisms and catalysis, 3, pp. 578-598.
- L. Guo, H. Jin, Y. Lu, 2015, Supercritical water gasification research and development in China, *The Journal of Supercritical Fluids*, 96, pp. 144-150.
- L. Oyedele, M. Regan, J. Meding, A. Ahmed, O. Ebohon, A. Elnokaly, 2013, Reducing waste to landfill in the UK: identifying impediments and critical solutions. *World Journal of Science, Technology and Sustainable Development*, 10 (2), pp. 131-142.
- M. Bermejo and M. Cocero, 2006, Supercritical Water Oxidation: A Technical Review, *Thermodynamics*, 52, 11, pp. 2933-3951.
- P. Azadi and R. Farnood, 2011, Review of heterogenous catalyst for sub- and supercritical water gasification of biomass and wastes, *International Journal of Hydrogen Energy*, 36 (16), 9529-9541.
- S. Syed-Hassan, Y. Wang, S. Hu, S. Su, J. Xiang, 2017, Thermochemical pro-cessing of sewage sludge to energy and fuel: fundamentals, challenges and considerations. *Renewable and Sustainable Energy Reviews*, 80, pp. 888-913.
- T. Nakakubo, N. Yoshida, Y. Hattori, 2017, Analysis of greenhouse gas emission reductions by collaboratively updating equipment in sewage treatment and municipal solid waste incineration plants. *Journal of Cleaner Production*, 168, pp. 803-813.
- V. K. Tyagi and S. L. Lo, 2013, Sludge: a waste or renewable source for energy and resources recovery? *Renewable and Sustainable Energy Reviews*, 25, 708-728.
- X. Tang, Y. Zheng, Z. Liao, Y. Wang, J. Yang, J. Cai, 2020, A review of developments in process flow for supercritical water oxidation, *Chemical Engineering Communications*, DOI: 10.1080/00986445.2020.1783537.
- Y. Liu, Z. Ni, X. Kong, J. Liu, 2017, Greenhouse gas emissions from municipal solid waste with a high organic fraction under different management scenarios. *Journal of Cleaner Production*. 147, pp. 451-457.

Design of Food-Energy-Water-Waste Nexus Systems in New York State under COVID-19 Pandemic

Ning Zhao, Fengqi You

Cornell University, Ithaca, New York, 14853, USA

nz225@cornell.edu

Abstract

This article addresses food-energy-water-waste nexus (FEWWN) optimization under the COVID-19 pandemic to alleviate the public health and environmental concerns from increasing food waste generation using waste-to-energy technologies. Food waste generation has noticeably increased during the pandemic across the globe. To alleviate the associated health and environmental concerns, food waste could be converted into electricity and heat through FEWWN systems using waste-to-energy facilities, such as anaerobic digesters and combined heat and power units in wastewater treatment plants and livestock farms. In this work, a multi-period multi-objective optimization model is proposed for the design of efficient nexus systems under various impacts of the pandemic. To illustrate the applicability of the proposed modelling framework, a case study for New York State is presented. The optimized nexus systems could potentially reduce the food waste disposal amounts by 38%. A clear trade-off between the objectives is revealed by the Pareto-optimal solutions. The minimum total cost for the FEWWN system is \$27.1 million; the optimal unit processing profit is \$11.9 per ton processed food waste. Spatial analyses illustrate a strong correlation between facility selections and their processing capacities. Sensitivity analysis revealed that electricity price and biogas yield are the most important factors for the economic objectives.

Keywords: COVID-19, waste-to-energy, nexus optimization, food waste.

1. Introduction

Coronavirus disease 2019 (COVID-19) has been detected in more than 200 countries resulting in tens of millions of confirmed cases and hundreds of thousands of deaths around the world in 2020. Under COVID-19, the food waste amount has increased substantially in the agriculture and residential sectors (Royte, 2020). If the increased amount of food waste is not managed properly, it may cause immense harm to both public health and the environment. In terms of public health, besides the risk of pathogen spread associated with food waste increases (Gerba et al., 2011), more caution and protection measures are required for the waste collectors and recyclers as suggested by the government, due to their potentially contaminating working environment (CDC, 2020). As for the environmental impacts, food waste is responsible for 6% of global greenhouse gas emissions before the COVID-19 outbreak (Poore et al., 2018). One

main cause is landfilling, a conventional method for handling food waste (Kiran et al., 2014), and it has the highest level of greenhouse gas emissions compared to other approaches (Schott et al., 2016). It is important to apply alternative processing technologies to the increased amount of food waste under the COVID-19 pandemic, to prevent the landfilling of increased food waste from further amplifying global warming.

To alleviate both the public health and the environmental concerns, the food waste can be processed through food-energy-water-waste nexus (FEWWN) systems using waste-to-energy technologies (Garcia et al., 2019), rendering it with the potential of serving as a valuable source of sustainable energy (Gong et al., 2015). For instance, anaerobic digestion (AD) has been practically used for the treatment and valorization of food waste (Chen et al., 2017). In AD, food waste is biologically degraded and converted into biogas that mainly consists of methane (60%-70%) and carbon dioxide (30%-40%) (Xu et al., 2018). The biogas produced can be subsequently used as fuel for energy production technologies (Tian et al., 2020), such as combined heat and power (CHP) (Garcia et al., 2016 and 2017) that produces thermal energy and electricity simultaneously. Considering the availability of waste-to-energy facilities represented by AD and CHP in livestock farms (Zhao et al., 2019) and wastewater treatment plants (Bora et al., 2020) across New York State, it is valuable to model and optimize the FEWWN to deal with the increased amount of food waste during the COVID-19 pandemic by including these facilities.

In this work, we address the FEWWN systems optimization on food waste valorization, while considering the impacts of the COVID-19 pandemic on the nexus systems. The impacts include the increasing food waste generation, the relatively short planning horizon, the limited availability of processing facilities, and potential transportation options using repurposed trucks. Under these impacts of the pandemic, the optimization problem focuses on addressing efficient nexus systems that convert the increased amount of food waste into electricity and heating using existing facilities, such as AD and CHP units in wastewater treatment plants and livestock farms. The nexus systems optimization is formulated as a multi-period multi-objective linear fractional program (LFP), with two objective functions of minimizing the total cost and minimizing the unit processing cost of the nexus systems. To illustrate the applicability of the proposed modeling framework, a case study on the FEWWN system for New York State under the COVID-19 pandemic is presented. The major novelties of this work include a novel FEWWN systems optimization modeling framework that considers various impacts of the COVID-19 pandemic on the nexus systems and the uses of existing waste-to-energy facilities and a unique application to a regional FEWWN system for New York State under the COVID-19 pandemic.

2. Problem Statement

The FEWWN system aims to convert the significantly increased amount of food waste due to the COVID-19 pandemic into electricity and heat, using existing waste-to-energy facilities, such as AD and CHP units on wastewater treatment plants (NYSDEC, 2019) and livestock farms (AgSTAR, 2020). In addition to facilitating energy production, such conversion could also alleviate the health and environmental concerns associated with the increasing food waste generation during the pandemic. Food waste feedstocks are collected from each collection site, and then they are transported to processing facilities. Notably, the feedstock could be transported using repurposed trucks, because the

pandemic has reduced the need for professional drivers, such as school bus workers (Guelpa, 2020). The feedstocks transported to a processing facility are converted by the available waste-to-energy technology into intermediate products (Nicoletti et al., 2019), such as biogas. Subsequently, the intermediate products are used as fuel for the energy production technologies in the facility (Ning et al., 2019). For instance, CHP could use the intermediate biogas from AD to produce heat and electricity, simultaneously. Additionally, the food waste that is not transported to waste-to-energy plants would be landfilled. We are given a planning horizon of multiple months, considering that a planning horizon of decades may not be appropriate to depict the COVID-19 pandemic. The planning horizon is partitioned into a set of time periods with identical intervals, and each time period is set to be one month in this study for the case of New York State. Considering the relatively short planning horizon due to the COVID-19 pandemic, the construction of new waste-to-energy facilities to deal with the sudden increases of food waste may not be feasible in the short timeframe, so only existing facilities are considered for food waste processing in this study. In this problem, two objective functions are considered, namely minimizing total cost and minimizing unit processing cost. The total cost depicts the overall cost to deal with the increased amount of food waste under the COVID-19 pandemic, including costs related to food waste processing using waste-to-energy facilities and costs associated with landfills. On the other hand, the unit processing cost represents the cost to acquire, transport, and convert a unit weight of food waste into electricity and heat. In terms of solution strategy, a parametric algorithm (Zhong et al., 2014) is adopted to deal with the fractional objective function, and an ε -constraint method is used to address the Pareto optimal solutions for the objective functions.

3. Results and Discussion

To investigate the potential of converting the increased food waste generation during the COVID-19 pandemic into electricity and heat, we present a case study on New York State, a state that has been influenced significantly by the pandemic. New York State has a 13% increase in monthly food waste generation compared to 2019 (DSNY, 2020). Existing AD and CHP units in wastewater treatment plants and farms in New York State are considered in this study to convert increased food waste into energy.

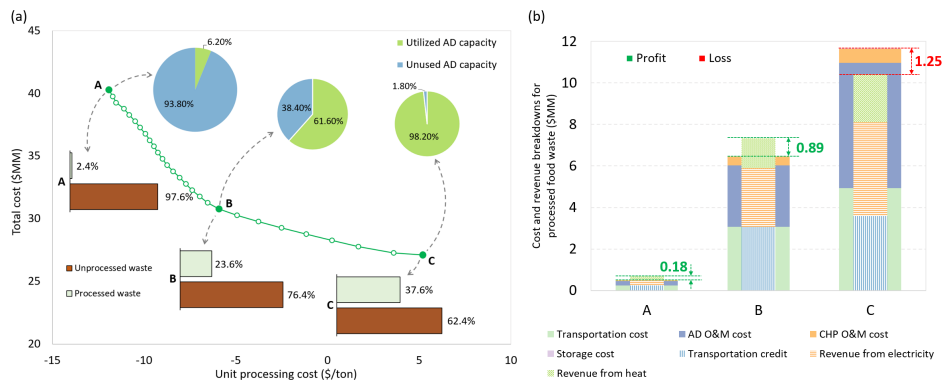


Figure 1. Optimization results for the nexus systems. (a) Pareto-optimal curve illustrating trade-offs between total cost and unit processing cost with the food waste processing level (column charts) and the extent of using existing AD capacities (pie charts). (b) Total economic breakdowns for the three representative Pareto-optimal solutions.

The Pareto-optimal solutions are shown in Figure 1(a), as well as the utilization of AD capacity and the processing levels of food waste. A clear trade-off between the total cost objective and the unit processing cost objective can be found from the Pareto-optimal curve. Notably, a negative unit processing cost indicates that the food waste processing system is profitable. Extreme point A minimizes the unit processing cost; it has the lowest unit processing cost of $-\$11.9/\text{ton}$, while extreme point C minimizes the total cost, so it has the lowest total cost of $\$27.1$ million. Point B represents a trade-off solution between the two extreme points. For the processing levels of food waste, the nexus systems have the potential of converting 37.6% of food waste produced in New York State, using 98% of the existing AD capacities in wastewater treatment plants and livestock farms. In Figure 1(b), the total economic breakdowns on cost and revenue are presented in wide columns and narrow columns, respectively, and the total profit and loss of waste processing are shown on the right-hand side of each column. Transportation costs and O&M costs for AD are the major contributors to the total processing cost for all three solutions, while O&M costs for CHP account for a small proportion of the cost. The profit increase from solution C to solution A could be $\$3.43$ million annually, due to the high transportation and production cost of solution C.

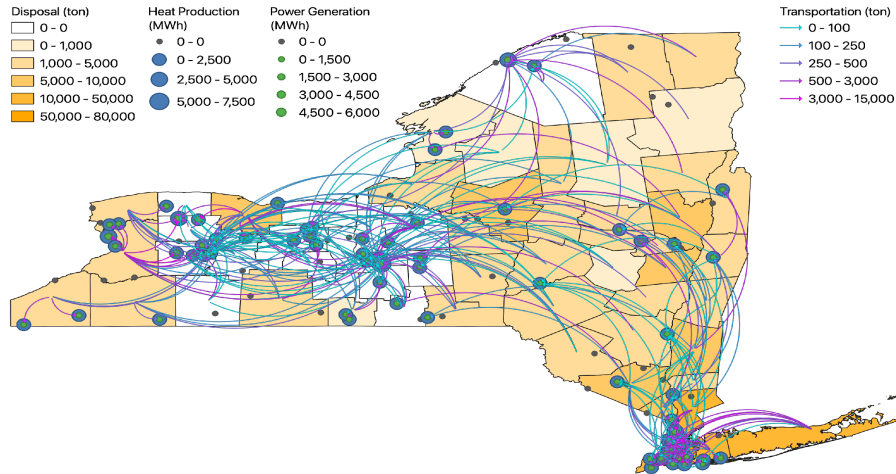


Figure 2. Facility locations, power generation, and heat production from solution C, as well as the corresponding total feedstock transportation and the county-level food waste landfill amounts within the planning horizon.

The spatial information of solution C that minimizes total cost is shown in Figure 2. In terms of facility selection, the number of active wastewater treatment facilities and farms is noticeably larger than those of solutions A and B. Notably, most plants with large capacities are active in solution C, while the small plants tend to be idle. This

trend can be further validated considering the utilization of existing AD capacities as presented in Figure 1(a). For solution C, although 98% of AD capacities are used for food waste conversion, several idle plants still exist as shown in Figure 2, indicating that these inactive plants are small ones that make less than 2% of the total state-level capacities. This is because using plants with large processing capacities could result in low unit O&M costs that can help reduce both the unit processing cost and the total cost, owing to the scaling factor of less than 1 (Ghafoori et al., 2006). In terms of transportation, feedstock collection sites near an active waste processing plant tend to transport more feedstocks to that facility. Moreover, feedstock transportations starting from collection sites with higher food waste generation levels are likely to be larger, and these trends are also observed for solutions A and B. As for landfill, most counties in New York State show considerable decreases in landfill amounts, mainly because of the vast increase of active waste-to-energy facilities for solution C. Taking advantage of these local small-scale waste processing facilities could both reduce the landfill cost and increase the energy production revenue, resulting in the minimum total cost for the nexus systems.

Sensitivity analyses are conducted in this study to quantify the impacts from the deviation of modelling parameters in a pre-optimization manner. In other words, while investigating the effects to vary one parameter, the entire nexus systems would be re-optimized under the updated value. Biogas yield, disposal cost, and electricity price are major factors affecting the total cost objective function, while the price of electricity is the most sensitive factor for the unit processing cost objective. This is presumably because the revenue from electricity generation is the largest category of revenue for the nexus systems that minimize unit processing costs as presented in Figure 1(b).

4. Conclusions

In this work, a multi-period multi-objective food-energy-water-waste nexus systems optimization model that considers various impacts of the COVID-19 pandemic on the nexus systems is proposed and applied to New York State. The model accounted for special features of the pandemic, including the increasing food waste generation, short planning horizon, limited availability of waste-to-energy facilities, and potential transportation options with repurposed trucks. Existing anaerobic digesters and combined heat and power units in wastewater treatment facilities and livestock farms were considered to process the increased amount of food waste under the pandemic and produce energy. The optimized systems showed considerable potential of reducing the disposal amounts of food waste by 38%, which could alleviate both the public health and environmental concerns. The Pareto-optimal solutions illustrated a clear trade-off between the objective functions. The minimum total cost was \$27.1 million, while the optimal unit processing profit was \$11.9 per ton processed food waste. Spatial analysis clearly revealed that facility selections were strongly correlated with their processing capacities. It also showed that the nexus systems could noticeably decrease the food waste disposal amounts in most counties in New York State. Finally, sensitivity analyses revealed that biogas yield and electricity price were the most sensitive factors for the total cost objective and the unit processing cost objective, respectively.

References

- AgSTAR, 2020, Livestock Anaerobic Digester Database. Available: <https://www.epa.gov/agstar/livestock-anaerobic-digester-database>.
- A. Bernstad Saraiva Schott, H. Wenzel, & J. la Cour Jansen, 2016, Identification of decisive factors for greenhouse gas emissions in comparative life cycle assessments of food waste management – an analytical review. *Journal of Cleaner Production*, 119, 13-24.
- R. R. Bora, R. E. Richardson, & F. You, 2020, Resource recovery and waste-to-energy from wastewater sludge via thermochemical conversion technologies in support of circular economy: a comprehensive review. *BMC Chemical Engineering*, 2(1), 8.
- CDC, 2020, What Waste Collectors and Recyclers Need to Know about COVID-19. Available: <https://www.cdc.gov/coronavirus/2019-ncov/community/organizations/waste-collection-recycling-workers.html>.
- T. Chen, D. Shen, H. Feng, Y. Long, & J. Yin, 2017, Comprehensive evaluation of environ-economic benefits of anaerobic digestion technology in an integrated food waste-based methane plant using a fuzzy mathematical model. *Applied Energy*, 208, 666-677.
- DSNY, 2020, DSNY Monthly Tonnage Data. Available: <https://data.cityofnewyork.us/City-Government/DSNY-Monthly-Tonnage-Data/ebb7-mvp5>.
- D. J. Garcia, F. You, 2016, The water-energy-food nexus and process systems engineering: A new focus. *Computers & Chemical Engineering*, 91, 49-67.
- D. J. Garcia, & F. You, 2017, Systems engineering opportunities for agricultural and organic waste management in the food–water–energy nexus. *Current Opinion in Chemical Engineering*, 18, 23-31.
- D.J. Garcia, B.M. Lovett, F. You, 2019, Considering agricultural wastes and ecosystem services in Food-Energy-Water-Waste Nexus system design. *Journal of Cleaner Production*, 228, 941-955.
- C. P. Gerba, A. H. Tamimi, C. Pettigrew, A. V. Weisbrod, & V. Rajagopalan, 2011, Sources of microbial pathogens in municipal solid waste landfills. *Waste Manag Res*, 29, 781-790.
- E. Ghafoori, & P. Flynn, 2006, Optimum Sizing for Anaerobic Digestion. Available: https://www.cesarnet.ca/biicap-archive/rif/report/Flynn_P.pdf.
- J. Gong, F. You, 2015, Sustainable design and synthesis of energy systems. *Current Opinion in Chemical Engineering*, 10, 77-86.
- P. Guelpa, 2020, 14,000 New York City school bus workers face furlough as city imposes cuts. Available: <https://www.wsws.org/en/articles/2020/05/12/sbus-m12.html>.
- E.U. Kiran, A.P. Trzcinski, W.J. Ng, et al., 2014, Bioconversion of food waste to energy: a review. *Fuel*, 134, 389-399.
- J. Nicoletti, C. Ning, F. You, 2019, Incorporating agricultural waste-to-energy pathways into biomass product and process network through data-driven nonlinear adaptive robust optimization. *Energy*, 180, 556-571.
- C. Ning, F. You, 2019, Data-driven Wasserstein distributionally robust optimization for biomass with agricultural waste-to-energy network design under uncertainty. *Applied Energy*, 255, 113857.
- NYSDEC, 2019, Current Descriptive Data of Municipal Wastewater Treatment Plants. Available: <https://data.ny.gov/Energy-Environment/Current-Descriptive-Data-of-Municipal-Wastewater-T/g5a2-qa6a>.
- J. Poore, & T. Nemecek, 2018, Reducing food's environmental impacts through producers and consumers. *Science*, 360(6392), 987.
- E. Royte, 2020, Food waste and food insecurity rising amid coronavirus panic. Available: <https://www.nationalgeographic.com/science/2020/03/food-waste-insecurity-rising-amid-coronavirus-panic/>.
- X. Tian, R.E. Richardson, J.W. Tester, et al., 2020, Retrofitting Municipal Wastewater Treatment Facilities toward a Greener and Circular Economy by Virtue of Resource Recovery:

- Techno-Economic Analysis and Life Cycle Assessment. ACS Sustainable Chemistry & Engineering, 8, 13823-13837.
- F. Xu, Y. Li, X. Ge, L. Yang, & Y. Li, 2018, Anaerobic digestion of food waste – Challenges and opportunities. Bioresource Technology, 247, 1047-1058.
- N. Zhao, F. You, 2019, Dairy waste-to-energy incentive policy design using Stackelberg-game-based modeling and optimization. Applied Energy, 254, 113701.
- Z. Zhong, & F. You, 2014, Globally convergent exact and inexact parametric algorithms for solving large-scale mixed-integer fractional programs and applications in process systems engineering. Computers & Chemical Engineering, 61, 90-101.

Assessment strategies of different molten salts for heat transfer in tri-generation systems

Rachid Klaimi ^{a,b}, Sabla Y. Alnouri ^{c,b*}, Mirko Stijepovic ^d

^a*Mechanical Engineering Department, American University of Beirut, P.O. Box 11-0236, Beirut, 1107-2020, Lebanon*

^b*The Baha and Walid Bassatne Department of Chemical and Petroleum Engineering, American University of Beirut, P.O. Box 11-0236, Riyad El-Solh, Beirut, Lebanon*

^c*Gas Processing Centre, College of Engineering, Qatar University, Doha, Qatar*

^d*Faculty of Technology and Metallurgy, University of Belgrade, Karnegijeva 4, 11000 Belgrade, Serbia*
sabla@qu.edu.qa

Abstract

Tri-Generation Systems (TGS) are designed for combined cooling, heating, and power production within a single process. In this paper, the integration of Concentrated Solar Power (CSP) together with TGS is studied. The tri-generation setup consists of a CSP facility that is integrated onto a desalination plant, producing three different levels of steam (very high pressure, low pressure and very low pressure), in addition to power and freshwater, simultaneously. Due to the complexity of the process, it was found imperative to investigate different factors that would improve the overall performance of the system, and ultimately decrease associated costs and thermal losses. Moreover, the use of molten salts offers many advantages, compared to more conventional fluid options, such as mineral and synthetic oils. Hence, a Mixed Integer Nonlinear Program (MINLP) that allows the selection of optimal heat transfer media, in TGS, has been formulated, and the performance of three different molten salts (solar salt, Hitec, and Hitec XL) has been analysed. The operating temperature of molten salts was found to have a significant effect on the steam quality generated in TGS-CSP systems. Moreover, solar salt resulted in the lowest energy production cost (at 28 cents/kWh) amongst all options. On the other hand, Hitec and Hitec XL were found to yield higher energy costs, estimated at 45 and 47 cents/kWh, respectively.

Keywords: Tri-generation systems, Concentrated solar power, Molten salts, Heat transfer, Power

1. Introduction

Concentrated solar power has been extensively proposed as an effective and economically viable renewable energy technology. CSP plants must be equipped with effective heat transfer media, and thermal energy storage systems. Molten salts have become an increasingly attractive option for high-temperature CSP applications (Fernandez et al., 2019). Several studies have investigated the performance of various molten salts from different perspectives, such as thermophysical properties, thermodynamic analysis, corrosion mitigation and system-level analysis. Such studies were found to be helpful for proposing the most convenient molten salt, to be used for commercial applications. For instance, Wang et al. (2021) conducted a system-level

analysis for $\text{MgCl}_2\text{-KCl}$ and showed that higher heat source temperatures result in slightly improved system efficiencies, due to the higher specific work associated with $\text{MgCl}_2\text{-KCl}$. Rong et al. (2021) succeeded in proving that ab-initio molecular dynamics of NaCl-CaCl_2 are effective for obtaining essential heat transfer information of molten salts, by investigating microstructures and thermophysical properties under various operating temperatures. Ibrahim et al. (2021) studied the performance of eight different molten salts, especially those that could withstand severe corrosion challenges in CSP systems. In their study, Rong et al. (2021) found that both MgNaK chloride and ZnNaK chloride possess high thermal stability, especially for temperatures above $800\text{ }^\circ\text{C}$. Additionally, low thermal energy storage costs were obtained for each. Hence, they recommended the use of molten halide salts for the next generation of CSP plants, since molten chloride salts are currently the most technologically mature. The performance of different molten salts in a tri-generation system has not been studied yet. As a result, this work develops an MINLP program that assesses the economic performance of different molten salts in TGS-CSP systems, for the production of different levels of steam, power and freshwater.

2. Methodology

The proposed model, based on the TGS-CSP system, is shown in Figure 1. First off, molten salt circulating in the CSP plant absorbs heat produced by the solar tower. The hot salt is then directed to a steam generator that produces very high pressure (VHP) steam. The temperature of VHP steam depends on the operating temperature of the studied molten salt, whereas the pressure is assumed constant at 100 bars. The generated LP steam can then be expanded into superheated steam. VLP steam may be converted to cooling water, which in turn may be used in the ejectors, and/or as a water source for VHP steam generation.

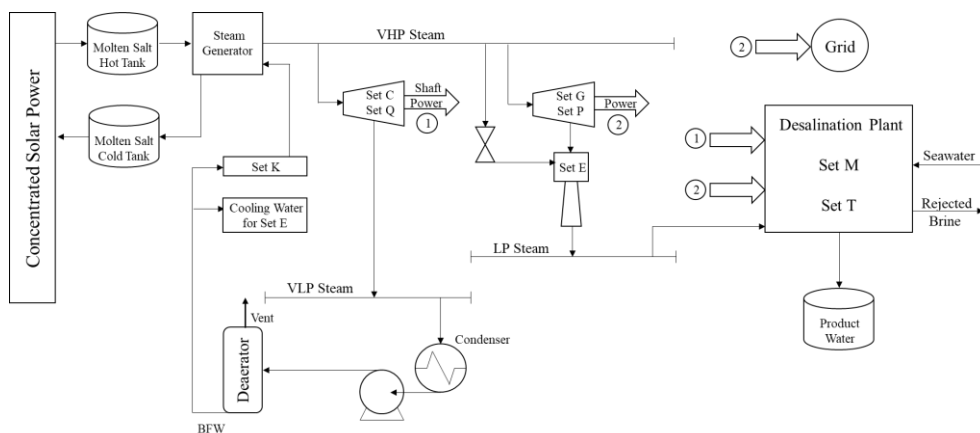


Figure 1. Tri-generation System Structure

On the other side of the system, inlet seawater can either be fed to a thermal and/or a membrane desalination option. The required energy demand for desalination may be covered either using LP steam produced from CSP, and/or shaft-electric power from

turbo-generators, and turbo-pumps. Thermal-based desalination mainly relies on LP steam for operation, since a very low electric power consumption (from turbo-generators) is required. However, membrane-based units consume shaft and electric power only, with no steam is required. Hence, any excess power produced can be sold to the grid, as a revenue stream. The molten salt used for heat transfer and energy storage in CSP is highly dependent on the required operating temperature. In fact, different salt temperatures would affect the temperature of VHP steam produced. Moreover, molten salt cost must be taken into account, especially when carrying out the assessment of different molten salt options. All of the above has been considered in the mathematical formulation of this problem.

3. Mathematical Formulation

A MINLP model, which aims to minimize the total cost of the TGS-CSP system has been developed. The objective function is shown in Eq. (1) below, where $cost^{therm}$ and $cost^{mem}$ are thermal and membrane desalination unit costs, respectively, $cost^{CSP}$ is the CSP unit cost, and rev^{elec} is the electricity revenue. All molten salt related costs are included within the capital cost of CSP plant. Based on the terminology shown in Figure 1 above, the following sets have been defined: set G as the back-pressure turbo generators, BPTG, set C as the condensing turbo-generators, CTG, set P as the back-pressure turbo-pumps, BPTP, set Q as the condensing turbo-pumps, CTP, set E as the ejectors, set K as the pumps (set K), Set T as thermal desalination units, and set M as membrane desalination units. The binary variables, Y_i , have been used to indicate the selection of technology i , whereas the decision variables of the model refer to: (1) input/output flowrates associated with each desalination technology as well as (2) mass flowrates of produced VHP steam into the different expansion units. Equality constraints have been used to ensure no violation of mass and energy balances around the different system units. For example, Eq. (2) and Eq. (3) represent the mass balance of VHP steam and molten salt, respectively, where G^{VHP} is the total mass flowrate of VHP steam produced, G_g^{in} , G_c^{in} , G_p^{in} , G_q^{in} and $G^{tv,in}$ are the inlet VHP steam flowrates to BPTG, CTG, BPTP, CTP, and throttling valve, respectively, while $G^{salt,in}$ and $G^{salt,out}$ are the inlet and outlet mass flowrates of the molten salt. Eq. (4) describes the energy balance around steam generator, where C_p^{salt} is the specific heat of salt, while T^{in} and T^{out} are the inlet and outlet temperature of molten salt to steam generator. Moreover, η_{SG} is heat transfer efficiency parameter and G^{SG} is the mass flowrate of water entering the steam generator, while $H^{out,s}$ and $H^{in,w}$ are the enthalpies for steam leaving and water, respectively.

$$Min.(cost^{therm} + cost^{mem} + cost^{CSP} - rev^{elec}) \quad (1)$$

$$G^{VHP} = \sum_{g \in G} G_g^{in} + \sum_{c \in C} G_c^{in} + \sum_{p \in P} G_p^{in} + \sum_{q \in Q} G_q^{in} + G^{tv,in} \quad (2)$$

$$G^{salt,in} = G^{salt,out} \quad (3)$$

$$G^{salt,in} C_p^{salt} (T^{in} - T^{out}) = \eta_{SG} G^{SG} (H^{out,s} - H^{in,w}) \quad (4)$$

The complete model includes similar mass and energy balance constraints on all remaining units shown in Figure 1. Moreover, performance limits on each unit have also been identified, and included as inequality constraints. For instance, the temperature of salt at the time of water boiling must ensure that no temperature cross-over is taking place in the steam generator, at any point. VHP steam inlet mass flowrate into the turbines were used to calculate the work produced by such units. Other inequality constraints ensure the minimum, (G_i^{min}, F_i^{min}) and maximum, (G_i^{max}, F_i^{max}) allowable mass flowrates of steam, G_i^{in} , and seawater, F_i , to the turbines and desalination units, respectively. In addition, the minimum water production recovery, χ_{pw} , and the minimum amount of electricity allocated to the grid, $P^{Grid,min}$, have also been accounted for. Eq. (5-6) below represent capacity constraints, whereas Eq. (7-8) refer to constraints on water production and electricity production, respectively, where PW^{total} is the product water flowrate, and P^{Grid} refers to amount of electric power allocated to the grid.

$$G_i^{min} Y_i \leq G_i^{in} \leq G_i^{max} Y_i \forall i \in G, C, P, Q \quad (5)$$

$$F_i^{min} \leq F_i \leq F_i^{max} \forall i \in T, M \quad (6)$$

$$PW^{total} \geq \chi_{pw} F^{in} \quad (7)$$

$$P^{Grid} \geq P^{Grid,min} \quad (8)$$

4. Case Study

In this case study, the proposed model has been implemented on a TGS producing steam and power from a CSP plant. Freshwater can also be produced onsite via desalination, using a seawater source with a salinity of 25 g/L. The desalination technologies considered in the study are: multi-stage flashing (MSF), multi-effect distillation (MED) and reverse osmosis (RO). Each technology is associated with its own set of specifications and constraints, which have been accounted for in the model. All the parameters required to estimate CSP plant performance have been obtained from Gunawan et al. (2019), whereas the parameters associated with each desalination technology have been obtained from Klaimi and Alnouri (2020). As previously mentioned, three different molten salts (solar salt, Hitec, Hitec XL) have been assessed for their heat transfer fluid abilities within the system. Each molten salt has its own set of specifications, such as operating temperatures, specific heat capacity and cost, all of which have been provided in Table 1. Since VHP steam temperature is greatly correlated with high operating molten salt temperatures, it was very important to observe this property factor while assessing the effect of those different salts onto the overall system design. Table 1 below summarizes the molten salt specifications, noting that the low operating temperature has been assumed to be 290°C in all cases. Moreover, all conducted cases assume an inlet seawater flowrate of 100,000 m³/d, with a total system recovery of 40 %.

Table 1: Molten salts specifications (Chang et al., 2015)

Molten Salt	High Operating Temperature (°C)	Specific Heat Capacity (kJ/kg K)	Cost (USD/kg)
Solar Salt	600	1.37	0.49
Hitec	535	1.56	0.93
Hitec XL	500	1.41	0.19

The MINLP optimization problem has been implemented on GAMS, using BARON as the global solver (GAMS Software). When solar salt is used, the optimal configuration of the TGS-CSP system obtained is provided in Figure 2. The results show that 11,961 m³/d of solar salt is required to generate steam and power. The optimal steam expansion technologies that have been chosen by the solver are one CTG and two CTP units. It was found that the turbo-generator produces 7.47 MW to cover the electric power required for desalination, while the rest is exported to the grid. On the other hand, the turbo-pumps produce 15.49 MW of shaft power for desalination. Both MED and RO have been selected for seawater desalination. The selection of these two technologies have been attributed to the fact that MED has a lower energy requirement than MSF, while RO is characterized with a higher water recovery than thermal-based technologies. It is important to note that 26 % of inlet seawater has been allocated to MED, while the amount of power allocated to the grid was 4.75 MW. Overall, this results in an energy production cost (EPC) that is equivalent to 28 cents/kWh.

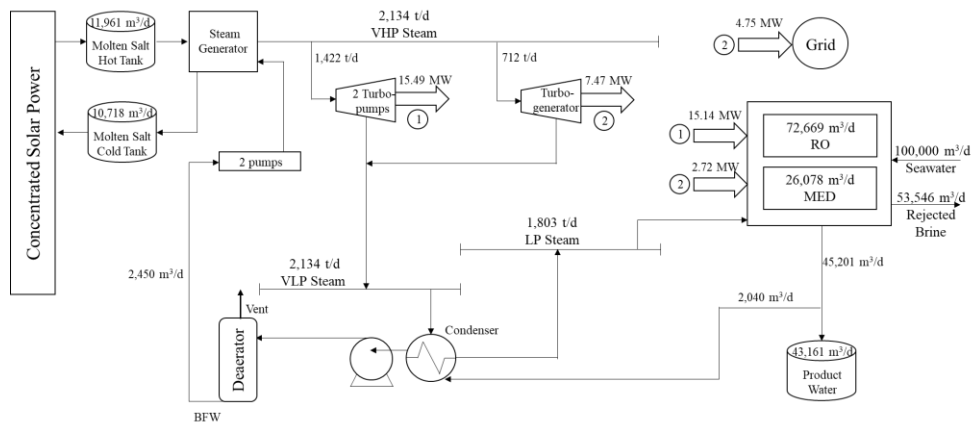


Figure 2. Optimal System Configuration for Case 1, using Solar Salt as the heat transfer medium

When Hitec salt has been used in place of solar salt, the lower operating temperature of this salt results in a higher required flowrate of 13,162 m³/d, when comparing the respective optimal solutions. The same desalination technologies as those reported in Case 1 have been selected, but a higher fraction of seawater has been allocated to MED. Therefore, it was found that an additional amount of LP steam is required. This resulted in the selection of a throttling valve for LP steam production, since it is a less expensive option than turbo-generators. As expected, the optimal solution resulted in a higher total

cost of the system, and a lower operating temperature of Hitec salt. A higher EPC estimated at 45 cents/kWh was obtained. When Hitec XL salt is used, the optimal configuration of the TGS-CSP system shows an increase in the total amount of seawater allocated to MED, which necessitates higher LP steam production. Therefore, besides CTG and CTP units, an additional BPTG unit has been selected in place of the throttling valve, due to its capability of generating electric power. This was found an attractive option economically, as a result of the extra revenue generated. Although the required flowrate of Hitec XL (estimated by 17,894 m³/d) is higher than that required for solar salt, the total cost of Hitec XL salt was lower. EPC was found to be 47 cents/kWh, which is even higher than the energy cost when using Hitec salt. A summary is provided in Table 2 below.

Table 2: Summary of main aspects in optimal results

Molten Salt	Energy	Water	EPC (cents/kWh)
Solar Salt	CTG, CTP	MED, RO	28
Hitec	CTG, CTP, TV	MED, RO	45
Hitec XL	BPTG, CTG, CTP	MED, RO	47

5. Conclusions

In this study, a mathematical model that assesses three different molten salts as heat transfer fluids in a tri-generation system producing steam, power and freshwater, has been presented. Solar salt was found to be the best option to be used in the TGS-CSP system, amongst all options, having resulted in a configuration with the lowest EPC.

References

- A. G. Fernandez, J. Gomez-Vidal, E. Oro, A. Kruiuzenga, A. Sole, L. F. Cabeza, 2019, Mainstreaming commercial CSP systems: A technology review, *Renewable Energy*, 140, 152-176
- A. Gunawan, R. A. Simmons, M. W. Haynes, D. Moreno, A. K. Menon, M. C. Hatzell, S. K. Yee, 2019, Techno-economics of cogeneration approaches for combined power and desalination from concentrated solar power, *Journal of Solar Energy Engineering*, 141, 2.
- A. Ibrahim, H. Peng, A. Riaz, M. A. Basit, U. Rashid, A. Basit, 2021, Molten salts in the light of corrosion mitigation strategies and embedded with nanoparticles to enhance the thermophysical properties for CSP plants, *Solar Energy Materials and Solar Cells*, 219.
- GAMS Software. Retrieved from: www.gams.com/latest/docs/S_BARON.html
- K. Wang, M. J. Lee, Z. D. Zhang, C. H. Min, P. Li, 2021, Evaluation of alternative eutectic salt as heat transfer fluid for solar power tower coupling a supercritical CO₂ Brayton cycle from the viewpoint of system-level analysis, *Journal of Cleaner Production*, 279.
- R. Klaimi, S. Y. Alnouri, 2020, GHG emission reduction assessment for desalination systems through carbon capture and renewable energy options, *Computer Aided Chemical Engineering*, 48, 1081-1086.
- Z. Chang, X. Li, C. Xu, C. Chang, Z. Wang, 2015, The design and numerical study of a 2 MWh molten salt thermozone tank, *Energy Procedia*, 69, 779-789.
- Z. Rong, G. Pan, J. Lu, S. Liu, J. Ding, W. Wang, D. J. Lee, 2021, Ab-initio molecular dynamics study on thermal property of NaCl-CaCl₂ molten salt for high-temperature heat transfer and storage, *Renewable Energy*, 163, 579-588.

Techno-economic Assessment of Load Following Operation for Super-critical Power Plants Equipped with Carbon Capture Feature

Calin-Cristian Cormos,^{a,*} Ana-Maria Cormos,^a Cristian Dinca^b

^a Babes-Bolyai University, Faculty of Chemistry and Chemical Engineering, Arany Janos 11, Cluj-Napoca, RO-400028, Cluj-Napoca, Romania

^b Politehnica University, Faculty of Power Engineering, 313 Splaiul Independentei, Bucharest, RO-060042, Romania

cormos@chem.ubbcluj.ro

Abstract

Greenhouse gas emission reduction from energy-intensive industrial applications (e.g., power generation) is of paramount importance today. The fossil-based power plants have to be operated in a load following (cycling) scenario to balance the time intermittent renewable sources (e.g., solar and wind). This paper is assessing the techno-economic implications of load following (dynamic) operation of decarbonized super-critical power plants to improve the main performance indicators. Reactive gas-liquid absorption method is evaluated as decarbonization technology. As evaluated super-critical power plants, one coal-based concept with 1,000 MW net output was considered as illustrative example. The overall carbon capture rate is set to min. 90 %. The proposed design was simulated in both base-load and dynamic-load conditions. The load following (dynamic) operation was considered for one-week duration considering the common pattern in European power generation sector. Relevant key design elements were assessed in details in term of main techno-economic implications (e.g., dynamic operation of carbon capture unit considering lean and rich solvent storage, integration of power block and carbon capture unit, power plant cycling etc.). The results show that the flexible decarbonized power generation has significant advantages in optimizing techno-economic performance and reducing the overall environmental impact.

Keywords: Flexible operation of decarbonized power plants, Reactive gas-liquid decarbonization system, Techno-economic and environmental assessment.

1. Introduction

Reducing CO₂ emissions from energy intensive sectors is a very important element in the fight against global warming and climate change. Several methods can be used to reduce the overall carbon footprint e.g., boosting renewable energy sources (e.g., solar, wind, biomass), improving energy conversion and utilization efficiencies, developing Carbon Capture, Utilisation and Storage (CCUS) technologies (European Commission, 2019). Currently, the share of renewable energy is significantly increasing in power generation sector but the time-variability of solar and wind technologies is putting an additional pressure on existing fossil base-load capacities which have to be operated in load following (dynamic) conditions. Since the fossil fuels are predicted to remain an

important element in the overall energy sector at least for short and medium term, suitable and flexible decarbonized power generation technologies are needed (Papadis and Tsatsaronis, 2020). This paper is assessing the techno-economic implications of load following (dynamic) operation of decarbonized coal-based super-critical power plants to improve the main performance indicators for better integration into future highly time-dependent energy systems. The key innovative element of this paper towards the current state of the art in the field is the quantification of techno-economic and environmental performance indicators for a decarbonized power plant operated in dynamic (load-following) conditions using solvent storage.

2. Design of decarbonized super-critical power plant, main model assumptions and thermal integration analysis

The decarbonization of super-critical power plant was assessed using the post-combustion chemical scrubbing by alkanolamines. Methyl-Di-Ethanol-Amine (MDEA) was considered as illustrative case. The chemical reaction for post-combustion CO₂ capture is the following:



The corresponded non-decarbonized super-critical power plant was also considered as benchmark case (noted as Case 1). The conceptual design of decarbonized super-critical power plant (noted as Case 2) is presented in Figure 1.

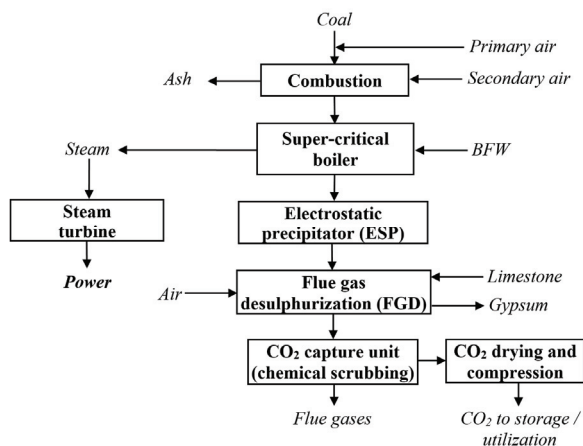


Figure 1. Design of decarbonized super-critical power plant (Case 2)

The most important model assumptions of evaluated super-critical power plants are presented in Table 1. These assumptions were used for modelling and simulation of the power plant using ChemCAD software. The model of CO₂ capture unit was then validated by comparison to experimental data (see Figure 2 left). In addition, the decarbonized design was thermally integrated using pinch method for optimization of thermal duty used in CO₂ capture unit for solvent regeneration. Figure 2 right presents the hot and cold composite curves for CO₂ capture unit.

Table 1. Design assumptions of evaluated gasification-based poly-generation systems

Power plant unit	Design assumptions
Coal composition and lower calorific value	Composition (dry): 72.30 % C, 7.45 % O, 1.69 % N, 4.11 % H, 0.56 % S, 13.89 % ash; Moisture: 8 % (as received) Lower heating value (LHV): 25.35 MJ/kg (as received)
Super-critical power block	Steam temperatures: 582 °C / 580 °C / 580 °C Steam pressures: 290 bar / 75 bar / 20 bar Steam turbine efficiencies: 88 % HP / 90 % MP / 94 % LP Final steam expansion pressure: 45 mbar Feed water pump efficiency: 85 %
Flue gas desulphurization unit	Wet desulphurization system using limestone slurry Sulphur removal efficiency: >98.5 %
CO ₂ capture unit	50 wt. % MDEA aqueous solution Absorption/desorption columns: 18 / 12 stages Carbon capture rate: 90 % Liquid to gas ratio: 3.75 Heat duty for solvent regeneration: 3 GJ/t
CO ₂ conditioning unit (drying and compression)	TEG gas-liquid absorption dehydration unit CO ₂ purity and pressure: >95 % (vol.) / 120 bar Multi-stage inter-cooling compressor with 85 % efficiency
Auxiliary units	Coal handling energy consumption: 0.5 % of thermal input Pump efficiency: 85 % Heat exchanger minimum temperature difference: 10 °C Heat exchanger pressure drops: 1 – 3 % inlet pressure

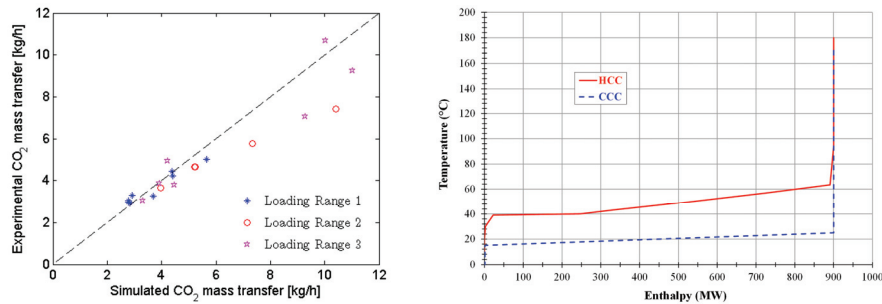


Figure 2. Validation of CO₂ capture unit (left); Thermal integration of CO₂ capture unit (right)

3. Base-load and flexible techno-economic and environmental assessment

Firstly, the techno-economic and environmental performances of both coal-based super-critical power plants with and without decarbonization were assessed in base-load conditions. The overall description of the techno-economic assessment methodology as well as the main economic assumptions are presented in details in a separate paper of the authors (Cormos and Cormos, 2017). Table 2 presents the key performance indicators of evaluated base-load coal-based super-critical power plant concepts. As can be noticed, the introduction of carbon capture feature (a carbon capture rate of 90 % was considered) induces significant modifications of all performance indexes e.g., the net energy efficiency decreases by about 9.6 percentage points (this is the energy penalty for carbon capture), specific CO₂ emission is reduced by about 88 %, specific investment cost is increasing by about 80 %, operation and maintenance costs are increasing by about 27 %, levelized cost of electricity is increasing by about 13.5 %.

Table 2. Main techno-economic and environmental performance indicators (base-load)

Performance indicator	UM	Case 1	Case 2
Coal flow rate	t/h	327.76	420.92
Coal lower heating value (LHV)	MJ/kg	25.35	
Coal thermal energy	MW _{th}	2,308.00	2,964.00
Gross power output	MW _e	1,057.75	1,140.10
Fuel processing consumption	MW _e	11.54	14.82
CO ₂ capture and conditioning unit	MW _e	-	71.05
Power block consumption	MW _e	46.21	54.23
Ancillary power consumption	MW _e	57.75	140.10
Net power output	MW _e	1,000.00	1,000.00
Net power efficiency	%	43.32	33.73
Carbon capture rate	%	0	90.00
Specific CO ₂ emissions	kg/MWh	800.12	93.95
Capital cost	M€	1,275.00	2,288.00
Specific capital investment cost	€/kW	1,275.00	2,288.00
Operational & maintenance cost	€/MWh	28.90	36.81
Levelized cost of electricity	€/MWh	66.00	74.90
CO ₂ removal cost	€/t	-	31.50
CO ₂ avoided cost	€/t	-	37.26

After evaluation of base-load performances, a dynamic (cycling) operation scenario of decarbonized power plant was considered. In this respect, Table 3 presents the considered weekly variation of the plant capacity. This time behaviour was selected considering the daily pattern load common in Central Europe (Astolfi et al., 2019).

Table 3. Time variation of plant capacity (weekly basis)

Monday – Friday		Saturday – Sunday
6 AM to 1 PM	10 PM to 5 AM	0 AM to 12 PM (all day)
7 PM to 9 PM	2 PM to 6 PM	
100 %	50 %	50 %

Considering this time variation of decarbonized power plant capacity, the flue gases flow in a flexible operation scenario compared to the base-load operation is about 73 % for the Monday to Friday period, 50 % for the weekend and about 66.5 % for the whole week. The nominal base-load flue gases flow for decarbonized power plant concept is about 4,100 t/h with 12.35 % vol. carbon dioxide and the corresponding plant capacity is 90 %. The power plant capacity has an important influence on levelized cost of electricity (Astolfi et al., 2019). Figure 3 presents the variation of levelized cost of electricity vs. plant capacity for both super-critical concepts with and without decarbonization. As can be observed, the plant capacity starts to have an important influence on levelized cost of electricity at values below 60 %. In addition, the reduction of the power plant load affects also the energy efficiency compared to the nominal steam cycle condition. In this respect, an off-design correction factor for the overall energy efficiency has to be used to consider the influence of part-load operation (Cormos, 2020).

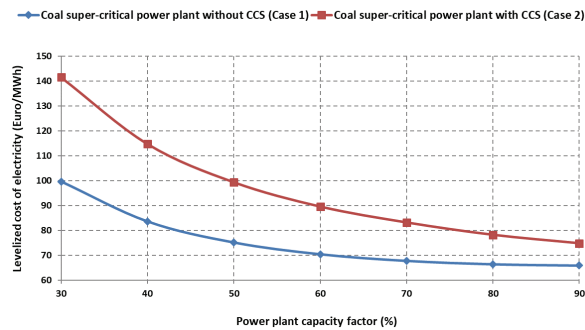


Figure 3. Validation of levelized cost of electricity vs. power plant capacity

For the time-flexible operation of the decarbonized super-critical power plant, when the power output is reduced, the flue gases flow is reduced with the same ratio. In order to keep the 90 % carbon capture ratio, the CO₂ capture gas-liquid absorption cycle has to be modified accordingly. In this respect, the flexible operation of decarbonized super-critical power plant is equipped with MDEA solvent storage capacities in both lean and rich CO₂ loadings. The solvent storage facility was sized considering the need for weekly operation of the power plant which implies that the absorption column follows the power plant operation and the desorption column operates at nominal load (74 % of nominal base-load) and the storage facility has enough capacity to store the both lean and rich forms of the MDEA solvent. Figure 4 presents the time variations of solvent storage and power plant loads in a weekly operation cycle. As it is shown, when power plant operates at nominal load (100 %), the rich solvent storage is gradually filling up and the lean solvent storage is discharged. When the power plant is operated at reduced loads (50 %), the process is reversed (regenerated lean solvent storage is filling up and rich solvent storage is discharged).

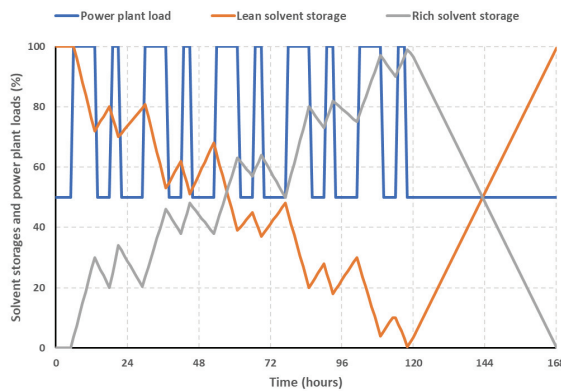


Figure 4. Weekly variations of lean / rich solvent storages and power plant load

The flexible operation of decarbonized power plant implies additional capital cost for solvent storage capacity. This capital cost was estimated based on its size with common cost correlations (Smith, 2016). But the capital cost of solvent storage is totally compensated by reducing the size of CO₂ capture unit, even further reduction of overall plant capital cost is achieved due to the fact the cost of liquid storage facility is lower than for a complex plant such as absorption – desorption cycle. Table 4 presents the variation of key performances in both nominal and part-load operation scenarios.

Table 4. Key performance indicators for nominal and part-load operation

Performance indicator	UM	Case 1	Case 2
Gas-liquid absorption cycle size factor	%	100.00	74.00
Specific capital investment cost	€/kW	2,288.00	2,140.00
Operational & maintenance cost	€/MWh	36.81	36.49
Levelized cost of electricity	€/MWh	74.90	72.25
CO ₂ avoided cost	€/t	37.26	33.26

It can be noticed that the flexible operation of the decarbonized power plant equipped with solvent storage facilities has better performance than nominal case e.g., reduced specific investment by about 6.5 %, reduced levelized cost of electricity by about 3.5 %.

4. Conclusions

This paper is evaluating the flexible operation of 1,000 MW super-critical power plant equipped with carbon capture feature using chemical scrubbing (90 % carbon capture rate was used). As innovative element, the load-following power plant operation considers MDEA solvent storage facility in both lean and rich forms. The evaluation shows that dynamic operation of decarbonized power plant has significant benefits e.g., reducing specific investment cost (by about 6.5 %), levelized cost of electricity (by about 3.5 %) and CO₂ capture costs (by about 10 %) compared to base-load operation.

Acknowledgments

The research leading to these results has received funding from NO Grants 2014 - 2021, under project contract no. 13/2020 and a grant of the Romanian Ministry of Education and Research, project number PN-III-P4-ID-PCE-2020-0032, within PNCDI III.

References

- M. Astolfi, E. De Lena, M.C. Romano, 2019, Improved flexibility and economics of Calcium Looping power plants by thermochemical energy storage, *International Journal of Greenhouse Gas Control*, 83, 140-155
- A.M. Cormos, C.C. Cormos, 2017, Techno-economic evaluations of post-combustion CO₂ capture from sub- and super-critical circulated fluidised bed combustion (CFBC) power plants, *Applied Thermal Engineering*, 127, 106-115
- C.C. Cormos, 2020, Techno-economic implications of flexible operation for super-critical power plants equipped with calcium looping cycle as a thermo-chemical energy storage system, *Fuel*, 280, 118293
- European Commission, 2019, The European Green Deal, COM(2019) 640, Brussels, Belgium
- E. Papadis, G. Tsatsaronis, 2020, Challenges in the decarbonization of the energy sector, *Energy*, 205, 118025
- R. Smith, 2016, *Chemical process design and integration*, 2nd ed., Wiley, Hoboken, USA

Utilisation of CO₂ in transnational LNG supply chain for the enhancement of jet fuel production in globally decentralised GTL industries

Ali Attiq Al-Yaeshi^a, Tareq Al-Ansari^{a,b*}

^a*Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*

^b*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*

*talansari@hbku.edu.qa

Abstract

Jet Fuel from the gas-to-liquid (jet fuel-GTL) process is a suitable alternative to that produced from conventional oil refining because of its combustion efficiency, and compatibility to both legacy and newly designed aircrafts engines. It contributes towards the diversification and expansion of fuel supply chain in the aviation industry due to its unique properties, as it is ultraclean releasing lower levels of Sulphur Dioxide, Nitrogen Oxides and particulate emissions upon burning. The State of Qatar, with its vast quantities of natural gas resources, can lead in the widespread adoption of jet fuel-GTL. The GTL technology also can contribute positively toward CO₂ emission reduction through the integration between GTL and carbon capture and utilisation (CCU) technology, as this will enhance the total production of refined products and reduce CO₂ emissions level. The concept model of integrating of CO₂ product with GTL plant has been discussed in (Al-Yaeshi *et al.*, 2020a) and the results are demonstrated positively in terms of techno-economic-environmental aspects. In fact, fuel production in Pearl GTL has prompted Qatar Airways (QA) to become the first commercial airliner to trial jet fuel-GTL. However, mainstream utilisation is challenged due to: (a) lack of availability of jet fuel-GTL for the purpose of re-fuelling at QA's worldwide destinations. As such, the presence of GTL plants is limited globally due to the availability of natural gas and the high capital costs for GTL plant construction; (b) volatility in jet fuel-GTL prices and the uncertainty in petroleum product markets. The objective of this study is to assess the techno-economic-environmental feasibility of a novel methodological framework based on decentralised LNG-to-GTL processes (or hubs) at optimal locations of LNG terminals. The objective is to enhance the availability of jet fuel-GTL and other GTL products in global aviation and other transportation sectors.

Keywords: Liquefied Natural Gas (LNG), Gas-To-Liquid (GTL), Jet Fuel, Aviation.

1. Introduction

The demand for energy is steadily increasing due to population growth and increased economic activity. Traditionally, crude oil has been used to satisfy various applications for energy in different forms such as gasoline, diesel, and kerosene (jet fuel). However, natural gas-based products have become increasingly competitive due to enhanced

exploration, availability and utilisation applications. Incidentally, greenhouse gas (GHG) emissions resulting from the utilisation of fossil fuels contribute towards global warming and as a result have propelled world leaders to commit to controlling the emissions generated from different sectors. A large portion of fossil fuel utilisation, especially oil derived products occurs in the transport industry which accounts for almost 66% of total global CO₂ emissions. The aviation sector forms approximately 8% from the remaining 16% of transportation sector globally (IEA, 2015). Global passengers have increased gradually from 1946-2016 reaching 3.7 billion in 2016 compared to 18 million in 1946. It is expected to continue to increase as networks and destinations increase especially to South America, China and India (Boeing, 2017). This implies that the manufacture of new planes is expected to increase as new routes are introduced and old models are replaced, where planes with a capacity of 100 seats or more are expected to increase to 42,500 from 20,500 aircrafts by 2036 (Boeing, 2017; Airbus, 2017). Since 2012, the International Civil Aviation Organisation (ICAO) has pledged to reduce GHG levels. However, the aviation sector continues to account for 4% of total global emissions, and may increase to 22 % by 2050. The IATA report stipulates that as part of efforts to remain below 2°C in a 2030 time horizon, emission levels should not exceed 39 % of levels recorded in 2005, and require a decrease by 41 % levels by 2050 compared with the reference year 2005 (Cames *et al.*, 2015).

Several options to reduce emissions from aviation exist, including the use of alternative fuels and other renewable energies. One of these fuels is the jet GTL which is derived from the Fischer-Tropsch technology to produce a synthetic lubrication oil and synthetic fuel from coal, natural gas or biomass. Introducing jet fuel-GTL as an alternative aviation fuel is especially attractive in this region due to low natural gas prices and extraction costs. However further optimisation is required to enhance the cost of extraction of natural gas and improve the cost and efficiency of GTL technology (Ramberg *et al.*, 2017). Dong *et al.* (2008) demonstrate the techno-economic analysis of GTL and LNG plants are comparable however the LNG technology is preferable attributed that the LNG is well established while the GTL is in the initial stage in the commercial market. The GTL has intending to expand widely due to the world energy demand and its clean refined products that classified as environment friendly, hence that will impact LNG markets. In terms of enhancing production of jet fuel derived from GTL, CCU can be considered a viable CO₂ reduction approach as discussed in chapter 4. Al-Yaeshi *et al.* (2020a) demonstrates the suitability of integrating CO₂ emissions into GTL plants in order to increase production rates of finished products, reduce rogue CO₂ emissions and to generate economic benefits. It is concluded that the overall process efficiency enhanced by increasing between 5-10% comparing with base reference of GTL plant. The objective of this study is to assess the techno-economic-environmental feasibility of a novel methodological framework. It considers decentralised LNG-to-GTL processes (or hubs) at optimal locations of LNG terminals promoting enhanced availability of jet fuel-GTL and other GTL products for the global aviation industry. Using Qatar which is rich in natural gas reserves as a case study, this study outlines the role of the natural gas derived fuel obtained via the gas-to-liquid (GTL) technology in the aviation sector in comparison to conventional aviation fuels obtained from crude oil.

Case Study: Qatar

The system considered comprises of an LNG supplier, Qatar Gas (QG), and various potential small-scale GTL plants worldwide that have been identified for the economic utilisation of natural gas, based on the Qatar Gas LNG customers destination zones and Qatar Airways (QA) flight destination zones. The integration model of CO₂ with GTL plant discussed in chapter 4 as a case study in Qatar and can be applied globally. The objective of the optimisation model is to develop a profitable allocation plan for proposed GTL plants and contribute towards reducing GHG emissions generated from the QA fleet and other airlines as they travel to and from Qatar generating techno-economic-environment benefits from all destinations. The State of Qatar is the world's largest exporter of Liquefied Natural Gas (LNG), driven by demands from Asia and Europe due to transitions from Crude oil and Coal to Natural gas. LNG exports from Qatar will continue to rise as new trains are developed expanding production by 40 % to 110 million tons. In terms of GTL, production from Qatar represent 70% of total global GTL production. The GTL process utilises Fischer–Tropsch technology to transform natural gas into synthetic liquid petroleum products which is a feedstock to other commercial applications and commodity markets. The national GTL output consists of an Oryx GTL plant commissioned in 2006 with capacity approximately 34,000 b/d. In addition, a second Pearl GTL plant, the largest global GTL project facility commissioned in 2011 produces approximately 140,000 bpd of refined products (Al-Yaeshi *et al.*, 2020b). Qatar Airways is one of the world's leading airlines, established on November 22, 1993, and its home, Hamad International Airport links over 150 international destinations across Africa, Central Asia, Europe, Far East, South Asia, Middle East, North America, South America and Oceania. The QA fleet consists more than 180 aircrafts and is expected to top 350 by 2022 (QA, 2016).

2. Methodology

The CCU application demonstrated in this study is discussed models developed by Al-Yaeshi *et al.* (2020a), and Al-Yaeshi *et al.* (2020b), where utilising CO₂ in GTL plant demonstrates a high potential in enhancing the GTL output products and contributes in reducing the CO₂ emissions to the atmosphere from the sources. In this study, the integral LP formulation considers optimising the GTL plant distribution based on the availability of QG LNG terminal destinations, QA route destinations, high aircraft traffic airports, high jet fuel consumption, and LNG purchase prices. The aim is to optimise the location of GTL plants upon a techno-economic-environment analysis and ensure life cycle CO₂ reductions of QA and HIA from well to wheel. In order to demonstrate significant improvement in the QA fleet performance, it is necessary for the fuel to be available in most of the destinations. As the such, the system proposed maximises the overall revenue of the Jet fuel-GTL utilisation and promotes Qatar as a hub for jet fuel-GTL.

The system described in Figure 1 consists of three main components, LNG terminal, GTL plant and distribution network. The operation starts by receiving the LNG cargo in the terminal where the unloading and regasification process begins. After which, the gas is compressed and transported by pipeline to the GTL plant where the gas is converted to the refined products such as Jet fuel, Gasoline and Diesel. The final stage is to transport the Jet fuel-GTL to the airport to refill the QA fleet and/or other aircrafts.

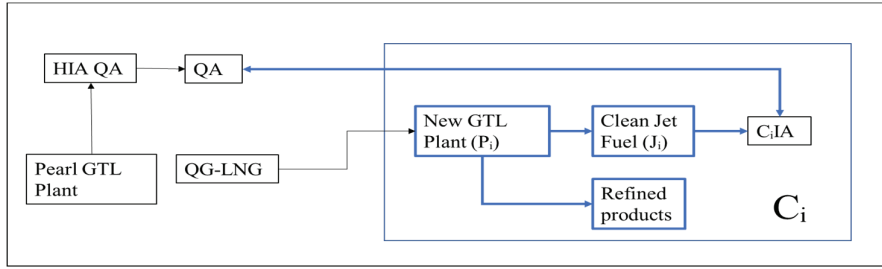


Figure 1. Model mapping the QA, LNG and GTL network.

The decision variables used in the Integer LP formulation are applied to solve the binary optimisation case study as indicated in Figure 1. The parameters model corresponds to the availability of LNG products, potential of natural gas resources, g and objects of QA connection q at the defined countries i . Each country i has an LNG price S_i and a number of QA flights F_i from Qatar to the same destination. QA consumes conventional jet fuel from country i . There is a potential to locate GTL plants P_i at country i to supply QA fleet with clean Jet fuel-GTL considering the high demand of jet fuel-GTL J_i at country i . However, this depends on the high demand of jet fuel in country i , hence the ratio R_i is defined:

$$R_i = \frac{J_s}{J_i}$$

Where J_s is the total of Jet fuel consumption at country i to the J_i Jet fuel-GTL plant produced at the same country.

The case study defines the specific GTL plant locations that will ensure a lifecycle reduction of CO₂ during return flights to Qatar using Jet fuel-GTL. The analysis is based on the 42 countries that Qatar supplies with LNG (or potentially GTL), in addition to QA routes. The LNG volume and prices are varied between (0.1 -12.2) MT/y and (\$3.5 – \$13.5) respectively (QG, 2019; Timera, 2019), while QA flights in the same countries are varied between 48–8,064 flight/y with total 51,108 flights per year (QA, 2018). The total jet fuel volume at the defined countries is 1,8750 million BBL/y while the proposed Jet fuel-GTL for each country is 4.380 million BBL/y (EIA, 2016) (The Global, 2018). A detailed nomenclature of model parameters is defined by set the formula to compute the maximum number of QA flights to determine the countries to mark the suggested GTL plant, hence access to clean Jet fuel-GTL for QA. Computation of the optimisation of QA flights is defined as;

Maximise $F(x) = \sum_{i=1}^n (F_i x_i)$, where n and F are the no. of countries and QA flights

The binary decision variables are as following;

$$(x_i) = \begin{cases} 1, & \text{if } C_i \text{ is selected} \\ 0, & \text{otherwise} \end{cases} \quad \text{where } i=1, \dots, n$$

Subject to the below conditional logics constraints;

$$\sum_{i=1}^n (S_i x_i) \leq \$25, \text{ where } S \text{ is the selling LNG price at the country } i$$

$$\sum_{i=1}^n (P_i x_i) \geq 1, \text{ where } P \text{ is the number of GTL plants at country } i$$

$$\sum_{i=1}^n (R_i x_i) \geq 10\% \text{ of } \sum_{i=1}^n J_i, \text{ where } R \text{ is the ratio of Jet production}$$

where; $S_i \geq 1, P_i \geq 1, R_i \geq 1$

3. Results

Table 1 indicates the optimum values for GTL plants locations when the total of QA flights is maximised in the defined countries based on constraints. Consequently, this reduces the annual rate of CO₂ emissions produced by the aircraft while travelling back to Qatar or during its operation in the HIA. Table 2 highlights the selected countries selected based in the integer LP formulation.

Table 1. The optimisation results.

Decision Variables	Results
Number of GTL plants	8
Number of flights/y	30960
LNG (\$)	24.11
Ratio of jet fuel (R _i)	192.4

Table 2. Country selection.

C _i	Country	C _i	Country
C ₁	Iran	C ₅	Spain
C ₂	KSA	C ₆	UAE
C ₃	Kuwait	C ₇	UK
C ₄	Oman	C ₈	USA

The modelling results illustrated in Figure 2 demonstrates that LNG prices at country *i* have the potential to make decisions in terms of the number of GTL plants. Although some countries have the highest rate for jet fuel consumption and record high QA flights however the LNG prices are the highest comparing with others in Europe and Middle east specially China, Japan and South Korea. A clear benefit is achieved in terms of increasing jet fuel utilising QA flights generating revenues in the selected countries, whilst reducing CO₂ emissions during return to Qatar. The total flights that benefit from the GTL plants exceed 60% of the total QA flights to the 42 countries resulting in emission reductions in the selected countries. The model indicates that the selected countries can establish a clean refined network consisting of LPG, and gasoline, diesel, wax to contribute effectively to CO₂ the emission reduction whilst remaining economically and environmentally attractive. It is important to note that selected countries have the potential for further optimisation based on NG accessibility.

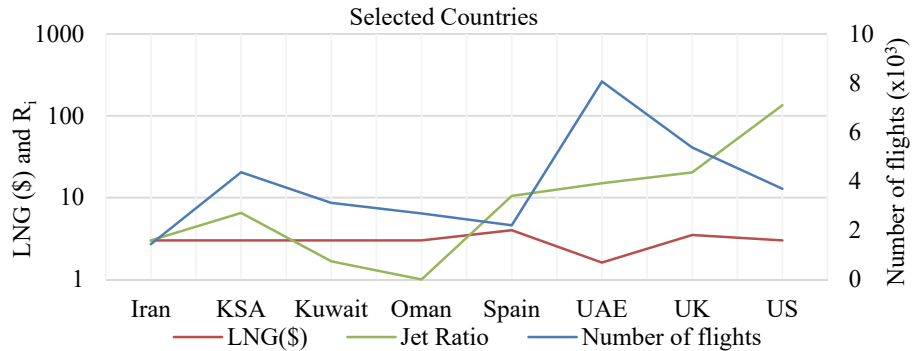


Figure 2. Results of variables parameters at selected countries for placing GTL plants.

Countries in the Gulf cooperation country (GCC) such as Kuwait, KSA, Oman, and UAE are located in close proximity implying that a large single GTL plant is sufficient to secure the airports with the required feedstock of Jet fuel-GTL. Iran also can be part of this supply chain despite possessing a large natural gas field, therefore implying that the list of GTL plants can be short listed to four.

4. Conclusion

The growth of the global aviation sector has increased dramatically in last few decades thus inducing significant carbon reduction challenges in the light of stricter regulations. Jet fuel-GTL is an optimal alternative for currently utilised Jet fuel produced from conventional Oil Refinery. Going forward, the state of Qatar can utilise its long-term LNG contracts to begin implementing LNG-to-GTL process. The modelling and analysis presented in this paper considers the novelty of LNG-to-GTL technology to secure clean refined products especially Jet fuel-GTL at the countries that have received LNG shipments, whilst reducing CO₂ emissions at both airports and flight routes. The approach demonstrates the potential economic and environment benefit for the selected countries to locate the GTL plants whilst considering the LNG prices, number of QA flights and Jet fuel ratio of proposed GTL plant to the total Jet consumption in the selected countries. Future work will consider a techno-economics assessment to locate the plants based on the CAPEX, OPEX and economic indicators such as NPV, IRR and PIR for each selected country.

5. References

- Al-Yaeshi, A.A., Govindan, R. and Al-Ansari, T., 2020a. Techno-economic-based dynamic network design for optimum large-scale Carbon Dioxide utilisation in process industries. *Journal of Cleaner Production*, p.122974.
- Al-Yaeshi, A.A., AlNouss, A., McKay, G. and Al-Ansari, T., 2020b. A simulation study on the effect of CO₂ injection on the performance of the GTL process. *Computers & Chemical Engineering*, 136, p.106768.
- Airbus, 2017, *Global Market Forecast Growing Horizons 2017/2036*, pp. 7–8. [accessed Novemembr 2017]
- Boeing, 2017, *Current Market Outlook 2017 - 2036* [accessed Novemembr 2017].
- Cames, M., Graichen, J., Siemons, A., & Cook, V., 2015, *Emission Reduction Targets for International Aviation and Shipping*. Study for the ENVI Committee. European Union, 53(9), 1689–1699.
- EIA, 2016, *The U.S. Energy Information Administration; Jet fuel consumption*
- IEA, 2015, *CO₂ EMISSIONS FROM FUEL COMBUSTION Highlights*. Iea, S/V(IEA - STATISTICS), 1–139.
- Qatar Airways Group, 2017, *Qatar Airways Group. Annual Report*. Retrieved from https://www.qatarairways.com/content/dam/documents/annual-reports/2017_Annual_Report_ENGLISH-WEB.pdf [accessed Novemembr 2017].
- QA, 2018, <https://www.qatarairways.com/en-qa/homepage.html> [accessed April 2018].
- Ramberg, D. J., Henry Chen, Y. H., Paltsev, S., & Parsons, J. E., 2017, *The economic viability of gas-to-liquids technology and the crude oil–natural gas price relationship*. *Energy Economics*, 63, 13–21.
- Karunanidhi, R.A.M.A.C.H.A.N.D.R.A.N., 2015. *The Aviation Fuel and the Passenger Aircraft for the Future-Bio Fuel, Synthetic Fuel* (Doctoral dissertation, Master Thesis, Department of Automotive and Aeronautical Engineering).
- Dong, L., Wei, S.A., Tan, S. and Zhang, H., 2008. *GTL or LNG: which is the best way to monetize “stranded” natural gas?*. *Petroleum Science*, 5(4), pp.388-394.
- Timera Energy, *European gas pricing dynamics*, 2011, <https://timera-energy.com/european-gas-pricing-dynamics/> [Accessed Novmembr 2017].
- The Global Economy, 2018, [Jet fuel consumption by country, around the world | TheGlobalEconomy.com](http://TheGlobalEconomy.com) [accessed January 2018].

River Water Quality Prediction in Malaysia Based on Extra Tree Regression Model Coupled with Linear Discriminant Analysis (LDA)

Danny Hartanto Djarum ^a, Zainal Ahmad ^{b,*}, and Jie Zhang ^c

School of Chemical Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal Penang, Malaysia

^b School Chemical Engineering and Advanced Materials, Merz Court, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom.

chzahmad@usm.my

Abstract

Poor river water quality as a result of rapid urbanization has led to various disease outbreak and destruction of wildlife ecosystem. Although continuous water quality monitoring has been proposed as the immediate solution, the complexity and expensive nature of such system often offset its purpose. Therefore, numerous studies have been made to forecast water quality utilizing different machine learning algorithms. In this paper, a comparative study was carried out to develop an efficient water quality index (WQI) prediction model based on more attainable monitoring parameters. Three different variation of ensemble decision tree models were analysed and compared, namely: Random forest regression (RF), extra tree regression (ETR), and decision tree + AdaBoost (BTR). These models were coupled with principle component analysis (PCA) and linear discriminant analysis (LDA) to reduce the dimensions of the input data. The results show that (ETR-LDA) model outperform the other ensemble tree models with an R^2 value up to 0.88. The ETR-LDA combination consistently score a higher R^2 values even when trained with a much smaller input dimension (i.e. 3) resulting in a faster training time. This model could positively contribute towards the long-term water quality management effort in a cost-effective manner.

Keywords: Water quality index, extra tree regression, linear discriminant analysis.

1. Introduction

Water utilized by domestic consumption and food production are primarily sourced from the surrounding river. Unfortunately, rapid urbanization and climate change has led to significant degradation in water quality which poses serious health risk (Imani et al., 2012). Poor water quality can give rise to multitude of health diseases such as polio, cholera, diarrhea, or typhoid which may directly affects human (Nawaz and Ali, 2018). Therefore, it is very important to continuously assess and monitor the river water quality. However, water quality measurement requires various biological, physical and chemical parameters, whereby some of the parameters are difficult and time consuming to measure. In Malaysia, the water quality measurement performed by DOE can be categorized into two groups. On site measurement, in which variable such as DO, SS, pH, NH₃-NL, Cl, Cd, Cr, Zn concentration conductivity, turbidity, and E coli can be easily recorded.

Whereas parameters like BOD and COD requires complex laboratory analysis and time consuming to calculate (5-7 days for BOD and few hours for COD) (Ahmad et al., 2016). The complexity and time constraints of BOD and COD measurements causes a major drawback in the overall WQI monitoring process, whereby it is impossible to alert the general public of the real-time WQI value. This has led to numerous studies being carried out to develop water quality prediction models based on various machine learning techniques. Furthermore, the size of water quality dataset fed into the machine learning model are often very large that it can cause enormous strain in the prediction model. High dimensional data could negatively impact the efficiency of the machine learning model, as the computational time increases exponentially with the dimensionality (Fan and Li, 2006). Therefore, performing dimensionality reduction (DR) is extremely crucial in order to improve the efficiency of the machine learning model. Mangai et al. (2020) developed a biological oxygen demand (BOD) prediction model coupled with principal component analysis (PCA) and correlation-features selection (CFS). Their work shows the two dimensionality reduction techniques, PCA and CFS did not compromise the performance of the BOD prediction models. In our study, only on-site monitoring parameters were utilized as the input features for the WQI prediction model. Three different variations of decision-tree based model were analysed and compared, namely random forest regression (FR), extra tree regression (ETR) and decision tree regression with AdaBoost (BTR). Additionally, principal component analysis (PCA) and linear discriminant analysis (LDA) were assessed to reduce the dimensions of the water quality dataset.

2. Methodology

2.1. Study Area and Data Collection

In this study, the data were collected from Malaysia's Department of Environment (DOE) in the span of four years from 2014 to 2018. The historical data represents water quality reading from 126 different monitoring sites within two major states in Malaysia namely Penang and Perak with an approximate area of around 10,800 km². The combined dataset has 2843 rows for 21 different parameters consisting of 4 location attributes, 3 date and time attributes, and 14 water quality attributes. The complete list of the water quality parameters can be observed from Table 1. The missing values from the dataset is relatively insignificant with highest observable percentage loss at only 1.20 % for WQI measurement and less than 1 % for all the other parameters. However, median value substitution was still performed to impute the missing values in order to fully utilize all 2843 rows of data. Out of the three location attributes, the states attributes were selected to train the water quality prediction model. As most machine learning algorithms prefers to work with numerical attributes, label encoding was performed to convert categorical value such as the states attribute to numerical value.

Table 1: List of input and output features for WQI prediction model.

List of all measured parameters	States, Basin, River, Station No, Days, Months, Years, DO (Dissolved Oxygen), BOD (Biochemical Oxygen Demand), COD (Chemical Oxygen Demand), SS (Suspended Solids), pH, NH ₃ -NL (Ammoniacal Solids), Conductivity, Turbidity, Cl, Cd, Cr, Zn concentration, E coli, WQI (water quality index)
Selected input feature	States, Days, Months, Years, DO, SS, pH, NH ₃ -NL, Conductivity, Turbidity, Cl, Cd, Cr, Zn, E coli

Selected output features	WQI (Water Quality Index)
--------------------------	---------------------------

Brief analysis on the WQI in each river basin shows that Beruas has the highest median WQI values at around 90, which falls under clean region of the Malaysia’s DOE water quality classification. The lowest median WQI values could be found in Perai river at 59 which falls under the polluted level. The rest of the locations falls under the slightly polluted range with WQI median values between 60 to 80.

2.2. Water Quality Index (WQI) Calculation

Water quality index (WQI) which was first introduced by Horton (1965), is a metric that represent multiple different weighted water quality parameters into a single numeric value. The main purpose of having such metric is to simplify the process of communicating water quality assessment to the general public and decision makers. Eq. (1) below shows how WQI is calculated by DOE in Malaysia (Mamun et al., 2009):

$$WQI = 0.22SIDO + 0.19SIBOD + 0.16SICOD + 0.15SIAN + 0.16SISS + 0.12SipH \quad (1)$$

Where: SIDO = Sub-index DO (% saturation); SIBOD = Sub-index BOD; SICOD = Sub-index COD; SIAN = Sub-index NH3-N; SISS = Sub-index SS; SipH = Sub-index pH.

As can be observed from Table 1, only on-site monitoring parameters that can be easily measured were selected as the model input parameters. Therefore, complex parameters such as BOD and COD which are time consuming to compute were excluded from the list of model input parameters. However, it could be observed from Table 2 that both BOD and COD are the third and the fourth most correlated variables towards WQI respectively (minus “-” sign depicts negative correlation). Excluding these parameters from the WQI model will inevitably affect the prediction performance. However, other highly correlated variables such as DO and NH3-NL would help minimize the performance impact.

Table 2: Pearson’s correlation analysis between input parameters against WQI.

Parameters	DO	States	Days	PH	Months	Years	Cd	Cr	Zn
Correlation	0.82	0.43	0.27	0.06	0.04	0.03	0.01	-0.07	-0.08
Parameters	Cl	Cond	Tur	SS	E. Coli	COD	BOD	NH3-NL	
Correlation	-0.10	-0.11	-0.12	-0.14	-0.14	-0.64	-0.64	-0.67	

2.3. Dimensionality Reduction

PCA is a very popular dimensionality reduction techniques, generally it performs a set of linear transformation on the original dataset into a smaller number of variables while explaining most of the variance. The transformed variables which are also called principal components (PC) are sorted in a descending order, whereby PC1 has the largest contribution of the explained variance. The equation for principal component is depicted as follow (Fahmi et al., 2011):

$$Y_{mn} = z_{m1}x_{1n} + z_{m2}x_{2n} + z_{m3}x_{3n} + \dots + z_{mi}x_{in} \quad (2)$$

Where y is the component score, z is the component loading, x is the measured value of a variable, m is the component number, n is the sample number, and k is the total number of variables.

On the other hand, LDA optimize the projection by preserving the discriminatory power of the dependent variable before transforming it into a lower dimensional space (Delac et al., 2005). This could be achieved as LDA seek to minimize the within class variance and maximize the between class variance which eventually lead to the greatest class discrimination. According to (Ngouna et al., 2020) water quality parameters can be divided into two main classes: compliant water quality and non-compliant water quality (which depicts water pollution). The between class scatter S_b matrix and within class scatter matrix S_w can be computed by eq. (3) and eq. (4) (Rupali et al., 2010):

$$S_b = \sum_{i=1}^C (u_i - u)(u_i - u)^T \quad (3)$$

$$S_w = \sum_{i=1}^C \sum_{j=1}^{M_i} (Y_j - u_i)(Y_j - u_i)^T \quad (4)$$

Where C is the number of distinct classes, M_i is the number of training samples in class i , u_i is the mean vector in class i , Y_j is the j th data of the set of samples in class i .

2.4. Prediction Model

As described in the earlier section, there are three ensemble decision tree-based regression models (FR, ETR, and BTR) that were analyzed and compared to predict WQI. Ensemble technique simply means that multiple decision trees are combined to improve the prediction performance. Two of the most widely known ensemble techniques are bagging and boosting. Bagging or bootstrap aggregation is generally performed to reduce the variance of the decision tree by generating multiple training dataset in parallel through random sampling. Each of the training set would be utilized to train their trees, resulting a more robust performance from averaging out the accuracy of all the trees. Boosting on the other hand generates series of decision trees sequentially with a goal of reducing the overall prediction error (Kotsiantis and Kanellopoulos, 2012). Random forest regression model (FR) is an example of decision tree model based on bagging technique. The advantages of FR compared to other machine learning models are: Robustness against parameters specification, powerful interactions against input variables, and tackling non-linear properties. ETR is similar to FR, but instead of subsampling the data with replacement, it utilizes the entire original dataset. Additionally, rather than selecting the optimum splitting point, it opted for randomization. Both FR and ETR compute the final prediction result as the average of all the generated decision trees (Geurts et al., 2006). BTR on the other hand is based on boosting ensemble technique, by which the decision tree is boosted using AdaBoost method. It works by building series of decision tree regressors in which larger weight is given to sample with higher degree of errors and vice versa. After numerous iterations, the resulting model would have a much lower overall prediction error (Drucker and Harris, 1997). The dataset was randomly distributed into 80 % of training set and 20 % of test validation set. The performance and efficiency of the prediction model was assessed by utilizing the R^2 , root mean square errors (RMSE) values, and training speed.

3. Results and Discussion

As shown in Table 3, the WQI prediction models were initially developed and compared without performing any dimensionality reduction techniques. The best model accuracy could be observed from both ETR and BTR models with an R^2 values 0.89 and RMSE scores of 4.632 and 4.688 respectively. FR model is not far behind at R^2 values of 0.87 and RMSE scores of 5.053. However, the efficiency of ETR is much better when compared to the FR model at 91 % faster training speed. This shows that ETR strategy of selecting random splitting point is much more efficient than that of FR.

Table 3: Prediction performance of different WQI prediction models for test dataset.

Model	Dimensions	Var	FR			ETR			BTR		
			R^2	RMSE	Speed	R^2	RMSE	Speed	R^2	RMSE	Speed
Base	-	1.00	0.87	5.053	1.3 s	0.89	4.632	91%>FR base	0.89	4.688	63%>FR base
Base + PCA	12	0.97	0.82	5.708	4%> FR base	0.86	5.151	10%> ETR base	0.87	4.921	3%> BTR base
	9	0.87	0.80	6.116	11%> FR base	0.82	5.780	26%> ETR base	0.83	5.672	8%> BTR base
	6	0.71	0.80	6.208	53%> FR base	0.82	5.856	54%> ETR base	0.82	5.819	51%> BTR base
	3	0.51	0.78	6.425	145%> FR base	0.79	6.323	113%> ETR base	0.77	6.570	135%> BTR base
Base + LDA	12	0.99	0.86	5.142	2%> FR base	0.88	4.860	10%> ETR base	0.88	4.811	1%> BTR base
	9	0.98	0.86	5.236	8%> FR base	0.87	4.918	31%> ETR base	0.87	4.887	5%> BTR base
	6	0.97	0.85	5.287	53%> FR base	0.87	5.147	62%> ETR base	0.86	4.929	43%> BTR base
	3	0.95	0.85	5.344	150%> FR base	0.85	5.339	127%> ETR base	0.85	5.396	129%> BTR base

Dimensionality reduction techniques were then performed to reduce the input dataset to 12, 9, 6, and 3 dimensions. In term of performance and efficiency, LDA reduced model consistently surpass that of PCA technique at any dimensional level. This could be observed from Table 3, whereby model trained with only 3-dimensional input data could still achieve an R^2 values of up to 0.85 and RMSE as low as 5.339 compared to that of PCA with the highest achievable R^2 values at only 0.78. The training speed for LDA reduced model is also much faster at up to 150 % faster compared to the respective base model without DR technique. This could be due to the fact that LDA has the ability to conserve majority of the discriminatory power from the original dataset when projecting the data into lower dimensional space, which allow LDA to maintain as much variance as possible. From Table 3, it could be observed that even at the lowest dimensional level, LDA could still explain 95 % of the variance in the original dataset. Whereas the cumulative variance explained by PCA technique dropped to as low as 51 %. Ideally, the final combination of prediction model and DR technique would be ETR + LDA (6-dimensions), where there is acceptable compromise between performance and efficiency with an R^2 values of 0.87 and 62 % faster training speed than base ETR model. After performing hyperparameter-tuning the performance and efficiency were further improved with an R^2 values of 0.88 and more than double the untuned training speed at 172 % faster than base ETR model.

4. Conclusion

In this study, three different type of decision tree-based regression model (FR, ETR, and BTR) were compared to predict WQI. The results of our study show that each of the decision tree model displayed satisfactory performance with R^2 values above 0.85 with

ETR being the most efficient model at up to 91 % faster training speed than the base FR model. Additionally, two dimensionality reduction techniques namely PCA and LDA were assessed. After performing hyperparameter-tuning, ETR + LDA (6-dimensions) was found to be combination with the best compromise between performance and efficiency with R^2 values of 0.88 and 172 % faster training speed than base ETR model. By utilizing only on-site measurements it is possible to develop a WQI forecasting system that could alert the general public almost in real-time. It could potentially be a more effective alternative than the existing WQI monitoring system.

Acknowledgment

This work was supported by Universiti Sains Malaysia (USM), special gratitude to Department of Environmental (DOE) Malaysia for providing the water quality data for this study and Kementerian Pendidikan Malaysia (KPM) through Fundamental Research Grant Scheme (FRGS) grant number PJKIMIA/6071414.

References

- M. A. Imani, Z. Kapelan, G. Fu, D. Butler, 2012, *Journal of Environmental Management*, 112, 9.
- Nawaz, S., Ali, Y. 2018. Factors Affecting the Performance of Water Treatment Plants in Pakistan. *Water Conservation Science and Engineering* 3, 191–203.
- Z. Ahmad, N. A. Rahim, A. Bahadori, J. Zhang, 2016, Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks, *International Journal of River Basin Management*, 15(1), 79–87.
- J. Fan, R. Li, 2006, *Proceedings of the international Congress of Mathematicians*.
- J. A. Mangai, B. B. Gulyani, 2020, *International conference on Modelling, Simulation and Intelligent Computing*, 255-263.
- R.K. Horton, 1965, An index number system for rating water quality, *J. Water Pollu. Cont. Fed.*, 37(3), 300-305.
- T. Abbasi, S.A. Abbasi, 2012, *Water Quality Indices*, Elsevier.
- Al-Mamun, A. Hafizah, M.Z. Alam, 2009, *Improvement Of Existing Water Quality Index In Selangor, Malaysia*.
- M. Fahmi, M. Nasir, M. Samsudin, M. Isahak, M. Roshide, A. Awaluddin, M. Mansor, H. Juahir, N. Ramli, 2011, River Water Quality Modelling Using Combined Principle Component Analysis (PCA) and Multiple Linear Regressions (MLR): A Case Study at Klang River, Malaysia. *World Applied Sciences Journal*. 14.
- K. Delac, M. Grgic, S. Grgic, 2005, Independent comparative study of PCA, ICA, and LDA on the FERET data set, *International Journal of Imaging Systems and Technology*, 15(5), 252–260.
- Ngouna, R.H., Ratolojanahary, R., Medjaher, K., Dauriac, F., Sebilo, M., Junca-Bourici, J., 2020. A data-driven method for detecting and diagnosing causes of water quality contamination in a dataset with a high rate of missing values. *Eng. Appl. Artif. Intell.* 95.
- D. Rupali, V. Bhupendra, 2010, Feature Reduction for Intrusion Detection Using Linear Discriminant Analysis. *International Journal on Computer Science and Engineering*. 2.
- S. Kotsiantis, D. Kanellopoulos, 2012, Combining bagging, boosting and random subspace ensembles for regression problems, *International journal of innovative computing, information & control: IJICIC*.
- P. Geurts, D. Ernst, L. Wehenkel, 2006, Extremely randomized trees, *Machine Learning*, pp.3–42.
- Drucker, Harris, 1997, Improving Regressors Using Boosting Techniques, *Proceedings of the 14th International Conference on Machine Learning*.

Synthesis and Design of Sustainable Integrated Process, Water Treatment, Energy Supply Networks and Carbon Utilization Networks Under Uncertainty

Yue Li,^a Fengqi You,^b Zhihong Yuan,^{a,*} Rafiqul Gani^{c,d}

^a*Department of Chemical Engineering, Tsinghua University, Beijing 100084, China*

^b*Cornell University, 318 Olin Hall, Ithaca, New York 14853, USA*

^c*PSE for SPEED company Ltd, Skyftemosen 6, DK 3450 Allerød, Denmark*

^d*Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea 34141*

**Corresponding author's E-mail: zhihongyuan@mail.tsinghua.edu.cn*

Abstract

This paper proposes a novel method for synthesis, design and innovation of sustainable integrated processes that combines water, energy, waste treatment and carbon capture and utilization (CCU) networks together with the overall objective to minimize carbon emissions and wastes under uncertainty. Compared to the earlier deterministic method (Li et al, 2020), in this method, uncertainty of parameters, CCU as well as extended models for the process, water and energy networks have been systematically addressed in this work. The applicability of the extended method is demonstrated through a case study that considers the xylitol production process together with water, energy and CCU networks. Raw material prices, product prices and demand are considered as stochastic. The results show that the multi-network synthesis and design method is able to determine the optimal solution with high economic benefits and zero or low carbon emission under uncertainty for different scenarios.

Keywords: process synthesis, multi-network optimization, uncertainty, CCU

1. Introduction

Chemical processes convert selected raw materials to desired chemical products, while requiring also utilities such as water and energy and produce wastes and greenhouse gases emissions, thereby cause negative environmental impacts. Different sources and generation processes of water and energy may result in different investment costs, operating costs, carbon and pollutant chemical emissions as well as the synthesis-design of the chemical production process. Different processes of waste treatment and carbon capture and utilization (CCU) may also be used to reduce waste and greenhouse gases emissions. These problems should also be considered along with the production process.

Although some publications consider the water utilization, power integration or carbon capture during the systematic chemical process synthesis and design, the sources of the water and power have yet to be simultaneously considered. A framework for general chemical process synthesis that comprehensively considers the overall water and energy supply, waste and carbon emission is rare. In recent work proposed a framework of multi-network problem for process synthesis-design integrated with water sub-network and power generation sub-network (Li et al. 2020). It proves that the multi-network optimization problem can be used to find a more sustainable alternative with low emissions and high benefits. However, feasible solutions with zero carbon emission usually cannot be found. In order to obtain sustainable solution with zero carbon emission, we introduced a carbon capture and collection network to the multi-network problem to construct an extended multi-network optimization problem. Also, for the multi-network integration problem, since the uncertainty of any network will radiate to other networks, when parameters such as market prices and demands change, the optimal solution of the original model may perform poorly. To hedge against the uncertainty (Ning et al., 2019), the MILP model of multi-network optimization problem is extended to a stochastic mixed integer linear programming (SMILP) model. Therefore, by setting the discrete distribution of uncertain parameters and solving the new two-stage stochastic programming model, the sustainable alternatives with the highest economic expectation and zero emission can be obtained. The applicability of this extended method is demonstrated through by a realistic conceptual case study.

2. Framework and Modeling

We propose a framework together with modeling methods for integrating process, water treatment, power generation, and CCU networks to obtain the more sustainable alternatives with environmental constraints under uncertainty. A brief description of the systematic framework and associated methods and tools are given below.

2.1. Framework

The framework of integration is shown in Figure 1. The new framework has four sub-networks: chemical process, energy generation, water treatment, and CCU process. Chemical process converts raw materials to target product. Power sub-network supplies energy to all sub-networks. Water sub-network supplies water to other sub-networks. The carbon dioxide generated by process, power and water sub-networks are captured and utilized by the CCU sub-network.

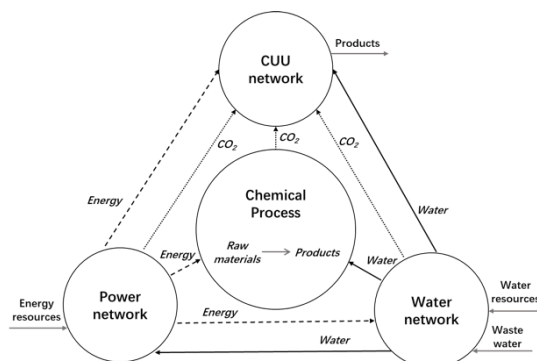


Figure 1. The integration framework of four networks.

2.2. Modeling

A superstructure-based optimization approach is employed to simultaneously synthesize and design a chemical process coupled with the power generation, water treatment and CCU sub-networks.

2.2.1. Superstructure

A superstructure is a visual representation to show all the feasible or allowed alternative routes of a decision-making design problem (Quaglia et al., 2012b). The superstructure for a process synthesis-design problem consists of columns, nodes, and arrows. They represent the processing steps, alternative processing techniques (intervals), and possible processing routes respectively. All sub-networks in Figure 1 will be represented using this process step-interval network superstructure (Bertran et al., 2017).

2.2.2. Generic interval model

The generic interval model gives the general form of the mathematical model for each interval to support the mathematical representation of the total network. As illustrated in Figure 2, the generic interval model has 5 operating units to convert the inlet flows to the outlet flows: a mixer, a reactor, a carbon emission calculation unit, a waste separator, and a product separation unit.

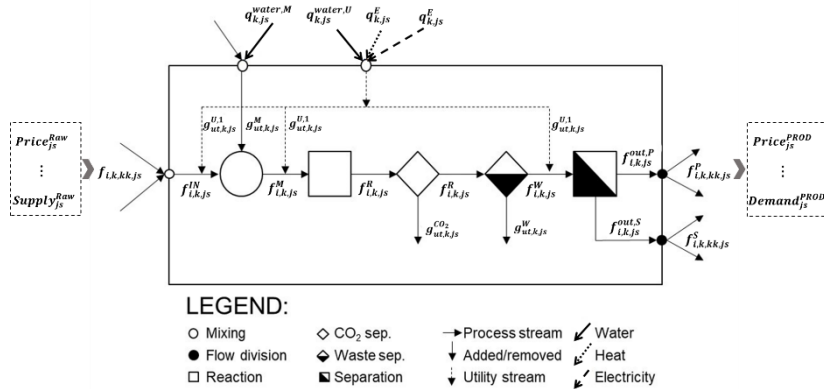


Figure 2. Generic interval model for multi-network stochastic programming.

2.2.3. Stochastic programming model

A two-stage stochastic programming approach (Birge et al., 1997) is considered to deal with the uncertainties of market prices of raw materials and products; and, market demand of products. The prices of raw materials and products will affect the selection of processes due to their different conversion and investment. On the other hand, market demand will affect the equipment scale. Here, the selection of production processing route and equipment scale (production capacity) are made “here-and-now” prior to the resolution of uncertainty for each-subnetworks, whereas the production rate of each sub-networks are made in “wait-and-see” mode after the uncertainties are revealed.

After setting the distribution of uncertain parameters, different scenarios js can be considered. In different scenarios, the general interval model has the following forms:

$$\text{Mixing} \quad g_{i,k,js}^M = \sum_{ii} f_{ii,k}^{IN} \mu_{i,ii,k,js} \quad (1)$$

$$f_{i,k,j_s}^M = f_{i,k,j_s}^{IN} + g_{i,k,j_s}^M \quad (2)$$

$$\text{Reaction} \quad f_{i,k,j_s}^R = f_{i,k,j_s}^M + \sum f_{react,k,j_s}^M \theta_{react,r,k} \gamma_{i,r,k} MW_i / MW_{react} \quad (3)$$

$$\text{CO}_2 \text{ separation} \quad g_{k,j_s}^{CO_2} = f_{co_2,k,j_s}^R \quad (4)$$

$$\text{Wastes separation} \quad g_{i,k,j_s}^W = f_{i,k,j_s}^R \delta_{i,k} \quad (5)$$

$$f_{i,k,j_s}^W = f_{i,k,j_s}^R - g_{i,k,j_s}^W \quad (6)$$

$$\text{Product-product separation} \quad f_{i,k,j_s}^{out,P} = f_{i,k,j_s}^W \sigma_{i,k} \quad (7)$$

$$f_{i,k,j_s}^{out,S} = f_{i,k,j_s}^W - f_{i,k,j_s}^{out,P} \quad (8)$$

$$\text{Utility consumption} \quad g_{ut,k,j_s}^U = \sum_i \beta_{ut,k}^1 f_{i,k,j_s}^{IN} + \sum_i \beta_{ut,k}^2 f_{i,k,j_s}^M + \sum_i \beta_{ut,k}^3 f_{i,k,j_s}^W \quad (9)$$

$$\text{Production rate} \quad f_{i,k,j_s}^M \leq F_{capacity} \quad (10)$$

The objective function of the multi-network optimization problem is determined as the expectation of the total economic benefits (Eq.11). P_{j_s} is the probability of each scenario. In addition, each scenario needs to meet environmental constraints (Eq.12, Eq.13).

$$Z = \sum_{n=1}^4 (P_{j_s} (S_{n,j_s}^{PROD} - C_{n,j_s}^{RAW} - C_{n,j_s}^C - C_{n,j_s}^E - C_{n,j_s}^H - C_{n,j_s}^{water}) - C_n^{CAP} / \tau) \quad (11)$$

$$\sum_{k_n} g_{k_n,j_s}^{CO_2} \leq CONS_{co_2} \quad (12)$$

$$\sum_{i_n} \sum_{k_n} g_{i_n,k_n,j_s}^W - Rec_{j_s} \leq CONS_{waste} \quad (13)$$

According to the above method, the SMINLP model of the multi-network optimization problem can be formulated and then simplified into the SMILP model through big-M method and capital piecewise linearization. Similar solution steps (Li et al. 2020) has been followed to obtain the optimal solution.

3. Case Study

3.1. Problem definition

The production of xylitol is considered as an illustrative example in this work. The raw material is corncob. Market price of corncob and xylitol, and market demand of xylitol are set as uncertain parameters. In this case, the multi-network optimization problem aims to find sustainable solutions with zero carbon emission, zero waste water emission and highest expected economic benefits.

3.2. Superstructure representation

The sub-networks of xylitol production, water treatment and power generation in this case follow the superstructure of previous work (Li et al. 2020). CCU sub-network is adapted from the case study of carbon dioxide utilization (Bertran et al. 2017). The superstructure of multi-network problem in this case is shown in Figure 3. The nodes (intervals) in each sub-network are set up using generic interval model.

3.3. Uncertain parameters distribution

Based on the market data collected in China, low, medium and high forecasts for prices of corncob and xylitol, and demands of xylitol are set as uncertain parameter values. The discrete distributions of uncertain parameters are listed in Table 1. Based on this distribution, 27 scenarios are generated.

Table 1. distributions of uncertain parameters.

Uncertain Parameters	Value	Unit	Probability
Price of corncob	97.5, 105, 112.5	\$/t	0.30, 0.40, 0.30
Price of xylitol	3900, 4200, 4500	\$/t	0.30, 0.40, 0.30
Demand of xylitol	10000, 12000, 14000	t/y	0.25, 0.5, 0.25

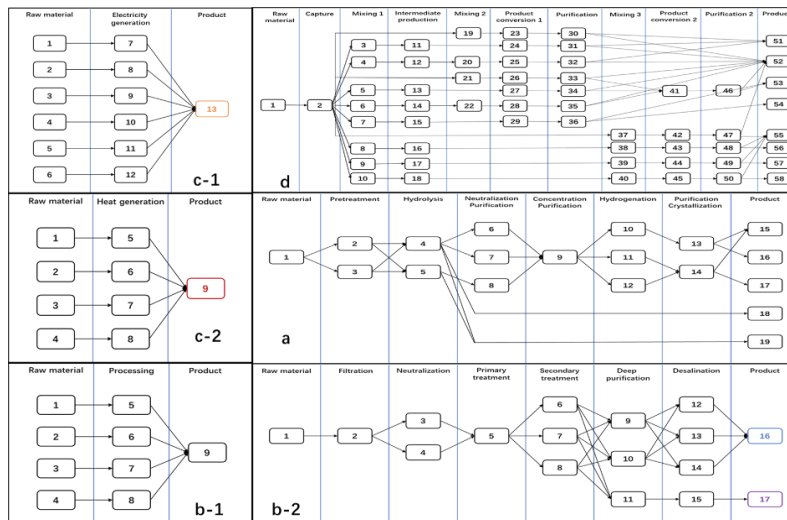


Figure 3. Superstructure representation of multi-network problem. a. Chemical process. b-1. Water plant. b-2. Waste water treatment. c-1. Electricity generation. c-2. Heat generation. d. Carbon capture and utilization.

3.4. Uncertain parameters distribution

After setting the uncertain parameter distribution, the SMILP model of multi-network optimization is established and solved simultaneously using the CPLEX optimizer. This model features 9,923,224 continuous variables, 10,178,711 equations and 381 binary variables whereas the deterministic model features 368,406 continuous variables and 378,062 equations. The optimal solution is obtained after 21186 iterations with 357 s CPU times. The results of the optimal routes are shown in Figure 4.

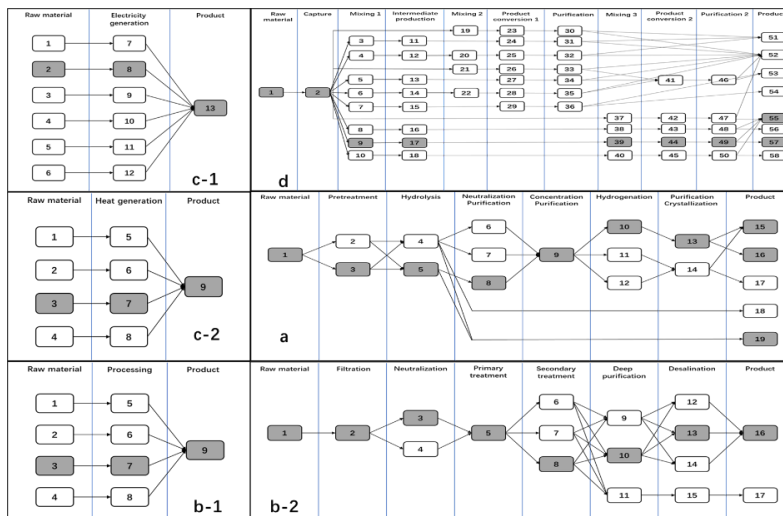


Figure 4. Result of optimal routes.

The optimal solution selects catalytic hydrogenation process to produce xylitol after hydrochloric acid hydrolysis. All waste water is deep purified and recycled. All carbon dioxide is captured by chemical absorption method and used to produce DMC. The optimal solution selects wind power and uses coal to produce steam. The mathematical results of the optimal solution and listed in Table 2.

Table 2. mathematical results of the optimal solution.

Term	Scenario	Value	Unit
Expected economic benefit	-	9,542,886	\$/y
Investment	-	172,892,655	\$
Output	1-9	10,000	t/y
	10-18	12,000	
	19-27	14,000	
CO ₂ capture	1-9	511,995	t/y
	10-18	614,394	
	19-27	716,793	
Steam consumption	1-9	853,325	t/y
	10-18	1,023,990	
	19-27	1,194,655	
Electricity consumption	1-9	172,566,400	kwh/y
	10-18	207,079,700	
	19-27	241,593,000	
Water consumption	1-9	2,341,356	t/y
	10-18	2,809,628	
	19-27	3,277,899	
Carbon emission	-	0	t/y
Waste emission	-	0	t/y

The results show that, in the face of market demand and price uncertainties, the multi-network stochastic programming model can help to obtain the optimal processing routes, equipment scale, and production rate for different scenarios with the highest economic benefit expectation. With the addition of the carbon capture and utilization sub-network, a comprehensive sustainable plan with zero carbon emissions can be obtained. However, in order to achieve zero carbon emissions, the scale of the CCU

sub-network is very large and leads to more investment, energy consumption and water consumptions, indicating the need for considering CCSU.

4. Case Study

An extended framework for integrating process, water treatment, power generation, and CCU networks has been proposed to obtain more sustainable alternatives with zero emission. In order to obtain the optimal solution under uncertainty, the stochastic programming of multi-network optimization problem has been implemented. The associated new methods are successfully applied to a multi-network problem of xylitol production under uncertainty and a sustainable optimal solution has been obtained.

References

- A. Quaglia, B. Sarup, G. Sin, R. Gani, 2012, *Computers & Chemical Engineering*, 38, 213-223.
- C. Ning, F. You, 2019, *Computers & Chemical Engineering*, 125, 434-448.
- J. R. Birge, F. Louveaux, 1997, *Introduction to Stochastic Programming*. New York: Springer-Verlag.
- MO. Bertran, R. Frauzem, A S. Sanchez-Arcilla, et al., 2017, *Computers & Chemical Engineering*, 106, 892-910.
- Y. Li, L. Zhang, Z. Yuan, R. Gani, 2020, *Computers & Chemical Engineering*, 141; 2020: 107041.
- Z. Yuan, M. R. Eden, & R. Gani, 2016, *Industrial & Engineering Chemistry Research*, 55(12), 3383-3419.

Optimising the Sustainability Performance of an Industrial Park: an Energy-Water-Food Nexus

Jamileh Fouladi^a, Ahmed AlNouss^a, Tareq Al-Ansari^{a,b*}

^a*Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, 34110, Doha, Qatar.*

^b*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, 34110, Doha, Qatar.*

talansari@hbku.edu.qa

Abstract

Rapid population growth has led to an increase in the demand for energy, water and food resources utilisation which results in environmental impacts and contributes to resource depletion. Therefore, it is necessary to design networks using various tools such as process integration and optimisation approaches, which have been applied to Energy-Water-Food (EWF) nexus studies. The main objective of the study is to design a systematic approach to capture the trade-offs between the economic and environmental metrics of sustainable design with industrial parks. It presents a novel superstructure and mathematical optimisation model that captures the synergies within EWF nexus considering interplants. Moreover, novel to this study, is the exergetic approach, which has not been previously applied to EWF nexus studies within industrial parks. The proposed nexus superstructure is applied to an eco-industrial park including wastewater treatment units, desalination plant, agricultural sub-systems, and a biomass gasification process for the recycling of biomass waste. The case study includes water-energy sources and sinks for chemical plants such as ammonia/urea and GTL processes in the State of Qatar which are simulated using the “what’sBest” Mixed-Integer Global Solver for Microsoft Excel by LINDO Systems Inc. Different cases with multiple optimisation objectives are evaluated in order to capture the trade-offs between the economic and environmental emission. Different indicators are used to assess the system resource efficiency such as exergy efficiency and global warming potential (GWP). The main focus is on capturing the synergic potential from utilising biomass from within the food sector and producing the syngas which decreases the natural gas consumption in other systems. Results of the study demonstrate clear benefits of water integration and biomass utilisation and contribute to the reduction of resource consumption. In the best scenario, the system demonstrates an 18 % reduction in the global warming potential, while the total annual cost of the design is increased by almost 9 %. The exergy efficiency of the system is enhanced reaching 53 %.

Keywords: Biomass, Wastewater, EWF Nexus, Optimisation, Integration.

1. Introduction

The growth in the global population affects consumption of energy, water, and food (EWF) resources. It has been estimated that the agriculture requires an increase of approximately 70 % in order to meet the population demand (Karan et al., 2018). Furthermore, the agriculture system consumes 70 % of the total global freshwater extraction, whilst food production accounts for nearly 30 % of the total energy consumed worldwide. Incidentally, there are significant interlinkages amongst EWF sectors, known as the 'nexus'. The integration and optimisation of the EWF nexus can lead to lower resource consumption and a more sustainable industry. Multiple studies have been conducted around the EWF nexus concept, for instance, a discussion on the decision making tools (Namany et al., 2019); a study based on a life cycle assessment (Al-Ansari et al., 2015); risk assessment studies within geospatial applications applied to food sectors (Haji et al., 2020); and studies on biofuels production (Alherbawi et al., 2020). The objective of this study is to apply the EWF nexus concept to the optimisation of industrial parks to support the transformation towards eco-industrial parks (EIP). The exchange of resources and wastes amongst various plants as advocated by industrial ecology can enhance the overall system efficiency. The majority of studies in this regard have focused on inter-plants integration, whilst others have adopted an intra-plants approach considering the integration of individual sectors such as energy and water. Lovelady and El-Halwagi developed a mathematical optimisation formulation to integrate the wastewater among number of plants (Lovelady and El-Halwagi, 2009). Alnouri et al. generated a water superstructure considering the direct and indirect water reuse for an EIP (Alnouri et al., 2014), where an optimisation model was developed to solve for the most cost effective and global optimum solution. In the concept of heat integration, different linear and non-linear programming models are proposed intra-plant and inter-plant (Nair et al., 2018). The energy-water integration is well studied within the EIP. Fouladi and Linke (2018) proposed an optimisation based model for a representative superstructure considering desalination, wastewater treatment, and cooling systems, and captures the trade-offs between the economic and environmental dimensions of the problem. Furthermore, Ghazouani et al. formulated a new optimisation model to integrate the mass and heat contents for multiple number of plants. Therefore, as mentioned previously, the concept of EIP within the EWF nexus needs to be further studied in order to explore the synergies between the two concepts which both aim to conserve resources. As such, the main objective of this study, is to generate and optimise a representative superstructure of EWF nexus which contributes towards the design of eco-industrial parks. A systematic approach is developed to search for the optimum design depending on economic and environmental objectives.

2. Network superstructure

In order to start the EWF nexus representation, a superstructure developed previously is used. The system boundaries are defined and sub systems are added to capture all the linkages between different sectors. In this work, some of the main sub-systems are considered in each sector, in which their integration and optimisation result in significant changes. The water sector consists of treatment units, desalination plants, and production units. Energy is considered to be in the form of a combined heat and power system, whilst the food sector includes the production of fertilisers and agriculture activities, all of which are connected to one another on a water-energy

sources-sinks basis. Figure 1 illustrates the systems and sub-systems of the proposed superstructure for the EWF nexus for a single plant.

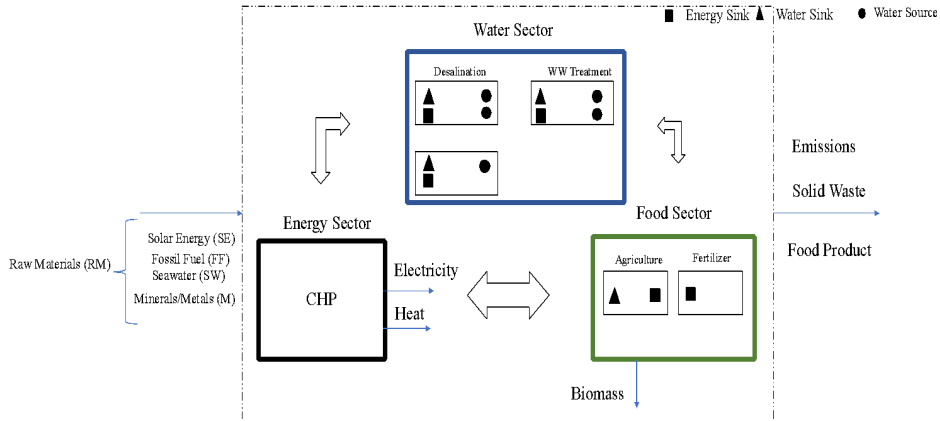


Figure 1: EWF Nexus representation systems and sub-systems

Moreover, Figure 2 captures the linkages among different sources and sinks.

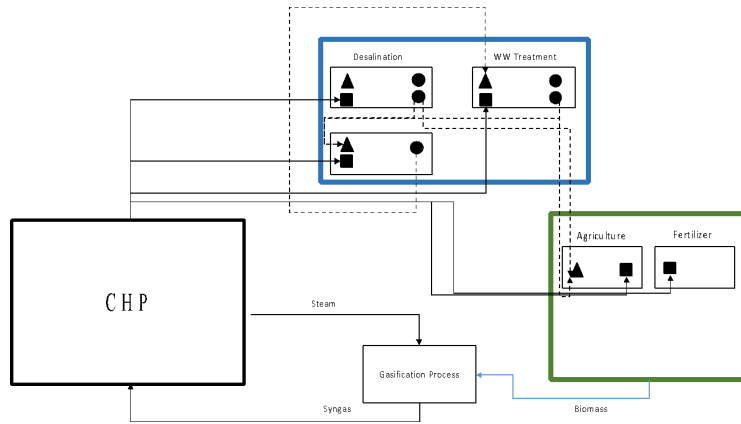


Figure 2: EWF Nexus with biomass utilization

An optimisation formulation is developed using network equalities (mass and energy balances), and inequalities (capacities, flow rates, and purities). The water network formulation is the expansion of an existing work (Alnouri et al., 2014). The “what’sBest” Mixed-Integer Global Solver for Microsoft Excel by LINDO Systems Inc is used to identify the global optimum solution of the network. The main objective of the problem is to minimise the total annual cost of the network as follows:

$$\text{Minimize: } C(\text{freshwater}) + C(\text{Treatment Units}) + C(\text{Fuel}) + C(\text{Biomass Utilisation})$$

Subject to:

$$g(x,y) < 0 \quad \text{Inequality constraints} \quad \square \quad \text{purities, flowrates, capacities}$$

$$h(x,y) = 0 \quad \text{Equality constraints} \quad \square \quad \text{mass balances, energy balances}$$

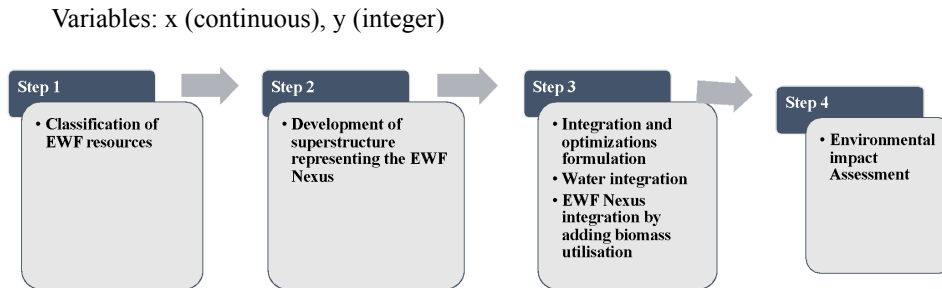


Figure 3: Modeling Procedures

Figure 3 illustrates the modelling procedures for the work which is all in terms of energy and water sources and sinks. Furthermore, to compare different scenarios and to support decision makers, a sustainability weighted return on investment metric (SWEIOM) developed by El-Halwagi (2017) is calculated using indicators such as energy, water, and waste reductions.

3. Results and discussion

A case study including two plants (GTL and Ammonia) with selected contaminants of TDS and TSS is solved to assess the proposed approach of EWF nexus. Water sources and sinks flowrates are taken from Aspen HYSYS results. HYSIS is a chemical process simulator which formulates the process mathematically. An air-stream gasification model developed previously to generate syngas is used for biomass utilisation scenarios (AlNouss et al., 2018).

Table 1: Simulation input data.

Flow rates (t/d)	GTL	Ammonia
Wastewater produced	9500	20000
Process water requirement	1625	31150

Four different cases are simulated and solved representing scenarios without any integration, with water integration, with water/biomass integration, and finally considering environmental objectives. In the first three cases, the optimisation objective minimises the cost, whilst in the final case the objective considers the minimisation of the environmental emissions. Table 2 summarises the main results from the optimisation solver.

Table 2: Results for the four selected cases.

Cases	Case 1 No Integration	Case 2 Water Integration	Case 3 Water-Biomass Integration	Case 4 Environmental Integration
Total Treated Water	0	19855	19855	19855
Total Annual Cost (M\$/year)	566	553	553	606

GWP (kg CO ₂ eq/year)	2.98E+10	2.98E+10	2.98E+10	2.43E+10
-------------------------------------	----------	----------	----------	----------

As illustrated from the results; the water network remains the same after case 1, which illustrates that wastewater treatment is chosen in both economic and environmental objective functions. The results in case 3 remains as case 2, which indicates that the biomass integration has been not selected as the optimum solution. For this reason, in the last case the objective function is changed. It can be noticed that the total annual cost is increased by approximately 9 %, while the global warming potential is decreased by 18%. The GWP remains constant until the final scenario since the energy integration is considered there by adding the sub-system which represents biomass utilisation.

Table 3 illustrates the exergy efficiency calculated for each optimisation case. It demonstrates that the exergy efficiency has the maximum value of 53 % with the biomass utilisation option and integrated wastewater sources.

Table 3: Exergy efficiency comparison for cases

Cases	Case 1	Case 2	Case 3	Case 4
Exergy Efficiency (%)	36.3	37	37	53

Moreover, to compare the cases and to select the most sustainable option, the SWROIM is calculated as illustrated in Table 3 below. Three different sustainability indicators are used; energy footprint, water footprint, and food waste footprint. The weighting factor for each indicator is varied in different scenarios to identify the best performing solution. As it is illustrated in Table 4, the findings demonstrate that for all indicators with different weighting factors, case 4 is selected as it has the largest SWROIM values.

Table 4: SWROIM results.

	Case 1	Case 2	Case 3	Case 4
SWROIM (%) (equal weights)	29	40	40	42.69
SWROIM (%) (0.5 water, 0.25 others)	31	53	53	54
SWROIM (%) (0.5 food, 0.25 others)	24	29	29	37
SWROIM (%) (0.5 energy, 0.25 others)	32	37	37	38

4. Conclusion

There are significant interactions amongst the EWF sectors which can be integrated to minimise consumption of resources. This study is novel as it considers an EWF Nexus approach within eco-industrial parks. It emphasizes on biomass utilization to capture the synergetic potential of integration of the EWF nexus concept with sustainability dimensions as applied to industrial parks. A case study to demonstrate this is developed accordingly and is evaluated from both economic and environmental perspectives. Evidently, biomass utilisation is an expensive process, however it can improve the environmental performance significantly. Finally, a sustainability metric is used to

compare different cases and to support decision makers in identifying the most sustainable configuration. As such, the results illustrate the significance of biomass utilisation as it enhances the exergy efficiency reaching 53 %. Generally, this work demonstrates the significance of the EWF nexus and its contribution towards resource management. The next goal is to study the seasonal effect on the optimised network and do some sensitivity analysis on different parameters. Moreover, adding more plants into the industrial park can also result in more significant outcomes where the pareto front curves are useful to capture the trade-offs between the economic and environmental dimensions.

5. Acknowledgment

This publication was made possible by GSRA grants no. GSRA6-1-0416-19014 from Qatar National Research Fund (a member of Qatar Foundation).

References

- A. AlNouss, G. Mckay, T. Al-Ansari, 2018, Optimum utilization of biomass for the production of power and fuels using gasification, *Computer Aided Chemical Engineering*, 43, 1481–1486.
- E. Karan, S. Asadi, R. Mohtar, M. Baawain, 2018, Towards the optimization of sustainable food-energy-water systems: A stochastic approach, *Journal of Cleaner Production*, 171, 662–674.
- E.M. Lovelady, M.M. El-Halwagi, 2009, Design and integration of eco-industrial parks for managing water resources, *Environmental Progress & Sustainable Energy*, 28, 2, 265–272.
- J. Fouladi, A. AlNouss, T. Al-Ansari, 2021, Sustainable energy-water-food nexus integration and optimisation in eco-industrial parks. *Computers & Chemical Engineering*, 146, 107229.
- J. Fouladi, P. Linke, 2018, Sustainable Industrial Water and Energy Nexus Integration for an Industrial Park, 13 International Symposium on Process Systems Engineering, 44, 1981–1986.
- M. Alherbawi, A. AlNouss, G. Mckay, T. Al-Ansari, 2020, Optimum Utilization of Jatropha Seedcake Considering the Energy, Water and Food Nexus, *European Symposium on Computer Aided Process Engineering*, 48, 229–234.
- M. Haji, R. Govindan, T. Al-Ansari, 2020, Novel approaches for Geospatial Risk Analytics in the Energy-Water-Food Nexus Using an EWF Nexus Node, *Computers & Chemical Engineering*, 106936.
- M. M. El-Halwagi, 2017, A return on investment metric for incorporating sustainability in process integration and improvement projects, *Clean Technologies and Environmental Policy*, 19, 2, 611–617.
- S.K. Nair, M. Soon, I.A. Karimi, 2018, Locating exchangers in an EIP-wide heat integration network, *Computers & Chemical Engineering*, 108, 57–73.
- S. Namany, T. Al-Ansari, R. Govindan, 2019, Optimisation of the energy, water, and food nexus for food security scenarios, *Computers & Chemical Engineering*, 129, 106513.
- S. Y. Alnouri, P. Linke, M.M. El-Halwagi, 2014, Water integration in industrial zones: a spatial representation with direct recycle applications, *Clean Technologies and Environmental Policy*, 16, 8, 1637–1659.
- T. Al-Ansari, A. Korre, Z. Nie, N. Shah, 2015, Development of a life cycle assessment tool for the assessment of food production systems within the energy, water and food nexus, *Sustainable Production and Consumption*, 2, 52–66.

A Study to Target Energy Consumption in Wastewater Treatment Plant using Machine Learning Algorithms

Akash Das, Piyush Kumar Kumawat, Nitin Dutt Chaturvedi

Process Systems Engineering Laboratory, Department of Chemical and Biochemical Engineering, Indian Institute of Technology Patna, Bihta, Patna, 801106, Bihar, India

Abstract

Wastewater treatment plants (WWTPs) are one of the world's largest municipal-level electricity users. Understanding the structure of the energy cost of WWTP is a challenge for planners, and its calculation could result in substantial saving capacity and cost estimation. The wastewater contains various organic materials, and the purpose of WWTP is to process these parameters to the appropriate parameter level. These input parameters are related to the demand for electricity but are challenging to model. The study utilizes some of the advanced machine learning models trained on real-world data to predict the energy consumptions of WWTPs. Four machine learning algorithms were leveraged on the available dataset to identify the model of best fit. The models' performance is evaluated based on standard metrics of mean absolute error (MAE), and root mean square error (RMSE). It shows that GRU is the model of best fit and is recommended for further future analysis.

Keywords: Wastewater treatment; Machine Learning; Data-Driven; Energy Consumption

1. Introduction

Wastewater treatment plants (WWTPs) may require about 1% to 3% of a country's total electrical energy output. Over 20% of public utilities' electrical energy consumption by municipalities is necessary for their operation (Capodaglio and Olsson, 2020). The energy requirements of water treatment and supply processes depend upon factors including the contamination of the water, water quality, geographical conditions, technology, and age of the industrial setup (Jian et al., 2014).

Energy is consumed at all levels of treating wastewater and is considered as a principal contributor to WWTPs expenses; to drive the influent pump, aerated grit chamber in the primary process, providing air in aeration process of biological treatment. Stillwell et al. (2010) stated that energy efficiency measures and treatment process modifications might reduce their energy cost. Machine learning applications are widely explored in recent years to carry out operations related to WWTPs. Guo et al. (2015) presented the study to predict effluent concentration in a WWTP using machine learning models. Li et al. (2019) used a radial basis function (RBF) neural network to forecast energy consumption.

In this study, machine learning algorithms are employed to estimate the energy consumptions from historical data. It is considered that the efficiency of WWTP is influenced by inflow loads, chemical oxygen demand (COD), and total nitrogen (TN). The available data set is trained using regression algorithms: support vector machines (SVM), artificial neural network (ANN), long-short term memory (LSTM), and gated

recurrent network (GRU). Further, error analysis has been performed to select the model with a minimum error corresponding to the data's best fitting. This study aims to find the best fitting model to estimate the approximate energy requirement for the process. This study assists the planner to intelligently select and utilize ML algorithms using the historical data of the influencing factors that can precisely model and estimate the energy consumption of WWTP given the influencing features.

2. Problem statement and data description

Given a set of historical data adapted from Li et al. (2019), each data point comprises influencing factors (inflow rate, COD removal, and TN removal) along with their subsequent energy consumption. A total of 360 data points are available over a period of one year. The influencing factors are the input parameters and energy consumption as an output parameter to develop and train the model.

The presented study's objective is to estimate energy targets using various ML regression algorithms using available historical data. These models were then compared to predict the best model to fit the model's data better.

3. Methodology development using machine learning algorithms

The proposed methodology uses a comparative approach among four machine learning algorithms. It compares them on standard metrics to conclude the algorithm that can best predict the data coming from WWTPs.

The input to RNN, LSTM, and GRU needs to be a 3D vector, typically consisting of the batch size, time steps, and features. Therefore, the data is segmented into a window size of 15 with three input features. The rolling window approach was utilized for this purpose. The input features were then normalized to get them within a definite range of values, enabling better training of the proposed models. After the datasets' preprocessing was completed, they were then divided into training sets (70%) and test sets(30%). The training datasets were used to develop the energy consumption models, and the test datasets were used to assess the accuracy and robustness of the developed models. The output predicted by the model was compared with that of the actual output on standard metrics of MSE and RMSE. The model that scores the least error in both the metrics was concluded to be the model of best fit. A graphical description of the proposed framework is provided in figure 1, and the machine learning models used are described in the following subsections.

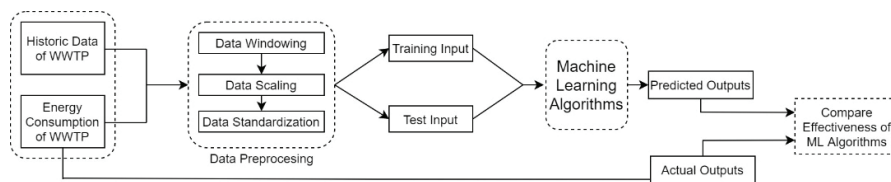


Figure 1. Proposed framework flow chart

3.1 Feedforward Neural Network Model

The Feedforward neural network is one of the simplest types of ANN devised. The information flows in the forward direction through the input layer of several hidden layers and a final layer of output nodes. The neurons are interconnected by weights, which form probability-weighted associations between input and output. The network trains itself by computing differences between the processed output by the network and the actual target

output. The network then adjusts the weights associated with each connection according to a set learning rate and the error values. After many such successive adjustments as the network loops over each time in the ANN, this way tends to produce outputs that are increasingly similar to the target output (results). The output from the j th node of an ANN is given by Eq. (1). Here O_j is the output vector, σ is the activation function, w_{ji} are the weight vectors and x_i is the output vector from the previous layer

$$O_j = \sigma\left(\sum w_{ji}x_i\right) \quad (1)$$

For this study, an ANN with two hidden layers and one output node has been used.

3.2 Recurrent Neural Network Model

Recurrent Neural Networks (RNNs) are an extension of the typical Feed Forward Neural Networks described above, using their internal state to process variable-length input sequences. The RNN handles variable-length sequences by having a recurrent hidden state whose activation at each time is dependent on the previous time. At a given time-step, each non-input unit calculates its current activation (output) as the nonlinear function of the weighted sum of all the activations of all the connected units (Lipton et al., 2015). For a given sequence $X = (x_1, x_2, \dots, x_n)$, the RNN will have activations $a^{<t>}$ and may optionally have an output $Y = (y_1, y_2, \dots, y_n)$, which may again be of variable length. Typically the activation and output of the recurrent hidden state are given by Eqs. (2) and (3). Where W_{ax}, W_{ay} are weight metrics and W_{aa} is the memory of the RNN that it has been maintaining from the previous layers.

$$a^{<t>} = g(W_{aa} a^{<t-1>} + W_{ax} x^{<t>} + b_a) \quad (2)$$

$$y^{<t>} = g(W_{ya} a^{<t>} + b_y) \quad (3)$$

This study uses a model with two bidirectional RNNs and one layer of simple RNN fed to a dense network of five hidden layers through an attention layer.

3.3 Long Short-Term Memory Model

The Long-Short Term Memory (LSTM) structure was motivated by an analysis of error flow in existing RNNs, which found that long time lags were inaccessible to existing architectures because the backpropagated errors either blows up or decays exponentially. Unlike the recurrent unit, which computes a weighted sum of the input signal and applies a nonlinear function, each j^{th} LSTM unit maintains a memory c_j^t at time t . The output a_j^t , or the activation of the LSTM unit is then given by Eq. (4),

$$a_j^t = \tau_j^t \tanh c_j^t \quad (4)$$

The memory cell c_j^t is updated by partially forgetting the existing memory and adding a new memory content c_j^t , as shown in Eq. (5),

$$c_j^t = f_j^t c_j^{t-1} + i_j^t c_j^t \quad (5)$$

where the new memory content is given by,

$$c_j^t = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c) \quad (6)$$

The extent to which the existing memory is forgotten is modulated by a forget gate f_j^t And the degree to which the new memory content is added to the memory cell is modulated by an input gate i_j^t . Thus unlike the traditional RNN, which over-writes its content at each step, an LSTM can regulate whether to keep the existing memory via the introduction of the gates, hence potentially capturing the long-term dependencies (Graves and Schmidhuber, 2005).

This research considers a model with two bidirectional LSTM layers and one layer of unidirectional LSTM, the output of which is fed to a dense network of five hidden layers.

3.4 Gated Recurrent Unit Model

A gated recurrent unit (GRU) was proposed by Cho et al. (2014) to make each recurrent unit adaptively capture dependencies of different time scales. GRU is similar to the LSTM, but it lacks the output gate and hence has fewer parameters. The activation $a^{<t>}$ of the GRU at time t is a linear interpolation between the previous activation $a^{<t-1>}$ and the candidate activation $a'^{<t>}$, given by Eq. (7), where an update gate τ_u decides how much the unit updates its activation or content.

$$a^{<t>} = (1 - \tau_u) a^{<t-1>} + \tau_u a'^{<t>} \quad (7)$$

The update gate is computed by Eq. (8), as,

$$\tau_u = \sigma(W_u [c^{<t-1>}, x^{<t>}] + b_u) \quad (8)$$

The reset gate is computed similarly to the update gate. This procedure of taking a linear sum between the existing state and the newly computed state is similar to the LSTM unit. The GRU, however, does not have any mechanism to control the degree to which its state is exposed but exposes the whole state each time. GRUs approximately maintain a similar performance and has fewer parameters that can help suppress overfitting.

The GRU for this study uses two bidirectional GRU layers and a unidirectional GRU layer passed on to a dense network with one hidden layer.

4. Results and Discussions

This study considers the influencing factors of inflow load, COD, and TN removal as feature vectors for predicting the energy consumption of the WWTP. Pearson's coefficient is calculated between the influencing factors and energy consumption is depicted in Table 1. It can be observed from the results that all the influencing factors display the least correlation with the target vector. This limits the developed models' performance and is considered a persistent problem in the data-driven approaches to model WWTP parameters. The table also lists the statistical details of the influencing factors and energy consumption.

Table 1. Details of the dataset preprocessing

	Pearson's Coefficient	Mean	Standard Deviation	Min	Max
Energy consumption	NA	22279.12	1884.362	5.39	9.52
Inflow load	0.391	7.642	0.624	16749	26913
COD removal	0.324	15.128	7.969	3.24	52.2
TN removal	0.385	1.856	0.543	0.52	3.42

The training process is carried out using ML algorithms with the training data. Each of the models' accuracy and robustness was then tested with the training data on standard metrics like MAE and RMSE. The model that yielded the least value of both the MAE and RMSE metrics was declared the model of best fit. Table 2 lists out the values that each of the models recorded when compared to the standard metrics.

The ANN model recorded the maximum MAE error of 0.77 results in the input data's poor fitting. Improved results are obtained with RNN, LSTM, and GRU models with MAE errors of 0.54, 0.54, 0.43, respectively. A similar trend is observed with RMSE errors presenting the edge they possess in modelling sequence data. The MAE error was reduced by 33% and RMSE error by 22% by changing the model architecture from the ANN layer to GRU layers. Analogous to the error values used to compare different models, Figure 2. represents the plot of energy consumption with the number of days. It compares energy predicted by the regression models trained in this study against the actual energy consumption by the WWTP. Figure 2(d) shows that the GRU model architecture could accurately predict the energy consumption of the WWTP as it fits the original energy consumption curve quite accurately. A lower precision of fitting is observed in the plots for the other ML models.

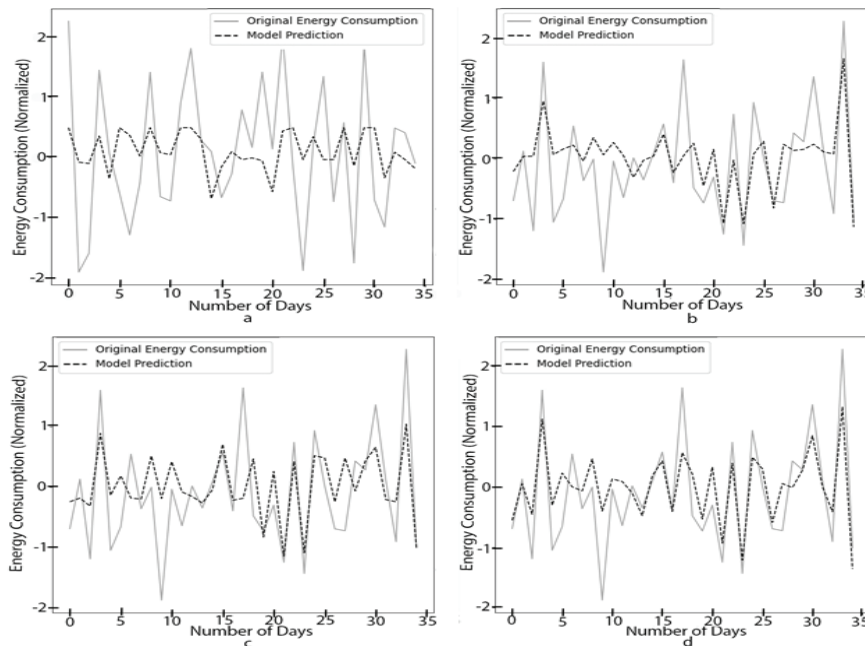


Figure 2. Energy predictions against actual energy consumption by the WWTP using (a) ANN, (b) RNN, (c) LSTM, (d) GRU

It can be seen from this study that all the sequence models (RNN, LSTM, and GRU) record a better performance from ANNs; this is because of the architecture of the sequence models that allows them to remember previous inputs while predicting the current ones. In RNN, it is difficult for it to learn to preserve the information over many timestamps due to vanishing gradient, which in turn affects the updated weights. Due to this, the network does not learn the effect of earlier input. GRU helps overcome this problem by the inclusion of update and reset gates in its architecture, where each gate has its weights and biases. This enables a better prediction of the target variable and achieves better accuracy than RNN.

The GRU model records the least error values on both metrics. Still, there is a noticeable difference between the output predicted by the model of best fit with that of the actual data. This difference can be reduced by using better quality data. However, it is challenging to model such data because of its increased non-linearity. Hence, an MAE error of 0.43 is acceptable even in the best fit model. Also, an increased number of available data would help in improving the accuracy of the models.

5. Conclusion

The energy consumption of a WWTP is a complex function and is influenced by several factors. In this paper, machine learning algorithms were used to create a suitable model that could predict the energy consumption of WWTPs. Historical data affecting energy consumption was taken into consideration for developing four different models. Developed models were compared based on MAE and RMSE for predicting energy consumption. GRU architecture gave the best results with the lowest error when tested on the entire dataset. The dataset's size and the data's complex characteristics became a significant hindrance in achieving greater accuracy. More parameters can be explored in future work, affecting energy consumption to predict outputs more accurately with a more extensive and well-correlated dataset. Also, significant developments in the model architecture, in terms of an ensemble model with a combination of ANNs, convolutional neural networks, and sequence models, will be explored as future research.

Acknowledgement

The authors would like to thank the Department of Science and Technology-Science and Engineering Research Board, India (DST-SERB) and Indian Institute of Technology, Patna for providing the research funding for this project under the grant no. ECR/2018/000197.

References

- A. Graves, J. Schmidhuber, 2005, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, *Neural Networks*, 18, 602–610.
- A. Stillwell, D. Hoppock, and M. Webber, 2010, Energy recovery from wastewater treatment plants in the united states: a case study of the energy-water nexus, *Sustainability*, 2, 945–962.
- A.G. Capodaglio and G. Olsson, 2020, Energy Issues in Sustainable Urban Wastewater Management: Use, Demand Reduction and Recovery in the Urban Water Cycle, *Sustainability*, 12, 266.
- H. Guo, K. Jeong, J. Lim, J. Jo, Y.M. Kim, J. Park, J.H. Kim, K.H. Cho, 2015, Prediction of effluent concentration in a wastewater treatment plant using machine learning models, *Journal of Environmental Sciences*, 32, 90-101.
- Y. Jian, W. Fu, L. H. Mao et al., 2014, Influence factors analysis of urban sewage treatment plant on energy consumption, *Journal of Beijing Jiaotong University*, 38, 33–37.
- Z.C. Lipton, J. Berkowitz, C. Elkan, 2015, A Critical Review of Recurrent Neural Networks for Sequence Learning, arXiv:1506.00019v4.

Development of sustainable integrated biorefinery networks in pulp and paper industries

Ghochapon Mongkhonsiri^a, Amata Anantpinijwatna^{b,*}, Pongtorn Charoensuppanimit^c, Amornchai Arpornwichanop^d, Rafiqul Gani^e, Suttichai Assabumrungrat^{a,f}

^a*Center of Excellence in Catalysis and Catalytic Reaction Engineering, Chulalongkorn University Biorefinery Cluster, Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand*

^b*Department of Chemical Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand*

^c*Control and Systems Engineering Research Laboratory, Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand*

^d*Center of Excellence in Process and Energy Systems Engineering, Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, 10330, Thailand*

^e*PSE for SPEED Company, Skyttemosen 6, Allerød DK 3450, Denmark*

^f*Bio-Circular-Green-economy Technology & Engineering Center, BCGeTEC, Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand 10330*

**amata.an@kmitl.ac.th*

Abstract

With the objective to obtain more sustainable production processes, the biorefinery network is integrated with traditional pulp and paper industries. A systematic framework with computer-aided tools consisting of synthesis, design and innovation stages has been applied to determine the biorefinery-integrated pulping process. An integrated network of succinic acid production and black liquor gasification for dimethyl ether (DME) production linked to an existing Soda pulping process is identified as the best option for increased profit, which also reduces pollutant emissions through integration of innovative CO₂ capture and utilization (CCU) steps to form the biorefinery-integrated-Soda-pulping network (BIS). This paper aims at also designing a sustainable biorefinery-integrated-Kraft-pulping network (BIK). Three integration scenarios are considered for further study: (I) the production of gasification-based dimethyl ether (DME); (II) the co-production of DME and succinic acid (SA); and (III) the co-production of DME and SA coupled with CCU. The best scenario is found to be Scenario II, which exhibits the best economic performance with 74% increase in profit compared to the conventional process. Scenario III achieves the highest energy efficiency at 39% and improved environmental performance, a 65% reduction of CO₂ emission compared to the conventional process, with only 0.7% profit reduction. The BIK option shows improved performance in terms of economic and environmental improvements compared to the BIS network, confirming that the integrated biorefinery network can transform the conventional Kraft pulping process to a more sustainable process with increased profit.

Keywords: Biorefinery network, Process synthesis and design, Pulp and paper industries, Systematic framework with computer-aided tool

1. Introduction

The existing infrastructure, logistics system and know-how of mature pulp and paper industry provide a great opportunity to commercialize the multiple-product network of biorefinery by transformation of low profitability of traditional pulping processes. Ekbohm et al. (2006) studied the black liquor gasification (BLG) to produce different transportation fuels. Dimethyl ether (DME) production has the highest efficiency of fuel energy output per unit of product. Moreover, the BLG-based biofuels and chemicals production in a Kraft pulping process can provide significant economic returns (Larson et al., 2006). Fornell et al., (2013) proposed an integrated kraft pulping process-biorefinery producing both ethanol and DME. Their results point to a feasible but electricity deficit system. The carbon capture is suggested but has not been considered in their study.

In an earlier work, Mongkhonsiri et al., (2018), applied a superstructure-based process synthesis to define an optimal network of a biorefinery integrated with an existing pulp mill. They also performed a scenario-based sensitivity analysis to study the effect of product price and uncertainty. Based on the generated superstructure, the network of integrated succinic acid production (SA) and black liquor gasification based dimethyl ether production (BLG-DME) were identified to have excellent potential for improved profitability. As further work (Mongkhonsiri et al., 2020), the biorefinery integrated-Soda pulping network (BIS) was designed and evaluated in terms of energy, economics and environmental impacts. The results show that the BIS improves both economic and environmental aspects. Moreover, a CO₂ capture and utilization unit (CCU) step was included to further reduce the environmental impact.

This paper aims to design and evaluate the biorefinery-Kraft pulping network (BIK) where biochemical, biofuel and bioenergy productions are combined with an existing Kraft pulping process. The integrated network is rigorously simulated. Moreover, the Integrated network is designed for energy self-sufficiency by utilizing biomass residues for steam and electricity generation. The process performance is evaluated in terms of energy, economics, and environmental impacts. The integration of two conventional processes with a Tomlinson recovery boiler and a Black-Liquor-Gasification-Combined cycle (BLGCC) have been used for comparison. The CCU design is also applied in the BIK network. Three scenarios, which are extended versions of those studied previously (Mongkhonsiri et al., 2020) compare different types of pulping processes.

2. Methodology

2.1. Process Design

Material flows for the designed network is illustrated in Figure 1. The biorefinery network includes SA, DME, and BGCC processes. The design aims for an energy self-sufficient network. Energy consumption in a Kraft pulping process is considered to provide the biomass fuel required in the BGCC operation. For this work, feedstocks for the BIK network are assumed to have different composition compared to the studied BIS network, which provides higher pulp yield and more BL's organic compound but less biomass fuel. The performances of the networks are discussed in section 3: results and discussion.

2.2. Process evaluation

Process performance is evaluated in terms of economic aspects, energy consumption and environmental impacts, based on the rigorous simulations. The operation period is assumed to be 8,330 h/y and the eucalyptus feed rate is 100 kt/y. The indicator for an

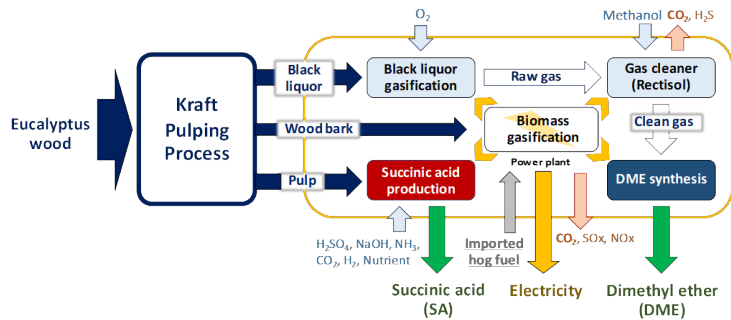


Figure 1. Material flow of integrated biorefinery-Kraft pulping network.

economic efficiency is the annual profit estimated from operating and capital costs based on 25 years of project lifetime. The capital cost includes only the new installation of integrated biorefinery process without the costs for the existing Kraft pulping facility. The environment impacts are evaluated by net emissions of pollutants: CO_2 , SO_x , NO_x and PM_{10} . The net gas emissions consider gas emissions from the processes estimated by emission factors (Larson et al., 2006) and the potential of gas reduction by the biomass utilization, the fossil-based products replacement, the diesel replacement by DME, the petroleum-based succinic acid and the fossil-based electricity generation based on the national grid of Thailand. Three different scenarios are considered.

Scenario I: Integration of the gasification-based DME production - This scenario considers only the integration of the BLG, the DME production and the biomass fuel utilization. Since the production of succinic acid was excluded; the unbleached pulp product was usually sold for paper making.

Scenario II: Integration of the DME and SA - This scenario includes the production of succinic acid (SA) from the unbleached pulp, together with the integrated network from Scenario I. The energy consumption of SA is included in the energy requirements.

Scenario III: Integration of innovative CCU - This scenario includes the CCU as part of the integrated network from Scenario II. CO_2 is captured from the flue gas by an existing Rectisol process without additional installation costs. CO_2 is utilized for methanol synthesis by the integrated CO_2 hydrogenation process. The performance of the methanol synthesis process has already been evaluated in terms of economic and environmental impacts by Frauzem (2017). The produced methanol is then supplied to the DME synthesis process; 30 ktons per year more of DME is acquired by the additional CO_2 -utilized methanol fed to the DME synthesis process.

3. Results and Discussion

3.1. Economic performance

The results of the economic assessment are reported in Table 1. The BIK network in Scenarios I, II and III point to profit over the conventional Kraft pulping process by 6, 40, 39 Million US dollars, respectively. Moreover, the excess power available in all integrated network scenarios provide additional income from the sale of electricity; as well as reduction of petroleum-based dependency. The integration of DME and SA processes (Scenarios II and III) do highly promote profitability of the existing Kraft pulping process. The CCU included in Scenario III decreases the profit by around 600 thousand US dollars per year or 1.1% compared to Scenario II. Although the integrated network in Scenario I yields half of the profit of Scenarios II and III, it still improves the

profit over the conventional Tomlinson boiler and BLGCC processes. Although capital cost of CO₂-utilized methanol production process and operating cost is increased by CCU (scenario III) comparing to scenario II, CO₂ capture and DME production process is an existing process for BLG-DME process. Compared to stand-alone process of CO₂ utilization, the integrated network of DME production from BL gasification and CO₂ utilization could reduce capital and operating costs.

3.2. Environmental impacts

The environmental impacts of the BIK network are calculated in terms of net pollutants: CO₂, SO_x, NO_x and PM10 emissions, reported in Table 2. The integrated DME and SA networks in Scenario II and the innovative CCU integration network in Scenario III discharge CO₂ and other pollutants by around 2 times more than Scenario I. Note that the conventional scenarios because of succinic acid production demands higher energy. Due to CCU integration, Scenario III achieves the highest net CO₂ emission reduction potential, thereby improving the potential of net CO₂ reduction by about 20% compared to BLGCC and 140% compared to Scenario II. Both Scenarios II and III with the integrated SA process can significantly reduce emissions of other pollutants due to the larger bio-based power generation replacing fossil-based fuel. Although, Scenario I has a smaller potential to reduce the net CO₂ than the BLGCC, it still could potentially lessen the total emissions of other pollutants with the produced DME substituting the use of diesel fuel. The net emissions of pollutants emissions for all scenarios of the BIK network are negative, indicating that the pollutions will be reduced in the global scale.

3.3. Comparison with the integrated biorefinery-Soda pulping network

The Soda pulping process has been previously considered as the receptor of the biorefinery-integrated network (Mongkhonsiri et al., 2020). The performance criteria of the BIK (Scenario I, II, III) and the BIS (Scenario IS, IIS, IIIS) networks are reported in Table 3. As can be seen from these results, the biorefinery concept is applicable to both types of pulping technology with comparable enhancement of the economic performance.

Table 1. Economic evaluation of integrated biorefinery scenarios.

Info/Process	Unit	Scenario I	Scenario II	Scenario III	Tomlinson	BLGCC
Raw material cost						
Eucalyptus wood	ton/y	100,000	100,000	100,000	100,000	100,000
Wood cost	\$/y	2,900,000	2,900,000	2,900,000	2,900,000	2,900,000
Purchased hog fuel	ton/y	42,500	102,500	102,500	-	-
Hog fuel cost	\$/year	977,500	2,357,500	2,357,500	-	-
Natural gas boost	MMBtu/y	-	-	-	-	15,100
Natural gas cost	\$/y	-	-	-	-	30,653
Total raw material cost	\$/y	3,877,500	5,257,500	5,257,500	2,900,000	2,931,653
Biochemical products sale						
Pulp	ton/y	39,510	-	-	39,510	39,510
Pulp sale	\$/y	21,493,440	-	-	21,493,440	21,493,440
DME	ton/y	5,920	5,920	27,500	-	-
DME sale	\$/y	4,025,600	4,025,600	18,700,000	-	-
Succinic acid	ton/y	-	17,800	17,800	-	-
Succinic acid sale	\$/y	-	53,400,000	53,400,000	-	-
Total product income	\$/y	25,519,040	57,425,600	72,100,000	21,493,440	21,493,440
Power income/cost						
Power production	MWh	88,000	185,000	185,000	20,600	63,800
Power use	MWh	82,200	115,000	143,000	61,200	65,000
Power export/import ^a	MWh	5,800	70,000	42,000	-40,600	-1,200
Power sale/buy	\$/y	580,000	7,000,000	4,200,000	-4,060,000	-120,000
Capital cost	\$/y	908,000	1,950,000	2,340,000	154,676	272,797
Operating cost	\$/y	908,000	3,150,000	15,270,000	154,676	272,797
Profit	\$/y	20,405,540	54,068,100	53,432,500	14,223,440	17,896,787

^a Negative value means power deficit and the power from other sources is required

Table 2 Environmental impact indicated by net CO₂, SO_x, NO_x and PM10

Info/Process	Unit	Scenario I	Scenario II	Scenario III	Tomlinson	BLGCC
CO₂ Emission						
Process CO ₂ emission	ton/y	158,000	350,000	289,000	47,900	63,200
CO ₂ from power import	ton/y	0	0	0	33,300	938
Total CO ₂ emission	ton/y	158,000	350,000	289,000	81,200	64,138
Air Pollutants Emission						
Total SO _x emission	ton/y	35.1	77.2	77.2	149	5.2
Total NO _x emission	ton/y	41	90.3	90.3	50.8	34.4
Total PM10 emission	ton/y	3.07	6.76	6.76	45.05	4.51
CO₂ Reduction						
Grid power production replacement	ton/y	72,200	151,000	151,000	16,900	52,300
Diesel replacement by DME	ton/y	3,320	3,320	6,150	-	-
Petro-based SA replacement	ton/y	-	33,600	33,600	-	-
Biomass consumption	ton/y	147,000	209,000	209,000	103,000	103,000
Total CO ₂ reduction	ton/y	222,520	396,920	399,750	119,900	155,300
Air Pollutants Reduction						
Total SO _x reduction	ton/y	354	693	732	72.2	224
Total NO _x reduction	ton/y	143	184	274	8.77	27.2
Total PM10 reduction	ton/y	87.9	172	181	17.9	55.4
Net CO₂ (Emission – Reduction)						
Net CO ₂ ^a	ton/y	-64,520	-46,920	-110,750	-38,700	-91,162
Net Air Pollutants (Emission – Reduction)						
Net SO _x ^a	ton/y	-318.9	-615.8	-654.6	76.8	-218.8
Net NO _x ^a	ton/y	-102	-93.7	-183.7	42.03	7.2
Net PM10 ^a	ton/y	-84.83	-165.24	-174.24	27.15	-50.89

^a Negative value means the consumption of CO₂ in the process

More biomass utilization in BIK leads to more reduction of CO₂ and pollutants because it provides bio-based production and energy by replacing the fossil-based products. A reduction of 30% in the net CO₂ emission is observed in Scenario II compared to Scenario IIS and 4% observed in Scenario III compared to Scenario IIS. Regarding integrated CO₂ utilization, Scenario III and IIS can reduce the net CO₂ and other pollutants together with a drop in profit. However, Scenario IIS shows higher profit than Scenario III because of higher DME production. In Scenario III, the higher sulfur content in the syngas is due to Na₂S (used as Kraft-pulping chemical), which causes a lower flow rate of the pure CO₂ stream captured for CO₂ utilization, but the amount of DME is less than in Scenario IIS despite more CO₂ emission. Although scenarios with integrated CCU in BIK and BIS shows profit drop, it could point to more benefit if a carbon tax is considered.

4. Conclusion

The biorefinery-integrated networks with pulping processes have been studied and compared with conventional stand-alone pulping processes. They point to more sustainable alternatives and enhance the biorefinery concept. Extension of the biorefinery concept has been successfully achieved through integration with the Kraft-pulping process. The integrated biorefinery-Kraft pulping process shows great potential for transformation of the conventional pulping process into a more sustainable biorefinery system with higher profitability, less fossil fuel dependence and less environmental impacts. Based on the results of this work, the biorefinery-integrated Kraft-pulping network has been found to be superior to the biorefinery-integrated Soda-pulping network in terms of economic and environmental impact considerations due to larger environmentally friendly activities. As extension of this work, it is suggested that energy and water integration be considered to further optimize and gain benefit in economics and

Table 3. Process performance comparison between integrated biorefinery-Kraft-pulping network and integrated biorefinery-Soda pulping network.

Info/Process	Unit	Biorefinery Kraft-pulping network			Biorefinery Soda-pulping network		
		Scenario I	Scenario II	Scenario III	Scenario IS	Scenario IIS	Scenario IIIS
Raw material	ton/y	100,000	100,000	100,000	100,000	100,000	100,000
	\$/y	2,900,000	2,900,000	2,900,000	2,300,000	2,300,000	2,300,000
Biomass fuel	ton/y	42,500	102,500	102,500	0	59,200	59,200
	\$/y	977,500	2,357,500	2,357,500	0	1,360,000	1,360,000
Pulp	ton/y	39,510	-	-	30,700	-	-
	\$/y	21,493,440	-	-	9,220,000	-	-
DME	ton/y	5,920	5,920	27,500	5,350	5,350	31,900
	\$/y	4,025,600	4,025,600	18,700,000	3,640,000	3,640,000	21,700,000
Succinic acid	ton/y	-	17,800	17,800	-	14,700	14,700
	\$/y	-	53,400,000	53,400,000	-	44,400,000	44,400,000
Power ^a	MWh	5,800	70,000	42,000	-3,240	83,100	56,700
	\$/y	580,000	7,000,000	4,200,000	-453,000	11,600,000	7,940,000
Capital & operating cost	\$/y	1,816,000	5,100,000	17,620,000	975,000	4,024,000	18,880,000
Profit	\$/y	20,400,000	54,100,000	53,400,000	9,300,000	52,000,000	51,200,000
Net CO₂ ^b	ton/y	-64,500	-46,900	-111,000	-61,300	-33,200	-107,000
Net SO_x ^b	ton/y	-319	-615	-654	-185	-561	-609
Net NO_x ^b	ton/y	-102	-93.7	-184	-101	-96.9	-207
Net PM10 ^b	ton/y	-84.8	-165	-174	-124	-150	-161

^a Negative value means power deficit and the power from other sources is required

^b Negative value means the consumption of CO₂ in the process

environmental impacts. If considered, carbon tax could improve the profit margin of integrated CO₂ capture and utilization processes.

Acknowledgement

The acknowledgement is made to the “Research Chair Grant” National Science and Technology Development Agency (NSTDA) and C2F Postdoctoral Fellowship of Chulalongkorn University.

References

- Ekbom, T., Berglin, N., Löegdberg, S., 2006. Black liquor gasification with motor fuel production - BLGMF II - A techno-economic feasibility study on catalytic Fischer-Tropsch synthesis for synthetic diesel production in comparison with methanol and DME as transport fuels. Technical Report, Swedish Energy Agency, Eskilstuna, Sweden.
- Fornell, R., Berntsson, T., Åsblad, A., 2013. Techno-economic analysis of a kraft pulp-mill-based biorefinery producing both ethanol and dimethyl ether. *Energy* 50 83–92.
- Frauzem, R., 2017. Sustainable process design with process intensification – Development and implementation of a framework for sustainable carbon dioxide capture and utilization processes. PhD. Thesis, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark
- Larson, E.D., Consonni, S., Katofsky, R., Campbell, M., Lisa K., Frederick, W.J., 2006c. A cost-benefit assessment of gasification-based biorefining in the Kraft pulp and paper industry. Princeton University and Politecnico di Milano, Volume 3: Fuel Chain and National Cost-Benefit Analysis.
- Mongkhonsiri, G., Charoensuppanimit, P., Anantpinijwatna, A., Gani, R., Assabumrungrat, S., 2020. Process development of sustainable biorefinery system integrated into the existing pulping process. *J. Clean. Prod.* 255, 120–278.
- Mongkhonsiri, G., Gani, R., Malakul, P., Assabumrungrat, S., 2018. Integration of the biorefinery concept for the development of sustainable processes for pulp and paper industry. *Comput. Chem. Eng.* 119, 70–84.

Optimization of biogas autothermal reforming integrated with PEM based on renewable energy

Jeongdong Kim^a, Meng Qi^a, Il Moon^{a*}

*^aDepartment of Chemical and Biomolecular Engineering, Yonsei University, 50
Yonsi-ro, Seodaemun-gu, Seoul 0722, Republic of Korea
ilmoon@yonsei.ac.kr*

Abstract

PEM is a mature technology to produce hydrogen using fluctuated renewable energy, but still utilization of oxygen from PEM has potential to integrate with reforming process. In this study, we proposed biogas autothermal reforming integrated with PEM and oxygen liquefaction system, to achieve continuous production of green hydrogen. The optimization is conducted to evaluate specific energy consumption. Conclusively, the proposed system produces green hydrogen of nominal 5596.7kmol/h in PEM section and constant 964.6kmol/h in ATR section based on fluctuated energy source.

Keywords: reforming, continuous operation, liquid oxygen storage, electrolysis, performance analysis

1. Main Text

Hydrogen is a promising energy source which has three times more heat than that of gasoline. Both electrolysis and reforming technologies could be considered as representative routes for hydrogen production. Integrating with renewable energy sources like wind, solar energy and biomaterial can make hydrogen close to complete clean fuel, the so-called “green hydrogen”.

In electrolysis, three different types of technology are studied for integrating with Renewable energy sources: alkaline electrolysis (AEL), polymer electrolyte membranes (PEM), and solid oxide electrolysis (SOEC). Of the technologies, PEM has fastest cold start time of seconds, and dynamic adjustment of operation load is possible. It has fast degradation and cost - ineffective as drawback, still it can be directly operated based on fluctuated renewable energy source, and most studies have been focused on operating PEM with renewable energy source. Clarke et al. (2010) proposed stand-alone electrolyser system consisted of PEM, hydrogen storage tank and renewable energy generator. In Barbir et al. (2005) study, grid independent and grid assisted PEM operation is compared conclusion revealed that PEM can be directly coupled with fluctuated energy source. However, in terms of economic feasibility, hydrogen production with PEM have potential to be integrated with other hydrogen production technologies as oxygen supplier.

Autothermal reforming (ATR), which is consisted of partial oxidation and steam reforming (SMR), requires oxygen supplier for hydrogen production, still its thermoneutral reaction is advantage compared with SMR required additional heat supply. In Timo et al. (2017), ATR combined with SMR has 27.7% more exergetic

efficiency than that of the SMR. Main source of inefficiency is at furnace of SMR which supplies heat and emits flue gas.

In this study, we proposed biogas ATR plant integrated with PEM and liquid oxygen storage plant (ATR-PEMLO) to produce hydrogen. ATR section in the proposed system is required to be operate steadily because of high operating temperature and pressure. To accomplish continuous oxygen supply from PEM to ATR, liquid oxygen production process in the proposed system acts as buffer from fluctuated renewable energy sources.

2. Process description

ATR-PEMLO system is comprised of three processes: (i) PEM stack operated by direct renewable energy source, (ii) cryogenic oxygen liquefaction system and liquid oxygen storage tank, (iii) ATR operated with biogas and oxygen continuously supplied from section (ii). Cold energy recovery in section (ii) and (iii) is accomplished by multi-stream heat exchanger (MSHE). Matlab R2017b is used to simulated PEM stack with governing equation of water electrolysis, and Aspen HYSYS V9.0 is used to simulate proposed system. Table 1 lists operating assumptions for the proposed system referred from Meng et al. (2008) for PEM, Sciacovelli et al. (2017) for cryogenic oxygen liquefaction system and conventional reforming operating range.

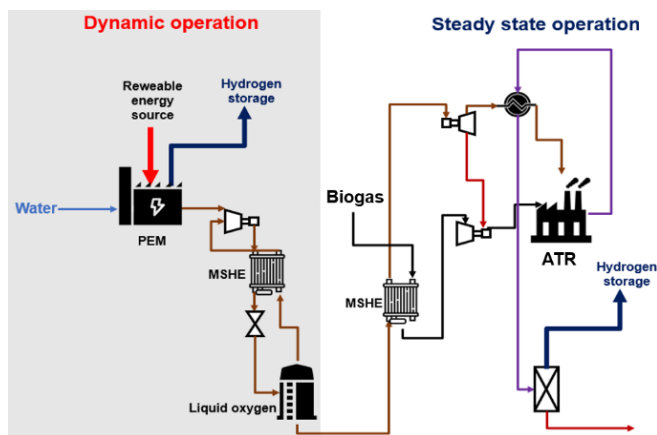


Figure 1. Schematic diagram of ATR-PEMLO system

Table 1. Operating assumption of ATR-PEMLO system

PEM current density	6000 A/ m ²
Thickness of the electrolyte membrane	50μm
PEM operating condition	80°C, 1bar
Compressor isentropic efficiency	85%
Turbine isentropic efficiency	85%
Cryogenic turbine isentropic efficiency	70%
Minimum temperature approach in MSHE	5°C

Minimum temperature approach in heat exchanger (HEX)	10°C
Liquid oxygen storage pressure range	1 to 18bar
Pressure range of reforming operation	7 to 10bar
Temperature range of feed in reformer	100 to 500°C
Carbon to oxygen (C/O) ratio range	1 to 3
Steam to carbon (S/C) ratio range	1 to 3

2.1. PEM modeling

During the water electrolysis operation, required power and heat can be determined by electro-chemical model.

$$E_{PEM} = JV \quad (1)$$

$$V = V_o + \eta_{act,a} + \eta_{act,c} + \eta_{ohm} \quad (2)$$

where V_o is the reversible potential determined by Nernst equation; η_{act} , η_{ohm} are activation overpotential of electrode, and ohmic overpotential of electrolyte, respectively.

In constant current density, activation overpotential can be calculated from Butler – Volmer equation expressed as;

$$\eta_{act,i} = \frac{RT}{F} \left(\frac{J}{2J_{o,i}} \right) = \frac{RT}{F} \ln \left[\frac{J}{2J_{o,i}} + \sqrt{\left(\frac{J}{2J_{o,i}} \right)^2 + 1} \right], \quad i = a, c \quad (3)$$

$$J_{o,i} = J_i^{ref} \exp \exp \left(- \frac{E_{act,i}}{RT} \right), \quad i = a, c \quad (4)$$

where J is current density; $J_{o,i}$ is exchange current density; J_i^{ref} is the pre-exponential factor; $E_{act,i}$ is the activation energy.

Heat supplied to PEM is calculated as;

$$Q_{PEM} = \frac{J}{2F} [T\Delta S - 2F(\eta_{act,a} + \eta_{act,c} + \eta_{ohm})] \quad (5)$$

Where ΔS is entropy increment of water electrolysis, and second term represents heat generation by irreversibility.

2.2. Cryogenic oxygen liquefaction process

Oxygen produced from PEM is sent to liquefaction section to store it as a liquid form. Heat recovery in two-stage compressor is conducted by pumped water with two HEXs. Thus, it enhances energy efficiency and heated water utilizes as feed in ATR section. After compressing, modified claud process is involved to liquify oxygen by renewable energy as electricity form. Two cryogenic turbines during MSHE produce cold energy

and remained cold energy is recycled to MSHE for complete liquefaction. Stored oxygen is continuously supplied to ATR section.

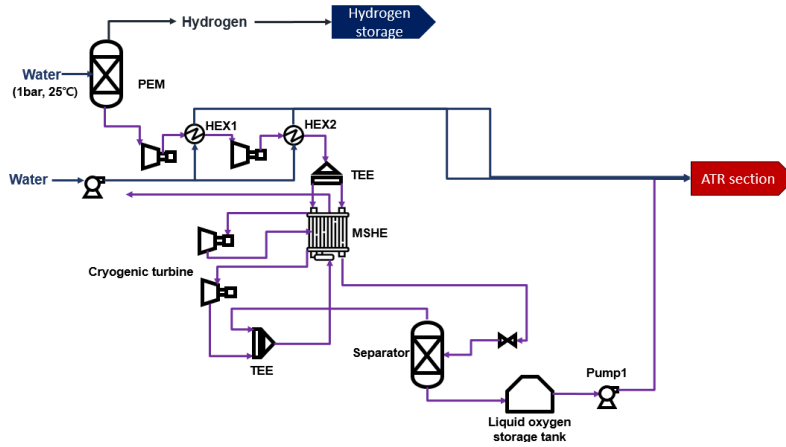


Figure 2. Cryogenic oxygen liquefaction process.

2.3. Biogas ATR process

ATR section is operated by cold energy stored in oxygen liquefaction section. High pressure liquid oxygen is vaporized by heat exchange with biogas and enters two-stage direct expand (DE) to meet operating condition of ATR. After reforming, high temperature product stream is decreased under 350°C for forward WGS reaction. This also means that thermal exergy of product stream can effectively heat turbine inlet temperature of DE process. Both cold energy utilization to biogas and DE process enables the system as complete standalone process. After purification, produced hydrogen is sent to storage facility with hydrogen from PEM section.

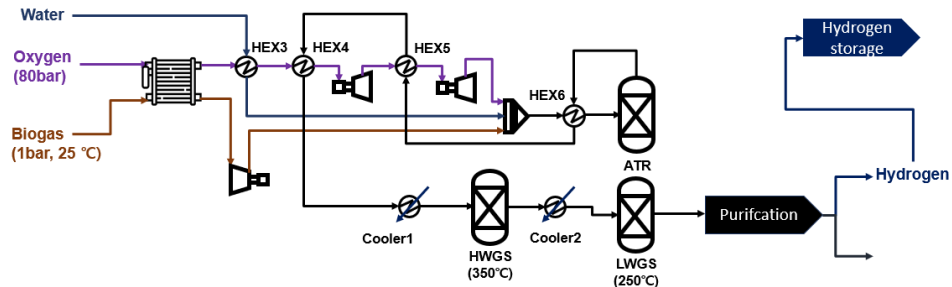


Figure 3. Biogas ATR process

3. Optimization of the proposed ATR-PEMLO system

To enhance energy efficiency of proposed system, the objective function is defined as specific power consumption with design variables as following equation.

$$\frac{P_{total}}{m_{H_2, total}} = \frac{P_{PEM} + P_{compressor} - P_{turbine}}{m_{H_2, PEM} + m_{H_2, ATR}} \quad (6)$$

where $X = (\gamma_{comp}, \chi_{oxygen}, T_{cryo1}, T_{cryo2}, T_{LO_2}, P_{storage}, \frac{C}{O}, \frac{S}{C}, P_{ATR}, T_{ATR})$

In the above variable set, γ_{comp} is compression ratio, and χ_{oxygen} is stream split ratio in liquefaction section, respectively; T_{cryo1} , T_{cryo2} are inlet temperature of first and second cryogenic turbine, and T_{LO_2} and $P_{storage}$ is storage temperature and pressure of liquid oxygen.; $\frac{C}{O}$ is carbon to oxygen ratio, $\frac{S}{C}$ is steam to carbon ratio, and feed stream composition is determined by these variables; P_{ATR} and T_{ATR} are operating pressure and temperature of ATR.

Subsequently, constraints determined by operating range in Table 1 as follows:

(1) The minimum temperature approach value of all MSHE should be higher than 5.0°C.
 $\Delta T_{min, MSHE} > 5.0^\circ\text{C}$

(7)

(2) The minimum temperature approach value of all HEX should be higher than 10.0°C.
 $\Delta T_{min, HEX} > 5.0^\circ\text{C}$

(8)

(3) The vapor fraction of the oxygen stream after MSHE should be lower than 1.
 $f_{vap, O_2} < 1.0$

(9)

(4) Because of safe operating range of oxygen storage tank and ATR, the temperature and pressure of corresponding unit should be placed in a specific range. Furthermore, S/C and C/O is also restricted in this study. Table 1 illustrated the constraints.

4. Result and discussion

For the optimization, `fmincon` in matlab which finds constrained minimum of function with constraints and interior-point algorithm is used to determine optimized variables. To evaluate objective function and constraints, HYSYS and matlab is linked during optimization procedure. Table 2 indicates the optimization results. In liquefaction section, 341.7kmol/h of liquid oxygen is stored with cold energy by consumption of renewable energy of 183.9kJ/mol. As a result of integration, ATR section is operated as standalone system because of power generation from DE process and utilization of cold energy from stored liquid oxygen.

Table 2. Optimization results for ATR-PEMLO system

Specific energy consumption for liquid oxygen	183.9kJ/mol
Liquid oxygen storage pressure and temperature	1000kPa, -153.6°C
Liquid oxygen production rate	341.7kmol/h
Hydrogen production in PEM	5596.7kmol/h
Specific energy consumption for ATR section	-1.8kJ/mol
C/O ratio	1.51

S/C ratio	2.53
Operating pressure and temperature	788kPa, 961°C
Hydrogen production in ATR	964.6kmol/h

5. Conclusions

A novel concept of continuous oxygen feed and cold energy using renewable energy, ATR-PEMLO system, is proposed in this study. To enhance the performance of hydrogen production, the independent variables in the entire system are optimized. As a result, ATR -PEMLO system shows 182.1kJ/mol specific energy consumption which is substantially low value compare to existing hydrogen production plant. Furthermore, the proposed system has advantage because direct coupling with renewable energy source is available without requirement of other facility to maintain energy source. By use of the proposed system, oxygen as byproduct of PEM can be effectively utilized as reforming agent. In conclusion, this study would contribute to efficient green hydrogen production from fluctuated renewable energy sources without additional energy input.

References

Integration of CO₂ capture to power plants: the effect of fuel and gas composition in process configuration

Ana Gabriela Romero-García,^a Nelly Ramírez-Corona,^b Eduardo Sánchez-Ramírez,^a Heriberto Alcocer-García,^a Juan Gabriel Segovia-Hernandez^{a*}

^a*Departamento de Ingeniería Química, Universidad de Guanajuato, Noria Alta s/n, Guanajuato, Gto., 36050, México.*

^b*Departamento de Ingeniería Química, Alimentos y Ambiental, Universidad de las Américas Puebla. ExHda. Santa Catarina Mártir s/n, San Andrés Cholula, Puebla, México, 72820*

**Corresponding author. Email: gsegovia@ugto.mx*

Abstract

In recent years, electricity demand has been increasing, according to the International Energy Agency, the energy sector is the biggest producer of greenhouse gas emissions (CO₂), having terrible environmental consequences. Various alternatives have been sought to reduce CO₂ emissions, highlighting the implementation of CO₂ capture and storage plants. In this work, it is shown an environmental and energetic analysis of a CO₂ capture plant coupled to a power plant. The analysis considered two different operating cases with different fuels in the power plant each case: biogas, coal, non-associated gas, and associated gas. The first one considers a constant fuel feed flow in the power plant. The second one, considers a constant energy demand. The results indicate that, for the first case, the fuel with the lowest environmental impact was the non-associated gas with 2.14 kEcopoints and 1083 MJ per kg of CO₂ recovered. For a constant energy demand, the fuel with the lowest environmental impact was the biogas with 0.57 kEcopoints and 193 MJ per kg of CO₂ recovered. For the systems here considered, those processes working with associated and non-associated gases remain the most efficient in terms of net energy produced.

Keywords: CO₂ Capture plant, biogas, coal, non-associated gas, associated gas.

1. Introduction

CO₂ is produced by diverse processes including the combustion of fossil fuels to produce electricity. According to the International Energy Agency, the 62% of the energy produced globally is obtained from burning fossil fuels, of which 27% comes from burning coal and 35% from burning natural gases such as: biogas, non-associated gas, and associated gas (IEA, 2019). In this way, electricity production is considered an unsustainable process according to the green chemistry principles and circular economy. However, current policies in both developed and developing countries still consider fossil fuel-based energies as the main source of energy. Several solutions have been sought to reduce CO₂ emissions, focusing on the implementation of CO₂ Capture and Storage plants (CCS) with post-combustion technology as the best alternative due to its high efficiency and easy implementation to existing power plants. During the capture, there are some technical aspects related to the process design and operation, for example, ensure enough

CO₂ capture, low energy requirements, low operating cost and low environmental load. Equally, the solvent selection for the CO₂ absorption, as well the required solvent gas ratio plays an important role on the process design. Pointing alkanolamines as a good solvent due to its high effectiveness, CO₂ capture also depends on the fuel used for electricity generation, because the flue gas composition greatly varies (Nagy, T. et al., 2013). Due to the growth of sustainability, green chemistry, and circular economy on the field of design and optimization processes, there is a clear need to analyze traditional processes. Even more the fossil fuel-based energies processes, focusing on the case of CO₂ capture plant coupled to power plant which it has been not reported yet. In this way, the intuitive benefit of CO₂ capture and reuse is apparently not so clear, and an analysis needs to be done in a critical and systematized manner.

In the present work, the environmental and energetic impact of coupling a CO₂ capture plant to power plant was analyzed according to an environmental perspective. Evaluating not only the possibility of reducing CO₂ emissions, but also turning electricity production in a sustainable process in accordance with the principles of green chemistry. It is important to highlight the environmental challenges involved to couple a CO₂ capture plant to a power plant. So, there are no work in the literature addressing the implementation of a CO₂ capture plant to a power plant from the perspective evaluated in this work. For the analysis made in this work, the use of four different fuels in the power plant was considered. Two operating scenarios were considered; in the first, the same fuel feed flow for all the plants and in the second, the same energy demand was specified. The design and simulation of the process plants were developed using the ASPEN Plus simulator, while the LCA was carried out with SimaPro software.

2. Study Cases and Methodology

In this work, two different scenarios of power generation plants coupled to a post-combustion CO₂ capture process are presented to analyze the environmental impact. The first case presents a constant fuel feed flow of 1 000 kmol/h and the second one presents a constant energy demand of 145 MW, each case was analyzed with the most used fossil fuels for power plants: biogas, coal, non-associated gas (NAG) and associated gas (AG). In both cases, for the simulation of the power plant coupled to a CO₂ capture plant the ASPEN PLUS process simulator is used (see Figure 1).

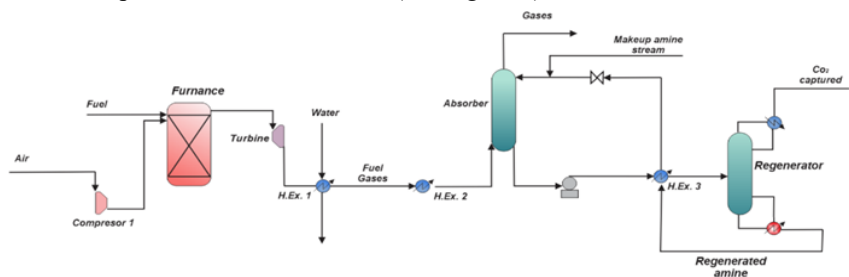


Figure 1. Representation of a thermoelectric power plant and the CO₂ capture plant in post-combustion using chemical absorption with monoethanolamine

For the power plant, the Peng-Robinson method is used to model the thermodynamic properties involved (Hasan, M. F. et al., 2012). For modeling the combustion chamber, it was considered the Gibbs free energy minimization by using a RGibbs block reactor. As operation specification it was considered a molar ratio of air to the fuel of 30: 1 with excess air ensuring complete combustion of each fuel and a pressure of 8 atm (Luyben,

2013). In first case the fuel feed flow it assigned to 1000kmol/h and for the second case the fuel feed flow was adjusted to achieve an energy production of 145 MW. For the simulation of the power plant the composition of the feed fuels was established as shown in Table 1.

Table 1. Flue composition in molar fraction

	CH ₄	C ₂ H ₆	C ₃ H ₈	i-C ₄ H ₁₀	N ₂	CO ₂
NAG	0.98	9E-3	1E-3	4E-4	4E-3	3E-3
AG	0.93	0.026	0.017	0.01	0.017	-
Biogas	0.8	-	-	-	0.015	0.185
	C	H	O	N	S	
Coal	0.51	0.41	0.066	7E-3	4E-3	

The flue gases obtained in the combustion were used as the feed flow to the absorption tower in the CO₂ capture plant. It was considered a chemical absorption using as solvent an aqueous solution of monoethanolamine (MEA) 30 % weight (Nagy, T., 2013). For modeling the absorption and regenerator columns it was considered the MESH equations by using a RadFrac equilibrium stage block. The reactions involved in the CO₂ capture involve the dissociation of the components involved in the absorption reaction, it is necessary to use a regenerator column to achieve high CO₂ recovery. For the regenerator column, the distillate flow and the reflux ratio were manipulated to achieve the biggest amount of CO₂ obtained in the combustion and thus reduce CO₂ emissions. For this reason, all the cases analyzed were standardized to a purity of 99 mol% CO₂ and a recovery of at least 95% of the CO₂ produced during the combustion. As a first step of the methodology applied, an optimization technique is employed to obtain the best design conditions for each case proposed in this work, then a LCA was made to the optimal design obtained in each case to analyze the environmental impact.

2.1. Optimization Methodology

For the optimization process, it was employed the stochastic optimization method of Differential Evolution with Tabu List (DETL) having as objective function the minimization of the energy requirement in the reboiler duty. The minimization of this objective was subject to the required recoveries and purities as is shown in Eq. 1.

$$\text{Min}(Q) = f(N_{TA}, N_{TR}, N_{FA}, N_{FR}, R_A, R_R, F_A, F_R, D_A, D_R) \quad (1)$$

Subject to:

$$\begin{aligned} \vec{y}_m &\geq \vec{x}_m \\ \vec{z}_m &\geq \vec{w}_m \end{aligned}$$

Where N_T are total number of column stages, N_R is the feed stage in column, R is the reflux ratio, F is the distillate/bottoms flux, and D is the column diameter. Where A and R represent the variables corresponding to the absorber and regenerator, respectively. Even though conventionally the diameter is calculated according to sizing equations, the main reason to consider the diameter as a decision variable lays on the fact of promoting a well hydraulic behavior. Eventually, the calculated diameter provokes flooding effects on the simulations which further may slightly influence the concentration profiles and the total mass balance. The multi-objective minimization considered 10 continuous and discrete variables. The optimization problem is subject to restrictions, the first restriction established is to achieve high CO₂ recovery, for this purpose \vec{y}_m represents the vector of CO₂ gases flow recovered in kg/h at the regenerator column, while \vec{x}_m represents 95 % of the flue gases fed to the capture section. The second restriction states that the

optimization problem is limited to ensuring CO₂ with a purity of 99 % molar, so \overline{z}_m represents the CO₂ purity in molar fraction at the exit of the regenerator column and \overline{w}_m represents a purity of 99 % molar fraction. The DETL method have shown being robust to optimize intensified separation systems. Srinivas & Rangaiah, (2013) showed that the use of some concepts of the metaheuristic tabu could improve the performance of differential evolution algorithm. The DETL algorithm is summarized in four steps: 1)Initialization, 2)Mutation, 3)Crossover and 4)Selection as is shown in Figure 2. The implementation of this optimization was made using a hybrid platform where the DETL method was coded in Microsoft Excel (ME). Initially, the method proposes a vector which is sent to Aspen Plus by means of dynamic data exchange (DDE). In there the separation process was rigorously simulated. For the optimization of process routes analyzed in this study, the following parameters for DETL method were used: 200 individuals, 500 generations, a tabu list of 50% of total individuals, a tabu radius of 0.0000025, 0.80 and 0.6 for crossover and mutation fractions, respectively.

2.2. Life Cycle Assessment

Once the optimal designs for each case were obtained it was applied a Life Cycle Assessment (LCA) to evaluate the environmental impact. According to ISO 14040, LCA has 4 phases: goal and scope definition, inventory analysis, impact assessment and interpretation. In the first one, the products to be studied shall be clearly defined in terms of the function that the product performs. The second, inventory analysis, involves data collection and calculation procedures to quantify the consumption of energy, raw material, air emissions, water discharges and solid wastes. The next phase is impact assessment, where the results of the inventory analysis are added up into environmental impacts using common equivalent units; for example, burning a fuel in a given process can be associated with effects on the impact category of global warming, which are measured by kilograms of CO₂ equivalent. The fourth phase of an LCA is interpretation, where the results obtained should be analyzed to establish understandable recommendations and decision arguments.

3. Results

In this section are presented the results of the two study cases. As mentioned earlier, the composition of the flue gases strongly depends on the type of fuel used during the combustion.

Table 2. Flue gas composition in molar fraction

	Case 1 – Fuel feed flow 1 000 kmol/h				Case 2 – Energy Demand 145 MW			
	Biogas	Coal	NAG	AG	Biogas	Coal	NAG	AG
N₂	0.767	0.786	0.767	0.766	0.749	0.781	0.765	0.766
O₂	0.168	0.189	0.144	0.139	0.135	0.162	0.139	0.139
CO₂	0.029	0.024	0.030	0.034	0.052	0.053	0.033	0.034
H₂O	0.035	0.002	0.059	0.062	0.064	0.004	0.064	0.062

Table 2 shows how the composition of the flue gases changes according to the fuel used during combustion. It is possible to observe that the CO₂ concentration increases in case two, where a specific energy demand must be achieved. This behavior can be explained as to achieve energy demand it is necessary to make an adjustment in the fuel flow. In case 1 the lower CO₂ concentration is obtained when coal is burned while the highest is obtained when Natural gases are burned. For case 2 is the opposite, the lower CO₂

concentration is obtained when natural gases are burned while the highest is obtained when coal is burned. These variations will have a direct influence on the capture effectiveness, energy requirements, amount of solvent as well as the CO₂ recovery (Nagy, T., 2013). Table 3 shows how CO₂ variations affect directly on operation conditions as the optimum ratios of solvent to flue gas (L/G), solvent to CO₂ recovered (L/CO₂ REC), reboiler duty to CO₂ recovered (QR/CO₂ REC), Net Energy per CO₂ recovered (E_{PP}/CO₂ REC) and environmental impact per CO₂ recovered (EI/CO₂ REC). In Case 1, it is needed more solvent to recover CO₂ in the capture process when CO₂ is obtained from coal, moreover when CO₂ is obtained from natural gasses the solvent to CO₂ recovery decreases. This behavior is consistent with the predicted from Table 2, as coal is the fuel with the lowest amount of CO₂ in case 1 will require more solvent during the capture and natural gasses will require less amount of solvent for CO₂ recovery. This trend changes in case 2, as coal and biogas are burned flue gases have a higher CO₂ concentration. This will require less solvent for CO₂ recovery. Moreover, reboiler duty in the regenerator column highly depends on the L/G ratio. In case 1 it is also coal the fuel that needs a higher energy requirement for the capture and natural gasses the fuels that require less energy. These results changed in Case 2, as the CO₂ concentration from burning coal increases, effectiveness also increase. The solvent and energy requirement involved in the capture decreases. For the net energy of the process per CO₂ recovered, in both cases the higher energy yield is when associated and non-associated gases are burned. Due to it is not an autonomous process, energy demand could be reduced, integrating heat from the combustion with the energy requirements of the regeneration column.

Table 3. Optimal operation conditions for both scenarios

	Case 1 – Fuel feed flow 1 000 kmol/h				Case 2 – Energy Demand 145 MW			
	Biogas	Coal	NAG	AG	Biogas	Coal	NAG	AG
L/G [kg/kg]	0.58	0.94	0.68	0.71	1.02	0.598	0.67	0.705
L/CO ₂ REC [kg/kg]	13.93	27.68	15.50	13.77	13.52	7.812	13.84	13.773
QR/CO ₂ REC [GJ/t CO ₂]	3.09	5.30	3.65	3.02	3.69	2.950	3.21	3.017
E _{PP} /CO ₂ REC [MJ/kg]	285	279	1083	990	193	311	1035	988
EI/CO ₂ REC [Ecopoint/kg]	4.11	354.2	4.25	9.14	0.22	2.34	4.45	4.96

The LCA analysis was conducted to evaluate the environmental impact during the generation of electricity, as well as the energetic and environmental implications of coupling a CO₂ capture process. As is shown in Figure 3, for case 1, the fuel with the lowest environmental impact is the non-associated gas with 2.14 kEcopoints. This can be explained as in this scenario, the non-associated gas is the fuel with the highest CO₂ concentration, improving capture efficiency. Also, is the one that requires less solvent and less energy for capture. These factors have a direct impact on environmental analysis. Iso, it is possible to see from Table 3 that the non-associated gas and the biogas are the fuels that have the lower environmental impact per kilogram of CO₂ recovered. For the case 2, the fuel with the lowest environmental impact is the biogas, with a score of 0.57 kEcopoints. In this case even though coal has better CO₂ capture, it is observed from the environmental analysis which is the fuel with the worst environmental indicators.

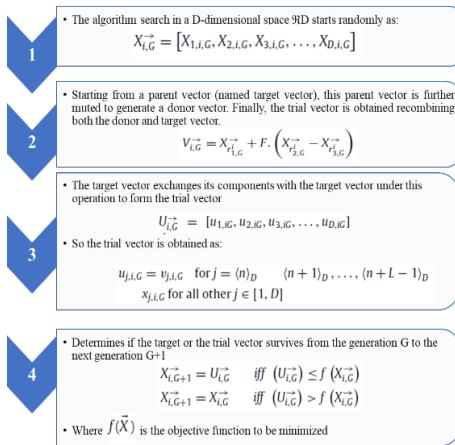
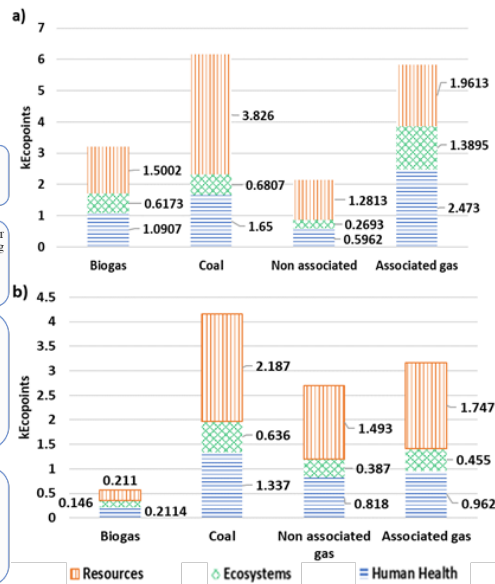


Figure 2. DETL methodology

Figure 3. Environmental impact in kEcopoint.
a) Best results for case 1, b) Best results for case 2

4. Conclusions

The implementation of the CO₂ capture process in power plants has been considered the most mature technology to reduce flue gas emissions associated to electricity production. Although capture plants help to reduce the environmental footprint of CO₂. However, before implementation it is necessary to make an analysis of the environmental implications that arise from the implementation, use of solvents and energy consumption. From the results obtained it can be concluded that the type of fuel and feed flow for power generation is a high weight variable during the capture process, impacting directly on solvent and energy requirements. As CO₂ concentration increases; CO₂ recovery and capture efficiency increases, so less solvent and energy requirements will be required during the capture process. Despite of the efforts in this field related to the efficiency of the capture process, it is essential to analyze the process considering not only the CO₂ capture as a strategy to reduce the negative effects of the power plant, but also by identifying new environmental effects due to the implementation of such capture process.

References

- Hasan, M. F., Baliban, R. C., Elia, J. A., & Floudas, C. A. (2012). Modeling, simulation, and optimization of postcombustion CO₂ capture for variable feed concentration and flow rate. 1. Chemical absorption and membrane processes. *Industrial & Engineering Chemistry Research*, 51(48), 15642-15664.
- International Energy Agency. (2019). World Energy Outlook 2019. International Energy Agency. Taken from: <https://www.iea.org/reports/world-energy-outlook-2019/electricity#abstract>
- Luyben, W. L. (2013). Chemical Process Engineering Principles of Combustion Turbines.
- Nagy, T., & Mizsey, P. (2013). Effect of fossil fuels on the parameters of CO₂ capture. *Environmental science & technology*, 47(15), 8948-8954.
- S., Sharma, and G.P. Rangaiah. "An improved multi-objective differential evolution with a termination criterion for optimizing chemical processes." *Computers & Chemical Engineering* 56 (2013): 155-173

Optimal Integration of Renewables, Flexible Carbon Capture, and Energy Storage for Reducing CO₂ emissions from Fossil Power Plants

Manali S. Zantye, Mengdi Li, M. M. Faruque Hasan*

*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843-3122, USA.
hasan@tamu.edu*

Abstract

Renewables can be used to supply the energy required for flexible CO₂ capture from fossil-fueled power plants, which in turn can act as an indirect energy storage to counter the intermittency of renewable energy. To that end, we propose a simultaneous design and operation scheme for the integration of renewables with dynamically varied flexible CO₂ capture connected to a fossil power plant. We develop a mixed integer linear program (MILP) to determine investment decisions for the integrated system in face of its high capital cost and the spatio-temporal variability of renewables. We also formulate a two-stage solution strategy to solve the resulting large-scale model by decoupling the design and time-dependent operational decisions. This framework is demonstrated through a nationwide case-study on coal power plants across the US.

Keywords: Clean Energy, Renewable Integration, Simultaneous Design and Operation.

1. Main Text

There is a significant push to adopt carbon-neutral energy sources to fulfill rising global energy demands. Although renewable energy is an inherently emission-free energy source, its intermittency and non-dispatchability pose several challenges in integration with electricity grids. To maintain grid reliability, cost-intensive infrastructure modifications are required including the installation of large-scale energy storage and increased cycling operations of fossil power plants. On the other hand, CO₂ capture and storage (CCS) can potentially reduce emissions from fossil power plants (Hasan et al., 2012); however, the high energy requirement for CO₂ capture, which can be 25–40% of the power plant output, restricts its widespread use. Renewables and CCS are different pathways for clean energy, and their limitations are traditionally addressed independently of each other leading to conservative integration costs and limited operational flexibility (Mac Dowell and Shah, 2014).

In this work, we propose that the synergy between the two technologies can be leveraged to counter their individual challenges. To this end, we evaluate the potential of a power generation system that integrates a coal power plant, a solar PV, a wind farm and a post-combustion CO₂ capture unit (Figure 1). The system delivers electricity to the grid using either fossil fuels, renewables, or a combination of both. The exhaust flue gas from the coal unit is directed to the capture system for CO₂ separation using a solvent-based absorption process. Steam is required to regenerate the solvent and can be extracted from

the steam turbines of the coal plant or using wind/solar generated power through an electric boiler. The capture system is flexible in a sense that it can dynamically vary the schedule and the steam demands for energy-intensive solvent regeneration, which is enabled via solvent storage (Zantye et al., 2019).

Such localized integration of renewables and CO₂ capture with existing coal plants presents several advantages as compared to the traditional grid-level integration. Firstly, renewable integration is facilitated without the need for costly grid modifications or the installation of additional transmission capacities. Secondly, time-varying flexible CCS operation using storage of the CO₂-rich solvent from absorption and the regenerated solvent from desorption improves the flexibility of the power plant without frequent cycling and steep ramping rates. CCS can utilize the excess renewable energy during off-peak demand periods, thereby keeping the power plant online for longer durations. Similarly, the capture operation can be turned down during peak load periods to increase the energy delivered to the grid. Thus, along with reducing the CO₂ emissions, flexible CCS operation acts as a form of energy storage to accommodate renewables.

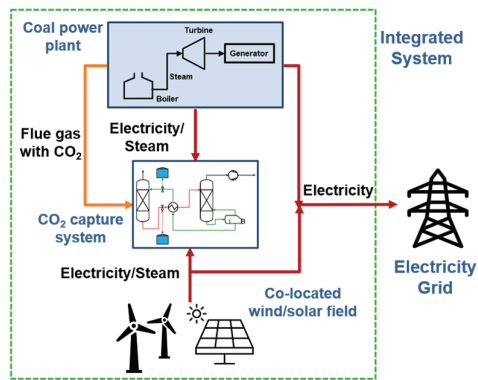


Figure 1: A schematic of the integrated system.

The concept of renewable-assisted CO₂ capture in fossil energy plants is relatively unexplored with the majority of the works in literature focusing primarily on the feasibility of the integration (Mokhtar et al., 2012). To meet the energy requirement of CO₂ capture using renewable energy, a medium to large-scale renewable farm is required. However, the high investment cost of renewables and its intermittency impacts the economic viability of the integration. To incorporate these trade-offs in optimal decision making, we describe a mathematical programming-based optimization model along with a computationally efficient methodology to solve the large-scale problem in Section 2. Lastly, we present a nationwide case-study and the potential for renewable-assisted CO₂ capture from coal power plants in the US in Section 3.

2. Model Formulation and Solution Strategy

The problem statement is as follows: for given time-varying electricity price, solar radiation intensity and wind speed profiles, determine the optimal design and operational schedules to integrate a CO₂ capture system with the coal-fired power plant and/or invest in a co-located renewable energy farm while maximizing the net present value (NPV) based on the projected costs and earnings from the overall system. The overall model (M0) is a large-scale mixed integer linear program (MILP) and has the general form as shown in Figure 2. On expansion of the general formulation M0, the design decisions x include: (i) the retrofit of the coal power plant with CO₂ capture given by the binary variable b^{ret} , (ii) the installed capacities of the co-located renewable field, denoted by

sz_{ren} , and (iii) the electric boiler capacity, designated as sz_b . The temporal variabilities in electricity price, wind speed and solar radiation are represented through deterministic scenarios, where $\omega \in \Omega = \{1, 2, \dots, NS\}$ denotes a scenario, and Ω is the set of all scenarios with hourly resolution. The objective function consists of two components: $f_d(\mathbf{x})$ representing the negative of the total investment cost which depends on design decisions \mathbf{x} alone, as shown in Eq. (2), and $f_o(\mathbf{y}_\omega, \mathbf{k}_\omega)$ representing the scenario-dependent operational revenue which depends on operational decisions \mathbf{y}_ω . Furthermore, $\mathbf{g}^{(d)}\mathbf{x}$ and $\mathbf{h}^{(d)}\mathbf{x}$ are the inequality and equality constraints on system design, while $\mathbf{g}^{(o)}(\mathbf{x}, \mathbf{y}_\omega, \mathbf{k}_\omega)$, $\mathbf{h}^{(o)}(\mathbf{x}, \mathbf{y}_\omega, \mathbf{k}_\omega)$ and $\mathbf{s}^{(o)}(\mathbf{x}, \mathbf{y}_\omega, \mathbf{y}_{\omega+1})$ are operational constraints.

The objective function of net present value, denoted by Eq. (1), is expressed as the difference between the system's net yearly earnings PF^{net} discounted for the time value of money and the current investment cost CC^{tot} . The NPV is determined assuming each year to be identical by using the time series data for solar radiation, wind speed and electricity price. This data is based on NREL studies where 30 years of data is condensed to a single representative or "typical" year for each location and is widely used for energy systems analyses. Cycling conditions on solvent storage are imposed to ensure that each year is identical. The discount rate r^{disc} and the project lifetime t^{lf} are considered to be 10% and 25 years respectively.

The first term in the total investment cost expression of Eq. (2) denotes the capital cost of the renewable technologies, the second term represents the cost of the electric boiler and the final term denotes the cost of the CO₂ capture system. The projected annual profit denoted by Eq. (3) is determined as the difference between the cumulative operational revenue and cost over all scenarios in a year. This corresponds to the second term of the objective function in the general formulation M0. The operational revenue given by Eq. (4) includes the system's earnings from long-term bilateral contracts with power distributors, participation in the spot electricity market and through the sale of the captured CO₂ to enhanced oil recovery (EOR) fields. The cost denoted in Eq. (5) comprises the operating costs of the coal power plant, tax on CO₂ emissions, cost of storing the captured CO₂ and transporting to end-use sites and finally, the ramping cost of the coal plant.

$$NPV = -CC^{tot} + PF^{net} \left(\frac{1}{r^{disc}} - \frac{1}{r^{disc}(1+r^{disc})^{t^{lf}}} \right) \quad (1)$$

$$CC^{tot} = \sum_{m \in ren} CO_m sz_m + CO_b sz_b + \left(CO^{capt} p_{cl}^{max} \frac{E_{cl}}{E_{cl}^{ref}} + N^{tank} CO^{tank} sz^{tank} \right) b^{ret} \quad (2)$$

$$PF^{net} = (1 - r^{tax}) \sum_{\omega \in \Omega} t_\omega (Rev_\omega - Cost_\omega) = \sum_{\omega \in \Omega} p_\omega f_o(\mathbf{y}_\omega, \mathbf{k}_\omega) \quad (3)$$

$$Rev_\omega = p^l \pi^l + (P_{g,\omega}^{in} - p^l) \pi_\omega^s + M_\omega^{capt} \pi^{eor} \quad \forall \omega \in \Omega \quad (4)$$

$$Cost_\omega = P_{cl,\omega}^{out} C_{cl}^{gen} + M^{em} C^{em} + M_\omega^{capt} C^{ts} + \frac{RC_{cl,\omega}}{t_\omega} \quad \forall \omega \in \Omega \quad (5)$$

The design constraints $\mathbf{g}^{(d)}\mathbf{x}, \mathbf{h}^{(d)}\mathbf{x}$ in the general formulation translate to capacity constraints for the renewable technologies. The operational constraints $\mathbf{g}^{(o)}(\mathbf{x}, \mathbf{y}_\omega, \mathbf{k}_\omega)$ and $\mathbf{h}^{(o)}(\mathbf{x}, \mathbf{y}_\omega, \mathbf{k}_\omega)$ include the overall energy balance, capacity and ramping constraints for the coal power plant and energy balance for the CO₂ capture system. The constraint

$\mathbf{s}^{(o)}(\mathbf{x}, \mathbf{y}_\omega, \mathbf{y}_{\omega+1})$ is a time-coupling constraint, which results from the mass balance for solvent storage, and further complicates the formulation.

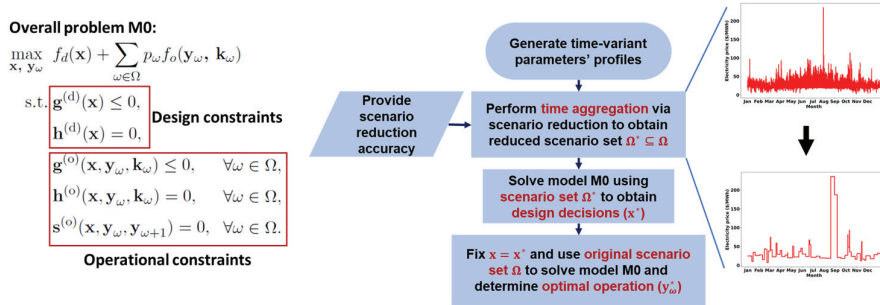


Figure 2: The general MILP formulation, and the proposed solution strategy.

The model M0, when used for large number of time points or scenarios, requires significant computation time to solve to determine the optimal long-term designs and short-term time-varying operating decisions. Therefore, we use a two-stage solution strategy as sketched in Figure 2. Specifically, the parameter profiles are first temporally aggregated using the SCENRED package (Dupacová et al., 2000) in GAMS environment. An example of such time aggregation is shown in Figure 2, where a yearly price profile with hourly resolution (8760 scenarios) is reduced to 53 scenarios while retaining a specified reduction accuracy of 94%. The reduced scenarios with reassigned frequencies are used to first solve the overall problem to obtain the optimal design decisions \mathbf{x} . The design decisions are then fixed and the problem is then solved using the original, unreduced scenario set to determine the optimal operational policy \mathbf{y}_ω . Although this strategy does not guarantee global optimality of the design decisions, it is effective in obtaining near-global solutions with a significant reduction in solution time (Arora et al., 2019).

3. Integration Case Study

The framework is demonstrated by a nationwide case-study across the US, through which we seek to identify coal power plants where it would be beneficial to invest in the integrated system to reduce CO₂ emissions. Data for 309 power plants with operational coal-based units shows a wide variation of capacities between 50 to 6000 MW, with 56% of the plants with capacities below 1000 MW. The CO₂ emission intensities of the standalone coal power plants vary between 0.5-2 ton MWh⁻¹, with the nationwide average being 1.11 ton MWh⁻¹. A full year of time-series data for wind speed and solar irradiation is obtained from NREL's NSRDB database.

Considering projected renewable energy prices of 300 \$ kW⁻¹ and a high carbon emission tax of 80 \$ ton⁻¹, the optimization results indicate that investment in a solar-assisted CO₂ capture system is beneficial for nearly one-third of the power plants across the US. Figure 4 shows the variation of the optimal co-located solar farm size. We observe a strong positive correlation between the co-located solar farm selection and the solar energy

availability represented by the capacity factor. Furthermore, the optimal PV farm size increases linearly with the power plant nameplate capacity, as can be seen from Figure 3a with the nationwide average size being 25.8% of the coal plant capacity. Figure 3b shows that the optimal capture system size is influenced by both the power plant nameplate capacity and its base-case emission intensity. The variation of the optimal levelized cost of electricity (LCOE) of the integrated system is displayed in Figure 4.

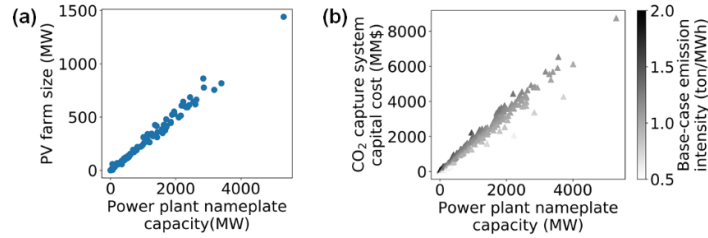


Figure 3: Optimal integrated (a) PV farm sizes, (b) CO₂ capture system capital costs.

We observe that the optimal LCOE is less than 40 \$ MWh⁻¹ or the cost of a new natural gas combined cycle (NGCC) plant, with higher LCOE reduction obtained for plants with solar integration. Furthermore, the reduction in CO₂ emission intensity obtained ranges between 87.5-91%. This indicates that it is economically profitable for the coal plant operators to invest in the integrated system as compared to replacing the plant with a natural gas unit to reduce emissions.

Figure 5a shows the distribution of solar output for a representative day in summer based on the full model optimization results. To evaluate the value of the CO₂ capture system in reducing renewable intermittency, we calculate the equivalent battery size required if the excess solar energy used in CO₂ capture is instead stored in a battery. Considering the high efficiency Li-ion battery type with a 2025 predicted total project cost of 362 \$ kW⁻¹, we observe the required battery capital cost to be linearly correlated with the co-located PV farm size as given in Figure 5b. Furthermore, the average battery capital cost to handle solar intermittency is found to be 4.4 times the solar PV farm size. The CO₂ capture system thereby avoids large investment while addressing renewable intermittency and directly reducing the CO₂ emissions of the coal power plant.

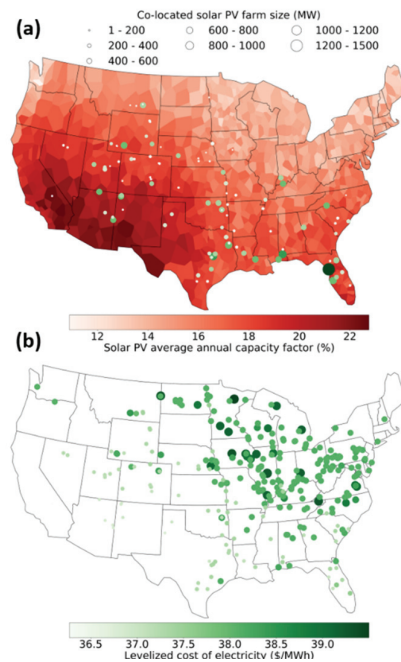


Figure 4: Nationwide integrated system design: (a) Selected co-located solar PV farms with optimal sizes, (b) Optimized LCOE of selected coal plants.

Figure 4: Nationwide integrated system design: (a) Selected co-located solar PV farms with optimal sizes, (b) Optimized LCOE of selected coal plants.

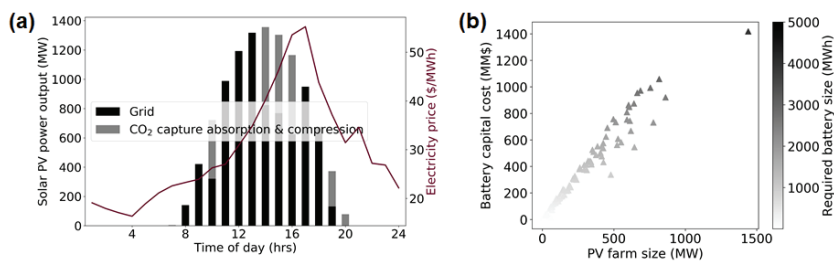


Figure 5: Equivalent battery size required to store excess solar energy. (a) Excess solar energy charged to battery, (b) Required battery size and capital cost for nationwide case.

4. Conclusions

A two-stage optimization framework using scenario-based time-aggregation was used to simultaneously determine the optimal design and operation decisions for clean energy. We found that the localized integration of renewables and CO₂ capture with coal power plants could compensate the CO₂ capture energy requirement as well as handle renewable energy variability. The results indicate that for a carbon tax above 80 \$ ton⁻¹ and renewable price below 300 \$ kW⁻¹, investment in a solar-assisted CO₂ capture is profitable for one-third of coal plants in the US. The resulting LCOE is lower than that of a new NGCC plant with 87.5-91% emission reduction. Thus, the integration can prove to be profitable despite the large capital investment. Furthermore, the CO₂ capture acts as an energy storage and avoids a battery investment of 4.4 times the capital cost of solar field to counter its intermittency.

5. Acknowledgements

The authors gratefully acknowledge support from U.S. Department of Energy (Grant number DE-FE0031771).

References

- A. Arora, J. Li, M. S. Zantye, M. M. F. Hasan, 2019. Process design frameworks for economic utilization of small-scale and unconventional feedstocks. In: *Computer Aided Chemical Engineering*. Vol. 47. Elsevier, pp. 83–88.
- J. Dupacová, N. Grüwe-Kuska, W. Römis, 2003. Scenario reduction in stochastic programming. *Mathematical Programming* 95 (3), 493–511.
- M. M. F. Hasan, R. C. Baliban, J. A. Elia, C. A. Floudas, 2012. Modeling, simulation, and optimization of postcombustion CO₂ capture for variable feed concentration and flow rate. 1. chemical absorption and membrane processes. *Industrial & Engineering Chemistry Research* 51 (48), 15642–15664.
- N. Mac Dowell, N. Shah, 2014. Optimisation of post-combustion CO₂ capture for flexible operation. *Energy Procedia* 63, 1525–1535.
- M. Mokhtar, M. T. Ali, R. Khalilpour, A. Abbas, N. Shah, A. Al Hajaj, P. Armstrong, M. Chiesa, S. Sgouridis, 2012. Solar-assisted post-combustion carbon capture feasibility study. *Applied Energy* 92, 668–676.
- M. S. Zantye, A. Arora, M. M. F. Hasan, 2019. Operational power plant scheduling with flexible carbon capture: A multistage stochastic optimization approach. *Computers & Chemical Engineering* 130, 106544.

Circular Economy Systems Engineering: A case study on the Coffee Supply Chain

Stefanos G. Baratsas^a, Efstratios N. Pistikopoulos^a, Styliani Avraamidou^{a,*}

^a*Texas A&M University, 400 Bizzell St, College Station, TX 77843, United States*

**styliana@tamu.edu*

Abstract

The current linear “take-make-waste” extractive models that have powered the tremendous growth of the last centuries, lead to the depletion of natural resources and environmental degradation. Circular Economy (CE) aims to address these impacts by building supply chains that are restorative, regenerative, and environmentally benign, through re-utilizing products and materials, using renewable energy sources, and closing any open loops. Process Systems Engineering (PSE) could play a critical role in this transition, providing the necessary analytical tools for a quantitative evaluation, and exploring alternative pathways for social and economic advancement. As such, we present here a novel framework for the modeling and optimization of CE food supply chains, along with a representative case study for the supply chain of coffee. First, the alternative pathways for the production of the desired product and the utilization of wastes are identified. Then, a Resource-Task-Network (RTN) representation that captures all these pathways is constructed. Since this analysis must be conducted holistically, a mixed-integer linear programming model that captures the entire supply chain, its objectives, and constraints is formulated, and it is solved to multi-objective optimality.

Keywords: circular economy, coffee supply chain, sustainability assessment, superstructure optimization, resource-task-network.

1. Introduction

Rising populations across the world seek to improve their standards of living, placing huge stresses on natural resources and supply chains. Energy and operational efficiency, improvement in manufacturing processes, and economic growth are vital to fulfill the increasing demand for goods, food and services. However, they still lead to natural resource degradation, substantial waste generation, water contamination, and surging greenhouse gas emissions. Thus, economic expansion shall be combined with sustainable development, ensuring the advancement of our societies while preserving the environment.

This requires a fundamental transformation of our economic model that promotes the “take-make-use-dispose-pollute” concept to a more “sustainable” one. Circular Economy (CE) has emerged as a potential solution for such a transition, with extreme emphasis being put towards improvement in reuse, remake, repair and recycling. CE aims to solve resource, waste, and emission challenges confronting society by creating a production - to - consumption total supply chain that is restorative, regenerative, and environmentally benign (MacArthur, 2015). The goals and key characteristics of CE (Reichel et al., 2016) are summarized as follows: a) minimization of material losses/residuals, b) reduction of input and use of natural resources, c) increase in the share of renewable resources and energy, d) reduction of emission levels, and e) increase the value durability of products.

CE can contribute to all dimensions of sustainable development, although the concept is relatively new, and there is little scientific guidance regarding its successful implementation and its effective evaluation. On top of that, significant challenges for the decision making are created from the interconnections among the diverse supply chain elements, stakeholders, and regulatory environments. Therefore, a holistic systems engineering approach is required to quantitatively navigate and thoroughly address the multi-scale, multi-faceted and interconnected CE supply chains (Avraamidou et al., 2020).

In this work, we present the foundations of a systems engineering framework and quantitative decision-making tool for the analysis and trade-off optimization of interconnected food supply chains. The framework utilizes mixed-integer modeling to establish the interconnections between different stages of the circular food supply chains as well as multi-objective optimization to consider all CE objectives and analyze trade-offs. The supply chain of coffee is used as a working example to illustrate the use and applicability of the proposed framework.

2. Coffee Supply Chain

Coffee is one of the most popular beverages globally with more than 167 million 60-kg bags of coffee being consumed yearly worldwide (ICO, 2020; Lombardini, 2017). The overall life cycle burdens of just one cup of spray dried soluble coffee (Humbert et al., 2009; Murthy and Naidu, 2012) are significant, as it is shown in Figure 1.

With the current linear supply chain model (Figure 2), only a small fraction of these burdens are reused or recycled, even though there are plenty of studies demonstrating sustainable alternatives (Murthy and Naidu, 2012; Chala et al., 2018; Blinová et al., 2017). Hence, as the coffee consumption increases, so does the amount of organic coffee waste and the amount of resources used, aggravating both the waste, water and energy management problems. A solution to this problem would be the transition to a circular supply chain (Figure 3) where renewable energy resources and more efficient processes will be utilized; water consumption will be reduced and wastewater treatment will be applied; waste generation will be minimized, and all wastes will be collected and utilized for the production of alternative products. The material flows ultimately shall be closed, while the energy flows shall be opened.

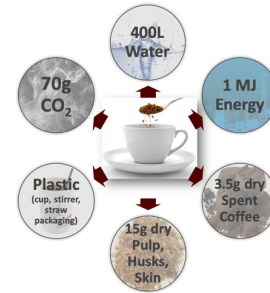


Figure 1: Life cycle burdens of 1 cup of spray dried soluble coffee

3. CE Systems Engineering Framework

The transition to a CE model while tackling the challenges from the multi-scale, multi-faceted and interconnected CE supply chains require a holistic systems engineering approach (Avraamidou et al., 2020). Therefore, and as an initial step, we have to identify and assess the alternative pathways for the production of the desired product, i.e. wet or dry method for the production of green beans (Chala et al., 2018). Coffee husk is the main pre-roasting by-product from the dry processing method, while coffee pulp, mucilage and parchment are the by-products of the wet processing method. Silver skin is the by-product from roasting the green beans, and spent coffee grounds are the final by-products from either brewing or extraction processes. Similarly, we have to identify and assess the alternative pathways for the waste utilization, i.e. utilization of pulp, husk, or spent coffee

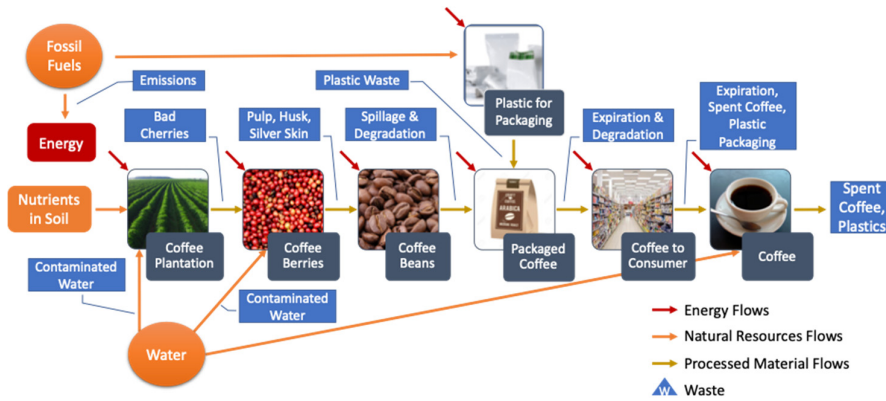


Figure 2: Simplified Linear Supply Chain of Coffee

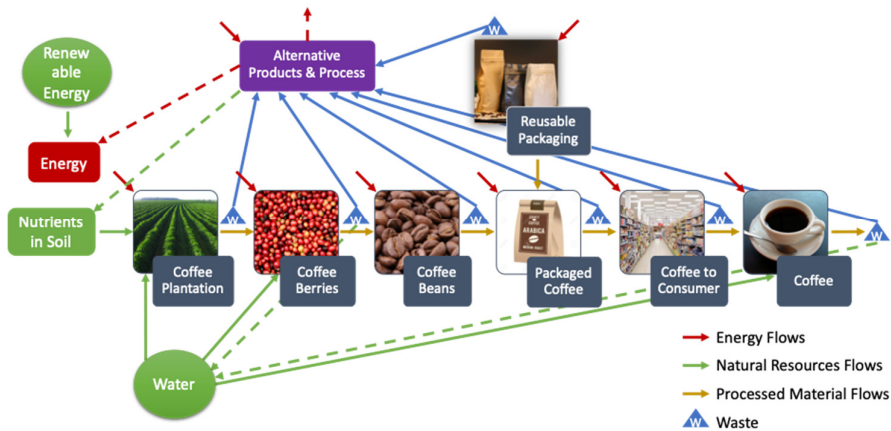


Figure 3: Simplified Circular Supply Chain of Coffee. Dashed lines refer to the energy and resource flows that return to the supply chain.

grounds for energy production (Murthy and Naidu, 2012; Gurram et al., 2016). This analysis, in conjunction with feedback from experimentalists, can help towards the identification of sustainable CE coffee supply chain alternatives. Then, a Resource-Task-Network (RTN) representation that incorporates all the alternative pathways is generated. Figure 4 illustrates such a RTN representation of the supply chain of coffee, capturing the sections in the production as well as in the consumption countries, the conversion coefficients, along with waste utilization alternatives for the production of energy. The next step is the formulation of a mixed-integer linear programming (MILP) model that reflects the entire supply chain, and includes the different objectives and constraints. Hence, binary variables are used for the selection or not of a process e.g. production of green beans through wet or dry method. Continuous variables are also defined for representing the amount of material used as input or output from each process as well as for the amount of material consumed or generated during each process. Moreover, the mass balances and the conversion equations are used to set the constraints.

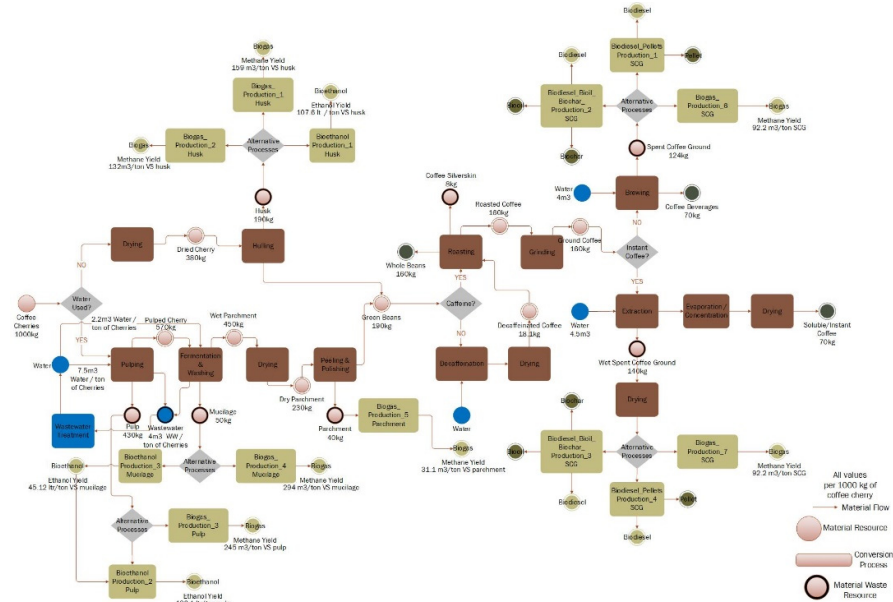


Figure 4: RTN of Coffee Supply Chain (Murthy and Naidu, 2012; Blinová et al., 2017; Ghosh and Venkatachalapathy, 2014; Braham and Bressani, 1979; Gathuo et al., 1991; ICO, 2011)

The final step involves a multi-objective optimization strategy to obtain trade-offs among multiple CE objectives, such as maximizing the energy output while minimizing the coffee cherries consumption, under different demand scenarios (Table 1) for the final products i.e. coffee beverages, instant coffee and whole beans. A variety of single or multi-objective functions can be optimized depending on the scenario under consideration. For the design of a CE supply chain, we select the objective functions by matching them with the CE goals and key characteristics. As such, by looking to increase the share of the renewable resources and energy, we will maximize the energy output of the supply chain, while by looking to reduce the input, we will minimize the coffee cherries consumption. Likewise, by targeting the reduction of natural resources, we will minimize the consumption of water, while the reduction of material losses and emission levels will be achieved through minimizing the waste generation and the CO₂ emissions respectively.

4. Discussion of Results

Utilizing the framework that was described in Section 3, we generate the pareto fronts (Figure 5) for the multi-objective optimization problem of maximizing the total energy output vs minimizing the consumption of coffee cherries, under five different demand scenarios (Table 1). The size of a bubble represents the normalized CO₂ emissions. Each colored line in the Figure 5 represents a different scenario. An upper bound of available cherries has been set at 100,000 tons for all 5 scenarios. Scenarios 1 and 2 require certain demands to be met for all 3 final products. As expected, the higher the consumption of cherries, the higher the energy produced. Once the target demands are

met, the excessive amount of cherries is converted to whole beans, while the excessive amount of coffee by-product is used to produce more energy. This is because even though whole beans can be sold as a final product to the market, they require further processing, either brewing or extraction, so as to yield the end coffee drink. Thus, producing just whole beans is favorable in terms of energy and environmental footprint.

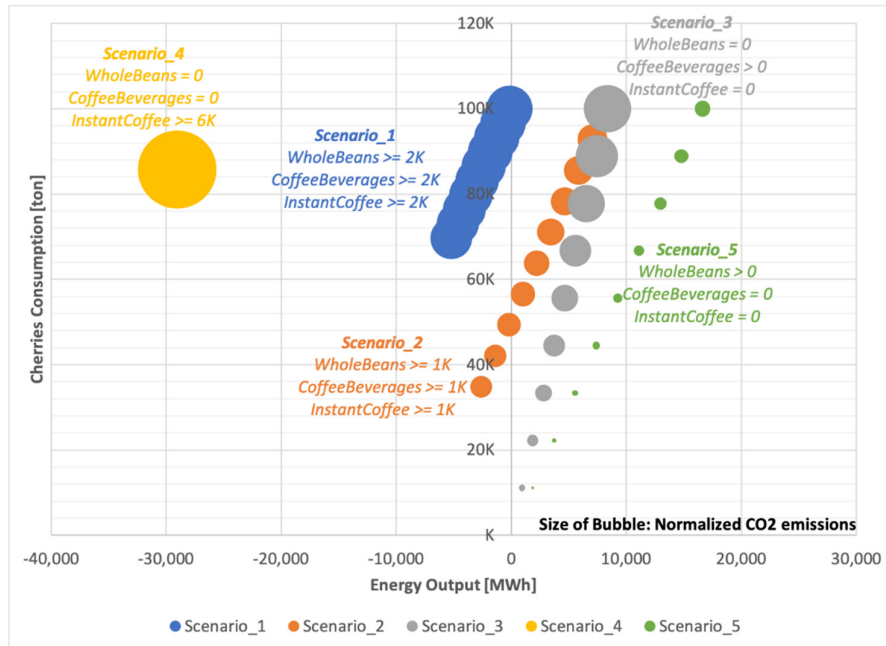


Figure 5: Pareto Analysis

Table 1: Final Coffee Products Demand Scenarios

	Units	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Whole Beans	ton	$\geq 2,000$	$\geq 1,000$	= 0	= 0	> 0
Coffee Beverages	ton	$\geq 2,000$	$\geq 1,000$	> 0	= 0	= 0
Instant Coffee	ton	$\geq 2,000$	$\geq 1,000$	= 0	$\geq 6,000$	= 0

The operating profiles of scenario 1 (with higher energy demands) are more to the left on the graph than those of scenario 2, because the extra demand of coffee beverages and instant coffee requires more brewing and extraction, which are both energy intensive processes. The rest of the scenarios represent instances where the demand of only one final product is specified. In particular, scenario 5, even rather unrealistic since it refers exclusively to demand of whole beans and requires further processing to produce the coffee drinks, it demonstrates the best-case scenario in terms of both energy and environmental performance. On the contrary, scenario 4, with only instant coffee as a deliverable, is the worst-case scenario both in terms of energy and environmental efficiency. Just one operating profile is produced since the optimal solution for both

objectives refers merely to the satisfaction of demand. Finally, in scenario 3 that describes the case of demand solely coming from coffee beverages, the operating profiles demonstrate a steeper slope in comparison to the ones from the other scenarios. This is attributed to the different energy and environmental benefits that coffee beverages have versus the other products.

5. Conclusions

The transition towards a circular economy model introduces a plethora of challenges. Here, we presented a CE systems engineering framework for the analysis and trade-off optimization of interconnected food supply chains. We demonstrated the effectiveness of the framework in analyzing different demand scenarios, considering some of the CE objectives simultaneously within a multi-objective optimization approach for a coffee supply chain case study.

6. Acknowledgments

We are grateful to Professor Leonardo Lombardini and his student Amanda Birnbaum for providing us with data for the coffee supply chain. This work is based upon work supported by the National Science Foundation under grant no. 1739977 [INFEWS]. Financial support from Texas A&M Energy Institute is also gratefully acknowledged.

References

- S. Avraamidou, S. G. Baratsas, Y. Tian, E. N. Pistikopoulos, 2020. Circular economy—a challenge and an opportunity for process systems engineering. *Computers & Chemical Engineering* 133, 106629.
- L. Blinová, M. Sirotiak, A. Bartošová, M. Soldán, 2017. Utilization of waste from coffee production. *Research Papers Faculty of Materials Science and Technology Slovak University of Technology* 25 (40), 91–101.
- J. E. Braham, R. Bressani, 1979. *Coffee pulp: composition, technology, and utilization*. IDRC, Ottawa, CA.
- B. Chala, H. Oechsner, S. Latif, J. Müller, 2018. Biogas potential of coffee processing waste in ethiopia. *Sustainability* 10 (8), 2678.
- B. Gathuo, P. Rantala, R. Määttä, 1991. Coffee industry wastes. *Water Science and Technology* 24 (1), 53–60.
- P. Ghosh, N. Venkatachalapathy, 2014. Processing and drying of coffee—a review. *Int. J. Eng. Res. Technol* 3 (12), 784–794.
- R. Gurrām, M. Al-Shannag, S. Knapp, T. Das, E. Singasaas, M. Alkasrawi, 2016. Technical possibilities of bioethanol production from coffee pulp: a renewable feedstock. *Clean Technologies and Environmental Policy* 18 (1), 269–278.
- S. Humbert, Y. Loerincik, V. Rossi, M. Margni, O. Jolliet, 2009. Life cycle assessment of spray dried soluble coffee and comparison with alternatives (drip filter and capsule espresso). *Journal of Cleaner Production* 17 (15), 1351–1358.
- ICO, May 2011. *Rules on statistics - statistical reports*. International Coffee Organization.
- ICO, 2020. *World coffee consumption 2019/2020*. International Coffee Organization.
- L. Lombardini, 2017. History of coffee and how it became a part of morning routines around the world. In: *2017 ASHS Annual Conference*. ASHS.
- E. MacArthur, 2015. *Circularity indicators: An approach to measuring circularity*. Methodology.
- P. S. Murthy, M. M. Naidu, 2012. Sustainable management of coffee industry by-products and value addition—a review. *Resources, Conservation and Recycling* 66, 45–58.
- A. Reichel, M. De Schoenmakere, J. Gillabel, J. Martin, Y. Hoogeveen, 2016.

Towards a Circular Economy Calculator for Measuring the “Circularity” of Companies

Stefanos G. Baratsas^a, Naseem Masoud^a, Valentini A. Pappa^a, Efstratios N. Pistikopoulos^a, Styliani Avraamidou^{a,*}

^a*Texas A&M University, 400 Bizzell St, College Station, TX 77843, United States*

**styliana@tamu.edu*

Abstract

Circular Economy (CE) aims to solve resource, waste, and emission challenges confronting society by creating a production-to-consumption total supply chain that is restorative, regenerative, and environmentally benign. A method for evaluating and comparing the “circularity” of different companies or scenarios is vital for effective decision making. A variety of metrics has been developed for measuring different aspects of CE. Despite the availability of metrics, CE has only been measured at the national/regional or product level with the main focus on material flows, while no metric is currently applicable at the company or product supply chain level. This work aims to address this research gap and provide i) a set of indicators and metrics with sector-specific dimensions, ii) a quantitative CE metric, iii) media for data visualization and analysis of CE indicators, and iv) an analytical tool to assess the multi-scale, multi-faceted and interconnected CE supply chains. Companies are able to track their transition towards CE, conduct temporal analysis, and compare their performance against their peers. The applicability and the capabilities of the developed CE methodology is demonstrated through a case study, where the Overall “Circularity” Index and the corresponding “Circularity” sub-indices of an industrial gas company are calculated over a period of 5 years.

Keywords: circular economy, circularity index, sustainability metric.

1. Introduction

The unprecedented economic development and the social advancement that occurred over the last centuries were inextricably linked to a “take-make-waste extractive” industrial model, which inevitably placed huge stresses on the natural resources and led to enormous environmental and socioeconomic impacts. The concept of Circular Economy (CE) has emerged as a potential solution to this challenging issue contributing to all dimensions towards sustainable development, promoting a transition to renewable energy sources, designing out waste and pollution, improving recycling while decoupling growth from the consumption of natural resources, and eventually regenerating natural systems (MacArthur et al., 2013). The effectiveness and the successful implementation of this transition on a global scale though, require systematic assessment of the alternative pathways and scenarios along with the development of holistic metrics to evaluate the different aspects of CE. A plethora of metrics have been proposed in the literature (Parchomenko et al., 2019; Roos Lindgreen et al., 2020; Elia et al., 2017; Cayzer et al., 2017; Saidani et al., 2017) to measure different aspects of CE; however, they capture separately the micro and macro level, without covering the meso level and without providing a holistic approach, whereas they rather focus on material flows (Avraamidou et al., 2020).

2. Circular Economy Metric Development

The proposed CE index takes into account a set of specific indicators and metrics based on the CE goals and key characteristics as well as the business sector that the company operates, and should not be confused with Life Cycle Assessment (LCA) which is an inherent attribute towards analysis, modeling, implementation and/or assessment of the CE (Avraamidou et al., 2020). Hence, five principal categories of indicators have been selected and matched with the goals and key characteristics of CE (Reichel et al., 2016) and are presented in Table 1.

Table 1: Matching of the Principal Categories with CE Goals

#	CE Goals & Key Characteristics		Principal Categories of Indicators
1	Reduction of material losses/residuals: Waste and pollutants minimization through the recovery and recycle of materials and products.	↔	Waste
2	Reduction of input and use of natural resources: The reduction of the stresses posed on natural resources through the efficient use of natural resources.	↔	Water, Procurement
3	Increase in the share of renewable resources and energy: Replacement of non-renewable resources with renewable ones, limiting the use of virgin materials.	↔	Energy
4	Reduction of emission levels: The reduction in direct and indirect emissions / pollutants.	↔	Emissions, Spillages
5	Increase the value durability of products: Extension of product lifetime through the redesign of products and high-quality recycling.	↔	Durability

An extra principal category, named “Organization” is also used for providing general information about a company’s business activity i.e. company’s revenue. As indicator, we define the information that must be measured and evaluated against a CE goal, e.g. total energy consumed within the company, while as metric we define the composite, normalized measure of an indicator against company’s level of business activity and productivity, e.g. total energy consumed within the company over company’s revenue. This is essential to capture the progress and efficacy of a company’s “circularity” year over year, as well as to conduct meaningful comparisons among companies (Herriott, 2016).

One or more indicators have been selected to evaluate the company’s performance against each CE goal or principal category. The indicators have been also matched with the Global Reporting Initiative (GRI) Standards (Global Sustainability Standards Board, 2016), ensuring uniformity in the reported results while providing a reference guide for those who want to use the proposed index. Similarly, one or more metrics have been chosen to standardize the indicators of each category. Different indicators and metrics are determined for each sector of the economy, in an attempt to reflect more accurately the specific characteristics and attributes of each sector. Tables 2 and 3 illustrate the indicators and metrics respectively, that are used for the analysis of the Industrial Sector for all principal categories.

The main sources of data are each company’s annual “Sustainability”, “Environmental-Social-Governance (ESG)” and “Financial” reports. Each company’s data from multiple years are collected and analyzed based on the proposed indicators and metrics, so as to calculate the annual Overall “Circularity” index and the annual “Circularity” index for each principal category (sub-indices) e.g. 2019 Overall “Circularity” Index, 2019

“Circularity” Index for Energy. Therefore, a company’s Overall “Circularity” as well as its “Circularity” versus every CE goal can be tracked on annual basis, and/or against its peers.

Table 2: CE Indicators for the Industrial Sector

Principal Categories	Indicators	GRI Standards Correspondence
Organization	Revenues [million \$]	GRI-201-1
	Products sold within the reporting period [# of products]	GRI-301-3
Waste	Waste generated - Hazardous [weight]	GRI-306-3
	Waste generated - Non-Hazardous [weight]	GRI-306-3
	Waste diverted from disposal [weight]	GRI-306-4
Water	Water withdrawal [volume]	GRI-303-3
	Fresh water discharge (<= 1,000mg/L TDS) [volume]	GRI-303-4
	Other water discharge (>= 1,000mg/L TDS) [volume]	GRI-303-4
	Water recycled or reused [volume]	GRI-303-3 (2016)
Procurement: Production & Packaging	Non-renewable materials used [volume or weight]	GRI-301-1
	Renewable materials used [volume or weight]	GRI-301-1
	Recycled Input material used [volume or weight]	GRI-301-2
	Packaging materials [# of products] reclaimed	GRI-301-3
Energy	Total energy consumed [joules or multiples]	GRI-302-1
	Renewable energy consumed [joules or multiples]	GRI-302-1
GHG Emissions	Direct GHG emissions (Scope 1) [tCO2e]	GRI-305-1
	Energy indirect GHG emissions (Scope 2) [tCO2e]	GRI-305-2
	Total use of products (Scope 3) [metric tons CO2 equivalent (tCO2e)]	GRI-305-3
	Emissions neutralized by carbon offset projects [tCO2e]	GRI-305-5
	Emissions of ozone-depleting substances (ODS) [metric tons of CFC-11 equivalent]	GRI-305-6
	Nitrogen oxides [NOx], sulfur oxides [SOx] & other significant air emissions [kg or multiples]	GRI-305-7
Spillages & Discharges	Volume of operational spills [volume]	GRI-306-3 (2016)
	No. of operational spills [#]	GRI-306-3 (2016)
	Environmental fines [\$]	GRI-307-1
	Notices of violation (NOVs) from spills [#]	-
Durability	Average lifespan of product or packaging [years]	GRI-306-2

Table 3: CE Metrics for the Industrial Sector

Principal Categories	Metric	Upper Bound	Formula Used
Waste	1a % of Hazardous waste over Total waste generated	100%	100%-1a
	1b % of Diverted waste over Total waste generated	100%	1b
Water	2a % of Recycled/reused water over Total water withdrawal	100%	2a
	2b % of Other water discharge over Total water discharge	100%	100%-2b
	2c % of Water consumed over Total water withdrawal	100%	100%-2c
Procurement: Production & Packaging	2d % of Recycled input materials used over Total material used	100%	2d
	2e % of Renewable material used over Total material used	100%	2e
	2f % of Reclaimed packaging materials over Total products sold	100%	2f
Energy	3a % of Renewable energy consumed over Total energy consumed	100%	3a
GHG Emissions	4a Net total emissions over Revenue [tCO2e over million \$]	10,000	1-4a
	4b Emissions of ODS over Revenue [metric tons of CFC-11 eq. over million \$]	0.05	1-4b
	4c NOx, SOx, and other significant air emissions over Revenue [metric tons over million \$]	0.5	1-4c
Spillages & Discharges	4d Notices of Violation (NOVs) from Spills [#]	50	1-4d
Durability	5a Average Lifespan of Product or Packaging [years]	20	5a

The target values of the Overall “Circularity” Index and its sub-indices are always 1 or 100%; thereby the formulas used for each metric have been designed to reflect to that target and are shown in Table 3. For example, metric 1a that captures the percentage of hazardous waste over total waste generated shall be preferably zero, but since our target is 1 then the formula $100\% - 1a$ is used.

The sub-index of each principal category is estimated as the weighted average of its metrics, using equal weights except from the GHG Emissions where the following weights are used: *4a* - 50%, *4b* - 20%, *4c* - 20%, *4d* - 10%. The overall index is calculated as the average of the sub-indices.

3. Case Study: “Circularity” Index of an Industrial Gas Company

The Overall “Circularity” Index along with the sub-indices of an industrial gas company, hereafter referred to as the “firm”, is calculated here. The data used for the subject calculations and analysis are taken from firms’ “Sustainability” reports (2016-2020) covering the period 2015-2019 (Air Products, 2016-2020). First, the annual metrics of each indicator are calculated using the methodology described in the previous section, and are summarized in Table 4.

Table 4: CE Metrics of the firm (2015-2019)

Principal Categories	Waste		Water			Procurement: Production & Packaging			Energy	GHG Emissions			Spillages & Discharges	Durability
	1a	1b	2a	2b	2c	2d	2e	2f	3a	4a	4b	4c	4d	5a
2015	0.604	0.228	0.0	1.0	0.0	0.0	0.85	0.0	0.017	0.678	1.0	0.560	0.84	1.0
2016	0.625	0.265	0.0	1.0	0.178	0.0	0.85	0.0	0.018	0.674	1.0	0.551	0.80	1.0
2017	0.653	0.282	0.0	1.0	0.179	0.0	0.865	0.0	0.069	0.682	1.0	0.578	0.72	1.0
2018	0.538	0.303	0.0	1.0	0.147	0.0	0.85	0.0	0.074	0.700	1.0	0.654	0.66	1.0
2019	0.649	0.201	0.0	1.0	0.186	0.0	0.85	0.0	0.071	0.695	1.0	0.649	0.74	1.0

The firm scores the maximum (100%) with regards to the durability metric (5a) throughout the years, since they supply most of their products in two-way bulk containers, semi-bulk containers or via pipelines with very long-life span. Even for small-scale supplies, they use returnable and reusable transportable pressure vessels with typical life spans of 20+ years. The energy metric (3a) demonstrated the highest improvement over this period (307.6%), but the renewable energy is still a tiny part of firm’s energy portfolio. A noticeable improvement of 15.9% was also shown in NOx/SOx emissions over Revenue (4c) with both parameters being improved over the years. On the contrary, the waste diverted as a percentage of the total waste generated (1b) and the number of violations from spills (4d) have deteriorated over the same period.

4. Results

The following figures illustrate firm’s annual “Circularity” Sub-Indices (Figure 1) and annual Overall “Circularity” Index (Figure 2) from 2015 to 2019, using the weights described in the previous sections. As it is shown, Durability is the best performing sub-index, having reached and maintained the target over the years, while in contrast, and despite the improvement over the years, Energy sub-index is still at the lower end of the scale. Thus, an increase in firm’s renewable energy footprint will boost the firm’s “circularity”. GHG Emissions & Spillages sub-index also improved slightly over the years, having reached 75.1% of the target. Waste and Water & Procurement sub-indices demonstrate a mixed picture, fluctuating year over year, but both have improved, even slightly, over this period. Consequently, firm’s “circularity” was at the highest level in

2017, after two consecutive years of improvement. Then, the index was barely reduced by 1.2% in 2018, mainly due to a 10.1% decrease in the Waste sub-index, before a marginal increase of 0.5% in 2019.

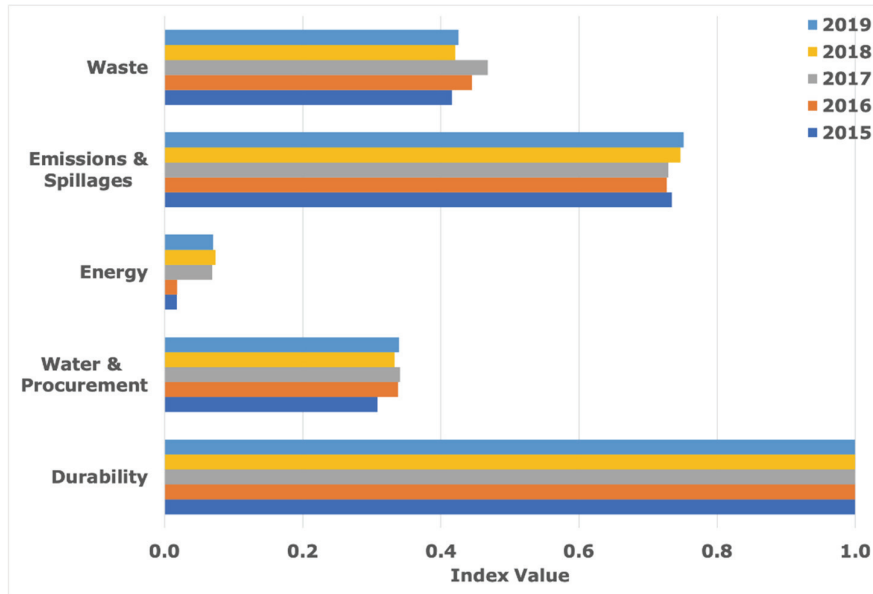


Figure 1: Firm's "Circularity" Sub-Indices (2015-2019)

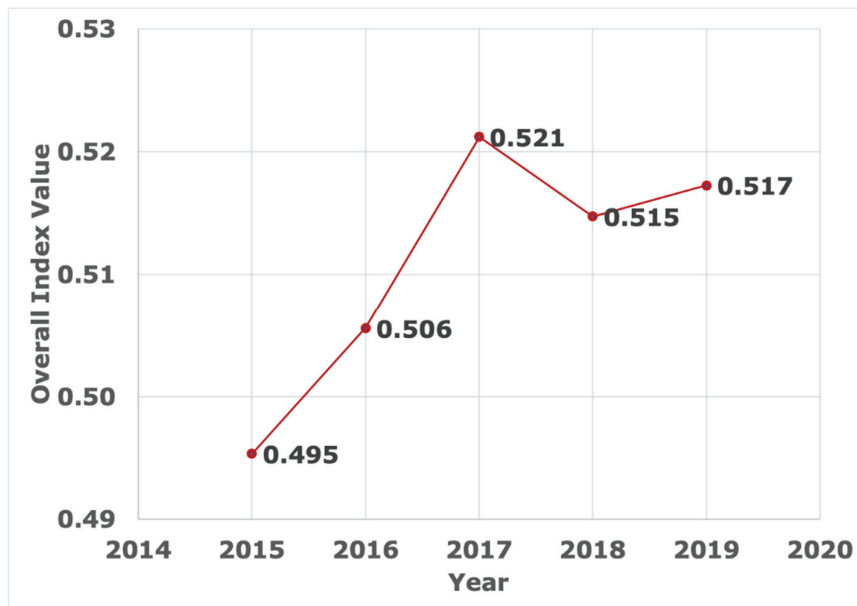


Figure 2: Firm's Overall "Circularity" Index (2015-2019)

5. Conclusions

A quantitative CE index for assessing a company's "circularity" has been presented, in an effort to accurately measure the various aspects of CE and identify potential areas of improvement towards the transition to a CE economic model. It is noted that various sector-specific indicators and metrics have been established in order to effectively and holistically capture all CE aspects. This enables year-over-year comparisons, benchmarking against peers in the industry, and identification of key areas that require improvement. Also, the aforementioned tool is an integral part of a CE systems engineering framework to quantitatively evaluate different multi-scale supply chain pathways, assess CE tradeoffs, and aid in informed optimal decision making. The capabilities and applicability of the subject methodology was demonstrated through a case study for a firm operating in the industrial sector.

The proposed methodology can act as an organization's internal CE and Sustainability assessment tool and will be incorporated into a web-based index calculator that is accessible online. As soon as the required data are available, the user will be able within 5-minutes to i) measure firm's "circularity", ii) track firm's periodic progress, iii) benchmark against firm's peers, and iv) visualize the analysis online.

6. Acknowledgements

Financial support from Texas A&M University and Texas A&M Energy Institute is gratefully acknowledged.

References

- Air Products, 2016-2020. Sustainability reports 2016-2020. <https://www.airproducts.com/company/sustainability> (Accessed: 11.12.2020).
- S. Avraamidou, S. G. Baratsas, Y. Tian, E. N. Pistikopoulos, 2020. Circular economy-a challenge and an opportunity for process systems engineering. *Computers & Chemical Engineering* 133, 106629.
- S. Cayzer, P. Griffiths, V. Beghetto, 2017. Design of indicators for measuring product performance in the circular economy. *International Journal of Sustainable Engineering* 10 (4-5), 289–298.
- V. Elia, M. G. Gnoni, F. Tornese, 2017. Measuring circular economy strategies through index methods: A critical analysis. *Journal of Cleaner Production* 142, 2741–2751.
- Global Sustainability Standards Board, 2016. GRI Standards. <https://www.globalreporting.org/> (Accessed: 11.12.2020).
- S. R. Herriott, 2016. Metrics for sustainable business: Measures and standards for the assessment of organizations. Routledge.
- E. MacArthur, et al., 2013. Towards the circular economy. *Journal of Industrial Ecology* 2, 23–44.
- Parchomenko, D. Nelen, J. Gillabel, H. Rechberger, 2019. Measuring the circular economy-a multiple correspondence analysis of 63 metrics. *Journal of Cleaner Production* 210, 200–216.
- Reichel, M. De Schoenmakere, J. Gillabel, J. Martin, Y. Hoogeveen, 2016. Circular economy in europe: Developing the knowledge base. *Eur. Environ. Agency Rep* 2, 2016.
- E. Roos Lindgreen, R. Salomone, T. Reyes, 2020. A critical review of academic approaches, methods and tools to assess circular economy at the micro level. *Sustainability* 12 (12), 4973.
- M. Saidani, B. Yannou, Y. Leroy, F. Cluzel, 2017. How to assess product performance in the circular economy? proposed requirements for the design of a circularity measurement framework. *Recycling* 2 (1), 6.

Economic Optimization of Qatar's Hydrocarbon-based Fuels for Sustainable Maritime Applications

Ahad Al-Enazi ^a, Eric C. Okonkwo ^b, Yusuf Bicer ^b, Tareq Al-Ansari ^{a,b*}

^a *Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*

^b *Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*

talansari@hbku.edu.qa

Abstract

Environmental regulations have always been an essential component of the natural gas supply chain. This study explores the challenges and opportunities associated with replacing heavy fuel oil used for maritime transportation with relatively cleaner and green fuels. The economic and environmental viability of alternative bunker fuels mainly liquefied natural gas, Ammonia, and hydrogen, is assessed. Additionally, the entire supply chain of natural gas is holistically evaluated by understanding the end-user requirements and exploring the feasibility of further processing the transported fuel (i.e., LNG) into the required form. The examination of supply-demand synergies constitutes the novelty of this study. With the aid of linear programming, the fuel-alternatives mentioned above are assessed on environmental and economic levels. A multi-objective optimisation model is developed considering the emissions constraints imposed on ship operators by regulatory authorities. The mathematical model also includes the possible transported fuel (i.e., LNG, Ammonia, and hydrogen) where both supply and end-user demand are factored into the formulation with the aim of achieving optimal value in an environmentally benign manner. The outcome of this study demonstrates the optimal combinations of transported and consumed fuels. The magnitude of each bunker fuel to be consumed indicates the best fuel to be used while considering both economic and environmental aspects. According to the result obtained from the developed model, Ammonia seems to be the most favourable, followed by HFO, LNG, and hydrogen, respectively. This is mainly attributed to the high cost associated with the consumption of hydrogen fuel. As for the transported fuel, the model suggests that LNG should be the main transported fuel followed by Hydrogen and Ammonia, respectively.

Keywords: LNG Supply Chain; Maritime Transportation; Clean Fuels; Bunker Fuels; Optimisation.

1. Introduction

In many businesses, a tradeoff is usually imposed between economic gain and environmental growth. Usually, a decision-maker will be in a situation that requires them to search for solutions/alternatives to reduce the environmental impact of their business yet still allow them to attain more monetary gains, thereby achieving sustainable development for their business. (Kumar, 2020) The shipping of hydrocarbons constitutes approximately 30% of the world's growing seaborne trade. (United Nations Conference on Trade and Development, 2018) According to LCA studies, the primary source of the

environmental impact during the life cycle of a ship is the operating stage. (Kameyama, Hiraoka, Tauchi) Moreover, traded natural gas is expected to increase from the recorded 25 % in 2017 to 28% of the global gas mix in 2022. (McKinsey. Global, 2018) Hence, efforts to reduce the environmental impact of oil and gas supply chains involve various measures, such as reduced flaring, process optimization, and the utilization of clean fuels, which have been extensively reviewed by many scholars in the fuel utilizing industries such as the aviation sector for over two decades. Today, policymakers and energy analysts worldwide believe that hydrogen has the potential to transform the world's energy outlook. Nevertheless, the complete transition to hydrogen as a replacement for fossil fuels is hindered by technical and economic challenges. (United Nations Environment Program, 2006) Meanwhile, Ammonia has been considered a potential alternative energy storage medium. It is a high-density carbon-free hydrogen carrier, providing a practical and clean alternative to fossil fuels. (Valera-Medina, Xiao, Owen-Jones, David, Bowen, 2018) A broad spectrum of applications in the industry entails the use of operation research to model and solve combinatorial optimization problems. Such applications include planning, scheduling, vehicle routing, and resource allocation. Moreover, to design an effective supply chain network, supply chain managers usually resort to using optimization modeling as it leads to determining the optimal mix of suppliers, production quantity/volume, and location of several facilities within the supply chain. Operation research practitioners normally use a modeling language to formulate and manage their developed mathematical models. Optimization Planning Language (OPL) is a computer-based modeling language concerned with solving optimization problems as it provides algorithms for mathematical integer programming, linear programming, and quadratic programming. (Hillier, Lieberman, 2001) The oil and gas industry have been utilizing the field of operation research as a scientific approach to help support refinery operations management. (Antunes, Gomes, 2008) In this particular study an optimization model was developed to look into the logistical aspect of the LNG supply chain. The objective function of this model is composed of two factors: (a). Economic: encompassing cost associated with bunker fuel in addition to transportation cost and revenue generated from selling the transported fuel. (b). Environmental: encompassing emissions generated by bunker fuel and emissions generated throughout the life cycle of transported fuel. This is to be governed by several restrictions such as the production capacity of the exporter and the emissions limits imposed by regulatory authorities (namely for CO₂, NO_x, and SO_x).

2. Problem Definition

Strategic level decisions are made when a natural resource is discovered and/or when producing a particular product is not as profitable as it used to be due to supply and demand interactions. Hence, exploring the possibility of making any particular modification to the pre-constructed processing facility becomes a necessary investment that enables the producer to maintain their position in the energy market. Also, from an operational perspective, ship operators are obligated to limit the emissions generated from their fleet in order to comply with emissions limits set by environmental regulatory authorities. This study is meant to address both of the aforementioned aspects while using the state of Qatar as a case study. Basically, the main questions to be answered by this study are: 1. Which bunker fuel is more economical and has the least environmental impact? 2. Is it more economically and environmentally viable for the state of Qatar to produce and ship another type of fuel instead of LNG? 3. Is there a possibility to create synergy between the bunker fuel consumed and the produced/transported fuel? Basically, the value of answering the first question lays in the fact that it will aid ship operators in

the fuel selection process while considering environmental and economic aspects. Moreover, on a strategic level, this model can be of use to decision-makers as it helps in evaluating current and future options for their fuel export. The model serves to assess the revenue gained from each Fuel and determine the strategy to be followed in terms of fuel production across a short and long timespan. The second question can be addressed by exploring the possibility of further processing natural gas into other forms of fuels required by customers. Lastly, the third question will examine the possibility of using the transported fuel as a bunker fuel, which will eliminate the need for ship bunkering and will significantly reduce operational costs.

The remaining sections of this paper illustrate the depiction of the problem in a mathematical language, which is then translated into a computer-based modelling language in order to facilitate the process of solving the model.

3. Mathematical Modelling

3.1. Governing Assumptions:

Due to the complex nature of the problem at hand, a set of governing assumptions had to be put in place in order to facilitate its representation by means of a mathematical language. These assumptions are listed below:

Table.1: A list of the pre-defined governing assumptions.

Governing Assumptions	
1.	Supplying LNG from the state of Qatar (i.e.one LNG loading port).
2.	One LNG-discharge port (Receiving terminal).
3.	Empty (return) voyages will not be considered.
4.	No partial unloading of ships. (i.e., a single loading port and a single discharge port per cargo).
5.	It is possible to have recurrent voyages (i.e., several cargos can be delivered to the same customer/ receiving terminal).
6.	All Ships are assumed to have the same capacity of 140,000 M3.
7.	All cargos are delivered using the operator fleet and sold at a fixed rate following a delivered ex-ship cargo transport system. A vessel is assumed to be available at the time of loading. (in this case, we need to assume an infinite fleet size).
8.	Both discharge and receiving ports are available at the time of loading/ unloading as unavailability of port can lead to an increase in total transportation costs as some delay penalties may apply. Cost of idle ships and cost of delays and demurrage costs are neglected.
9.	Fuel consumed for auxiliary system is constant and is neglected.
10.	All ship engines are assumed to be compatible with all of the suggested bunker fuels (i.e.no additional engine retrofitting measures are considered).
11.	Cost of receiving terminal regasification slot and canal charges are disregarded.
12.	All Ships are assumed to travel at a constant speed throughout the entire voyage.

3.2. Input Parameters

Table.2: Produced and Consumed Fuel related data.

Fuel i	Selling Price P_i	Consumed Fuel Cost $C_i^{Consumed}$	Produced Fuel Cost $C_i^{Produced}$	Produced Fuel Demand D_i	Fuel Production Capacity Cap_i	Specific Fuel Consumption $Consumption_i^{Specific Fuel}$	Fuel Density $Density_i$
Unit	\$/kg	\$/kg	\$/kg	Million kg/year	Million kg/year	Kg/km	kg/m ³
1.NH ₃	0.275	0.25	0.225	1980000	1275	200.941	696
2.LNG	0.385	0.35	0.315	700000000	9000	43.855	455

3.H ₂	0.88	0.80	0.72	28400000	450	3.177	71
4.HFO	0.308	0.28	0.252	0	0	124.611	905

Table.3: CO₂, NO_x, and SO_x emissions related data.

Emission Type (j)	GHG Emissions Limit (Per Voyage)			Fuel GHG Emissions Factor $Emissions Factor_i^j$			
	1.NOx	2.SOx	3.CO ₂	1.NOx	2.SOx	3.CO ₂	
Fuel (i)	Unit	Ton	Ton	Ton	Kg/Km	Kg/Km	Kg/Km
1.NH ₃	637	637	637	0.657277006	0	0	
2.LNG	99.4	99.4	99.4	0.061449948	0	135.0188511	
3.H ₂	974.4	974.4	974.4	0.011174968	0	0	
4.HFO	1267	1267	1267	4.1811	1.15×10 ⁻¹¹	5.1×10 ⁻¹⁰	

$Distance_{Qatar\ to\ k}$: Distance between exporting port located in Qatar and port k in km (in this case, only one case is considered Himeiji Terminal located in Japan where the assumed distance is 11693.528 km).

$Speed_{Ship}$: The assumed speed is 27.87 km/hr.

Cap_{Ship} : Capacity of the ship used for transportation in m³ is assumed to be 140,000 m³.

$Power_{Engine}$: Main engine power of the ship is assumed to be 15,006 KW.

3.3. Scenario Development:

The adopted approach of developing this model entails beginning with a very simple model then gradually increasing the level of complexity of the model until it nearly reflects the complexity of the real problem. (Hillier, Lieberman, 2001). Only the first scenario will be portrayed in this paper. This scenario is aimed at maximizing profit were bunker fuel type is independent of transported fuel type excluding receiving terminal regasification slot cost and canal charges.

Decision Variables:

N_i : Number of shipments delivered from fuel (i)

F_i : Amount of fuel (i) consumed per voyage as bunker fuel in kg

Parameters:

P_i : Selling price of fuel (i) in \$/kg

$C_i^{Consumed}$: Bunker fuel cost for fuel (i) in \$/kg

$C_i^{Produced}$: Transported fuel cost for fuel (i) in \$/kg

D_i : Demand for fuel (i) in kg

$Limit_i^j$: Limit of GHG emissions of type (j) for fuel (i) in kg

$Emissions\ Factor_i^j$: Emissions factor for GHG emissions of type (j) for fuel (i)

Cap_i : Production capacity of fuel (i) in m³

$Density_i$: Density of fuel (i) in kg/m³

$Distance_{Qatar\ to\ k}$: Distance between exporting port located in Qatar and port (k) in KM

$Speed_{Ship}$: Ship speed in km/hr

Cap_{Ship} : Capacity of the ship used for transportation in Kg

$Power_{Engine}$: Main engine power of the ship in kW

$Consumption_i^{Specific\ Fuel}$: Specific fuel consumption for Fuel (i)

where,

$i = 1,2,3,4$ refers to Ammonia, LNG, Hydrogen, HFO respectively.

$j = 1,2,3$ refers to NO, SOx, CO₂ respectively.

$k = 1$ refers to Himeiji Terminal located in Japan.

Objective:

$$\text{Max } \sum_{i=1}^4 (P_i \times \text{Cap}_{\text{Ship}} \times N_i \times \text{Density}_i - C_i^{\text{Produced}} \times \text{Cap}_{\text{Ship}} \times N_i \times \text{Density}_i - C_i^{\text{Consumed}} \times F_i) \quad \forall i \in i = 1, \dots, 4 \quad \text{eq.(1)}$$

Subject to:

$$\text{Cap}_i \geq \text{Cap}_{\text{Ship}} \times N_i \times \text{Density}_i \quad \forall i \in i = 1, \dots, 4 \quad \text{eq.(2)}$$

$$\text{Cap}_i \times 0.8 \leq \text{Cap}_{\text{Ship}} \times N_i \times \text{Density}_i \quad \forall i \in i = 1, \dots, 4 \quad \text{eq.(3)}$$

$$D_i \geq \text{Cap}_{\text{Ship}} \times N_i \times \text{Density}_i \quad \forall i \in i = 1, \dots, 4 \quad \text{eq.(4)}$$

$$N_i \times \text{Emissions Factor}_i^j \times \text{Distance}_{\text{Qatar to } k} \leq \text{Limit}_i^j \quad \forall i, j, k \in i = 1, \dots, 4, j = 1, 2, 3, k = 1 \quad \text{eq.(5)}$$

$$F_i \geq \text{Consumption}_i^{\text{Specific Fuel}} \times \text{Distance}_{\text{Qatar to } k} / \text{Speed}_{\text{Ship}} \quad \forall i, k \in i = 1, \dots, 4, k = 1 \quad \text{eq.(6)}$$

$$N_i, F_i, P_i, C_i^{\text{Consumed}}, C_i^{\text{Produced}}, \text{Limit}_i^j, \text{Emissions Factor}_i^j, \text{Cap}_i, \text{Distance}_{\text{Qatar to } k}, \text{Speed}_{\text{Ship}}, \text{Cap}_{\text{Ship}}, C_i^{\text{Transportation}} \geq 0 \quad \forall i, j, k \in i = 1, \dots, 4, j = 1, 2, 3, k = 1 \quad \text{eq.(7)}$$

4. Results

Based on the given input data, the model suggests that the optimal combination of Fuel resulting in the best revenue to be attained is as follows:

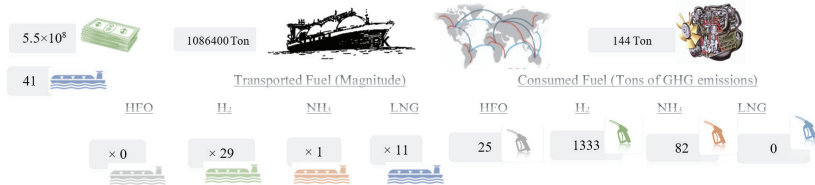


Figure.1: Summary of the obtained results.

5. Conclusions

This study presents Mixed Integer Linear Programming (MILP) as a means to provide an objective assessment of multiple potential maritime fuels. Based on the assumed input data, the results imply that Ammonia is the most favourable bunker fuel given the fact that it is an environmentally friendly fuel which is less costly than hydrogen. Moreover, the results also show that the most favourable Fuel to be transported is Hydrogen, given its relatively high selling price. However, any turbulence in the energy market might lead to a change in this result. Moreover, LNG was deemed as the least favourable bunker fuel due to its high CO₂ emissions. The use of CO₂ abatement technology while consuming LNG as bunker fuel can lead to a significantly different result. This can be further studied by conducting a thorough sensitivity analysis. Results and conclusions attained from this model can be of value to decision-makers when good demand and price forecasts are used.

as an input to this model. However, it should be noted that the quality of the forecasts used will dictate the reliability of the results obtained from this model. Future work involves extending the model to include a set of additional scenarios, which will, in turn, develop the complexity of the model.

6. Acknowledgment

This publication was made possible by GSRA grant, ID# GSRA6-2-0615-19090, from the Qatar National Research Fund (a member of Qatar Foundation). The authors also acknowledge the support provided by Hamad Bin Khalifa University, Qatar Foundation, Qatar.

7. References

- A. Gomes, C. Antunes, 2008, Operational research models and methods in the energy sector, *European Journal of Operational Research*, Editorial, 1-2.
- A. Valera-Medina, H. Xiao, M.Owen-Jones, W.I.F. David, P.J. Bowen, 2018, Ammonia for power, *Progress in Energy and Combustion Science*, 69, 63-102.
- F. Hillier, G. Lieberman, 2001, *Introduction to Operations Research*, 1-31.
- McKinsey, 2018, *Global Gas & LNG Outlook to 2035*, *Energy Insights 2018*, September 2018, 1-31.
- M.Kumar, 2020, *Application of Operational Research in Sustainable Environmental Management and Climate Change*, ZBW – Leibniz Information Centre for Economics, 1-16
- M. Kameyama, K. Hiraoka, H.Tauchi, 2007, *Study on Life Cycle Impact Assessment for Ships*, 1-11.
- United Nations Conference on Trade and Development, 2018, *Review of Maritime Transport*, 1-116.
- United Nations Environment Program (UNEP), 2006, *The Hydrogen Economy*.

Unmanned aerial vehicles in precision agriculture towards circular economy: a process system engineering (PSE) assessment

Mohammed Yaqot,^a Brenno C. Menezes^{a,*} Tareq Al-Ansari,^{a,b}

^a*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar Foundation, Qatar*

^b*Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar Foundation, Qatar*

bmenezes@hbku.edu.qa

Abstract

Farming is a complex and responsive business that considers types of soils and crops, topography, climate, etc., while minimising inputs like land and water for an optimal yield. Typically, a farm to be managed involves planting, irrigation, spraying of nutrients and pesticides, manual inspection when looking for signals of harvesting stress or pest infestation, etc. When these activities are unreliable, yields of crops fluctuate so do the enterprise margins. However, to reach higher levels of performance, today's agribusiness moves towards more sustainable and efficient operations within the so-called precision agriculture. From such a scenario, a process system engineering assessment is proposed to evaluate the impacts of industry 4.0 (I4) applications under circular economy (CE) concepts in the agricultural field. Particularly, we highlight drones or unmanned aerial vehicles as a tool in the agroindustry since numerous researchers and industries are designing and testing different principles and technologies that could be applied by replacing manpower with drones in engineering fields. A sensitivity risk analysis compares the I4 adoption and its CE impacts against an outdated agricultural-based production structure for corn production. It has been evaluated using subjective ratings for twenty sub-process components such as irrigation, monitoring, among others. The findings indicate that investments in agricultural drones' capabilities will modify sub-processes under CE principles, resulting in a process system re-engineering in this field. Agricultural drones can increase the economy of the processes coupled with environmentally friendly applications, whereby the impact on the social pillar of the CE is still debatable.

Keywords: Precision agriculture, circular economy, industry 4.0, society 5.0.

1. Introduction

Modern society has gone through three previous industrial revolutions, when workloads were still highly labour-intensive in comparison with the ongoing high-tech societies in the advent of the 4th industrial revolution, so-called industry 4.0. Industries 1.0 and 2.0 have brought human activities from focusing on hunting and agriculture (society 1.0) to industrial societies, when the combined inventions of mechanical movement (industry 1.0

and society 2.0) and electricity (industry 2.0 and society 3.0) replaced manpower from artesian/craft work to a mass production. Less than a century, industry 3.0 upgraded human activities to an extensive information age (society 4.0) of an automated and programmable production. Nowadays, industry 4.0 (I4) is reshaping communities into the so-called society 5.0 (S5) and enabling the creation of a new industrial society (within the augmented information age) where artificial intelligence (AI) will sense, calculate, and actuate to optimally support automated decision-making (Menezes et al., 2019). The I4-S5 combination within the smartness of everything is represented by the increasingly connected societies that are supported by automation of things (AoT) using high-performance computing and communication technologies in a cyber-physical system engine, whereby augmented information and pervasive communication technologies facilitate innovative industrial solutions that will completely reshape production systems.

However, these transformational industrial societies are facing severe social, economic, and environmental challenges to adhere to the concepts of the so-called circular economy (CE) within today's highly constrained sustainable process mandate. Moreover, the pressure in increasing levels of automation has been pushing the termination of a huge number of jobs. Therefore, such industrial societies are still not prepared for the onset of CE's disruptions on well-established processes of work and ways of well-being since there are unbalances between I4 technologies and CE strategies to be solved.

Nevertheless, despite the equilibrium of the I4-S5 to be determined for the next generation of industries and societies, the needs of improvements in the agricultural production go beyond economy. World population growth is expected to reach 9.7 billion in 2050 (UN, 2019) challenges the search for increased production of crops to meet food demands without significant impact on the environment and society. In such direction, as a case of I4 in agriculture, we evaluate the impacts in the CE by using unmanned aerial vehicles (UAV) or *agricultural drones*. These apparatuses have been providing sustainable values in increasing crop production, monitoring plant health, and collecting vital insights of soil and vegetation, since the speed and accuracy these data are gathered using drones allow tangible improvements in farming management (Al-Ansari and Chawla, 2020). From such scenario, we propose a process system engineering assessment in the corn value chain considering a quantitative operational-risk evaluation over the application of drones under CE's principles.

2. Unmanned aerial vehicle impacts in precision agriculture

Today's move in academia and industry towards researching, developing, and deploying novel apparatuses and operations within so-called precision agriculture is modifying farming businesses (Menezes et al., 2020). In this field, UAV or agriculture drones are one of the best I4 solutions to support the next generation of agriculture as it revolutionises operations and reduces its environmental impacts. It becomes enormously popular over the last five years due to its inexpensive rapidly improving technology in terms of battery, motor, and scientific know-how stabilisation. Several UAV models have been developed over the past years varying in purpose, size, weight, and design. Each design has its own benefits and drawbacks based on target application usage. UAV allows the sensing of the physical world by capturing aerial data to generate accurate maps and 3D models of their surroundings. By analysing these maps and 3D models, different industrial applications are enabling faster decisions that increase efficiency, improve safety, and reduce operational costs (Zhai et al., 2020). With such benefits, the UAV

usage rises across industries including engineering, mining, environment, agriculture, surveying, humanitarian, etc.

The modern technologies in agricultural fields facilitate and generate a significant amount of data that can be processed remotely enabling a variety of advanced analytics and solving real world problems for an efficient agribusiness. Collected data is stored and mined using supercomputers providing the foundation for developing data-driven applications and solutions utilising artificial intelligence (AI). The supercomputers store data from multiple sources such as sensors on the ground for soil and plant analysis, mobile sensors or images attached on drones, and satellite imaging. The streams of data are the basis for extracting key operational information about soil condition, crop nutrient uptake, crop yield forecasting, estimation of optimum harvesting times, gas emissions for monitoring, and soil contamination.

Many inputs and outputs to be considered for a precision agriculture in farming depend on the widespread application of I4. Traditional methods for planting, irrigation, inspection, spreading of pesticides, etc., can potentially waste time and resources as well as affect the environment negatively by misusing of soil, water, etc. These conventional farming activities are going to be replaced by autonomous engines to uniformly apply water, nutrients, and pesticides, to precisely plant the seeds within a perfect distance, and so forth. In terms of inspections, besides satellite imaging that became a revolutionary tool several years ago in the agriculture field, UAV offers an effective way to survey small to large scale operations, inspecting fields for signs of stress and pest infections as well as assessing crops management. Then, data analytics can be used to pinpoint areas of crop stress to determine when, where, and how much water, fertiliser, and pesticide is needed to sustainably produce healthy crops with upmost efficiency facilitated by I4 capabilities.

Considering the sustainable evolution of the industry and society, the interplaying of I4 technologies and CE fundamentals is a new frontline faced by engineering communities for their expansion. Although the underlying principles of CE have been included as an additional constraint in academia and industry since the last few decades, today's theories and applications of CE are systematically encapsulated in a wider and a detailed approach. In the broader, the relationships among economic, environmental, and social pillars are being considered from now to design, operate, and control manufacturing systems and supply chains. Furthermore, CE's details on re-do (or do-not) activities, behaviours, beliefs, etc., are transforming traditional industry and society towards their next generation. Hence, both CE concepts (pillars and R's) account for the expansion of the sustainable industry interrelating environmental and social concerns that involve I4 technologies within the society 5.0 (S5). Ritz et al. (2019) demonstrated that training in agricultural technologies in the AgTech program in France increased the profit in agricultural businesses among I4-trained farmers and defend that this should become a new prerequisite for smart farming. In such a context, we cover a case of I4 and CE interconnections in the agriculture industry that accesses operations and risks of I4 adoption and impacts in the CE in a PSE assessment.

3. Agricultural production operations and risks towards circular economy

We consider in the I4-CE operational-risk assessment the upstream or supply side of the system boundary in the agribusiness value chain seen in Figure 1. Although the business

model describes the full range of activities comprising all steps that involve agricultural processes and products from harvesting to retailing, covering inbound logistics, operations, outbound logistics, marketing, sales, and service, we focus on the production step only where I4 applications under CE principles can be assessed and evaluated.

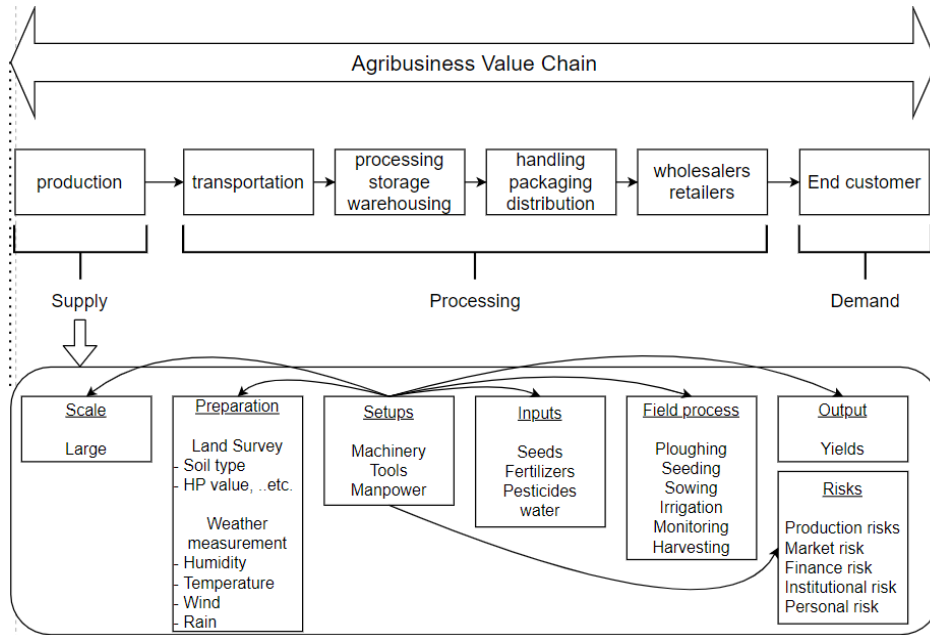


Figure 1: Agribusiness value chain.

According to Manitoba agriculture and resource development (Manitobia, 2020), the total cost for production of crops takes into account many setups, stocks and flows related to operations, fixed costs, and labour expenses. Operating costs are composed by seeds, treatment, fertiliser, herbicide, fungicide, insecticide, fuel, machinery operating, machinery lease, rental, custom, crop insurance, hail insurance, land taxes, drying costs, interest rate, and other minor costs. Fixed costs include land costs, machinery costs, storage costs. Labour cost is the hourly wage of the total number of needed labours per acre.

To evaluate the impact of I4 and CE in manpower towards precision agriculture, the type of I4 technology used in each production step is investigated and categorised into unmanned aerial vehicles (UAV), manned aerial vehicles (MAV), unmanned ground vehicles (UGV), and manned ground vehicles (MGV). In the methodology, every production step, associated with each I4 adoption, is subjectively evaluated using CE's detailed principles using the 9R model (refuse, rethink, reduce, reuse, repair, refurbish, remanufacture, repurpose, recycle, and recover), which is the extended form of the 3R on reduce, reuse, and recycle (Kirchherr et al., 2017). Then, the I4-CE impact in manpower is identified per each step, whereby labour involvement is clearly shown in land preparation, fertilising, irrigation, disease control, monitoring fields, and harvesting. As a labour-intensive industry, our proposition is to evaluate the impacts of I4 applications, under CE concepts, to manpower in agribusiness, in which drones participate as a sustainable and efficient manpower replacement tool in this field.

The operational-risk matrix methodology quantifies the possible risk level of each I4 option under CE pillars (economic, environment, and social) considering I4 adoption into CE impact risk values of 0.1, 0.3, and 0.6 for low, medium, high risks, respectively. According to the department of agriculture in the United States (Dohlman, 2020; Harwood et al., 1999), agricultural risks that impact farming are classified into five categories: production, market, financial, institutional, and personal. Production risks are factors that affect both the quantity and quality of produced commodities. It is driven by the uncertain natural growth processes of crops and livestock. Market risks refer to the uncertain prices that producers must pay for inputs used to grow different types of commodities. Financial risks are related to loans, rising interest rates, restricted credit availability, etc. Institutional risks have a major impact on agribusiness that results from uncertainties surrounding government actions such as price levels, waste disposal rules, taxes, chemical use regulations, etc. Personal risks include job profiles, human health, relationships that affect the agribusiness, illness, accidents, death, divorce, and any social crises that can threaten the agribusiness.

Pooling these risks in the I4-CE adoption-impact, we use the notation n for the number of processes, x denotes the CE pillar (economic, environment, social), P refer to the process itself. In the sensitivity risk analysis, the risk sum (Eq.1), its average (Eq.2) and the risk pooling (Eq.3) are defined to evaluate the environmental, economic, and social effects by adopting I4 in the agriculture processes.

$$\sum_{i=1}^n P_i^x \quad (1), \quad \frac{\sum_{i=1}^n P_i^x}{n} \quad (2), \quad \sum_{i=1}^x \frac{\sum_{i=1}^n P_i^x}{n} \quad (3)$$

4. Results

An integrated I4-CE sensitivity risk analysis against the base model has been evaluated based on subjective ratings for twenty agricultural subprocesses and summarised in Table 1. The summation and averages of risks for each CE principle (Eco, Env, Soc) for all agricultural subprocesses are calculated and then pooled. Results show that UAV risk pool value is the lowest with 0.49, which is close to the theoretical minimum risk pool. MAV score 1.25 and UGV 1.20, both around the theoretical medium risk pool. MGV yield 1.44 in the direction to the theoretical high-risk pool. The theoretical minimum, medium, and maximum risk pool values are 0.3, 0.9, and 1.8, respectively. Some subprocesses are kept unrated (as zero risk) when there is no effect by the adoption of I4.

Table 1: Summary results.

I4	UVA			MAV			UGV			MGV		
CE	Eco	Env	Soc	Eco	Env	Soc	Eco	Env	Soc	Eco	Env	Soc
Risk sum	2.3	1.5	3.1	4.9	7.2	3.4	5.4	6.9	4.5	7.9	8.7	5.6
Average	0.16	0.1	0.2	0.4	0.56	0.3	0.38	0.48	0.3	0.51	0.56	0.4
Risk pool	0.49			1.25			1.20			1.44		

5. Conclusion

The results in Table 1 show that the UAV adoption promotes the most efficient utilisation of resources since the agricultural subprocesses will be re-executed efficiently allowing CE principles to be reached while maximising yields, although the effects on manpower would result in either replacing or reducing of it. More precision predictive and prescriptive analytics facilitated by agricultural drones, with the use of machine learning and artificial intelligence, creates results to be delivered through several channels, including farmers' smartphones, farming servers, etc. Such augmented information enables farmers to make decisions or allows automated controls that manage drones to execute routine tasks such as fertiliser application, herbicide spraying, inspection, and monitoring in upmost efficient way. Drones facilitate fast and cheap inspections (descriptive and diagnostic data) for any potential issues such as lack of nutrients in certain areas of the land assets, indication of disease outbreaks, readiness of the crops to be harvested, etc. Moreover, collected data from drones can be visualised, interpreted to surface trends, and provide insights on the optimum solutions and segmentation of the tasks best carried out by either human labour or autonomous vehicles.

Farmers are inclined to be more receptive to the use of technology which supports to meet sustainable and efficient production pathways without significant perturbation on traditional practices as well as when this is not compromising quality of life, and the environment. Agricultural drones have been considered to bring economic prosperity coupled to environmentally friendly applications, whereby the impact on social equity is still debatable. Although a side effect of the automation wave in precision agriculture is closing jobs on the upstream side (production), it generates standard manpower's personal skills that will be upgraded and evolved in the so-called society 5.0 (S5).

Acknowledgment

The authors are grateful to Hamad Bin Khalifa University, Doha, Qatar, for supporting this research.

References

- T. Al-Ansari, S. Chawla, 2020, Qatar harnesses AI to make food output more resilient, Accessed: 26/11/2020 at <https://www.gulf-times.com/story/678810>.
- A. Bechar, C. Vigneault, 2016, Agricultural robots for field operations: concepts and components, *Biosystems Engineering*, 149, 94–111.
- E. Dohlman, 2020, Risk in agriculture, Economic Research Service United States Department of Agriculture, Accessed: 14/11/2020 at <https://www.ers.usda.gov/topics/farm-practices-management/risk-management/risk-in-agriculture>.
- J. Harwood, R. Heifner, K. Coble, J. Perry, S. Agapi, 1999, Managing risk in farming: concepts, research, and analysis, Economic Research Service, U.S. Department of Agriculture. *Agricultural Economic Report*, 774, 49–50.
- J. Kirchherr, D. Reike, M. Hekkert, 2017, Conceptualizing the circular economy: an analysis of 114 definitions, *Resources, Conservation and Recycling*, 127, 221–232.
- Manitobia, 2020, Estimated cost of crop production, Agriculture and Resource Development. Manitobia, Canada, 13, Accessed: 14/11/2020. <https://www.gov.mb.ca/agriculture/farm-management/production-economics/cost-of-production.html>.
- B.C. Menezes, J.D. Kelly, A.G. Leal, G.C. Le Roux, 2019, Predictive, prescriptive and detective analytics for smart manufacturing in the information age, *IFAC-PapersOnLine*, 52(1), 568–573.

- B.C. Menezes, J.D. Kelly, T. Al-Ansari, 2020, Livestock production planning with batch-lines in the agriculture industry, *Computer Aided Chemical Engineering*, 48, 465-470.
- S. Ritz, D. Rizzo, J. Dantan, M. Dubois, F. Fourati, A. Belloy, A. Combaud, 2019, Training in agricultural technologies: a new prerequisite for smart farming, 3rd RDV Techniques AXEMA, SIMA, France.
- UN, 2019, World population prospects, Department of Economic and Social Affairs, 141, Accessed: 14/11/2020. <http://www.ncbi.nlm.nih.gov/pubmed/12283219>.
- Z. Zhai, J.F. Martínez, V. Beltran, N.L. Martínez, 2020, Decision support systems for agriculture 4.0: survey and challenges. *Computers and Electronics in Agriculture*, 170, 105256.

Utilisation of Carbon Dioxide and Gasified Biomass for the Generation of Value Added Products

Ahmed AlNouss,^a Gordon Mckay,^a Tareq Al-Ansari^{a,b*}

^a*Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.*

^b*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

*talansari@hbku.edu.qa

Abstract

Renewable fuels generation and the elimination of fossil-fuel reliance are an important part of sustainable growth. As petrochemicals are a key economic cornerstone of contemporary society, an impetus exists to find potential supply routes so as to maintain scarce resources. In addition, waste utilisation to manufacture value-added products is essential to mitigate emissions of greenhouse gases. Biomass, which is a neutral CO₂ fuel, acts as a possible basis for the generation of energy and fuels. In addition, biomass gasification creates a syngas rich in hydrogen that can be used in the petrochemical industry for the production of value-added materials such as ammonia and methanol. Based on the existing developed biomass-treating infrastructure, incorporating large-scale biomass value-added products into the existing petrochemical generating network can give significant benefits. In addition, there is high possibility to utilise other undesired by-products such as CO₂ which exists within the same infrastructure network. In this research, Aspen Plus model is built based on the oxygen/steam gasification configuration for the production of H₂-rich syngas by different feedstock available in the State of Qatar. The resulting syngas is used with an external combined CO₂ source in the processing of methanol and urea. The results showed that the addition of the external CO₂ source to the biomass gasification system could save around 3% of the emitted stream and increase the production of methanol and urea by around 4% and 10%, respectively. Moreover, the economics of the integrated system are optimised to maximise the utilisation of CO₂ and minimise the capital and operating costs along with net environmental emissions.

Keywords: Biomass Gasification, CO₂ Utilisation, Economic Assessment, Optimisation

1. Introduction

Efficient utilisation of biomass, especially the landfilled waste biomass presents opportunities for significant reductions in global greenhouse gas (GHG) emissions, if used efficiently. Biomass is expected to contribute 120 to 155 exajoule (EJ) of the annual primary energy supply by 2050, which is substantial in comparison to the current total energy supply of 475 EJ (IPCC, 2011). Thermochemical conversion of biomass utilises catalysts and heat to convert solid materials into electric power, chemicals, or fuels. This is in contrast to biochemical biomass processing, that utilises microorganisms and enzymes for the same function. In reality, humanity has employed both thermochemical and biochemical methods for millennia. Interestingly, the prevalence for well over a century of fossil resources thermochemical conversion into energy, chemicals and electricity may be the cause for ignoring the thermochemical processing of biomass

(Zhang and Brown, 2019). Gasification offers variety of advantages as a commercial proven technology with a uniform intermediate product generation and diverse biomass feedstock options. The conversion of carbonaceous feedstock into liquid fuels has a long tradition of gasification. All these processes include the thermal oxidation of biomass to syngas at temperatures up to 1500 °C, producing a gas mixture of carbon monoxide, carbon dioxide, hydrogens, methane and small amounts of light hydrocarbons (National Academies of Sciences and Medicine, 2018). Up to 90 percent of the CO₂ emitted from fossil fuel installations can be permanently deposited with advanced carbon capture and sequestration (CC&S) technologies. Efficient CC&S demonstration and introduction will almost remove GHG emissions and contamination related to coal, boost the atmosphere, health and the economy. Techniques such ocean alkalinity and soil amendment using biochar, bioenergy with carbon capture and storage (CCS) and direct air capture are instrumental to achieve this goal. CCS is considered to have some drawbacks, including information gaps in cost, life-cycle impacts, storage capacity and efficiency between production and usage. However, it remains a valuable method for minimizing CO₂ pollution that is being captured and stored (Parvez et al., 2020). The method of carbon capture and utilisation (CCU) is similar where CO₂ is used instead of storage. As a means of lowering CO₂ emissions and gaining economic benefits at the same time, CCU strategies have recently gained interest. Accordingly, carbon capture utilisation and storage (CCUS) technology is known to be an economically viable means of minimizing GHG emissions before more intensive use is made of renewable energies. By CCUS, CO₂ waste streams become feedstock of marketable product synthesis. The CO₂ can be used as a source of carbon and an oxidant in many chemical reactions, replacing conventional feedstock products in the chemical, pharmaceutical, polymer and automotive industries. The gasification atmosphere of carbon and biomass has been used recently as complete or partial steam substitution. The H₂/CO ratio of generated syngas can be regulated by the CO₂ supply, which is a special feature in the use of CO₂ in the atmosphere. The CCU techniques were studied in the production of multiple value-added materials through fixed network architecture. The technological and economic study revealed a possible revenue of \$0.48 to \$4.35 billion from CO₂ use in the Gas-to-Liquid (GTL) process and a maximum revenue in the methanol processing process for 1.62 to 6 MT/y CO₂ use with a maximum utilisation level (Al-Yaeshi et al., 2020b; Al-Yaeshi et al., 2020a). As for CO₂ used as a gasifying agent in thermochemical conversion of biomass, a number of papers have been published but the knowledge on the basics of use of CO₂ in the thermochemical conversion reaction or the resulting biomass-based syngas is still missing in the processing of useful products. In an attempt to contribute to the reduction of GHG and recycling of biomass waste, this study investigates the integration of CO₂, which is primarily captured and utilised from fossil-based existing production facilities, within the biomass recycling and gasification plant is proposed as sustainable solution for the two problems. This integration can provide an added advantage for the production of chemicals and reducing the associated environmental impacts. In addition, the possible production volumes for fuel and green chemical products can also be improved, and potential beneficial results can thus be accomplished by addition of multiple biomass feedstock and refining the gasification process to yield high-quality syngas under working conditions, feed blending and increased parallel units. In this study, the steam/oxygen gasification of multiple biomass feedstock including food waste, date pits, sludge and manure is considered for the generation of hydrogen-rich syngas that is utilised along with captured CO₂ in the generation of methanol and urea products. The integrated system is then evaluated by means of economic and environmental assessments and sensitivity analysis to allocate the optimal utilisation of CO₂ stream.

2. Model development

Advanced process models are useful to derive accurate energy efficiency and cost-effective assessments of combined biomass gasification and CO₂ utilisation routes. A model of biomass gasification plant should be capable of imitating true plant behaviour, regardless of ability. For simulating a variety of plant implementations, the commercial Aspen Plus platform is appropriate. The base simulation established in this study is focused on a steam/oxygen gasification configuration of various feedstock available in Qatar for the development of H₂-rich product gas used in the production of methanol and urea with captured CO₂ flows. In an earlier work to improve waste to hydrogen generation through the optimisation of gasification processes, the base simulation model using steam and oxygen was established (AlNouss et al., 2020). The base model has then been expanded to explore other gasification mechanisms (AlNouss et al., 2020c), to assess the potential integration between the gasification combined cycle and post-combustion CO₂ removal (Ghiat et al., 2020), to optimise the superstructure representation of application sinks and biomass sources while incorporating energy-water-food nexus approach (AlNouss et al., 2019a; AlNouss et al., 2019b). In this paper, the steam/oxygen gasification model is integrated with downstream application of methanol and urea production while utilising an external CO₂ source to increase the production volume of the end products. Figure 1 illustrates the block flow diagram displaying the various processing steps of the integrated system. Aspen Plus simulation models are developed based on the assumptions of zero-dimensional simulation, neglected formation of tar with char is totally carbon, atmospheric operation with neglected pressure drop, steady-state and isothermal operation and instantaneous pyrolysis and drying. Three property packages, Peng Robinson with Boston-Mathias modifications, RKS-BM and NRTL are used to simulate the nonpolar and real species presented in the model. The multiple biomass feedstock with ultimate and proximate analyses presented in Table 1, is fed to the decomposition unit where they are converted from unconventional to conventional components. The impurities associated with biomass such as Ash and sulphur are separated before the actual gasification take place. The main carbon stream is sent to the gasification unit simulated as Gibbs reactor, where carbon reacts with steam and oxygen, at atmospheric pressure and a temperature of 850 °C, are carried out. Production lines for the processing of urea and methanol are further used with effluent syngas rich of H₂ and an additional CO₂ stream.

Classified as petrochemical plants; urea and methanol are produced in a series of reaction and separation sections yield fertilising, anti-freeze and solvent used materials. Urea production line starts with biomass gasification to yield syngas that is further purified from water and CO₂ before entering the loops of ammonia generation and urea production as illustrated in Figure 1. The urea generation loop assumes a total of 80% CO₂ and ammonia conversion to urea and a recovery of 90% for CO₂ and ammonia (Urea). Whereas, an ammonia separation of 75% and a 7% purging of unreacted recycle H₂ and N₂ stream are assumed in the processing loop of ammonia in general with a 19 percent conversion from N₂, and a 1:3 N₂ to H₂ ratio (Arora et al., 2016; Ammonia, 2. Production processes). The external CO₂ stream is injected in the urea production loop as an addition to the normal CO₂ utilised in the process in the ratio of 0.74:0.57 tonnes of CO₂:ammonia to produce 1 tonne of urea (Al-Yaeshi et al., 2020b).

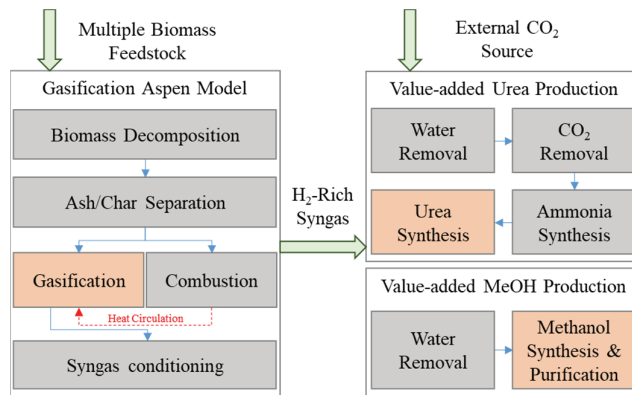


Figure 1: Block flow diagram of the GTL process.

On the other hand, methanol processing steps are simulated with the assumptions of a 9% purge of unreacted recycle gas stream and forward reaction mechanism. It utilises the hydrogen rich gas generated in the biomass gasification unit after removal of CO₂ and Water impurities to achieve the compulsory (CO₂+ CO):(H₂-CO₂) ratio of 2 as illustrated in Figure 1. Before entering the methanol convertor loop, a series of compressors and coolers bring the purified syngas to the required conditions which is then reacts with the external CO₂ injection to yield crude methanol. Further, methanol is purified from water and leftover gases using light ends and heavy ends distillation columns.

Table 1: Proximate and ultimate analyses of biomass feedstock as received basis.

	Date pit	Food Waste	Manure	Sludge
Mass Flowrate (t/y)	7.60x10 ³	7.0x10 ³	5.27x10 ⁵	3.65x10 ⁴
Ultimate analyses (wt %)				
Ash	1.0	6.2	21.4	71.8
S	0.0	0.0	0.5	0.1
Cl	0.0	0.0	1.0	0.0
O	37.9	37.4	31.4	5.7
C	49.8	46.4	37.1	19.1
H	6.8	6.9	5.1	2.3
N	4.5	3.1	3.7	1.1
Proximate analyses (wt %)				
Fixed carbon	17.2	7.7	13.5	19.4
Volatile matter	81.8	86.1	65.0	8.8
Ash	1.0	6.2	21.6	71.8
Moisture	5.0	75.1	27.4	8.3
LHV (dry basis) (MJ/kg)	34.07	19.12	19.40	20.50

To ensure optimal viability and enforcement, it is important to ensure the feasibility of any downstream development route by environmental and economic assessments. The different production lines under review are measured from an environmental and economic perspectives by first quantifying the optimum injection of CO₂ and calculating the net environmental pollution and the expense and income of the investment and operation. The combined economic and environmental research methods are used to analyse and accumulate relevant costs and pollutants of processes (AInouss et al., 2020).

3. Results and discussion

Aiming to quantify the optimal CO₂ utilisation, sensitivity analysis has been conducted by varying the amount of CO₂ injection and observing the trend of methanol and urea production as illustrated in Figure 2. The optimal CO₂ is approximated at 14,800 and 11,992 kg/h for the cases of methanol and urea production, respectively.

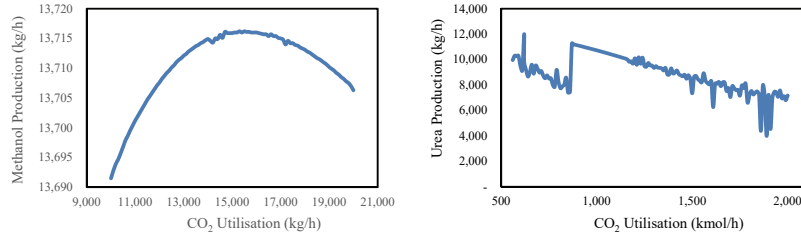


Figure 2: Trends of methanol and urea production with the increase in CO₂ injection.

Following the identification of optimal CO₂ injection, the material and energy balances associated with the optimal configurations are evaluated using Aspen Plus software. The results in Table 2 demonstrates higher requirement of steam in the methanol production route compared to the urea production line, necessary to generate the required (CO₂+CO):(H₂-CO₂) ratio along with lower oxygen requirement.

Table 2: Material and energy requirements of the integrated utilisation techniques.

Utilization technique	Urea	Methanol
Circulated Combustion Heat (MW)	104.2	68.4
Biomass Carbon content (kmol/h)	1.45 x 10 ³	1.45 x 10 ³
Air (kmol/h) (95 O ₂ , 1.6 N ₂ , 3.4 Ar)	1.00 x 10 ³	0.80 x 10 ³
Steam (kmol/h)	1.64 x 10 ³	4.00 x 10 ³
N ₂ (kmol/h)	2.37 x 10 ²	3.73 x 10 ²
Syngas (kmol/h)	5.45 x 10 ³	7.94 x 10 ³
H ₂ (kmol/h)	7.97 x 10 ²	1.19 x 10 ³
CO (kmol/h)	2.66 x 10 ²	2.20 x 10 ²
CO ₂ (kmol/h) after CO ₂ removal	0	2.49 x 10 ²
Product Mass Flow (t/y)	1.05 x 10 ⁵	8.61 x 10 ⁴
Production reactor (MW)	-0.17	-19.38

Moreover, the techno-economic and environmental assessment has been conducted to benchmark the addition of CO₂ to the normal production base cases. The results illustrated in Table 3 demonstrate an increase in the operating and capital costs with the addition of CO₂ stream. This is justified given the higher operating capacity of the subsequent units and the additional utility requirement. This increase is accompanied with an increase of 4% and 10% additional methanol and urea production for the optimal CO₂ utilisation case.

Table 3: The economic and environmental results of the integrated utilisation techniques.

Parameter	Methanol Production			Urea Production		
	Without CO ₂	Optimal CO ₂	Higher CO ₂	Normal CO ₂	Optimal CO ₂	Higher CO ₂
Total Capital Cost [\$]	2.50x10 ⁷	2.55x10 ⁷	2.58x10 ⁷	3.20x10 ⁷	3.22x10 ⁷	3.22x10 ⁷
Total Operating Cost [\$/y]	4.42x10 ⁷	4.57x10 ⁷	4.68x10 ⁷	3.40x10 ⁷	3.42x10 ⁷	3.43x10 ⁷
CO ₂ Utilisation [kg/h]	0	14,800	30,000	11,145	27,286	38,729
Methanol Production [kg/h]	13,163	13,634	13,568	10,912	11,992	11,192
Net CO ₂ Emissions [kg/h]	45,849	44,598	44,126	45,255	45,254	45,843
Total Annualised Cost [\$/y]	4.94x10 ⁷	5.09x10 ⁷	5.21x10 ⁷	4.06x10 ⁷	4.08x10 ⁷	4.09x10 ⁷
Revenue [\$/y]	5.65x10 ⁷	5.85x10 ⁷	5.82x10 ⁷	4.78x10 ⁷	5.28x10 ⁷	5.08x10 ⁷
Net Profit per kg Output [\$/kg]	0.062	0.064	0.052	0.076	0.113	0.096

From the optimal CO₂ injection results, it seems inefficient to raise the CO₂ injection by such a large amount for a negligible increase in methanol production. However, the net profit per kg output has increased by 4% and 50% for the methanol and urea production cases at optimal CO₂. In addition, a reduction in the net environmental emissions approximated at 3% for the case methanol production along with the fact that biomass and CO₂ are utilised as waste streams. These changes in the economic and environmental results are inverted at a higher CO₂ injection rates.

4. Conclusions

Based on the existing developed biomass-treating infrastructure, incorporating large-scale biomass value-added products into the existing petrochemical generating network can contribute significant benefits. Furthermore, there is a high scope for utilising the same infrastructure network by-products such as CO₂. In this research, the integrated system of the oxygen/steam gasification configuration and CO₂ utilisation is evaluated for the generation of H₂-rich syngas by different feedstock available in the State of Qatar. The resulting syngas is used with an external combined CO₂ source in the processing of methanol and urea. The results showed that the addition of the external CO₂ source to the biomass gasification system could save around 3% of the emitted stream and increase the production of methanol and urea by around 4% and 10%, respectively. Moreover, the economics of the integrated system are optimised to maximise the utilisation of CO₂ and minimise the capital and operating costs along with net environmental emissions.

References

- A.A. Al-Yaeshi, A. AlNouss, G. McKay, and T. Al-Ansari, 2020a, A simulation study on the effect of CO₂ injection on the performance of the GTL process, *Computers & Chemical Engineering*, 136, 106768.
- A.A. Al-Yaeshi, R. Govindan, and T. Al-Ansari, 2020b, Techno-economic-based dynamic network design for optimum large-scale carbon dioxide utilisation in process industries, *Journal of Cleaner Production*, 275, 122974.
- A. AlNouss, G. McKay, and T. Al-Ansari, 2019a, Superstructure Optimization for the Production of Fuels, Fertilizers and Power using Biomass Gasification, *Computer Aided Chemical Engineering*, 46, 301-06.
- A. AlNouss, G. McKay, and T. Al-Ansari, 2020, Enhancing waste to hydrogen production through biomass feedstock blending: A techno-economic-environmental evaluation, *Applied Energy*, 266, 114885.
- A. AlNouss, S. Namany, G. McKay, and T. Al-Ansari, 2019b, Applying a Sustainability Metric in Energy, Water and Food Nexus Applications; A Biomass Utilization Case Study to Improve Investment Decisions, *Computer Aided Chemical Engineering*, 46, 205-10.
- Ammonia, 2. Production processes, *Ullmann's Encyclopedia of Industrial Chemistry*.
- P. Arora, A.F.A. Hoadley, S.M. Mahajani, and A. Ganesh, 2016, Small-scale ammonia production from biomass: A techno-enviro-economic perspective, *Industrial & Engineering Chemistry Research*, 55, 22, 6422-34.
- I. Ghiat, A. AlNouss, G. McKay, and T. Al-Ansari, 2020, Biomass-based integrated gasification combined cycle with post-combustion CO₂ recovery by potassium carbonate: Techno-economic and environmental analysis, *Computers & Chemical Engineering*, 135, 106758.
- IPCC, 2011, IPCC special report on renewable energy sources and climate change mitigation.
- E. National Academies of Sciences, and Medicine, 2018, *Bioenergy with Carbon Capture and Storage Approaches for Carbon Dioxide Removal and Reliable Sequestration: Proceedings of a Workshop—in Brief* (The National Academies Press: Washington, DC).
- A.M. Parvez, M.T. Afzal, T.G. Victor Hebb, and M. Schmid, 2020, Utilization of CO₂ in thermochemical conversion of biomass for enhanced product properties: A review, *Journal of CO₂ Utilization*, 40, 101217.
- Urea, *Ullmann's Encyclopedia of Industrial Chemistry*.
- X. Zhang, and R.C. Brown, 2019, Introduction to Thermochemical Processing of Biomass into Fuels, Chemicals, and Power, *Thermochemical Processing of Biomass: Conversion into Fuels, Chemicals and Power*, 1-16.

Synthesis of the EU Supply Networks for the Gradual Transition to 100 % Renewable Energy Production

Sanja Potrč,^a Lidija Čuček,^a Zdravko Kravanja^{a,*}

^a*University of Maribor, Faculty of Chemistry and Chemical Engineering, Smetanova ulica 17, 2000 Maribor, Slovenia*

zdravko.kravanja@um.si

Abstract

This contribution represents the synthesis of sustainable EU-27 energy supply networks, taking into account the production of renewable heat, alternative electricity, biofuel (first, second and third generation), hydrogen, food and other bio-products. A multi-period, mixed-integer linear programming (MILP) model is formulated, with the Sustainability Net Present Value (SNPV) as an objective. A gradual energy transition to 100 % renewable production of heat, electricity and fuels over a period of 40 years is examined. Efficient large-scale heat pumps, biomass cogeneration plants (CHP) and geothermal systems are considered for the heat generation. Taking these technologies into account, it is proposed that 58.2 % of renewable heat could be produced by biomass CHP plants, 40.3 % by large-scale heat pumps and a further 1.5 % using geothermal energy. However, storage technologies need to be integrated into the supply network to satisfy peak electricity and heat demand.

Keywords: supply network optimization, renewable energy transition, renewable heating, alternative electricity, biofuels

1. Introduction

To achieve the long-term reduction of greenhouse gas emissions and the goals of Paris Agreement, a major effort is required, as the consumption of renewable energy is still low in some sectors. At the end of 2018, the share of renewable energy in the EU in the transport sector was about 8.3 %, the share of renewable energy in the heating sector 21.1 %, while the share of renewable energy in total energy in electricity generation was 32.2% (Eurostat, 2020). Clearly, there is a large untapped potential, especially in the transport and heating and cooling sectors. Transition from fossil to 100 % renewable energy supply (Hansen et al., 2019; Bogdanov et al., 2021) requires an assessment of the local availability of renewable energy sources (RES), the use of appropriate renewable energy technologies and the system that can adequately integrate RES to meet daily, monthly and yearly demand (Østergaard et al., 2020). Moreover, the integration of storage technologies will play a very important role (Child et al., 2018). Biomass, wind, solar, geothermal and hydro energy are among the key resources in the transition to renewable production of heat, electricity and fuels. However, emphasis should also be placed on energy savings. There is great potential for increasing energy efficiency in the buildings, improving the linkages between electricity and heating systems and encourage the reuse of waste heat to ensure sustainable, secure and competitive energy supply at the EU level (European parliament, 2020).

This work is an extension of the production and supply of food, fuels and electricity (Potrč et al., 2020) to the 100 % production and supply of heat from renewable energy sources by 2060. In heat production, coal and natural gas are gradually being replaced by alternative electricity, using efficient large-scale compression heat pumps, biomass CHP and geothermal energy combined with heat pumps. To the best of the authors' knowledge, no study has so far examined the long-term transition to 100 % renewable production and supply of heat, electricity and fuels at EU level without compromising food production, and taking into account the sustainability objective. The aim of the study is to examine the economic viability, environmental friendliness and social benefits of a 100 % renewable heat supply over the next 40 years in addition to the supply of electricity and fuels from renewable energy sources, and to assess the associated production capability of the EU.

2. Approach to the sustainable synthesis of renewable supply networks

The synthesis of 100 % renewable energy supply networks was based on a four-layer superstructure (Čuček et al., 2010). The first layer contains the potential locations of the resources. The second layer considers the production of electricity and heat from RES, together with electricity storage, pre-treatment facilities (oil extraction, drying, etc...) and storage of the resources that are further used for biofuel production. On the third level biorefineries are considered and on the fourth layer the end users (demand), including logistics within and between the layers. For the production of electricity, wind farms, solar photovoltaics (PV) and geothermal power plants (GT plant) are taken into account, together with biomass cogeneration technologies. Renewable heat can also be produced with large-scale compression heat pumps (HP) using electricity generated from renewable sources, and geothermal systems combined with large absorption heat pumps. In the case of HP, the average practical coefficient of performance (COP) has a value between 3.5 and 4.5 (David et al., 2017), and this assumption has been used in this study. The use of intermittent sources (wind, solar) could be a bottleneck to meet peak electricity demand. Therefore, pumped hydro storage (PHS) systems with 80 % efficiency were integrated into the network to manage the supply at higher demand. Corn grain and wheat are primarily considered as raw materials for food supply, but the surplus can also be used for the production of biofuels, together with wheat straw, corn stover, forest residues, miscanthus, algae and waste cooking oil. Technologies used for the production of gasoline substitute (green gasoline, bioethanol), diesel substitute (FT-diesel, biodiesel), hydrogen and other bioproducts are: i) gasification and syngas fermentation, ii) gasification and catalytic synthesis of biomass, iii) Fischer Tropsch (FT) synthesis, iv) gasification and lignocellulosic hydrogen production, v) dry grind process and vi) production of biodiesel from waste cooking oil (WCO) and algal oil, with methanol or ethanol as catalysts (Martin and Grossmann 2013).

The optimization was performed using multi-period mixed-integer linear programming (MILP) model. It should be noted that non-linear investment terms were linearized with piecewise linear approximation. The model considers yearly, monthly, daily and hourly time periods. In the case of biomass and waste availability, and demand for biofuels and food, monthly periods are taken into account, while wind and solar energy availability, electricity demand, heat demand and electricity storage are modelled on an hourly basis. Yearly time periods are included to assess the gradual, steadily increasing transition to sustainable energy systems. In the present case study, the maximization of the Sustainability Net Present Value (*SNPV*) is used as an objective (Zore et al., 2018), which

is composed of the Environmental NPV (NPV_{Eco}), the Economic NPV ($NPV_{Economic}$) and the Social NPV (NPV_{Social}), as described by Eq. (1).

$$SNPV = NPV_{Economic} + NPV_{Eco} + NPV_{Social} \quad (1)$$

$NPV_{Economic}$ is calculated as the sum of annual net cash flows (the sum of revenues and salvage value reduced by expenditures and investments) over the project lifetime, with each annual net cash flow discounted to the present time. NPV_{Eco} is defined as the difference between eco-benefits and eco-costs, where the burdening and unburdening effects of raw materials, waste, technologies, transport, energy, products and additional land use are determined using the eco-cost coefficients (Delft, 2016). NPV_{Social} is determined as the sum of the social security contributions paid by newly employed workers and employers, and social unburdening effect from the creation of new jobs, reduced by the social costs incurred by the social support provided by the state and the company to the employees. The model includes energy and mass balances, conversion and production constraints, economic, environmental, and social constraints, and constraints on investment, pretreatment, operating, storage, and transportation costs. The restrictions for the model are also the area in each zone for food, biofuel, heat and electricity production, minimum and maximum capacities for facilities and storage, and maximum allowable distances between source sites, production facilities, and demand sites. Best available renewable heating technologies have been selected to be used and their specifications (costs, capacities, lifetime) were taken from (IRENA, 2013). All different kinds of data on the total area of each zone, yields per hectare and the availability of intermittent sources in each zone, conversion factors, costs and prices, demand (Eurostat, 2020), coordinates of each zone, coefficients used to determine the environmental impact, wages, transport losses, etc., were inserted into the optimisation model in a data-independent form, allowing the model to be applied to other large-scale case studies.

3. EU-27 case study

The methodology is applied to the European Union case study, in which the EU is divided into 47 zones. Larger Member States are divided into several zones to take account of differences in the availability of resources within a country. To achieve 100 % renewable energy supply, the following gradual transition is assumed: a) for 2020, the current share of renewable energy in electricity, heat and fuel production in each Member State is optimized, b) by 2040, a minimum of 50 % renewable energy should be achieved in each sector, c) by 2050, 75 % renewable energy should be produced, and d) 100 % renewable energy supply must be achieved in all sectors by 2060. Gradual electrification of the transport sector is assumed, up to 50 % by 2060, leading to increased demand for renewable electricity. The share of nuclear energy in total electricity generation is assumed to gradually decrease from the current 26 % to 15 % by 2060. The multi-period model that synthesizes sustainable renewable energy supply networks is formulated in GAMS modelling interface (version 27.2.0) and consists of 4,619,882 single equations, 146,827,714 single variables, and 25,286 discrete variables. The model is solved in about 10 hours with the Gurobi solver on the HPC server DL580 G9 CTO (4 processors - 32 core, Intel® Xeon® CPU E5-4627 v2 @ 3.30 GHz, 768 GB RAM).

3.1 Results and discussions

The main optimization results are presented in Table 1 and show the dynamics of the use of different technologies from 2020 to 2060, where the demand for renewable energy will increase from the current share to 100 %. The main technologies proposed for heat

generation in the early years are large-scale heat pumps (HP) (73 % of renewable energy demand), that use the surplus electricity from wind turbines and solar photovoltaic systems. The share of heat produced from biomass CHP is increasing over the years; for 2040, when the heat sector must reach 50 % renewable energy, 27.7 % of renewable energy needs will be covered by biomass CHP, while for the 100 % transition, it is proposed to produce 58.2 % of renewable heat using biomass CHP and 40.3 % from HP. In addition, it is proposed that 1.5 % of heat should be generated by geothermal systems using HP to increase the temperature of the heat. Wind farms are the most promising technology for alternative electricity generation, producing 64.1 % of the total renewable electricity demand in 2060. The share of solar PV will increase with the growing demand for renewable electricity, reaching 31.3 % when a complete transition is required. Small parts of the RES electricity demand are covered by biomass CHP and geothermal systems (4.6 %).

Table 1. Main results of the transition towards 100 % renewable energy production by 2060.

Items	Year 2020	Year 2040	Year 2060
Heat (renewable energy target, %)	22.1	50.0	100.0
Large HP (TWh/y)	872 (73.0 %)	1,955 (72.3 %)	2,179 (40.3 %)
Biomass CHP (TWh/y)	323 (27.0 %)	749 (27.7 %)	3,147 (58.2 %)
Geothermal+HP (TWh/y)	/	/	81 (1.5 %)
Electricity (renewable target, %)	34.1	50.0	100.0
Wind farm (TWh/y)	965 (97.3 %)	1,405 (78.7 %)	2,493 (64.1 %)
Solar PV (TWh/y)	10 (1.0 %)	337 (18.9 %)	1,217 (31.3 %)
Biomass CHP (TWh/y)	17 (1.7 %)	43 (2.4 %)	167 (4.3 %)
GT plant (TWh/y)	/	/	12 (0.3 %)
Biofuel (share in transport sector, %)	8.9	30.0	50.0
Bioethanol (kt/y)	8,483 (43.1 %)	25,398 (35.7 %)	49,088 (41.4 %)
Green gasoline (kt/y)	1,358 (6.9 %)	10,173 (14.3 %)	10,173 (8.6 %)
Biodiesel (kt/y)	9,841 (50.0 %)	10,600 (14.9 %)	13,280 (11.2 %)
FT-diesel (kt/y)	/	24,971 (35.1 %)	46,005 (38.8 %)
NPV_{Sustainability} (1.0•10⁶ €)	/	2,979,482	8,378,409
Area used (% of total)	11.10	12.90	24.86

With 50 % electrification of the transport sector expected by 2060, the demand for biofuels is expected to reach half of today's consumption of diesel and gasoline. Biodiesel is proposed as a diesel substitute in the near future, while for the long-term transition, it is proposed to produce FT-diesel in larger quantities. On the other hand, bioethanol is proposed as the most suitable replacement for fossil-based petrol in the coming years. Note that the use of wheat and corn and the associated production of first-generation bioethanol was not selected due to competition with the food supply chain. The results showed that the sustainability NPV increases over the years and amounts to 8,378,409 M€ by 2060, with the environmental and economic pillars making the highest contribution to the overall sustainability. Compared to our previous model, which considers the production of fuel and electricity from RES, the value of SNPV has almost

doubled due to the integration of heat into the renewable energy supply network. It should be noted that the integration of the energy storage system into the network has significantly contributed to this. The results indicate that it is economically advantageous to store excess electricity (most of the surplus electricity is generated at night and at weekends) using pumped hydro storage plants, and feed it back into the grid when demand for electricity and the selling price of the electricity are higher. The area dedicated to the production of food, renewable fuels, electricity and heat will increase over the years. To achieve 100 % renewable energy supply within the EU, almost 25 % of the total area is used, twice as much as is needed to meet 50 % of demand. Figure 1 shows the distribution of renewable heat generation between 2020 and 2060. For the first year (Figure 1a, optimized current demand), mainly large-scale HPs are proposed; in each zone it is proposed that the surplus electricity generated by intermittent sources is used for heat production. In addition, biomass CHP plants are proposed mainly in Finland and to a lesser extent also in Slovakia, Bulgaria, Spain, Estonia and Lithuania, as these zones have higher availability of forest residues compared to other sources.

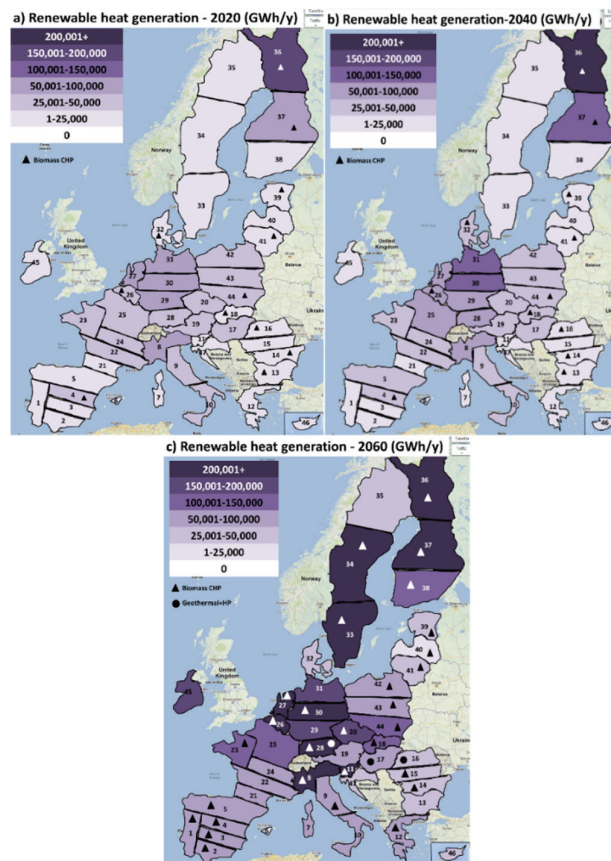


Figure 1. Distribution of renewable heat generation within EU-27 for: a) year 2020, b) year 2040 - 50 % renewable heat target, and c) year 2060 - 100 % renewable heat target.

For the year 2040 (Figure 1b), when renewable heat will have to cover 50 % of heat demand, higher capacities at existing sites are proposed. For the 100 % renewable heat supply (Figure 1c), biomass CHP plants are proposed to be installed additionally in Finland, Slovenia, Portugal, central Spain, France, northern Italy, Czech Republic, Romania, Poland and Germany. Geothermal systems are also suggested in Hungary, Romania and southern Germany. The production of renewable heat using large HP is proposed to be used in all zones in all periods considered.

4. Conclusions

A multi-period MILP model for the synthesis of sustainable renewable energy supply networks has been developed. Efficient renewable energy sources are considered, such as biomass, waste, residues, geothermal energy and intermittent sources (solar and wind energy). The results show under which conditions it would be possible to generate renewable heat in addition to renewable fuels and electricity within the EU in order to reach the zero-carbon emissions target. The most promising solution in the first years are large-scale heat pumps using an excess of alternative electricity, while in the following years biomass CHP plants will be very important for achieving the 100 % transition to renewable heat supply; it is proposed to achieve 58.2 % of renewable heat supply by biomass cogeneration plants and 40.3 % by large compression heat pumps. However, in order to achieve a complete supply of renewable energy in the electricity, heating and cooling and transport sectors, a large part of the area would have to be dedicated (almost 25 % of the total area). As we cannot afford to double the area dedicated to the renewable energy production, the emphasis should be oriented toward increasing energy efficiency, e.g. through cross-sectoral energy integration. The study will be further expanded to include uncertainties in costs, prices, yields, efficiencies, readiness of technologies, and other uncertainty issues to obtain an even more realistic design.

References

- D. Bogdanov, A. Gulagi, M. Fasihi and C. Breyer, 2021, Full energy sector transition towards 100% renewable energy supply: Integrating power, heat, transport and industry sectors including desalination, *Applied Energy*, 283, 116273.
- M. Child, D. Bogdanov and C. Breyer, 2018, The role of storage technologies for the transition to a 100% renewable energy system in Europe, *Energy Procedia*, 155, 44-60.
- L. Čuček, H. L. Lam, J. J. Klemeš, P. S. Varbanov and Z. Kravanja, 2010, Synthesis of regional networks for the supply of energy and bioproducts, *Clean Technologies and Environmental Policy*, 12, 6, 635-645.
- A. David, B. V. Mathiesen, H. Averfalk, S. Werner and H. Lund, 2017, Heat roadmap Europe: large-scale electric heat pumps in district heating systems, *Energies*, 10, (4): 578.
- Delft, 2016, Eco-Costs/Value Ratio, <www.ecocostsvalue.com>, accessed on: 10.11.2020.
- European Parliament, 2020, Energy efficiency, <europarl.europa.eu/factsheets/en/sheet/69/energy-efficiency>, accessed on: 8.11.2020.
- Eurostat, 2020, SHARES 2018 summary results (Short Assessment of Renewable Energy Sources), <ec.europa.eu/eurostat/web/energy/data/shares>, accessed on: 3.11.2020.
- K. Hansen, B. V. Mathiesen and I. Ridjan Skov, 2019, Full energy system transition towards 100% renewable energy in Germany in 2050, *Renewable and Sustainable Energy Reviews*, 102, 1-13.
- IRENA, 2013, Technology Briefs, <irena.org/publications/2013/Jan/IRENA-IEA-ETSAP-Technology-Briefs>, accessed on: 10.11.2020
- M. Martin and I. E. Grossmann, 2013, On the systematic synthesis of sustainable biorefineries, *Industrial & Engineering Chemistry Research*, 52, 9, 3044-3064.

- P. A. Østergaard, N. Duic, Y. Noorollahi, H. Mikulcic and S. Kalogirou, 2020, Sustainable development using renewable energy technology, *Renewable energy*, 146, 2430-2437.
- S. Potrč, L. Čuček, Ž. Zore and Z. Kravanja, 2020, Synthesis of Large-Scale Supply Networks for Complete Long-term Transition from Fossil to Renewable-based Production of Energy and Bioproducts, *Chemical Engineering Transactions*, 81, 1039-1044.
- Ž. Zore, L. Čuček, D. Širovnik, Z. Novak Pintarič and Z. Kravanja, 2018, Maximizing the sustainability net present value of renewable energy supply networks, *Chemical Engineering Research and Design*, 131, 245-265.

Simultaneous optimization and heat integration of a macroalgae-based biorefinery

Fernando D. Ramos, Andrés I. Casoni, Vanina Estrada, M. Soledad Diaz*

Planta Piloto de Ingeniería Química- PLAPIQUI (UNS-CONICET), Camino la Carrindanga km. 7, Bahía Blanca (8000), Argentina

sdiaz@plapiqui.edu.ar

Abstract

In this work, we address a simultaneous optimization and heat integration of a macroalgae biorefinery based on a previous study. As regards heat integration, it considers minimization of heat exchanging total annual cost, assuming isothermal mixing of the streams and overall heat balance within each stage. The resulting MINLP model has 1236 continuous variables, 339 binary variables and 1338 constraints. Hot and cold utilities were 43 % reduced signifying 53 % of savings. Moreover, it favored the environmental aspects of the process which disregarded though it may be, we consider it crucial.

Keywords: brown macroalgae, heat integration, MINLP.

1. Introduction

Macroalgae or seaweeds are photoautotrophic multicellular organisms grouped according to their photosynthetic pigments in green (Chlorophyta), red (Rhodophyta) and brown (Phaeophyta) algae. The main current applications of macroalgae, with a global market of USD 6 billion per year (FAO, 2018), are human food, hydrocolloids production (alginates, agar and carrageenans), fertilizers, biomass for fuel and wastewater treatment (Casoni et al. 2020).

The need of green processes has led to the study of the suitability of macroalgae as feedstock for sustainable production of fuels, chemicals, and materials within an integrated biorefinery concept (e.g. Ingle et al. 2018, Casoni et al. 2020). With 4867 km of coastline and the appropriate climatic conditions for macroalgae growth, Argentine Patagonia constitutes a key place for the macroalgae-based sustainable industry development.

In this work, we propose a mixed-integer nonlinear programming (MINLP) model for the optimization and heat integration (Nemet et al., 2020) of a macroalgae-based biorefinery based on two Argentinean brown algae (*Macrocystis pyrifera* and *Lessonia* sp.) to produce sustainable production of chemicals, materials, and biofuels. To our best knowledge, this specific kind of biorefineries have not been deeply studied. Therefore, incorporating heat integration aspects, contributes to constitute a useful model to evaluate the viability of brown macroalgae biorefineries.

Process Description

The proposed framework is based on a macroalgae-based biorefinery model developed in an early study (Casoni et al., 2020) which has been extended to handle and address heat

integration aspects. The optimum pathway was the production of isosorbide dinitrate from *M. pyrifera* depicted in Figure 1.

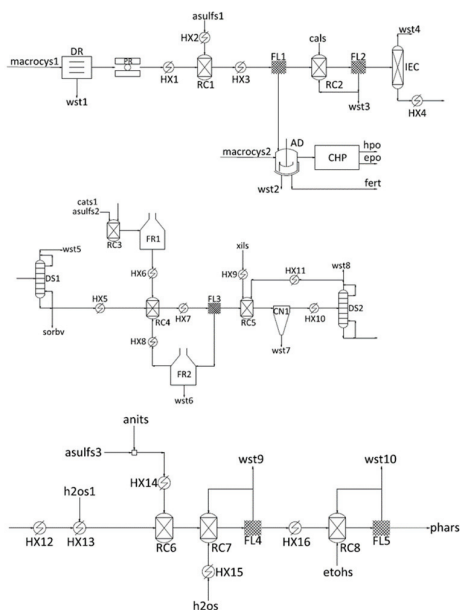


Figure 1. From top to bottom: Section 1, 2 and 3 showing the optimum pathway for the production of isosorbide dinitrate from *M. pyrifera*.

The process begins with drying the macroalgae up to obtain 20 % of moisture in the unit DR, followed by a press unit (PR) for macroalgae milling. Then, the process stream is heated up to 120 °C before feeding it into the reactor RC1 where the biomass is hydrolyzed. Afterward, the resulting stream is cooled down to 25 °C and filtered in FL1 separating the hydrolyzed solid and an aqueous stream containing hydrolyzed sugars. The hydrolyzed solid is fed into an anaerobic digester (AD), which is associated with a CHP unit to produce heat and electric power. Besides, AD produces fertilizers as by-products. The outgoing stream of FL1, constituted by aqueous sugar, is fed into the reactor RC2 where is mixed with calcium carbonate (in a 0.043:1 calcium carbonate-to-acid ratio) to neutralize the remaining acid. After that, the stream is filtered in FL2 to recover the calcium carbonate, which is redirected to RC2. Finally, the process stream is fed into a column containing a polymeric ionic exchange resin (IEC unit) to isolated sorbitol (70 wt.% aqueous solution). From the top of this column, a stream containing other sugars is considered as waste while the bottom one is led to section two: isosorbide production. Sorbitol stream from the first section is preheated up to 150 °C and enters to distillation column (DS1) to obtain pure sorbitol. The resulting stream is fed into the reactor RC4 and is mixed with sulfated zirconium as a catalyst in a 0.02:1 catalyst-to-sorbitol ratio. The sulfated zirconium is obtained in reactor RC3 by mixing zirconium with sulfuric acid (98 wt.%) in a 25:1 zirconium-to-acid ratio. Then this stream is calcinated in the furnace FR1 to obtain the catalyst, which is led to reactor RC4. RC4 outlet stream is cooled down to 120 °C and then is filtered in FL3 to separate the catalyst that is calcinated and

regenerated into FR2 and led back to RC4. The process stream outgoing FL3 unit, which contains isosorbide and other impurities, enters reactor RC5 where xylene is added in a 9.61:1 xylene-to-isosorbide ratio for impurity precipitation which are removed within centrifuge CN1. The outlet stream is heated up to 144 °C and fed into distillation column DS2. The top stream containing gaseous xylene is cooled down to 120 °C and redirected to reactor RC5. The bottom stream that is composed by isosorbide is cooled down to 25 °C and sent to section three as substrate of the final products: Isosorbide dinitrate (ISDN). This final section begins with diluting isosorbide with water in a 3.267:1 isosorbide-to-water ratio and cooling this stream down to 50 °C. Afterward, the main stream enters to reactor RC6 where isosorbide reacts with a mixture of nitric (63 wt.%) and sulfuric acid (98 wt.%) in a 7.315:1 sulfuric acid-to-nitric acid ratio. Acids are added to RC6 considering a 1.36:1 sulfuric acid-to-isosorbide ratio. The RC6 outlet stream is fed into RC7, where water is added to precipitate isosorbide dinitrate. The water is added in a 0.197:1 isosorbide dinitrate-to-water ratio. The resulting stream is filtered to recover isosorbide dinitrate while water is recirculated to RC7. The isosorbide dinitrate stream is heated up to 40 °C and mixed with ethanol in a 1.15:1 ethanol-to-isosorbide dinitrate ratio within the reactor RC8. Once this step is completed, the stream is filtered recirculating ethanol to RC8 and obtaining the final product.

2. Mathematical modeling

Firstly, the process units are modeled using mass and energy balances, equipment design equations, capital cost correlations, integer constraints and a process sustainability metric (RePSIM) is set as the objective function and maximized. Secondly, heat integration and the optimal heat exchanger network is performed using a modified SYNHEAT (Yee & Grossmann, 1990). In this sense, a new optimization problem is solved simultaneously with the heat exchanger network synthesis.

3.1 Mass balances

Non-reactive units mass balances are formulated as follows:

$$\sum_k^K f_{\theta,j}^k = \sum_r^R f_{r,j}^\theta \quad \forall \theta \in \vartheta \quad \forall j \in J \quad (1)$$

where $f_{\theta,j}^k$ is component j mass flowrate from the inlet stream k to unit θ in kg/h and $f_{r,j}^\theta$ represents component j mass flowrate from unit θ to outlet stream r in kg/h.

Reactive units mass balances are formulated as Eq. (2) describes.

$$f_{r,j}^\theta = \sum_k^K f_{\theta,j}^k + \sum_{\substack{h \\ \in J \quad \forall sh \in SH}}^H \xi_{j,sh}^h \cdot C_h \cdot \sum_k^K f_{\theta,sh}^k \quad \forall \theta \in \vartheta' \quad \forall r \in R \quad \forall j \quad (2)$$

where $f_{r,j}^\theta$ is the mass flowrate of component j from reactive unit θ to outlet stream r in kg/h. $f_{\theta,j}^k$ is component j mass flowrate from the inlet stream k to unit θ in kg/h. sh represents the limiting reactant while C_h is the conversion of the limiting reactant for reaction h . $\xi_{j,sh}^h$ denotes the mass coefficient between j and sh in kg/kg, which is positive

for products and negative for reactants. Finally, $f_{\theta,sh}^k$ is the component sh mass flowrate from the inlet stream k to unit θ in kg/h.

3.2 Energy balances

Linear relationships are assumed for energy consumption. In this sense, thermal and electrical energy consumption are calculated as follows:

$$\begin{aligned} \text{Thermal energy consumption: } Q_{\theta} & \\ &= \sum_k^K \sum_j^J f_{\theta,j}^k \cdot h_k - \sum_r^R \sum_j^J f_{r,j}^{\theta} \cdot h_r \end{aligned} \quad (3)$$

$$\forall \theta \in \vartheta \cup \vartheta'$$

where Q_{θ} corresponds to the heat exchanged in unit θ in kJ/h, h_k and h_r correspond to the enthalpies of inlet and outlet streams K and R of unit θ . $f_{\theta,j}^k$ and $f_{r,j}^{\theta}$ are the mass flowrates entering and leaving unit θ in kg/h.

$$\begin{aligned} \text{Electrical energy consumption: } EC_{\theta} &= ECR_{\theta} \cdot \left(\sum_k^K \sum_j^J f_{\theta,j}^k \right) \end{aligned} \quad (4)$$

$$\forall \theta \in \vartheta \cup \vartheta'$$

where EC_{θ} represents electrical energy consumption in kJ/h, ECR_{θ} is the energy consumption ratio per unit of mass flowrate relative to unit θ in kJ/kg.

3.3 Integer and mixed integer constraints

Potential units proposed in the former superstructure (Casoni et al. 2020) are associated with binary variables in order to determine the optimal technological pathway. These variables are included in the model using propositional logic and Big M formulation, and are represented by inequalities as it is shown in Eq. (5).

$$\sum_k^K \sum_j^J f_{\theta,j}^k - BM \cdot y_i \leq 0 \quad \forall \theta \in \vartheta \cup \vartheta' \quad \forall i \in I \quad (5)$$

where $f_{\theta,j}^k$ denotes the mass flowrate of component j in stream k of unit θ while BM is a large number that makes the constraint become redundant when $y_i = 1$. On the other hand, the mass flowrate is forced to be null when $y_i = 0$. It is worth mentioning that BM choice is made considering the maximum mass flowrate involved in the corresponding i technological alternative. Furthermore, binary variables are involved in the heat exchanger network design denoting the matches between hot and cold process streams and utilities.

3.4 Design and economic constraints

The nonlinearities of the problem appear in the equipment design equations and economics constraints, which were formulated following Ramos et al. (2019).

3.5 Objective function

RePSIM Eq. (6) is used as the objective function to evaluate the sustainability of the project. This metric aims at positive values, so that it is maximized. For further details please see Martin (2016).

$$RePSIM = -P + (M + B - F - RM - E - W)C_{CT} - I(C_{annual}) + JS + (F\&F_A + F\&F_B) \quad (6)$$

On the other hand, heat integration of the flowsheet is performed minimizing the annual cost for the network as it is shown below (Biegler et al., 1997).

$$NetCost = HU_{cost} + CU_{cost} + FC_{cost} + A_{cost} \quad (7)$$

Where HU_{cost} and CU_{cost} are the hot and cold utility cost, respectively. FC_{cost} is the fixed cost for heat exchanger and A_{cost} is the area cost of the exchanger.

3. Numerical Results

The MINLP model presented 5398 continuous variables and 13 binary variables and 5336 constraints and it was solved with DICOPT (CONOPT and CPLEX) (Grossmann et al., 2003). Besides, the energetic integration model possesses 1236 continuous variables, 339 binary variables and 1338 constraints. The heat integration reduced the number of cold and heat utilities (Figure 2) which directly impacted on the environmental pillar of RePSIM. Before the heat integration, the process required 828 kW while after heat integration this value was 470 kW (43 % less). These results are in agreement with the published literature (Chong et al., 2020; Song et al., 2018), which mention a 32.1 % and a 42.8 % of energy reduction in a macroalgae waste biorefinery and a cellulosic bioethanol plant, respectively. In term of cost, this represented a 54 % of savings. As regard RePSIM environmental pillar, it showed an improvement of 14 %. Finally, it is worth noting that not only does heat integration improve economic aspects of a process, but it also improves the environmental one.

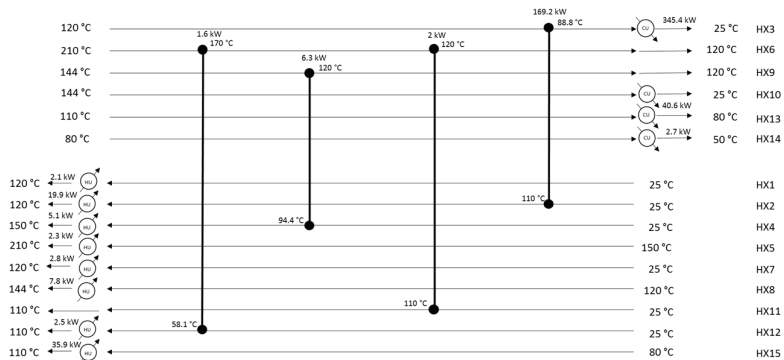


Figure 2. Heat exchanger network

4. Conclusion

In this work, we propose a mixed-integer nonlinear programming (MINLP) model for the optimization and heat integration of a macroalgae-based biorefinery to produce chemical compounds from brown macroalgae biomass. The heat integration was effective in reducing heat requirements denoting how beneficial this type of analysis results in process design. Future work would involve a sensitivity analysis study and waste water treatment in order to keep focusing on environmental aspects.

References

- FAO, 2018. Global status of seaweed production, trade and utilization. Rome Globefish Res. Programme 124, 120. Licence: CC BY-NC-SA 3.0 IGO.
- A. I. Casoni, F. D. Ramos, V. Estrada, M. S. Diaz, 2020. Sustainable and Economic analysis of Marine Macroalgae Based Chemical Production – Process Design and Optimization. *Journal of Cleaner Production*. 276, 122792.
- K. Ingle, E. Vitkin, A. Robin, Z. Yakhini, D. Mishori, A. Golberg, 2018. Macroalgae biorefinery from *Kappaphycus alvarezii*: conversion modeling and performance prediction for India and Philippines as examples. *BioEnergy Res.* 11 (1), 22-32.
- A. Nemet, T. G. Walmsley, E. Ahmetovic, Z. Kravanja. 2020. Process synthesis and simultaneous heat and electricity integration to reduce consumption of primary energy sources. *Comp. Aided Chem. Eng.*, 48, 901-906.
- T. F. Yee, I. E. Grossmann. 1990. Simultaneous optimization models for heat integration – II. Heat exchanger networks synthesis. *Comp. Chem. Eng.*, 28, 1165-1184.
- F. D. Ramos, C.A. Delpino, M.A. Villar, S. M. Diaz, 2019. Design and optimization of poly(hydroxyalkanoate)s production plants using alternative substrates. *Bioresour. Technol.* 289, 121699.s
- M. Martin, 2016. RePSIM metric for design of sustainable renewable based fuel and power production processes, *Energy*, 114, 833-845.
- L. T. Biegler, I. E. Grossmann, A. W. Westerberg. 1997. Systematic methods of chemical process design. Prentice Hall PTR.
- T. Y. Chong, S. A. Cheah, C.T. Ong, L.Y. Wong, C. R. Goh, I. S. Tan, H. C. Y. Foo, M. K. Lam, S. Lim. 2020. Techno-economic evaluation of third-generation bioethanol production utilizing the macroalgae waste: A case study in Malaysia, *Energy*, 210, 118491.
- C. Song, Y. Qiu, Q. Liu, N. Ji, Y. Zhao, Y. Kitamura, X. Hou. 2018. Process intensification of cellulosic ethanol production by waste heat integration, *Chem. Eng. Res. Des.*, 132, 115-22.

Integrated optimization of Multi-Energy System operation and thermal comfort management for buildings

Lavinia Marina Paola Ghilardi^a, Alessandro Francesco Castelli^a, Luca Moretti^a,
Mirko Morini^b, Emanuele Martelli^{a,*}

^a*Politecnico di Milano, Department of Energy, Via Lambruschini 4, Milano, 20154, Italy*

^b*University of Parma, Department of Engineering and Architecture, Parco Area delle Scienze 181/a, Parma, 43125, Italy*

**emanuele.martelli@polimi.it*

Abstract

The growing interest in decarbonization promotes the spread of distributed generation and efficient load control approaches (demand-side management). In this framework, the definition of mathematical tools to tackle the integrated optimal management of energy resources and flexible loads is fundamental. This study addresses the optimal scheduling problem of a Multi-Energy System (MES) and thermal comfort management for buildings in a unified approach. Linearized simplified models of the buildings and district heating network (DHN) are included in the algorithm. The optimization can exploit thermal inertia of buildings and of DHN to increase the flexibility of the system. The proposed method is applied to a group of buildings of the Campus of University of Parma.

Keywords: Multi Energy Systems, demand-side management, MILP, district heating network.

1. Introduction

The current energy transition towards more environmentally sustainable solutions has boosted the employment of distributed energy resources, encompassing both distributed generation and controllable loads. Local energy production has to cope with fluctuating renewable energy sources (RES) availability and therefore the need of flexible solutions is of major importance. The building sector shows great potential in this direction, as thermal mass - readily available - can be exploited to shift loads. Moreover, buildings represent a large fraction of final energy consumption, accounting for approximately 40% in Europe (Rousselot, 2018). This work addresses the concurrent optimization of the thermal comfort management in buildings, operation of the district heating network and scheduling of the Multi-Energy System (MES) supplying their electrical and thermal power. An extension of the MILP (Mixed Integer Linear Programming) scheduling problem formulation proposed by Bischi et al. (2019) for MESs is proposed, by including a linearized dynamic thermal model of the buildings and a simplified linearized model of district heating network (DHN). The formulation allows optimizing not only the operation of the units (unit commitment and economic dispatch) but also the profiles of thermal energy supplied to each building, the dynamic evolution of their indoor temperature and

delivery and return temperatures of the DHN. This enables the possibility of exploiting the heat capacity of the system to shift energy consumption and find a more economic operational planning for the generation units. The proposed method is applied to the Campus of University of Parma considering two different MES designs (supplying heating, cooling and electricity) and the optimized strategies are compared with the conventional management strategy.

2. Problem statement

The optimization addresses the weekly scheduling of the generators with the possibility to exploit the heat capacity of the buildings and of the DHN in order to increase the operational flexibility of the units and reduce energy consumption (or increase revenues from electricity sales to the grid). Constraints on indoor temperature variations are included in occupancy hours, to guarantee that thermal comfort is not jeopardized. The general operation planning problem can be formulated as follows:

“Given:

- a set of dispatchable generation units \mathcal{M} , with performance curves and technical limits;
- a set of non-dispatchable units \mathcal{ND} ;
- a set of buildings \mathcal{K} , with defined construction parameters;
- forecasted electricity prices, forecasted electrical demand profile, forecasted occupancy profiles, forecasted outdoor air temperature and irradiance,

determine (for each time step t within the time horizon \mathcal{T}):

- the dispatchable units to be turned on and their load;
- the power exchanged with the grid;
- heating or cooling power supplied to buildings;
- the buildings indoor temperature;
- the delivery and return temperatures of the primary and secondary circuits;

which minimizes the operating cost subject to the following constraints:

- dynamic thermal balance of the building;
- heating, cooling and electrical balance of the whole system;
- technical limits of generating units;
- thermal comfort requirements (desired temperature range within occupancy hours and ASHRAE dynamic comfort constraints);

thermal propagation delay in pipes, limits on heat transfer in heat exchangers and heat balances across the heat distribution network.

3. Energy system and buildings modelling

Several studies have developed MILP models for the optimal operational planning of MES, where binary variables are required to include start-up/shut-down operation of units. However, a considerable number of papers deal separately with the energy system and the buildings supplied, either considering energy demand as an exogenous input or mainly focusing on buildings' perspective. The decoupling of these two aspects results in underuse of the load shift capability of the buildings and sub-optimal performance of the generating units (Darivianakis et al., 2019). For this purpose, a linearized differential model of the building is included in the algorithm, accounting for power input, heat exchange through walls, air infiltration losses and the contribution of air forced ventilation. When dealing with the management of buildings within an energy district, the thermal propagation delay in pipes cannot be neglected and therefore the linearized

dynamics of the district heating network are included in the MILP. In the following, we present the main features of the model.

3.1. Performance curves

The part-load performance curves (power input - output relationship) of the generating units have been linearized with minor approximations (<5%), as shown in Zatti et al. (2018). The effect of the outdoor air temperature T_{ext} is computed for the heat pump (HP) and the compression chiller (CC), through a linear function whose value affects the different load rates independently. As a matter of fact, the power output of these units is multiplied to the coefficient k_T , resulting from the best-fit of the performance data found in Yu et. al (2007) and Johnson (2013):

$$k_T = (m_T \cdot T_{ext} + q_T) \quad (1)$$

Where: m_T and q_T are the coefficients of the linear best-fit. Moreover, a penalty for start-up is included in the model as a reduction of machines useful output. To this end, we introduced a binary variable $\Delta(i, t)$ to state whether the unit has been switched on at a given timestep. This variable is linked to the unit commitment binary variable $z(i, t)$ by a set of inequality constraints, as described by Bischi et. al. (2019).

3.2. Building model

The model adopted in this work is a single state grey-box, derived from Gambarotta et al. (2017), where the coefficients of buildings are found trough best-fit of TRNSYS white-box model, by activating sequentially the contributions:

$$C_j \frac{dT_j}{dt} = UA \cdot (T_{ext}(t) - T_j(t)) + mc_{p,nat} \cdot (T_{ext}(t) - T_j(t)) + mc_{p,force} \cdot (T_{UTA}(t) - T_j(t)) + Q_{irr}(t) + Q_{occ}(t) + Q_j(t) \quad (2)$$

Where: $T_j(t)$ is indoor temperature, $T_{ext}(t)$ is outdoor air temperature, $T_{UTA}(t)$ is the temperature of the air coming from the air treatment unit, C_j is the total thermal capacity of the building, UA is the mean thermal transmittance of the envelope multiplied by its surface, $mc_{p,nat}$ and $mc_{p,force}$ are respectively thermal capacity of infiltration air and air forced ventilation, $Q_{irr}(t)$ and $Q_{occ}(t)$ are solar and internal gains, while $Q_j(t)$ is heat supplied to end-users. In particular, $T_j(t)$ and $Q_j(t)$ are the variables optimized by the model, in compliance with comfort requirements. In order to be implemented in the MILP, the differential equation is approximated using forward finite differences. We compared the approximated solution with the analytic one found for some test input profiles with a time step of 7.5 minutes (adopted for the optimization). The resulting absolute error in assessing the internal building temperature was lower than 0.05°C, sufficiently good for the purposes of the operational planning optimization.

3.3. District heating network model

The general scheme of DHN is depicted in Figure 1. The proposed approach can be applied to DHNs with radial configuration, where each building has a dedicated primary circuit connecting its internal secondary loop with the power generation site. The DHN is assumed to operate with a constant mass flow rate, while the heat transferred across the system is adjusted by changing the primary water delivery temperature T_p^D (equal for all buildings). Moreover, each building can decrease its input thermal power by opening the bypass valve of the heat exchanger. Assuming that the flow with minimum heat capacity rate is the primary loop during the heating season and the secondary loop in the cooling

season, the thermal power provided to each building j can be written as a linear function of the delivery temperature of the DHN (see Eq. (3)).

$$Q(j, t) \leq \varepsilon(j) \cdot c_w \cdot \dot{m}_{\min}(j) \cdot (T_H^{IN}(j, t) - T_C^{IN}(j, t)) \quad (3)$$

Where: ε is nominal effectiveness of the heat exchanger, c_w is heat capacity of water, \dot{m}_{\min} is the mass flow rate of the fluid with minimum thermal capacity, $T_H^{IN}(j, t)$ is the inlet temperature of the hot fluid and $T_C^{IN}(j, t)$ is the inlet temperature of the cold fluid. A constraint is included in the MILP to guarantee that the minimum thermal capacity is always on the same side of the heat exchanger. In order to consider the delay in heat propagation, we related the inlet $T_{in}(t)$ and outlet $T_{out}(t)$ temperature of the pipe, following the approach of Dobos et al. (2014). Assuming one dimensional flow, neglecting heat conduction in the axial direction and heat capacity of pipes, the solution of the energy conservation equation in the pipes is:

$$T_{out}(t) = T_{\infty} + (T_{in}(t - \Delta t) - T_{\infty}) \cdot (1 - \frac{4 \cdot k_p}{D_p \cdot c_w \cdot \rho_w} \cdot \Delta t) \quad (4)$$

Where: Δt is time delay, ρ_w is water density, k_p is the heat transfer coefficient between the pipes and the surroundings, D_p is the pipe diameter, T_{∞} is the temperature of the surroundings. This equation can be employed in the MILP model as long as the time discretization is sufficiently tight to capture the time delay in heat propagation (i.e. the time resolution has to be lower than the time the water needs to reach the closest building from the power plant). For this purpose, we considered a time resolution of 7.5 minutes in the case study.

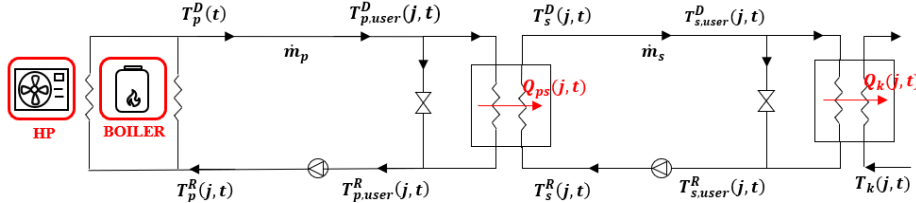


Figure 1 – Scheme of the primary and secondary heating loops serving building j .

3.4. District heating network model

In the conventional management strategy, the thermal power provided to the buildings is adjusted to keep a set point of $20^{\circ}\text{C} \pm 0.2^{\circ}\text{C}$ during the heating season, and of $25^{\circ}\text{C} \pm 0.2^{\circ}\text{C}$ in the cooling season. The water delivery temperature of DHN is increased linearly as a function of the outdoor air temperature. The conventional optimization approach (without thermal comfort management where the heat demand profile is fixed) can be reproduced by imposing these binding constraints on the internal building temperatures. Instead, the Thermal Comfort Management (TCM) strategy, the building indoor air temperatures, the DHN water delivery and return temperatures and the thermal power supplied to each building are optimized in the MILP, in compliance with comfort requirements of the users during occupancy hours. In particular, we assumed that the users (students and university staff) accept to participate to a demand side management program, allowing a quality band of $\pm 2^{\circ}\text{C}$ around the summer and winter setpoints (18°C - 22°C in the cold season and 23°C - 27°C in the hot season). To avoid abrupt changes of temperature, ASHRAE standards (2004) constraints on ramps and drifts are included: the variation of temperature within an hour must be lower than 2.2°C , while the limit in 15 minutes is of 1.1°C . During non-occupancy hours, the building temperatures are free to vary outside the quality ranges.

4. Case study

The presented method has been applied to a group of 12 buildings of Campus of University of Parma. The area to volume ratio of the buildings varies from 33% to 55%, while time-constant (the ratio between thermal capacity and dispersions) ranges from 34 h to 158 h. Given the diversity of these features, the resulting indoor temperature profiles will differ between buildings. Thermal comfort requirements must be respected during working hours, i.e. from Monday to Friday from 8 a.m. to 6 p.m. The scenarios investigated encompass 3 different weeks of the year, representative of winter, mid-season and summer. The purchase price of electricity is derived from historical data of 2018 non-domestic users in Italian protected market (tri-hour tariff), while the purchase price of natural gas is defined on the basis of yearly average consumption of the university campus. We considered 2 different MES configurations: Design D1 features a boiler and a compression chiller, while design D2 comprises a HP - covering half of the heat peak - auxiliary boilers, a CC and PV panels.

5. Results

The optimizations have been formulated with Pyomo and the MILP has been solved with Gurobi solver. The solution of the large scale MILP (250160 continuous variables, 6725 binary variables, 384059 constraints) needs approximately 5 hours for optimizing the weekly operation with a time resolution of 7.5 minutes. If the optimization horizon is shortened to three days, the computational time reduces to only 20 minutes. With TCM, the algorithm can either decide to reduce the heat requirements of buildings or to store heat in their mass (i.e. when there is availability of RES). Figure 2 highlights the different thermal behavior (in design D2) of two buildings of the campus B0 and B1, respectively featuring a high and a low thermal capacity. The MILP can also exploit DHN thermal inertia, by varying water delivery temperature. As such, this temperature rises in the presence of RES (Figure 2) and it drops in the rest of the day, reducing the share of excess PV production sold to the grid at low price (from 84% in the reference strategy to 38% with TCM) and increasing the HP production (Figure 3). When there is a mismatch between thermal demand and low electricity price periods (i.e. with the CC), TCM enables production shifting. Furthermore, the flexibility provided by TCM slightly enhances machines performance, reducing boiler switching on/off and operation at partial loads, and advancing chillers start up in order to benefit from lower outdoor temperature (mean COP improves from 3.7 to 4.2). All these sources of flexibility explained so far operate concurrently and positively affect the total operating costs, as shown in Table 1, where the savings of TCM compared to the reference scenario are outlined in three weeks of the year (winter, spring and summer).

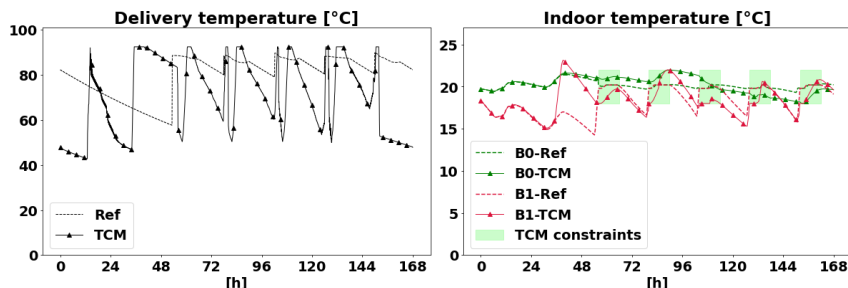


Figure 2. Temperature profiles (Saturday-Friday) of Design D2, in the reference strategy and with TCM. Right: indoor temperature profiles of buildings B0 and B1, with respectively high and low thermal capacity.

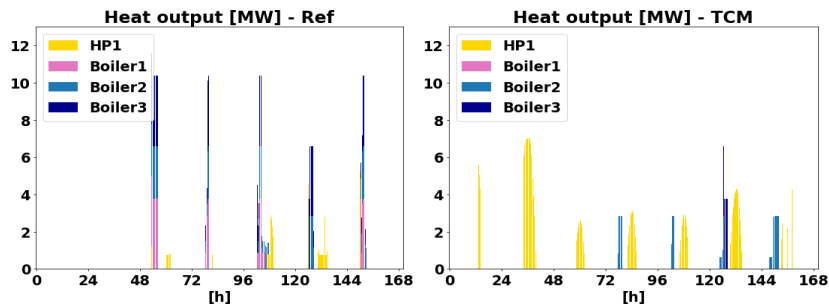


Figure 2. Unit commitment (Saturday-Friday) of Design D2, in the reference strategy and with TCM.

Total operating costs			
	W1	W2	W3
D1	-14.4%	-48.0%	-26.4%
D2	-2.8%	-21.6%	-7.6%

Table 1. Variation of total operating costs with TCM compared to the reference strategy in MES Design D1 and D2, in three weeks of the year (W1 = winter; W2 = spring, W3 = summer)

6. Conclusions and future works

The presented MILP approach appears to be effective not only for operational planning, but also to manage comfort in buildings. Its limited computational time allows solving weekly basis problems with very fine time resolution. It can be easily integrated into rolling horizon algorithms today used for optimizing the energy management strategy of multi-energy systems and energy districts.

Future extensions of the methodology may include more detailed models of the buildings and thermal energy comfort, as well as extending the optimization approach to cope with forecast uncertainty.

References

- M. Rousselot, 2018, ODYSSEY. Energy Efficiency Trends in Buildings, Pages 1–8, 2018.
- G. Darivianakis, A. Georghiou, R. S. Smith, and J. Lygeros, 2019, The Power of Diversity: Data Driven Robust Predictive Control for Energy-Efficient Buildings and Districts, *IEEE Trans. Control Syst. Technol.*, 27, Pages 132–145
- M. Zatti, M. Gabba, M. Rossi, M. Morini, A. Gambarotta, and E. Martelli, 2018, Towards the optimal design and operation of multi-energy systems: The ‘efficity’ project, *Environ. Eng. Manag. J.*, 17, Pages 2409–2419
- A. Bischì et al., 2019, A rolling-horizon optimization algorithm for the long term operational scheduling of cogeneration systems, *Energy*, 184, Pages 73–90
- F. W. Yu and K. T. Chan, 2007, Optimum load sharing strategy for multiple-chiller systems serving air-conditioned buildings, *Build. Environ.*, 42, , Pages 1581–1593
- R. K. Johnson, 2013, Measured Performance of a Low Temperature Air Source Heat Pump
- A. Gambarotta, M. Morini, M. Rossi, and M. Stonfer, 2017, A Library for the Simulation of Smart Energy Systems: The Case of the Campus of the University of Parma, *Energy Procedia*, 105, Pages 1776–1781
- L. Dobos, J. Jäschke, H. Mordt, S. Skogestad, and S. Narasimhan, 2014, Dynamic Model and Control of Heat Exchanger Networks, *Nt.Ntm.No.*, 37, Pages 37–49
- ASHRAE STANDARD 55-2004. Thermal Environmental Conditions for Human Occupancy. GURUBI. <http://www.gurobi.com>.

Towards a Rational, Quantum-Chemistry-Based Selection and Screening of Green Solvents for Liquid-Liquid Phase Transfer Catalysis

Abhimanyu Pudi,^{a,b} Adam P. Karcz,^a Rofice Dickson,^a Suojiang Zhang,^{b,c} Seyed Soheil Mansouri,^{a,b} Martin P. Andersson^{a*}

^a*Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søtofts Plads 228A, Kgs. Lyngby, DK-2800, Denmark*

^b*Sino-Danish College, University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing, 100049, China*

^c*Beijing Key Laboratory of Ionic Liquids Clean Process, State Key Laboratory of Multiphase Complex Systems, CAS Key Laboratory of Green Process and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, 1 North 2nd Street, Zhongguancun, Haidian District, Beijing, 100190, China*
martan@kt.dtu.dk

Abstract

Coupling of reaction and extraction, along with the ability to significantly increase reaction yields at milder conditions, makes liquid-liquid phase transfer catalysis an attractive intensified solution for many industrially important processes. Although there has been some work in terms of mechanistic modeling, solvent selection for such processes has remained a largely untouched topic by academia and industry. This work provides a path towards a systematic solvent screening or design framework by providing a quantum-chemistry-based evaluation method to find suitable green solvents for toluene in the process of hydrogen sulfide valorization from natural gas. From a pool of eight candidate alternatives, three are chosen to be suitable substitutes for toluene based on technical, economic, and environmental measures.

Keywords: green chemistry, sustainable engineering, reactive extraction, COSMO-RS, reaction path analysis.

1. Introduction

Liquid-liquid phase transfer catalysis (LL-PTC) belongs to an important class of intensified processes called multiphase reactive extraction. It has been reported to enable novel synthesis routes, higher yields, and faster reactions, while also facilitating the separation of certain species. Pudi et al. (2020) have shown a promising application of LL-PTC to offshore gas sweetening process as a replacement for the expensive and unsustainable triazine-based hydrogen sulfide (H₂S) removal method. In that work, the advantage of integrating quantum chemistry and Conductor-like Screening Model for Real Solvents (COSMO-RS) into process design has been clearly established. However, the emphasis of such studies has been put mostly on catalytic and mechanistic considerations with little attention paid to the choice of organic solvent. Due to the increasingly tight regulations and strong public opinion, a thoughtful selection for process-efficient and sustainable solvents in industrial activities has become an immediate requirement for chemists and chemical engineers alike. In an effort to meet

this need, this work shifts the focus towards the rational selection of green solvents to perform the biphasic production of benzyl mercaptan (BnMer) from benzyl chloride (BnCl) using tetrabutylammonium bromide (QBr) as the phase transfer catalyst. The specific chemical system and the reactions are reported by Pudi et al. (2020).

2. Methodology

2.1. Computational tools

The quantum chemistry program package, TURBOMOLE version 7.4, is employed for electronic structure calculations using density functional theory (DFT) and COSMO. The calculations are performed using the Becke-Perdew functional with the Ahlrichs basis set of triple-zeta valence plus polarization (BP-TZVP) to obtain the optimized structures, ground-state electronic energies, vibrational frequencies, thermal corrections to the free energy, and optimized reaction path. Based on the optimized DFT-COSMO molecular structures, the COSMO-RS solvation model implemented in COSMOtherm Release 19 is used to obtain liquid-phase thermodynamic properties, such as partition coefficients and free energies of solvation. In addition, the ProPred toolbox in the Integrated Computer-Aided System (Gani et al., 1997) is used to evaluate the environmental impact factors.

2.2. Solvent (pre-)selection based on published guides

Using the immiscibility of a solvent in water as a necessary requirement for the biphasic system, eight potential alternatives to toluene are selected from the vast pool of solvent candidates assessed in the CHEM21 selection guide (Prat et al., 2015) and Sigma-Aldrich's Greener Solvent Alternatives: 1-butanol, 2-methyl tetrahydrofuran (2-MeTHF), anisole, benzyl alcohol, butyl acetate, ethyl acetate, isopropyl acetate, and methyl isobutyl ketone (MIBK).

2.3. Reaction path optimization and free energies

Based on the approximation of a locally quadratic potential energy surface, reaction path at each step is optimized with the sole constraint of equally spaced structures using the reactants and products as the path boundaries (Plessow, 2013). An initial path is generated using a slight variation of the linear synchronous transit method (Halgren & Lipscomb, 1977). This iterative process is continued until one or more good initial guesses are obtained for the transition state (TS) optimization. Vibrational frequency analysis is performed on these initial guesses to find the closest approximation to the TS. This structure is then optimized to find the TS structure. The free energy of activation is equal to the free energy of reaction with the TS as the product of the reaction. Therefore, both free energies are calculated using the same procedure as reported by Hellweg and Eckert (2017).

2.4. Environmental impact

Four environmental impact factors (EIFs) are evaluated for each solvent: human toxicity potential by ingestion (HTPI), human toxicity potential by exposure (HTPE), aquatic toxicity potential (ATP), photochemical oxidation potential (PCOP). Using Eqs. (3)-(6), these EIFs for a chemical j are calculated based on the properties listed in Table 2. In the absence of experimental values for the properties, ProPred estimates them based on the combined group contribution and atom connectivity index method.

$$EIF_{j,HTPI} = \frac{1}{(LD_{50})_j} \quad (1)$$

$$EIF_{j,HTPE} = \frac{1}{(PEL)_j} \quad (2)$$

$$EIF_{j,ATP} = \frac{1}{(LC_{50}^{FM})_j} \quad (3)$$

$$EIF_{j,PCOP} = \frac{1}{(PCO)_j} \quad (4)$$

Table 1 Environmental properties obtained from ProPred to calculate the EIFs

Property	Unit	Description
-Log(LC ₅₀) ^{FM}	Log(mol/L)	50% lethal concentration for Fathead Minnow after 96 h.
-Log(LD ₅₀)	Log(mol/kg)	50% lethal oral dose for rat
-Log(PEL)	Log(mol/m ³)	Permissible exposure limit as defined by OSHA-TWA
-Log(PCO)		Photochemical oxidation potential from the reaction of NO _x and volatile organic compounds

Due to the differences in units and scales, EIFs are then standardized for each impact category using the following equation:

$$EIF_{j,k}^s = \frac{EIF_{j,k}}{\mu_k + 2\sigma_k} \quad (5)$$

where μ is the mean value, σ is the standard deviation, k is the impact category, j is the solvent, and s stands for standardized. The standardized EIFs are then weighted and summed to obtain the environmental impact of each solvent as a single score (SS) as shown in Eq. (8). The weights are chosen based on the offshore application, thus the weight for ATP was set the highest. Human exposure due to handling is important, so it followed closely, and photochemical smog is placed for a lower scaling.

$$SS_j = EIF_{j,PCOP}^s + 2EIF_{j,HTPE}^s + 3EIF_{j,ATP}^s + 4EIF_{j,ATP}^s \quad (6)$$

3. Results and Discussion

Choosing the best solvent for a specific application depends on several different factors, such as process performance, ease of product recovery, price, and environmental impact. For example, the initial selection criterion of water immiscibility among the green solvent candidates reported by the selection guides allows for higher yield, higher selectivity, etc., conforming to the advantages of a biphasic system in terms of process performance (Mąkosza & Fedoryński, 2003). In this context, the solvents are evaluated based on several important screening criteria relevant to the specific chemical system under consideration: partition behavior, reaction energy barrier, reaction free energy, organic product separation, solvent price, and environmental impact.

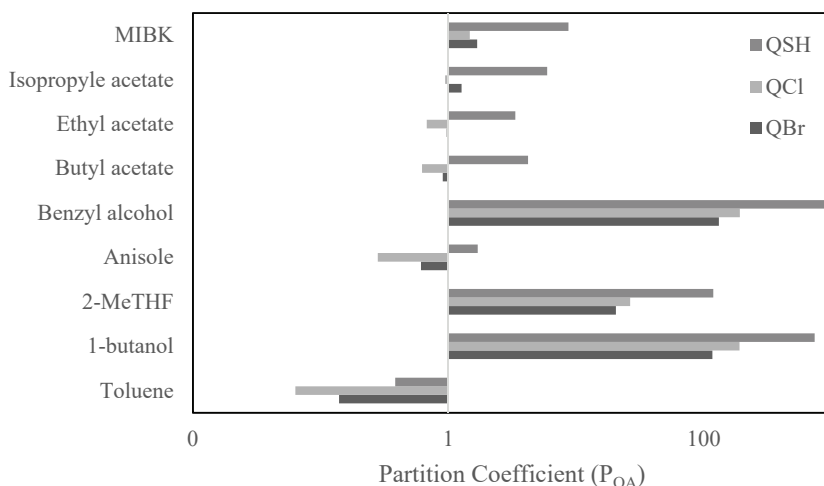


Figure 1 Partition coefficients (organic/aqueous) of the three catalyst species in all solvents

In LL-PTC, the partition behavior of the reactants, products, and the catalyst species is an important consideration. Partition coefficients for all the species are calculated in COSMOtherm with a value greater than 1 implying a preference for the organic phase and a value less than 1 implying a preference for the aqueous phase. A good organic solvent would result in high partition coefficients for BnCl and BnMer. Since all the considered solvents exhibit this behavior, this becomes irrelevant for screening. Proper partition behavior of the active and inactive catalyst forms is crucial for an efficient and effective LL-PTC. The active catalyst form, tetrabutylammonium hydrogen sulfide (QSH), must transfer to the organic phase to form BnMer, while the inactive catalyst forms, QBr and tetrabutylammonium chloride (QCl), must simultaneously transfer to the aqueous phase to form QSH. Ideally, the partition coefficient of QSH should be greater than 1 while that of QBr and QCl should be less than 1. Figure 1 shows (in log-scale) how the solvent candidates affect the partition coefficients of the three catalyst forms. All the three species have high partition coefficients in 1-butanol, 2-MeTHF, and benzyl alcohol, making them less favorable in terms of process performance. On the other hand, ethyl acetate, butyl acetate, and anisole cause the desired partition behavior.

Faster kinetics lead to a smaller reactor volume, while better thermodynamic favorability results in higher yield. The free energies of activation and reaction are a measure of kinetics and thermodynamic favorability, respectively. Figure 2 shows that, except for 1-butanol and benzyl alcohol, all the candidates result in an improved reaction performance in comparison to toluene with 3.4–7.8 % reduction in the reaction energy barrier. The two solvents with much poorer reaction kinetics also exhibit less favorable thermodynamics.

Ease of product recovery affects both the economics and the total environmental footprint of the process. Organic solvents with lower volatilities and higher boiling points correspond to lower emissions of volatile organic components. However, such solvents require a more expensive product recovery since the most common method for product recovery is distillation or evaporation of the solvent. In terms of the specific chemical

system under study, the boiling point of BnMer is 195 °C. Therefore, an easier product recovery requires the boiling point of the solvent to be lower to a satisfactory extent. However, it can't be too low since the reaction temperature is about 60 °C. Figure 3 shows that anisole, benzyl alcohol, and butyl acetate make bad candidates due to their high boiling points, while 2-MeTHF and ethyl acetate pose a slight risk since their boiling points are higher than the reaction temperature only by about 20 °C.

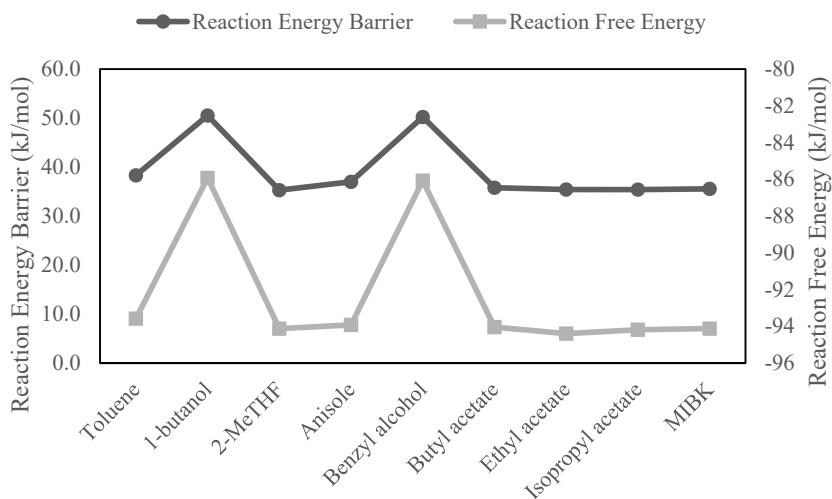


Figure 2 Free energies of activation and reaction in all the considered solvents

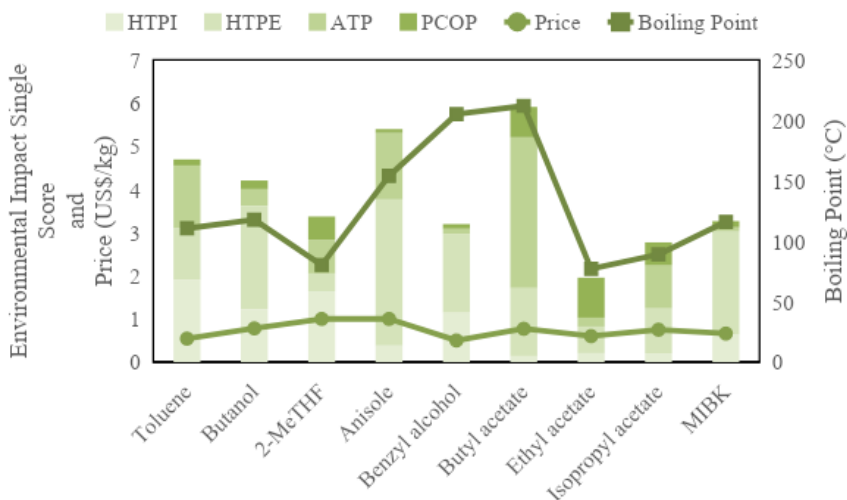


Figure 3 Comparison of toluene against the selected alternatives in terms of their environmental impact single scores (broken down by category), prices, and boiling points

Figure 3 also compares the prices and environmental impact scores (SS) of all the solvents. Although there is no considerable difference in the prices of the solvents,

environmental impact varies quite widely among the candidates. Anisole and butyl acetate have worse single scores than toluene, making them the only candidates that cannot be considered “greener” substitutes for this process. Ethyl acetate and isopropyl acetate perform significantly better than the other alternatives. Considering all the screening factors, ethyl acetate comes out as the best solvent with isopropyl acetate and MIBK doing quite well to warrant a further look into their capability to replace toluene.

4. Conclusion

Solvents are a major source of energy usage and waste generation in many chemical, biochemical, and petrochemical processes, including LL-PTC. In previous works, LL-PTC has been shown to substantially improve the economics and sustainability of offshore natural gas sweetening processes by replacing the existing triazine-based H₂S removal. In this work, the process is further improved by providing a quantum-chemistry-based screening method to find and evaluate green and sustainable alternatives to toluene in terms of technical, economic, and environmental factors. Based on the results, the three best solvents to replace toluene are ethyl acetate, isopropyl acetate, and MIBK. This work paves the way for a more systematic, optimization-based solvent screening or design framework in the future.

Acknowledgements

This work is partly funded by the Danish Hydrocarbon Research and Technology Center and the Sino-Danish Center for Education and Research.

References

- Gani, R., Hytoft, G., Jakssland, C., & Jensen, A. K. (1997). An integrated computer aided system for integrated design of chemical processes. *Computers and Chemical Engineering*, 21(10), 1135–1146. [https://doi.org/10.1016/S0098-1354\(96\)00324-9](https://doi.org/10.1016/S0098-1354(96)00324-9)
- Halgren, T. A., & Lipscomb, W. N. (1977). The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chemical Physics Letters*, 49(2), 225–232. [https://doi.org/10.1016/0009-2614\(77\)80574-5](https://doi.org/10.1016/0009-2614(77)80574-5)
- Hellweg, A., & Eckert, F. (2017). Brick by brick computation of the Gibbs free energy of reaction in solution using quantum chemistry and COSMO-RS. *AIChE Journal*, 63(9), 3944–3954. <https://doi.org/10.1002/aic.15716>
- Mąkosza, M., & Fedoryński, M. (2003). Phase Transfer Catalysis. *Catalysis Reviews - Science and Engineering*, 45(3–4), 321–367. <https://doi.org/10.1081/CR-120025537>
- Plessow, P. (2013). Reaction path optimization without NEB springs or interpolation algorithms. *Journal of Chemical Theory and Computation*, 9(3), 1305–1310. <https://doi.org/10.1021/ct300951j>
- Prat, D., Wells, A., Hayler, J., Sneddon, H., McElroy, C. R., Abou-Shehada, S., & Dunn, P. J. (2015). CHEM21 selection guide of classical- and less classical-solvents. *Green Chemistry*, 18(1), 288–296. <https://doi.org/10.1039/c5gc01008j>
- Pudi, A., Karcz, A. P., Shadravan, V., Andersson, M. P., & Mansouri, S. S. (2020). Modeling of liquid-liquid phase transfer catalysis: process intensification via integration of process systems engineering and computational chemistry. In *Computer Aided Chemical Engineering* (Vol. 48, pp. 43–48). Elsevier B.V. <https://doi.org/10.1016/B978-0-12-823377-1.50008-2>

A Model-Centric Multivariate Analysis for Reducing H₂S Emissions in Wastewater Treatment Plants

Wilfredo Angulo^a, Yris Gonzalez^a, Santiago D. Salas^a, Dany De Cecchis^a and Alexander Espinoza^a

^a *Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Naturales y Matemáticas, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador*

Abstract

Validated mathematical models were studied applying a multivariate analysis of relevant outputs in a wastewater treatment plant. Monte-Carlo based simulations permit to obtain enough responses of the outputs when varying relevant inputs. Uncertainty is incorporated into the flow rates of wastewater and the wind velocity. Results show the strengths and limitations of the models as well as the impact towards the reduction of H₂S. The dynamic variability is captured using a covariance analysis at different time intervals.

Keywords: hydrogen sulfide, mathematical modelling, parameter sensitivity.

1. Introduction

Wastewater treatment plants (WWTP) generate unpleasant odors due to the biological degradation of organic waste. The main influential compound is hydrogen sulfide (H₂S), and it has the potential to provoke the corrosion of wastewater infrastructure, odor nuisance in the community and possible health impact on the personnel. H₂S emissions are easily perceived by the human olfactory system because it detects odors at low concentrations over short time intervals. Growing environmental awareness and the search of better quality of life in urban areas drives the role of environmental control agencies. New policies or the revision of existing ones intend increase the control of odor emissions. In this sense, mathematical models developed for estimating the emission rates of WWTP, represent important decision-making tools when adequately exploited (Santos, et al., 2009)

Mathematical models require the definition of parameters and initial conditions for achieving a solution. Nevertheless, this information is not always available or there is a degree of uncertainty due to natural variations, measurement errors or simply the inability of measuring. Uncertainty and multivariate analyses permit to realistically assess the inherent variations of processes (Marino et al., 2008; Salas et al., 2007; Gonzalez, et al., 2020)

In this work, four validated models are studied applying a dynamic multivariate analysis in a WWTP. Initially, a Monte-Carlo based simulation permits to obtain samples of the outputs when changing five inputs. Uncertainty is incorporated to the flow rate of wastewater and the wind velocity. The results permit to observe the strengths and

limitations of the models as well as the impact towards the reduction of H_2S . Finally, the dynamic variations are captured using a covariance analysis at different time intervals.

2. Background

2.1. System description

The studied system corresponds to the experimental WWTP introduced by (Santos, et al.,2009). This unit treats the wastewater of a town in Brazil, with a fixed daily flow. It consists of a pumping, screening and grit removal station, an upward flow anaerobic sludge mantle reactor (UASB), three secondary biofilters (BF) and a tertiary biofilter, UV disinfection and sludge drying beds as illustrated in Figure 1. Regarding the mathematical model, (Santos, et al.,2009) proposed that the WWTP is treated as an homogeneous reactor, regardless the fluid flow and the mass transfer responsible of the distribution of the pollutant (H_2S) concentration inside. The H_2S can be accumulated, can passed through the reactor, or be affected by any of the following removal mechanism; volatilization of the wastewater-atmosphere interface, stripping because the diffusion into air bubbles, biodegradation, adsorption of solid particles or biomass, absorption for insoluble liquids and by chemical oxidation due to reactor aeration.

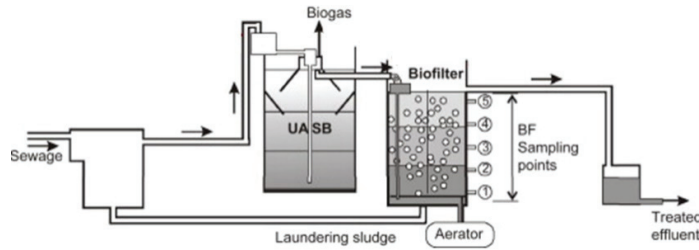


Figure 1: Experimental WWTP. Source: (Santos, et al.,2009)

2.2. Mathematical Model

The mathematical model under study (Corsi et al., 1992; Santos, et al.,2009) considers the WWTP as a single isothermal continuous stirred tank reactor. In this sense, the model is obtained through the mass balance of H_2S . Here, $[H_2S]$ represents the contaminant concentration after being mixed with the fluid inside the unit. The balance provides the initial value as a model: Find the function $[H_2S]: [0, +\infty) \rightarrow \mathbb{R}_+$ such that

$$V \frac{d[H_2S]}{dt}(t) = Q\{[H_2S]_0 - [H_2S](t)\} + \sum_{i \in I} R_i\{[H_2S](t); \theta_i\}; \forall t \in (0, +\infty) \quad (1)$$

subject to an initial condition $[H_2S]_0$ (at $t = 0$) that represents the contaminant concentration in the liquid phase entering the unit. Here, V is the volume of the unit, and Q is the sewage flow rate. R_i represents the constitutive relationship corresponding to the i -th removal mechanism that belongs to the set $I = \{v, s, b, ad, ab, q\}$, where v is the volatilization, s is the stripping, b is the biodegradation, ad is the adsorption, ab is the absorption, and q is the chemical oxidation. On the other hand, θ_i is a vector contains the parameters corresponding to the i -th removal mechanism, which has as many components as parameters appear in R_i . These relationships are taken from (Santos, et al.,2009).

Table 1 enlists the expressions of the chemical removal mechanisms that provide the four different mathematical models. Here, k is the specific rate, k_0 is the constant for oxidation of H_2S , k_1 is the constant for oxidation of HS^- , K_1 is the first dissociation constant for

H_2S , $[\text{O}_2]$, $[\text{S}]$ and $[\text{H}^+]$ are the oxygen, sulfur and hydrogen concentrations respectively, and T is the temperature in $^\circ\text{C}$. The model described in Eq. 1, together with each removal mechanism, is re-scaled and simplified using dimensionless variables and parameters. This provides an adequate handling of the model from the numerical perspective without the effects due to the scale differences. Also, it helps the multivariate analysis on the physical parameters.

Chemical removal mechanisms R_q	Source
$-Vk[\text{S}]^{0.8}[\text{O}_2]^{1.19}$	(Jolley et al., 1985)
$-Vk[\text{S}]^{0.38}[\text{O}_2]^{9.21}$	(Wilmot et al., 1988)
$-Vk[\text{S}]^{0.41}[\text{O}_2]^{0.39\log[\text{S}]}$	(Buisman et al., 1990)
$-\frac{k_0[\text{H}^+] + k_1K_1}{[\text{H}^+] + K_1}[\text{S}]^{0.9}[\text{O}_2]^{0.2}1.06^{(T-20)}$	(Nielsen et al., 2004)

Table 1: Mathematical expressions for the chemical removal mechanisms.

3. Multivariate analysis for a wastewater treatment plant

For a better insight regarding the dynamics of the aforementioned models, Monte-Carlo based simulations are carried out. First, the relationship between the operational variables and the outputs of the models is studied in order to identify those with greater influence on key variables. Second, we perform an analysis including uncertainty at the operational level by varying the possible disturbances of the WWTP, i.e., the wind velocity and the inlet wastewater flow rate. The proposed model-centric framework for the analysis of the different options for wastewater treatment are developed in a Python environment because it allows a versatile implementation. Libraries such as *Numpy*, *SciPy*, *Pandas*, *Matplotlib* and *Seaborn* facilitate obtaining and analyzing the results.

3.1. Monte-Carlo based analysis

To explore particularities of the relationships between the variables involved in the different models, Monte-Carlo based simulations are carried out. We generate a total of random uniform 1,000 samples varying the initial values of the following five variables: the gas phase temperature (T Gas Phase), the liquid phase temperature (Liq Phase), the level of pH, oxygen concentration ($[\text{O}_2]$), and the concentration of sulphur ($[\text{S}]$). The ranges of the initial values used in each variable are T Gas Phase $\in [17.8, 29.8]$, T Liq Phase $\in [13.7, 21.5]$, pH $\in [6.9, 7.5]$, $[\text{O}_2] \in [0.1, 8.5]$, $[\text{S}] \in [5, 300]$. The simulation of the models were run using the *solve_ivp* function available in the library module *scipy.integrate*. The variables and the parameters are properly dimensionless re-scaled, and the simulations run until the dimensionless time of 5.0. This value of time is chosen because the exploratory space of the process variables do not exhibit significant variation on its dynamics.

Once the results are obtained, a multivariate analysis of the output is performed by selecting the time steps $t = 1, 2, \dots, 5$. In this analysis, we look for insights about the relationship between the five operational variables and the reduction of the concentration of H_2S through the time stepping. Correlation matrices between the process variables for each model output are obtained at each time step. The dynamic multivariate analysis permits to observe the influence of certain variables along the dynamic process

simulation, and to determine which of these variables has a greater influence in the outputs.

3.2. Dynamic uncertainty analysis

In this study, uncertainty was incorporated to variables that appear to perform as disturbances and influence the dynamic response of the system. While the operational variables are set as T Gas Phase = 23.8, T Liq Phase = 17.6, pH = 7.2, $[O_2] = 4.3$, and $[S] = 152.5$, the disturbances, including uncertainty, correspond to the wind velocity (Wind V) and the inlet wastewater flow rate (Flux). An uncertainty class of 25% is considered as reasonable based on (Sin et al., 2009). Thus, Wind V = $4.025 \pm 25\%$ and Flux = $5.5 \times 10^{-4} \pm 25\%$. A total of 2,000 samples are randomly uniform generated.

4. Results

4.1. Monte-Carlo based analysis

For the multivariate analysis, 1,000 simulations run for each of the chemical oxidation removal models. The concentration value of H_2S for each operational variable is obtained at each of the chosen time steps.

Figure 2 portrays a series of the plots for the most significant variables, $[O_2]$ and $[S]$ for the model using the chemical removal model by (Jolley et al., 1985) at the times $t = 1, 2, \dots, 5$. The plots indicate a very strong inverse correlation between the $[O_2]$ and the concentration of $[H_2S]$ in each time step, and a direct inverse correlation between $[S]$ and $[H_2S]$.

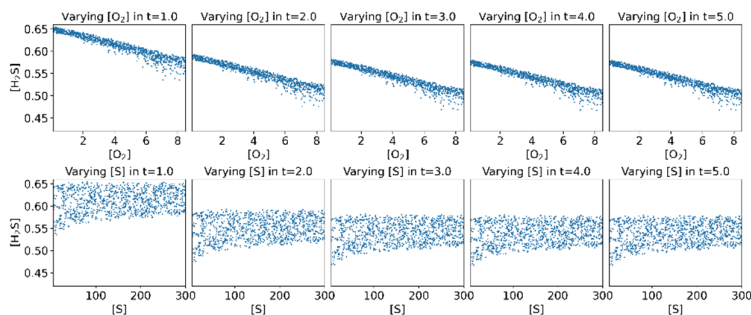


Figure 2: Relationship between $[H_2S]$ and $[O_2]$ (upper plots) and the $[S]$ (lower plots) at the dimensionless time steps 1.0, 2.0, ..., 5.0, using the chemical oxidation removal model by (Jolley et al., 1985).

Table 2 enlists the correlation values between the operational variables at each time step for each chemical oxidation removal model. The values greater or equal to 0.1 are boldfaced. Figure 3 depicts graphically the information given in Table 2. These results permit to identify the variables that have more influence in the reduction of $[H_2S]$ when using a particular chemical removal model. For the utilization of the chemical removal model given by (Buisman et al., 1990) the variable $[O_2]$ and the output $[H_2S]$, in each time step, show a strong inverse correlation (~ -0.78) while $[S]$ and the output $[H_2S]$, in each time step, exhibit a very weak direct correlation. Regarding the use of the model given by (Wilmot et al., 1988) the scenario is similar, excluding that the correlation of $[H_2S]$ with $[O_2]$ becomes weaker while the correlation with $[S]$ stronger. When using the model by (Jolley et al., 1985) the trend is similar to the first one. The correlation gets

inversely stronger (~ -0.96) with respect to $[O_2]$ and directly weaker with $[S]$. Finally, regarding the model introduced by (Nielsen et al., 2004) the results show that the only significant variable in the output is the T Gas Phase, which depicts a very strong inverse correlation with $[H_2S]$.

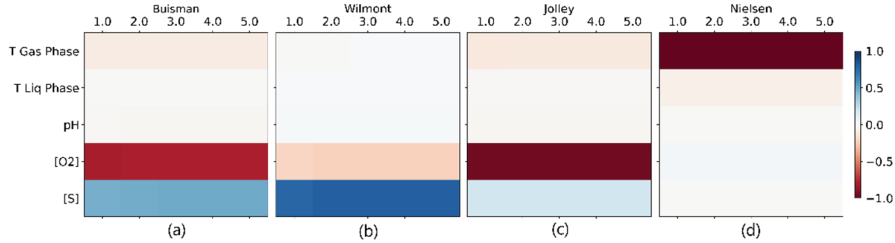


Figure 3: Correlation matrix between parameters values of T Gas Phase, T Liq Phase, pH, $[O_2]$, $[S]$ and the dimensionless time steps 1.0, 2.0, ..., 5.0; for each of the different models used for oxidation chemical removal.

Chemical removal	Buisman					Wilmont					
	Parameters	1.0	2.0	3.0	4.0	5.0	1.0	2.0	3.0	4.0	5.0
T Gas Phase		-0.085	-0.081	-0.08	-0.08	-0.079	-0.013	-0.009	-0.008	-0.007	-0.007
T Liq Phase		-0.013	-0.011	-0.011	-0.011	-0.011	0.001	0.001	0.002	0.002	0.002
pH		-0.023	-0.024	-0.024	-0.024	-0.024	0.009	0.009	0.009	0.009	0.009
$[O_2]$		-0.799	-0.789	-0.784	-0.783	-0.783	-0.21	-0.22	-0.222	-0.222	-0.222
$[S]$		0.464	0.476	0.481	0.482	0.483	0.775	0.798	0.801	0.801	0.801

Chemical removal	Jolley					Nielsen					
	Parameters	1.0	2.0	3.0	4.0	5.0	1.0	2.0	3.0	4.0	5.0
T Gas Phase		-0.103	-0.1	-0.098	-0.098	-0.098	-0.999	-0.999	-0.999	-0.999	-0.999
T Liq Phase		-0.022	-0.022	-0.021	-0.021	-0.021	-0.062	-0.062	-0.062	-0.062	-0.062
pH		-0.028	-0.028	-0.028	-0.028	-0.028	-0.015	-0.015	-0.015	-0.015	-0.015
$[O_2]$		-0.954	-0.956	-0.956	-0.956	-0.956	0.028	0.028	0.028	0.028	0.028
$[S]$		0.176	0.174	0.174	0.174	0.174	-0.009	-0.009	-0.009	-0.009	-0.009

Table 2: Correlation values between the values of parameters of the T Gas Phase, T Liq Phase, pH, $[O_2]$, $[S]$ and the dimensionless time steps 1.0, 2.0, ..., 5.0; for each of the different models used for oxidation chemical removal.

4.2. Dynamic uncertainty analysis

The uncertainty analysis aims to capture the dynamics of the outputs when incorporating variability to process inputs. Figure 4 depicts the results of stochastic simulations, using the four models and by varying $\pm 25\%$ the wind velocity and the wastewater inlet flow. The plots show the minimum and maximum values. The darker region corresponds to the percentiles 25 and 75. The results of the simulations show that by varying 25% the inputs a maximum variation of the $[H_2S]$ of 2.73%, 3.06%, 2.66%, 2.66%, for the simulations using the chemical oxidation removal models of (Buisman et al., 1990), (Jolley et al., 1985), (Wilmot et al., 1988) and (Nielsen et al., 2004), respectively.

5. Conclusions

A model-centric framework permits to exploit the capabilities of mathematical representations of dynamic systems. In this work, a WWTP was studied in detail using different modelling approaches towards the reduction of H_2S emissions. Depending on the final application or available data, the modeller could select a suitable system. The main influential variables correspond to the oxygen concentration, sulfur concentration, and the gas temperature. The oxygen concentration appears to be a reasonable control

variable through the aeration mode, capable of reducing the H_2S emissions. Additionally, the analysis under uncertainty with the selected class portrays important information regarding possible dynamic variations in the outputs. The output variations could be significant to the process and require special attention, specially when satisfying quality standards is desired. In terms of decision-making this provides insightful guidance to select the modelling strategy as well as further considerations for an adequate design and operation.

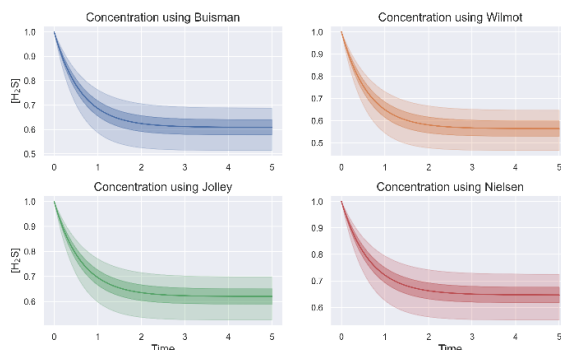


Figure 4: Plots of $[\text{H}_2\text{S}]$ with the variation of the wind velocity and the inlet flux for each of the different models used for oxidation chemical removal.

References

- C. Buisman, P. Uspeert, A. Janssen, G. Lettinga, 1990. Kinetics of chemical and biological sulphide oxidation in aqueous solutions. *Water Research* 24 (5), 667 – 671.
- R. Corsi, D. P. Y. Chang, E. D. Schroeder, 1992. A modeling approach for voc emissions from sewers. *Water Environment Research* 64 (5), 734–741.
- Y. Gonzalez, W. Angulo, D. De Cecchis, M. Lucena, S. D. Salas, 2020. Optimization of a cyclone reactor for biomass hydrolysis through global sensitivity analysis and stochastic optimization. *Computer Aided Chemical Engineering* 48, 667–672.
- A. Haaning Nielsen, J. Vollertsen, T. Hvitved-Jacobsen, 2004. Chemical sulfide oxidation of wastewater - effects of pH and temperature. *Water Science and Technology* 50 (4), 185–192.
- R. A. Jolley, C. F. Forster, 1985. The kinetics of sulphide oxidation. *Environmental Technology Letters* 6 (1-11), 1–10.
- S. Marino, I. B. Hogue, C. J. Ray, D. E. Kirschner, 2008. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology* 254 (1), 178 – 196.
- S. Salas, A. Geraili, J. Romagnoli, 2017. Optimization of renewable energy businesses under operational level uncertainties through extensive sensitivity analysis and stochastic global optimization. *Industrial & Engineering Chemistry Research* 56 (12), 3360–3372.
- J. M. Santos, E. S. Lopes, N. C. Reis Junior, L. M. de Sá, N. J. Horan, 2009. Mathematical modelling of hydrogen sulphide emission and removal in aerobic biofilters comprising chemical oxidation. *Water Research* 43 (14), 3355 – 3364.
- G. Sin, K. V. Gernaey, M. B. Neumann, M. C. van Loosdrecht, W. Gujer, 2009. Uncertainty analysis in wwtp model applications: a critical discussion using an example from design. *Water Research* 43 (11), 2894–2906.
- P. D. Wilmot, K. Cadee, J. J. Katinic, B. V. Kavanagh, 1988. Kinetics of sulfide oxidation by dissolved oxygen. *Journal (Water Pollution Control Federation)* 60 (7), 1264–1270.

Gas-to-liquid strategy for liquid fuel production from industrial residue gases: Process development and technoeconomic analysis

Thai Ngan Do,^a Chanhee You,^a Jiyong Kim^{a,*}

^a Department of Energy and Chemical Engineering, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon, 22012, South Korea

jkim77@inu.ac.kr

Abstract

This study aims to develop and evaluate new gas-to-liquid technologies for production liquid fuels, such as methanol, synthetic fuels, and dimethyl ether, using industrial residue gases. To resolve global energy and environmental issues, the upcycling of a coke oven gas (COG) which contains high-caloric chemicals of CH₄, CO, H₂ to value-added chemicals is one of the most promising strategies. In achieving the goal, we first developed for the upcycling processes of the residue gases to value-added chemicals and fuels (e.g., methanol, dimethyl ether, and Fischer-Tropsch fuels) through the gas-to-liquid framework. The COG-to-fuel processes involve three sub-processes: gas pre-treating, liquid fuels synthesis, and separation. In particular, the combined steam and CO₂ reforming (CSCR) was adopted to reform COG into H₂ and CO which is further adjusted to the optimal ratio for target products before entering the following synthesis stage. Here, the H₂/CO ratio of 2 is useful for methanol synthesis and Fischer-Tropsch synthesis while the ratio of 1 is better for dimethyl ether synthesis. We then developed process models using Aspen Plus simulator and performed a techno-economic evaluation to obtain technical (mass and energy flows) and economic (capital and operating costs) parameters. As a result, we identified the most energy-efficient, economic, and eco-friendly gas-to-fuel process under different market scenarios such as chemical prices and CO₂ reduction policies.

Keywords: Gas-to-liquid, Coke-oven gas, Methanol, Fischer-Tropsch, Dimethyl ether, Technoeconomic.

1. Introduction

The depletion of conventional energy resources, the growth of energy demand and climate change are the critical energy and environmental issues. To resolve such issues, many alternative technological routes for energy products from alternative feedstocks have been investigated. Recently, the reuse of industrial residue gas (e.g., coke-oven gas, Lintz-Donawitz) for heating or fuel products have received a lot of attention. A number of studies have been found for the upcycling of residue gases to high value-added chemicals and fuels (Kim et al, 2020). Therein, coke-oven gas (COG), which is one of the residue gases from the iron and steel making industry with high-caloric chemicals of CH₄, CO, H₂ can be utilized as a raw material for chemicals and fuels. In this study, we developed and evaluated the upcycling processes of coke-oven gas (COG) to liquid fuels such as methanol, dimethyl ether, and Fischer-Tropsch fuels through the gas-to-liquid

framework. The COG-to-fuel processes were firstly developed with three main sections of combined steam and CO₂ reforming for syngas, fuel synthesis, and separation. Therein, the combined steam and CO₂ reforming (CSCR) is flexible to adjust H₂/CO ratio for the specific following fuel synthesis (the ratio of 2 for methanol synthesis and Fischer-Tropsch synthesis, and 1 for dimethyl ether synthesis). We then evaluated a techno-economic (e.g., carbon and energy efficiency, unit production cost) and environmental performance (net CO_{2eq} emission). Consequently, the most energy-efficient, economic, and eco-friendly fuel production process were identified with the insight of major cost- and energy-drivers of COG-to-fuels to provide practical R&D targets and strategies.

2. Process simulation and analysis method

2.1. Process development

In this study, three process models were developed using Aspen Plus V10.0 according to the targeted products: COG-to-MeOH, COG-to-DME, and COG-to-FT. The proposed gas-to-fuel processes include three main stages of: 1) the pre-processing of a raw gas for syngas production, 2) synthesis to liquid fuels from syngas, and 3) separation and purification. Firstly, the combined steam and CO₂ reforming (CSCR) was adopted to reform CH₄ in COG into H₂ and CO, in which the H₂/CO ratio is flexible by adjusting the competitive steam reforming and CO₂ reforming reactions. Particularly, the methanol synthesis and Fischer-Tropsch synthesis prefers the syngas ratio of 2, while the favourable ratio between H₂ and CO for maximizing the dimethyl ether synthesis is assumed to be 1. The produced syngas is then fed into a reactor for the targeted fuels (MeOH, DME, or FT fuel):

- Methanol synthesis: catalyst of Cu/Zn/Al/Zr, at 250 °C, 50 bar (Do et al., 2019)
- Dimethyl ether synthesis: bi-function catalytic of CuO/ZnO/Al₂O₃ and γ -Al₂O₃, at 260 °C, 50 bar (Mevawala et al., 2017)
- Fischer-Tropsch synthesis: Ni/CaO/A₂O₃, at 210 °C, 38 bar (Han et al., 2019)

In the separation section, the outlet stream from reactors was cooled and separated the main product while the unreacted gas was recycled. For COG-to-MeOH, the simple flash tank and distillation was used to separate MeOH from water and recycle unreacted CO, H₂, CO₂. In COG-to-DME, the Restisol absorption unit was firstly operated to separate CO₂, which is produced together with DME. Then, the distillation column was used to purify DME from a small by-product MeOH. Finally, the flash tank and distillation were also adopted in COG-to-FT to separate liquid fuels (e.g., gasoline, diesel, and wax), from the water and light gas (sent to the reformer).

2.2. Analysis method

The COG-to-fuel processes were evaluated the techno-economic and environmental performance using many criteria. Firstly, the carbon efficiency (*CEF*) and energy efficiency (*EEF*) show how much carbon and energy were captured and utilized in the produced fuels (Do et al., 2020) as formulated in Eqs. (1) and (2).

$$CEF(\%) = \frac{\text{Carbon in product}}{\text{Carbon in feed}} \times 100 \quad (1)$$

$$EEF(\%) = \frac{\text{Heat of product}}{\text{Heat of feed} + \text{Utility}} \times 100 \quad (2)$$

For economic analysis, unit production cost (*UPC*) was calculated for fuels as expressed in Eq. (3). Here, the total production cost includes the capital expenditures (*CAPEX*) and operating expenses (*OPEX*) (Do el et al, 2020). The environmental performance of processes is performed in net CO₂ equivalent emission (*NEC*) which is expressed as the amount of CO₂ equivalent (CO_{2eq}) emit per unit of produced fuel, as presented in Eq. (4).

$$UPC(\$/\text{kg}) = \frac{CAPEX + OPEX}{\text{Amount of product}} \quad (3)$$

$$NCE(\text{kg}/\text{kg}) = \frac{\text{Direct emission} + \text{Indirect emission} - \text{Feed Inventory}}{\text{Amount of product}} \quad (4)$$

3. Techno-economic analysis

3.1. Process simulation and technical performance

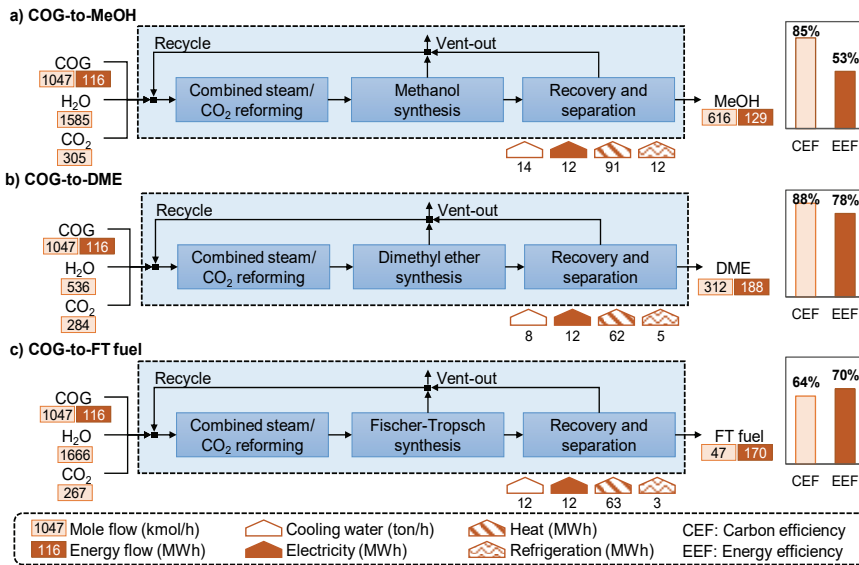


Figure 1. Block diagram with carbon efficiency and energy efficiency of COG-to-fuel processes

The process simulation of MeOH, DME, and FT fuel production was targeted to operate with the same feed rate (1,047 kmol/h of COG) and additional CO₂ and water, which are at different rates for adjusting the optimal syngas ratio, corresponding to the capacity of 158, 114, and 51 kt/y. Figure 1 presents the simplified block diagram with the mass and energy flows of each process as well as the carbon and energy efficiency. As shown in Figure 1, MeOH and DME production processes show relative high CEF of 85% and 88%

respectively, while the DME and FT fuel production routes was analysed to show high energy efficient of 78% and 70%, respectively.

3.2. Unit production cost

In the economic evaluation, the fuel production plants were assumed at annual 8,000-hours operability, 20 years of plant-life, and 8% of interest rate. The other assumptions and parameters for material and utility are presented in Table 1. Based on the obtained mass and energy flow information, the capital and operating costs were estimated using Aspen Process Economic Analyzer. The total production cost and UPC of fuels were estimated and presented in Table 2. The COG-to-MeOH is the highest expense plant. But due to the highest production rate (114 kt/y) compared to COG-to-DME and COG-to-FT plants, the MeOH production plant results in the lowest UPC for MeOH (0.46 \$/kg). To compare three different types of fuels, the UPC was further estimated into the unit of energy as gasoline gallon equivalent (\$/GGE). Consequently, the DME production process performs the lowest UPC at 1.92 \$/GGE.

Table 1. Assumption of raw material and utility price and CO_{2eq} inventory

	Unit price		CO _{2eq} inventory	
Coke-oven gas	2.5	\$/MMBTU	2.05	kgCO _{2eq} /MMBTU
CO ₂	35	\$/ton	1.00	kgCO _{2eq} /kg
Process water	1	\$/ton	0	kgCO _{2eq} /kg
Cooling water	0.03	\$/ton	0	kgCO _{2eq} /kg
Electricity	0.07	\$/kWh	0.62	kgCO _{2eq} /kWh
Fired heat	0.03	\$/kWh	0.34	kgCO _{2eq} /kWh
High pressure steam	14.5	\$/ton	0.19	kgCO _{2eq} /kWh
Low pressure steam	10.5	\$/ton	0.19	kgCO _{2eq} /kWh
Refrigeration	0.067	\$/kWh	1.50	kgCO _{2eq} /kWh

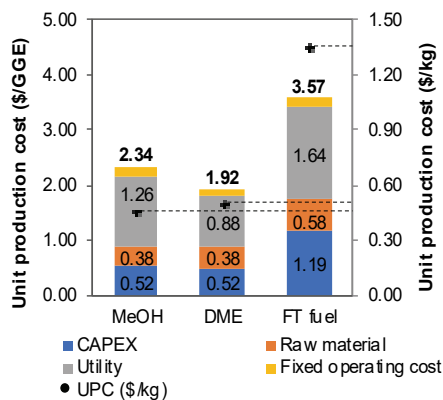


Figure 2. Breakdown of fuel unit production cost

Table 2. Total production cost and unit production cost of fuels

	COG-to-MeOH	COG-to-DME	COG-to-FT
<i>Capital expenditure (M\$/year)</i>			
(CAPEX=DC+IDC+WC)	16,169,041	15,548,946	22,866,501
Direct plant cost (DC)	9,441,496	9,079,408	13,352,307
Indirect plant cost (IDC)	4,816,310	4,631,601	6,811,298
Working capital (WC)	1,911,234	1,837,937	2,702,896
<i>Operating expenditure (M\$/year)</i>			
(OPEX=VOC+FOC)	56,277,886	42,300,327	45,620,557
Variable operating cost (VOC)	50,928,699	38,129,049	42,443,363
Raw material (RM)	11,892,086	11,481,053	11,059,361
COG	7,902,695	7,902,695	7,902,695
CO ₂	3,761,020	3,501,092	2,916,526
Process water	228,371	77,266	240,140
Utility (UT)	39,036,614	26,647,996	31,384,002
Cooling water	3,580,254	2,090,380	3,175,849
Electricity	6,358,806	6,109,776	4,141,934
Fired heat	20,820,159	13,048,426	21,369,119
High-pressure steam	0	2,192,372	1,136,703
Low-pressure steam	1,473,290	0	0
Refrigeration	6,804,105	3,207,042	1,560,397
Fixed operating cost (FOC)	5,349,187	4,171,278	3,177,194
Total production cost (M\$/year)	72,446,927	57,849,272	68,487,058
<i>Main Product</i>			
	<i>MeOH</i>	<i>DME</i>	<i>FT fuel</i>
Capacity (kg/y)	158,341,168	114,267,039	50,633,972
UPC (\$/kg)	0.46	0.51	1.35
Capacity (GGE/y)	30,946,586	30,125,292	19,160,394
UPC (\$/GGE)	2.34	1.92	3.57

Figure 2 presents the cost breakdown of the three fuel production plants with CAPEX, raw material cost, utility cost, and fixed operating cost. The lowest UPC of the DME production plant is primarily due to the lower utility cost which is the dominant component at around 45% over the others. Then, CAPEX is the second largest factor with 27-33% of UPC, followed by raw material cost at 16-20%. There are opportunities to lowering UPC of fuel with lower utility and raw material cost. Since COG feedstock is one of industrial residue gases, it is potential to receive incentive at zero cost.

3.3. Net CO_{2eq} emission

Table 3. Environmental performance of COG-to-fuels

	COG-to-MeOH	COG-to-DME	COG-to-FT
Feed inventory (kgCO _{2eq} /kg _{fuel})	0.72	0.93	1.77
Indirect CO _{2eq} emission (kgCO _{2eq} /kg _{fuel})	2.49	2.03	5.12
Direct CO _{2eq} emission (kgCO _{2eq} /kg _{fuel})	0.01	0.09	0.05
NCE (kgCO_{2eq}/kg_{fuel})	1.78	1.19	3.40

The net CO_{2eq} emission estimated to one unit of fuels as presented in Table 3. The indirect emission is the largest contribution to NCEs, which corresponds to the high consumption of utility in the processes. COG-to-DME process emits lower CO_{2eq} during its operation over COG-to-MeOH and COG-to-FT.

4. Conclusions

This study proposed new residue gas upcycling processes, and evaluated the technical, economic and environmental capability for the MeOH, DME, and FT fuel production. The COG-to-DME process achieved the highest technical metrics at 88% carbon efficiency, 78% of energy efficiency with the lowest cost at 1.92 \$/GGE, and the best environmental performance with NCE at 1.19 score. In addition, the major cost- and energy-drivers of COG-to-fuels was identified through the techno-economic evaluation. The strategy to lower utility cost and residue gas price is very helpful to achieve improved economic benefits of COG-to-fuels processes. As a future work, the sensitivity analysis using various scenarios, such as different market prices and CO₂ tax, can be performed to determine the debottleneck strategies.

Acknowledgment

This work was supported by...

References

- S. Kim, J. Kim, 2020, The optimal carbon and hydrogen balance for methanol production from coke oven gas and Linz-Donawitz gas: Process development and techno-economic analysis, *Fuel*, 266.
- J. Kim, T. A. Johnson, J. E. Miller, E. B. Stechel, C.T. Maravelias, 2012, Fuel production from CO₂ using solar-thermal energy: System level analysis. *Energy Environ. Sci.* 5, 8417–8429.
- T. N. Do, J. Kim, 2019, Process development and techno-economic evaluation of methanol production by direct CO₂ hydrogenation using solar-thermal energy, *J CO₂ Util*, 33, 461-472
- C. Mevawala, Y. Jiang, D. Bhattacharyya, 2017, Plant-wide modeling and analysis of the shale gas to dimethyl ether (DME) process via direct and indirect synthesis routes. *Appl. Energy* 204, 163–180.
- S. Han, S. Kim, Y. T. Kim, G. Kim, J. Kim, 2019, Optimization-based assessment framework for carbon utilization strategies: Energy production from coke oven gas, *Energy Convers Manag*, 187, 1–14.
- T. N. Do, J. Kim, 2020, Green C2-C4 hydrocarbon production through direct CO₂ hydrogenation with renewable hydrogen: Process development and techno-economic analysis. *Energy Convers Manag*, 214, 112866.

Multi-Objective Optimization of Lignocellulosic Ethanol Production Based on Regional Land and Water Availability

Rashi Dhanraj^a, Varun Punnathanam^a, Yogendra Shastri^{a,*}
^a*Indian Institute of Technology Bombay, Mumbai 400076, India*
yshastri@che.iitb.ac.in

Abstract

The production design for lignocellulosic biofuels should incorporate both environmental and economic objective. This work suggests a multi-objective optimization model design for ethanol production, which includes the economic goal balanced by limited regional land and water availability. The allocation of agricultural land to selected crops is the key decision variable. The model has four objectives, namely, minimization of ethanol cost, minimization of the water footprint of ethanol, minimization of irrigation water and maximization of farmers' profit. The resulting multi-objective mixed-integer linear programming problem was applied to a case study of 33 districts in Maharashtra, India. Results indicated a significant trade-off between ethanol cost and irrigation water requirement. Sorghum, wheat, and cotton were the prominent crops recommended by the model. Ethanol cost varied from ₹ 47/L (\$0.63/L) to ₹ 81/L (\$ 1.10/L), while water footprint varied from 674 L/L to 301 L/L, irrigation water varied from 43 m³/ha to 2768 m³/ha, and farmers' profit from ₹ 21059/ha (\$ 281/ha) to ₹ 4,770 /ha (\$ 64.51/ha).

Keywords: Optimization, lignocellulosic biofuels, land availability, water availability.

1. Introduction

The annual surplus agricultural residue produce in India is approximately 234.5 million Mg (Hiloidhari et al., 2014). The potential in India to produce lignocellulosic bio-fuels by using agricultural residue has been studied in context with socio-economic impact (Purohit and Dhar, 2018). The studies of lignocellulosic biofuel were further extended to include environmental impact. Formulation of multi-objective optimization problems gives a better understanding of the trade-off between environmental and economic impact (Mandade and Shastri, 2019). The agricultural phase in the life cycle of biofuel is an important phase in land use and water footprint. The land and water resource availability are limited, thus incorporating them in the life cycle assessment of bio-fuel production could give better understanding of the long-term sustainability of bio-fuel production. However, the water footprint of agricultural residue does not give the actual water withdrawal in the region; similarly, the economic interest of bio-refinery is not enough. Thus, water consumed by crops and the economic interest of farmers has to be considered in agricultural planning. This work formulates a mixed-integer linear programming (MILP) problem considering four objectives and regional water availability constraints. The four objectives of the model are based on the factors mentioned above. The bio-refinery model considered in this work was developed earlier in the literature (Punnathanam and Shastri, 2020), and has been modified so that the model would suggest an optimum agricultural pattern. Water consumed at the

biorefinery is negligible compared to water consumed during the agricultural phase, and thus has been neglected. The article is arranged as follows. The second section details the optimization problem formulation. The third section describes the case study, followed by results and discussion in Section 4, and conclusion in Section 5.

2. Modeling framework

The scope of the model has been represented in the schematics in Figure 1. The model considers the balance between water consumed by crops and the regional availability of water during different seasons, as well as the economic aspect of ethanol production and agriculture. The time period is 1 year, and the water availability varies with the season; thus, the time period was divided into three seasons: kharif, rabi and zaid (summer) season. The constraints of the model are bio-fuel demand, and regional availability of land and water resources. These constraints are discussed in detail later. Rainfall and groundwater were the only sources of water considered in the model, while surface water was excluded.

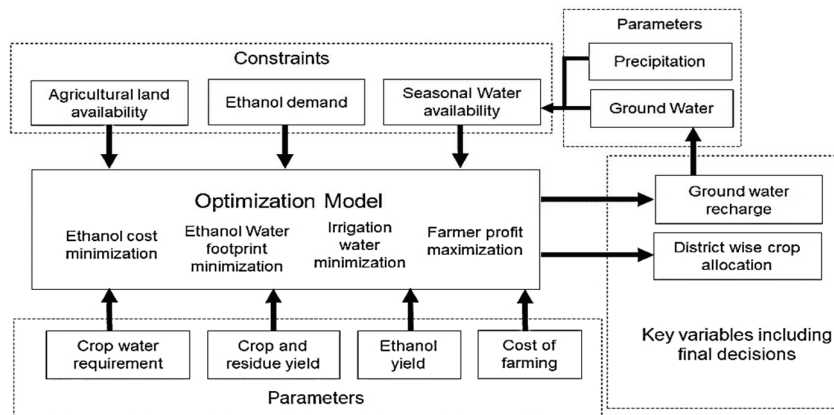


Figure 1: Schematics representing the scope of the model.

The water balance was done at district level (equivalent to a county). The crop water requirement (CWR) was calculated before the formulation of the model. CROPWAT 8.0, model developed by FAO (Clarke et al., 2000) was used for calculation of CWR. Rain is the primary source of water for crops, in case the rainwater is not sufficient irrigation is used to meet the water requirement. The part of rainfall consumed by crops is effective rainfall, rest of the rainwater either percolates deep in the ground or flows away as runoff (Allen, R. G. et al., 1998). The groundwater is replenished every year through rainfall and seepage to the canal, reservoir and dams. The groundwater recharged through sources other than rainwater was considered constant in the model. Data on soil moisture, runoff and groundwater recharge was obtained from the WaterBalance app developed by PoCRA moisture (Wankhade et al., 2019). It was assumed that the groundwater recharge water from one district doesn't flow to another district. Recharge from agricultural land was the only recharge considered in this work. The data generated for crop water requirement from CROPWAT 8.0 and groundwater recharge data from Waterbalance was used in the optimization model. The optimization model formulation is depicted in Subsection 2.1.

2.1 Optimization model formulation

The mixed-integer linear programming problem was formulated to solve the multi-objective optimization problem in General Algebraic Modelling System (GAMS) version 24.9.2 and solved using the CPLEX® MILP solver (Robichaud, 2010). This section discusses the constraints, decision variable and the objective functions of the model.

2.1.1 Constraints and decision variable

The primary decision variable in this model is $x(i, j, k)$, which the agricultural land is allocated to crops. Here index i , j , and k denotes the location of farm, crops and season, respectively. The first constraint in the model is ethanol demand. The ethanol demand is calculated by multiplying population of the region under study with per capita petrol consumption and the percentage blending of petrol with ethanol. The amount of agricultural residue converted to ethanol is denoted by $Re s_b(j, k, l)$, where index l indicates the location of bio-refinery.

$$E_D(k, l) = \sum_j Re s_b(j, k, l) \times x(i, j, k) \times E_c(j) \quad (1)$$

Here, $E_D(k, l)$ is the ethanol demand in of a bio-refinery at the location l , during the season k . $E_c(j)$ is the volume (L) of ethanol produced per unit mass (t) of feedstock. The constraint on land allocation is formulated as follows:

$$\sum_j x(i, j, k) \leq A_L(i, k) \times A_p(i, k) \forall i, k \quad (2)$$

Here, $A_L(i, k)$ is agricultural land available in the district i during the season k . The agricultural land available in a district is used for growing vegetables and fruits as well. Thus, $A_p(i, k)$ the is introduce in the equation. $A_p(i, k)$ Is fraction of total agricultural land allocated to crops. The water availability varies with the season, so does the water demand as crops are planted in different seasons and have different growth period. To capture this trend, water balance is done for different seasons. The water balance during kharif season is as follows:

$$\sum_j ER(i, j, k_n) \times x(i, j, k_n) \leq \{R(i, k_n)(1 - \delta)\} - \sum_j GWR(i, j, k_n) \times x(i, j, k_n) \quad (3)$$

Here, $ER(i, j, k_n)$ is the rainwater consumed by crop j in the district i during the kharif season (k_n), the subscript n denotes kharif. The rainwater is represented by $R(i, k_n)$, here δ represents the fraction of runoff, while $GWR(i, j, k_n)$ is the groundwater recharge. The groundwater balance is shown in Eq. (4)

$$\sum_j IR(i, j, k_n) \times x(i, j, k_n) \leq GW(i, k_n) + N(i, k_s) \quad (4)$$

Here, $IR(i, j, k_n)$, is irrigation water required by crops during the kharif season, $GW(i, k_n)$, is the groundwater available is kharif season. The groundwater which was not used during the kharif season and the groundwater recharge were passed down to the consecutive season. Here, $N(i, k_s)$ represents the water passed down from season before kharif season. Similarly, water balance is done for rabi and zaid seasons.

2.1.2. Multi-Objective optimization

There are four objectives in this multi-objective optimization model. The first objective is minimization of ethanol cost, and is formulated as follows:

$$\text{Min } J_C = \sum_{i,j,k} C_E(i, j, k) \quad (5)$$

Here, $C_E(i, j, k)$ is the cost of ethanol production. Cost of ethanol includes the cost of residue, cost of transport, cost of ethanol production process and storage cost. The objective of maximization of profit earned by farmers is formulated as follows.

$$\text{Max } J_P = \sum_{i,j,k} P_F(i, j, k) \quad (6)$$

The profit earned by farmers $P_F(i, j, k)$ the difference between the price of agricultural product and residue to the cost of cultivation. The third objective, minimization of water footprint of ethanol, is formulated as follow:

$$\text{Min } J_{WF} = \sum_{i,j,k} WF_R(i, j, k) \quad (7)$$

Here, $WF_R(i, j, k)$ is the water footprint of ethanol produced from different agricultural residue. The fourth objective, minimization of irrigation water consumption is formulated as follows:

$$\text{Min } J_{IR} = \sum_{i,j,k} IR(i, j, k) \times x(i, j, k) \quad (8)$$

Here, $IR(i, j, k)$ is the irrigation water required by the crop j in m^3/ha , which is multiplied by land allocated to the crop. The multi-objective optimization was solved using the ε -constraint method. In this method except for one objective function, all the objective functions are formulated as constraints, as shown in Eq. (9).

$$\begin{aligned} & \text{Min } J_C \\ \text{Subject to } & J_P \geq \varepsilon_1 \\ & J_{WF} \leq \varepsilon_2 \\ & J_{IR} \leq \varepsilon_3 \end{aligned} \quad (9)$$

Here, ε_1 , ε_2 , and ε_3 are the bound set on various objective functions. This single objective optimization problem is solved for various combinations of these bounds to obtain the trade-offs between different objective functions.

3. Case study of Maharashtra, India

The model was applied to the 33 districts of Maharashtra state (Mumbai city, Mumbai Suburban and Palghar districts were excluded from the study). The Maharashtra state has intensive water demand from the agricultural sector and suffers from water scarcity (Aayog et al., 2017). The crops chosen for the study were sugarcane, wheat, rice, cotton, and sorghum. The groundwater recharge data for the Solapur district was calibrated for different crops. Based on the data for the Solapur district, the correlation between rainfall and groundwater recharge was obtained using the linear regression method. This correlation was used for all districts. Results are discussed in Section 4.

4. Result and discussion

The model was run for each single-objective optimization followed by multi-objective optimization. For each single-objective optimization run, the number of constraints is 37,894, variables are 37,063, and the binary variables are 33. The values of the water footprint of biofuel, farmer's profit and irrigation water obtained from single-objective optimization runs was used to obtain the 20 values of ε_1 , ε_2 , and ε_3 . Thus, 8,000 independent single-objective optimization problems are solved to get the Pareto-front of multi-objective optimization. In order to make it easy to understand the results of the multi-objective optimization, the trade-off between irrigation water requirement and farmer's profit was explained at water footprint bound to various values. The model gave feasible solution for water footprints ranging from 674.28 to 398.24 L/L.

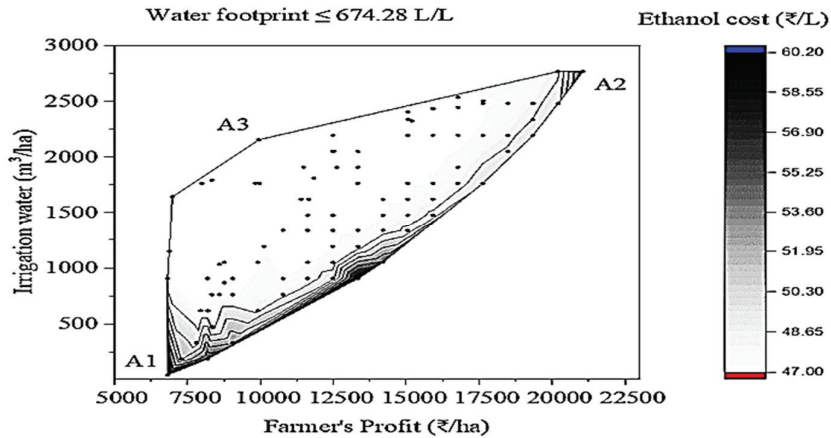


Figure 2. Plots of trade-offs between irrigation water utilization, farmers' profit and ethanol cost when water footprint ≤ 674.28 L/L

In this paper, we had discussed the Pareto-front when water footprint was bound to 674.28 L/L, as shown in Figure 2. The irrigation water was minimum at point A1, farmer's profit was maximum at point A2, and the ethanol cost was minimum at A3.

When the irrigation water requirement was minimum, 17% and 12% more land was allocated to sorghum than what was allocated when the farmer's profit was maximum and the ethanol cost was minimum, respectively. The profit earned by farmers depends on several factors such as the cost of cultivation, the price of the main product, as well as the price of residue. Wheat stalk and sorghum grain have relatively high prices, and thus wheat and sorghum accounted for about 28% and 47% of the total agricultural land utilized at point A2. The land was allocated to cotton to minimize the ethanol cost, but the cotton stalk is not enough to meet ethanol demand; thus, wheat stalk and sorghum stalk are used in the biorefinery. Wheat stalk reduces ethanol cost, while sorghum allows groundwater recharge, which passes down to the consecutive season, thus providing water for wheat production.

5. Conclusion

Amongst all the crops model showed favourability towards sorghum, stress on groundwater was less, also, ethanol yield and cost of sorghum stalk are moderate. The incorporation of regional water and land resource availability gave a better understanding of energy and water nexus. The model showed economic benefits earned by farmers comes at the cost of water stress in the region; thus, a balance has to be created between the two. The challenges associated with land allocation could further be studied by including flow across the district and surface water in the model.

Reference

- Aayog, N., Energy, T., Delhi, N., 2017. Study on Assessment of Water Foot Prints of India ' s Long Term Energy Scenarios 1–101.
- Allen, R. G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration guidelines for computing crop water requirements., FAO Irrigation & drainage Paper 56. FAO, Food and Agriculture Organization of the United Nations, Roma.
- Clarke, D., Smith, M., El-Askari, K., 2000. CropWat for Windows : User guide. FAO. Roma 43.
- Hiloidhari, M., Das, D., Baruah, D.C., 2014. Bioenergy potential from crop residue biomass in India. *Renew. Sustain. Energy Rev.* 32, 504–512. <https://doi.org/10.1016/j.rser.2014.01.025>
- Mandade, P., Shastri, Y., 2019. Multi-objective optimization of lignocellulosic feedstock selection for ethanol production in India. *J. Clean. Prod.* 231, 1226–1234. <https://doi.org/10.1016/j.jclepro.2019.05.311>
- Punnathanam, V., Shastri, Y., 2020. Efficient optimization of a large-scale biorefinery system using a novel decomposition based approach. *Chem. Eng. Res. Des.* 160, 175–189. <https://doi.org/10.1016/j.cherd.2020.05.023>
- Purohit, P., Dhar, S., 2018. Lignocellulosic biofuels in India: Current perspectives, potential issues and future prospects. *AIMS Energy* 6, 453–486. <https://doi.org/10.3934/energy.2018.3.453>
- Robichaud, V., 2010. An Introduction to GAMS 1. Matrix 1–14.

Deep Learning and AutoML for Dynamic Modeling of LNG Regasification Process Using Seawater

Yongbeom Shin,^a Dongil Shin^{a,b,*}

^a *Department of Chemical Engineering, Myongji University, Yongin, Gyeonggido 17058, Korea*

^b *Department of Disaster and Safety, Myongji University, Yongin, Gyeonggido 17058, Korea*
dongil@mju.ac.kr

Abstract

Around the world, 70% of the NG gasification volume of LNG regasification terminal is processed through ORV. First principle-based modeling studies have been conducted to improve the heat exchange efficiency and optimize the operation, but ORV is equipment in which the heat transfer coefficient is irregular according to time and space, and it undergoes a complex modeling process. In this study, FNN, LSTM, and AutoML-based modeling that predicts dynamic changes in NG and seawater discharged temperatures according to changing operating condition in ORV using actual operation data were performed to confirm the effectiveness of data-based modeling for complex systems. The MSE-based prediction accuracy was insignificant between LSTM and AutoML, and FNN showed twice the error in prediction. The performance of AutoML was superior to the manually developed FNN, and the training time took 27 times longer than the LSTM, but the total time required for model development was 1/25, which demonstrates the possibility of automated model design using AutoML, without manual intervention. The prediction of the NG and seawater discharge temperatures using LSTM and AutoML showed an error of less than 0.5K. The real-time optimization of the amount of LNG vaporization that can be processed through ORV in winter is performed with the LSTM model. It is possible to additionally process 19.7% of LNG compared to the previous operation, and by the effect of reducing operation cost, it is suggested that the optimal operation guideline of ORV can be provided using the developed dynamic prediction model.

Keywords: Machine learning, dynamic modeling, operations decision support, AutoML.

1. Introduction

The open rack vaporizer (ORV) of the LNG regasification process uses seawater as a heat source and is the most widely used due to its advantages such as low operating cost, ease of maintenance, and safety. In winter, however, when seawater temperature is below 5°C, which is the minimum requirement for ORV operation, a submerged combustion vaporizer (SCV) with a high operating cost is used to achieve the target NG production. If real-time optimization for more amount of LNG that can be vaporized through ORV in winter is possible, it reduces process operating cost by minimizing the use of SCV.

First principle-based modeling studies have been performed to improve the heat exchange efficiency of ORV and optimize the operating parameters. Wang et al. (2013) and Jin et al. (2014) performed heat exchange modeling reflecting the icing phenomenon of the tube

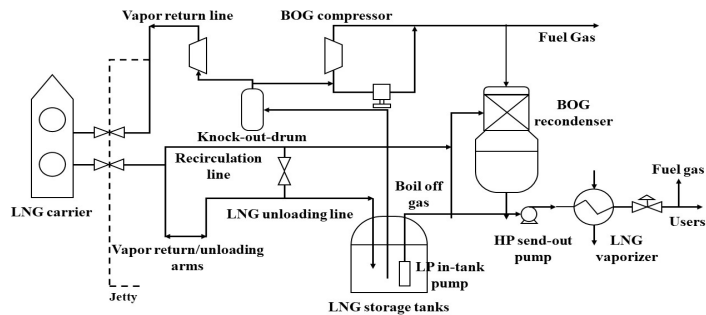


Figure 1. Onshore LNG regasification process (adapted from D'alessandro et al. 2016).

surface to present Super ORV design standards. Cheng et al. (2019) modeled the heat exchange phenomena to present the criteria for ORV design and analyzed the effect of operating parameters. However, first principle-based heat exchange modeling in ORV experiences a complex process because of the phenomenon of freezing seawater on the tube surface by -160°C LNG flowing and forming an irregular film on the tube surface, resulting in irregular heat transfer coefficients according to time and location.

In the case of a complex system, the first principle model contains many number of equations for high accuracy, which makes it tough to predict promptly and accurately. However, an artificial neural network, which reflects nonlinearity and has a low computational load, with the desired level of accuracy (Singh et al. 2013).

In this study, we developed a prediction model based on Feed-forward neural network (FNN), Long short-term memory (LSTM), and AutoML that predicts dynamic changes in NG and seawater discharge temperatures using actual operation data. The purpose of this study is to evaluate the feasibility of selecting an algorithm suitable for processing time series data by comparing the performance, including AutoML, automatically generating an optimal machine learning model for field engineers who are non-data science experts. Through real-time optimization of more amount of LNG vaporization that can be processed with ORV based on the suggested predictive model, especially for winter when the minimum operating conditions of ORV are not satisfied, we evaluate the applicability of the dynamic prediction model to reduce the operating cost of the LNG regasification process and provide operations decision support.

Table 1. Dynamic modeling results of FNN, LSTM and AutoML.

	Number of datasets	Network structure (L: LSTM layer, B: batch normalization layer)	Number of parameters	Test MSE	Training time (s)
FNN	1	320(input)-30-B-30-B-10(output)	10,990	0.000315	117
	2	320(input)-30-B-30-B-10(output)	10,990	0.000485	109
	3	320(input)-30-B-30-B-10(output)	10,990	0.000263	111
	4	320(input)-30-B-30-B-10(output)	10,990	0.000315	108
	5	320(input)-30-B-30-B-10(output)	10,990	0.000624	111
	Average			0.000400	111
LSTM	1	8(input)-4L-50-B-10(output)	1,068	0.000187	296
	2	8(input)-4L-50-B-10(output)	1,068	0.000183	300
	3	8(input)-4L-50-B-10(output)	1,068	0.000211	299
	4	8(input)-4L-50-B-10(output)	1,068	0.000218	283
	5	8(input)-4L-50-B-10(output)	1,068	0.000236	273
	Average			0.000207	290
AutoML	5	320(input)-256-32-64-10(output)	93,803	0.000253	7,918

2. LNG Regasification Process

Onshore LNG regasification process is a process of producing and supplying NG, the final product, by receiving LNG from an LNG carrier and storing it in a cryogenic liquid state and then vaporizing LNG through a vaporizer. Figure 1 is a diagram of the LNG regasification process. Various dynamic modelings of the widely used ORV based on artificial neural networks are performed during this research.

3. Dynamic Modeling Using ANN

3.1. LNG Regasification Process Dataset

Real operation data were obtained for the LNG regasification process. A prediction model was constructed using a data set with a total of eight features, such as seawater flow rate, LNG flow rate, incoming seawater temperature, discharge seawater temperature, LNG temperature, NG temperature, seawater pressure, and NG pressure.

3.2. Data Preprocessing

As data preprocessing, Min-Max normalization was performed to obtain the effect of reducing the prediction error and training time, and for the objective performance evaluation of the developed model, the entire data set was applied k-fold cross validation, a method of repeating training and validation by dividing into k of the same size. In this study, model training and validation were repeated 5 times (k=5).

3.3. Dynamic Modeling

The proposed ORV dynamics prediction model consists of 8 input variables and 2 output variables. Using process data in 40 time steps, from the past (t=-39) to the present (t=0), a model was designed to predict the dynamic changes of the future 5 time steps (t=1,...,5) of the discharge NG and seawater temperatures, which are constrained by environmental and process factors, according to changes in ORV seawater flow rate, seawater temperature, and LNG flow rate.

A network was constructed using Pytorch 1.5.0, and data were trained using GPGPU processors. FNN, LSTM and AutoML were applied, and the optimal values of the hyperparameters were determined using grid search method, which is the process of searching through a manually specified subset.

3.4. Deep Neural Network Modeling

To find the optimal network structure of the FNN, the initial structure was set as epoch=500, batch size=250, learning rate=0.1, and number of neurons=100. The optimal values were determined in the order of the number of hidden layers, the number of neurons in each layer, batch size, learning rate, and epoch. For LSTM, the initial structure was set to epoch=300, batch size=250, learning rate=0.1, hidden dimension=10, and number of neurons=50. The optimal values were determined in the order of the number

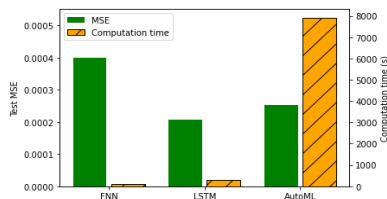


Figure 2. Model performance of FNN, LSTM and AutoML.

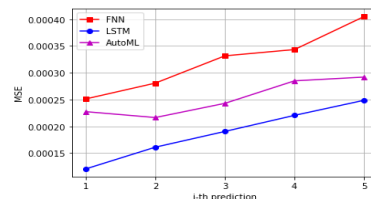


Figure 3. MSE according to the output sequence

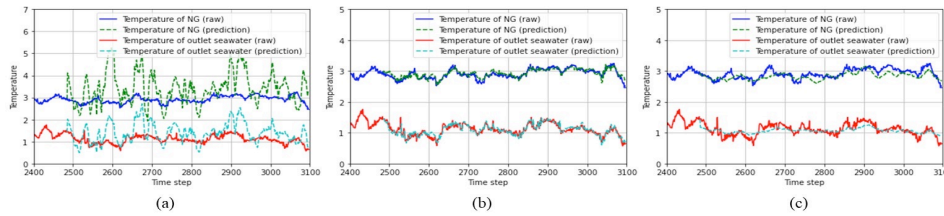


Figure 4. Prediction results of (a) FNN, (b) LSTM and (c) AutoML.

of LSTM layers, hidden dimension, number of neurons, batch size, learning rate, and epoch. Table 1 shows the results of FNN and LSTM modeling for five datasets.

3.5. AutoML

Automatic design of a machine learning model using AutoML was performed. An automatic model generated by using the AutoKeras Python library (developed by Texas A&M University) has 3 hidden layers, and the total number of neurons is 352 (256-32-64), which is about 6 times more than the previous FNN model. Since the process of finding the optimal structure by Bayesian optimization is included, the training required 7918 seconds, which is 27 times longer than LSTM, but the test MSE showed better performance than the FNN with 0.000253. Table 1 shows the proposed structure by AutoML.

4. Performance of Dynamic Prediction Models

4.1. Model Performance Comparison

For the performance evaluation of the optimal FNN, LSTM and, AutoML, the test MSEs were compared. In the case of FNN, the average MSE was 0.000400, for LSTM, 0.000207, for AutoML, 0.000253. Figure 2 shows the model performances of FNN, LSTM and, AutoML. Figure 3 shows the prediction performance according to the future time step: LSTM performs the best with the lowest error in all predictions, more suitable for predicting process time series data. The performance of AutoML was superior compared to the FNN model, and especially the total development time was reduced to 1/25 compared to the LSTM model with a complex structure that required more than a week of work. Development of predictive models using AutoML only required domain knowledge at the level that can be designed by field engineers and shows similar performance to models designed by experts, thus the model development using AutoML was concluded as applicable.

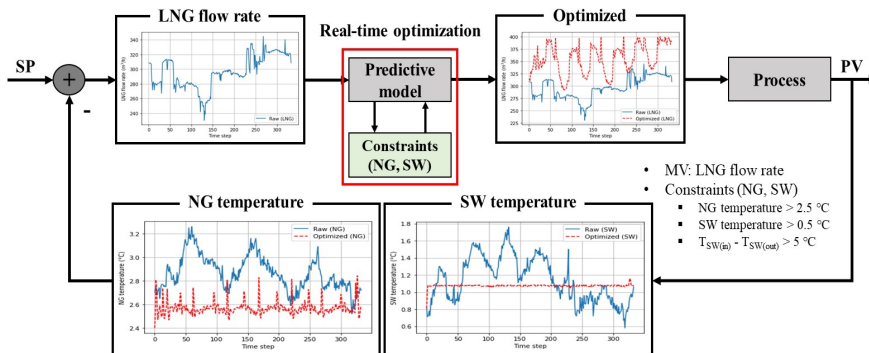


Figure 5. Result for real-time optimization of LNG flow rate.

4.2. Prediction Results

The real operation data were applied to the FNN, LSTM, and AutoML models to compare the prediction results. Figure 4 shows the prediction results for time step 2400 to 3100. FNN showed an error of 0-2.5K, LSTM of 0-0.42K, and AutoML of 0-0.38K. The prediction error of FNN was small, but it did not reflect the trend of the actual data. The model proposed through LSTM and AutoML well reflected the trend of actual data; and data-based modeling was outstanding for complex systems, which is smaller than the prediction error (0-6K) of the first principle-based ORV model proposed by Su et al. (2014).

5. Application of Developed Dynamic Model

5.1. Real-time Optimization

To present a guide for manual operation, especially for winter, the maximum amount of LNG that can be vaporized through ORV is calculated in real-time while effectively satisfying the constraints, to confirm the effect of reducing operating cost using the predictive model. The real-time optimization of the amount of LNG that can be processed through ORV was performed by the following equations.

$$\max \quad L(x) = \left(\sum_{i=0}^n x_i \right) / n \quad (1)$$

$$\text{s.t.} \quad F_{NN(NG)}(x_i) \geq 2 \quad (2)$$

$$F_{NN(sw)}(x_i) \geq 0.5 \quad (3)$$

$$T_{sw(in)} - F_{NV(NG)}(x_i) \leq 5 \quad (4)$$

$$0 \leq x_i \leq 450 \quad (5)$$

where x_i is the LNG flow rate in ORV at the time i , n is the total operating time, T_{sw} is seawater temperature, $F(x)$ is the output value of the prediction model.

The objective function, Eq. (1), is to maximize the average flow rate of LNG while satisfying all the constraints. The discharged NG temperature should always be higher than 1°C, which is the design condition of ORV. In the case of discharged seawater that is released to the sea after use, a constraint on the temperature difference between intake and drainage is enforced due to the influence of the marine ecosystem. For constraints, Eqs. (2)-(5), the discharged NG temperature was set to higher than 2°C, and the discharged seawater temperature and the temperature difference were set to more than 0.5°C, below 5°C.

5.2. Scenario: Reduction in LNG flow rate under 5°C of seawater

When the temperature of seawater decreases, the current (table-based) operation method excessively decreases the LNG flow rate to satisfy the constraints, as a result, the NG and seawater discharge temperatures rapidly increase. By performing real-time optimization through the LSTM model, it was intended to prevent a rapid rise in the discharged temperatures of NG and seawater and to avoid excessive LNG flow rate reduction. Figure 5 shows the results of real-time optimization. The LNG flow rate that can be processed through ORV increased by 19.7% per hour from the previous average of 296.62 to 355.19 m³/h. It vaporizes 24 tons of additional LNG per hour; operation cost of \$216 per hour can be reduced by maximizing the use of ORV instead of SCV usage.

6. Conclusions

To overcome the weaknesses of first principle modeling of ORV, complicated and requiring a lot of time, we developed a prediction model based on FNN, LSTM, and AutoML using real operation data. The test MSE of the LSTM model showed the best performance of 0.000207, and the AutoML-based model of 0.000253, which was higher than the FNN model of 0.0004. Besides the process of finding the optimal model through Bayesian optimization, it took 7918.42 seconds to train, but when comparing the time taken to manually develop the entire model with the LSTM through the complex modeling process, the whole design time demand was greatly reduced to 1/25. AutoML-based modeling requires domain knowledge only at the level of field engineers and shows similar performance to the models developed by experts, proposing that AutoML is effective in practical application.

In the actual data prediction, the FNN showed an error of less than 3K. The LSTM and AutoML models indicate high accuracy with an error of less than 0.5K, and the data-based model showed superior prediction accuracy compared to the first principle-based model. The LSTM model with high accuracy was used to predict the maximum amount of LNG vaporization in real-time within the range that satisfies the constraints of the NG and seawater discharge temperatures in winter. By real-time optimization of the LNG flow rate in ORV, it improved by 19.7%, and operating cost of \$216 per hour can be reduced by decreasing the SCV usage. Accordingly, by recommending the maximum LNG flow rate in real-time, the possibility of supporting the operations decision using the predictive models was validated.

References

- H. Cheng, Y. Ju and Y. Fu, 2019, Thermal performance calculation with heat transfer correlations and numerical simulation analysis for typical LNG open rack vaporizer, *Applied Thermal Engineering*, **149**, 1069-1079.
- A.A. D'alessandro, E.M. Izurieta and S.M. Tonelli, 2016, Decision-making tools for a LNG regasification plant siting, *Journal of Loss Prevention in the Process Industries*, **43**, 255-262.
- T. Jin, M. Wang and K. Tang, 2014, Simulation and performance analysis of a heat transfer tube in SuperORV, *Cryogenics*, **61**, 127-132.
- A.K. Singh, B. Tyagi, V. Kumar, 2013, First Principle Modeling and Neural Network-Based Empirical Modeling with Experimental Validation of Binary Distillation Column, *Chemical Product and Process Modeling*, **8**(1) 53-70.
- H.-D. Su, S.-R. Yu, J.-L. Fan and X. Ling, 2014, Numerical Simulation of Gasification Process on Rib-tube of Open Rack Vaporizer, *Journal of Computers*, **9**(2), 301-307.
- M. Wang, T. Jin, K. Tang and G. Chen, 2013, Numerical simulation analysis of a heat transfer tube in SuperORV, *Natural Gas Industry*, **33**, 102-107.

Acknowledgements

This research was partly supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE)(P0008475, Development Program for Smart Digital Engineering Specialist) and Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government(MOTIE)(20207200000070, Development of performance standardization and operation risk estimation for renewable energy-linked alkaline water electrolysis hydrogen production systems using digital twins).

Carbon or Nitrogen-based e-fuels? A comparative techno-economic and full environmental assessment

Diego Freire Ordóñez ^{a,*}, Thorsteinn Halfdanarson ^a, Caroline Ganzer ^a, Gonzalo Guillén-Gosálbez ^b, Niall Mac Dowell ^c, Nilay Shah ^a

^a *Department of Chemical Engineering, Imperial College London, SW7 2AZ, UK*

^b *Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, Switzerland*

^c *Centre for Process Systems Engineering, Imperial College London, SW7 2AZ, UK*
dmf15@ic.ac.uk

Abstract

The increasing energy demand for mobility services and the growing concern about global warming have become significant drivers for these services' decarbonisation. In this regard, the production and use of fuels obtained from just water, air and renewable energies instead of conventional fossil fuels have caught much attention within the research community. Recently, nitrogen-based e-fuels have been praised for their potential to satisfy mobility and transportation services with a reduced carbon footprint compared to their carbon-based analogues, given their carbon-neutral nature. To evaluate this hypothesis, we conducted a location-based, techno-economic and cradle-to-grave environmental assessment for solar methanol (MeOH) and ammonia (NH₃) based on an optimisation model. Methanol and ammonia were considered for this study due to their relative ease of manufacture and lower production costs than complex fuels, *e.g.*, FT-fuels, and the growing interest in using them as transportation fuels. From this analysis, we concluded that ammonia could have similar production costs, *ca.*, 300 USD/GJ, but better environmental performance than methanol regarding global warming potential (GWP) and the three endpoint impact categories of the ReciPe 2016 LCA damage model, *i.e.*, human health, ecosystems and resources. These results are highly dependent on the hydrogen storage options available; their costs and carbon footprints.

Keywords: e-fuels, techno-economic assessment, life-cycle assessment, optimisation, carbon and nitrogen-based fuels

1. Introduction

The non-renewable nature of fossil fuels and the increasing demand for energy services, especially for power generation, heating, transportation, and rural energy services, have contributed to the depletion of oil reserves and global warming and its underlying effects. Recent studies (Dominković *et al.*, 2018; Brynolf *et al.*, 2018) affirm that the transport sector accounts for more than 20% of worldwide greenhouse gas emissions mainly due to the consumption of petroleum-derived fuels. Given the importance of sustainably meeting current and future demand for mobility services, recent research has focused on solutions that promote renewables' increasing utilisation. In this regard, e-fuels, *i.e.*, fuels produced from carbon (Brynolf *et al.*, 2018) or nitrogen (Ganzer and Mac Dowell, 2020) feedstocks, and "green" hydrogen, emerge as promising alternatives to tackle the environmental problems associated to the use of fossil fuels. As global warming is a

leading environmental concern, the study of carbon-based e-fuels has been prioritised owing to the possibility of employing atmospheric CO₂ as carbon feedstock. However, the potential of nitrogen-based e-fuels in deep decarbonisation scenarios may have been underestimated, given the carbon-neutral nature of these kinds of fuels. In this context, we conduct a thorough techno-economic and environmental evaluation of carbon and nitrogen-based e-fuels aimed at decarbonising the road transportation sector. Our assessment framework includes a broad range of tools, including process modelling and optimisation, heat integration, net production cost (NPC) calculation and cradle-to-grave LCAs. This framework was applied for the continuous production of MeOH and NH₃ in London, considering the hourly-based solar radiation of 2018, and targeting the same amount of energy output from both e-fuels. Based on the model results, we found that the production and utilisation of MeOH as fuel would have less environmental impacts than NH₃, even though their production costs are approximately the same. However, when we neglected the environmental impacts associated with the H₂ storage, *i.e.*, composite tanks, it was noted that NH₃ performed better than MeOH in all the environmental impact categories assessed. Summing up, we have shown the importance of including a detailed assessment of the intermittency of renewable energies and the corresponding energy storage means when evaluating the economic and environmental performance of e-fuels, which is often underestimated in this kind of studies.

2. Methodology

2.1. Model description

To evaluate the production of e-methanol (MeOH) and e-ammonia (NH₃), we followed the framework proposed by Ganzer and Mac Dowell (2020). This model takes the fluctuating nature of solar energy into account and allows determining the minimum net production cost (NPC) of the e-fuels based on optimal production routes, equipment sizing, heat integration and energy storage levels of chemicals and electricity. The implementation and solution of this MILP model were made via GAMS 32.1.0 with CPLEX 12.10.0.0 as the solver. We run this model for London with the hourly-based PV capacity factors taken from renewables.ninja for 2018.

2.2. Economic assessment

For the economic assessment, the total cost only includes the CAPEX of the processes and the storage options given that the cost of utilities, *i.e.*, heating, cooling and electricity, raw materials, *i.e.*, solar radiation, water and air, and labour costs could be considered negligible based on this model approach. The NPCs of the e-fuels were estimated considering the working capital to be 10% of the total capital investment (TCI); an interest rate of 10%, a plant lifetime of 20 years and an annual operation of 8600 h.

2.3. Environmental assessment

The environmental assessment was implemented in Simapro v9.0 interfacing with Ecoinvent 3.5, based on a cradle-to-grave approach and the life cycle impact model ReCiPe 2016, as described in a previous publication (Freire Ordóñez and Guillén-Gosálbez, 2020). The foreground system's LCA inventories were obtained from the model of Ganzer and Mac Dowell (2020) while the background data, *i.e.*, subprocesses involved in the provision of raw materials and utilities to the main processes, was retrieved from Ecoinvent. The composite tanks required to store H₂ were modelled based on the work of Benitez, *et al.* (2021), the Li-ion (LFP) batteries for the electricity storage according to (Majeau-Bettez *et al.*, 2011), and the utilisation-phase emissions of both e-fuels based on the study of (Al-Breiki and Bicer, 2021).

3. Results and Discussion

The optimal production routes for the given conditions are shown in Fig. 1 and Fig. 2.

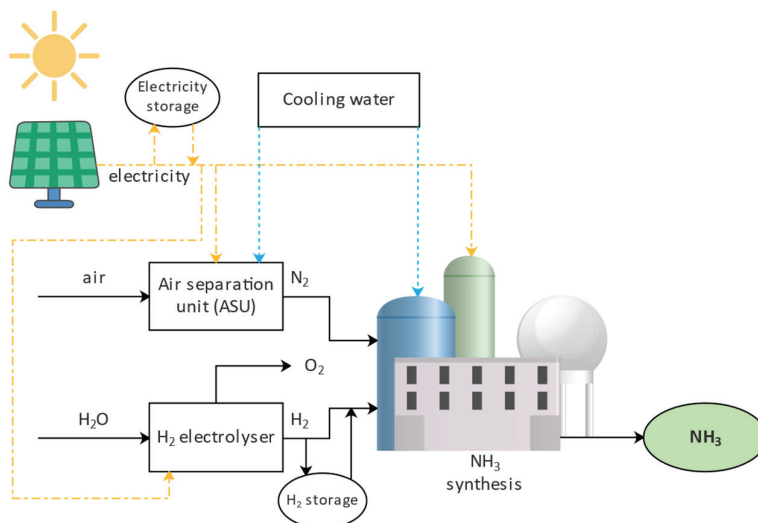


Fig. 1. NH₃ production process from solar electrolytic H₂ and N₂ from air separation. Only a cooling utility, *i.e.*, cooling water, is required after heat integration

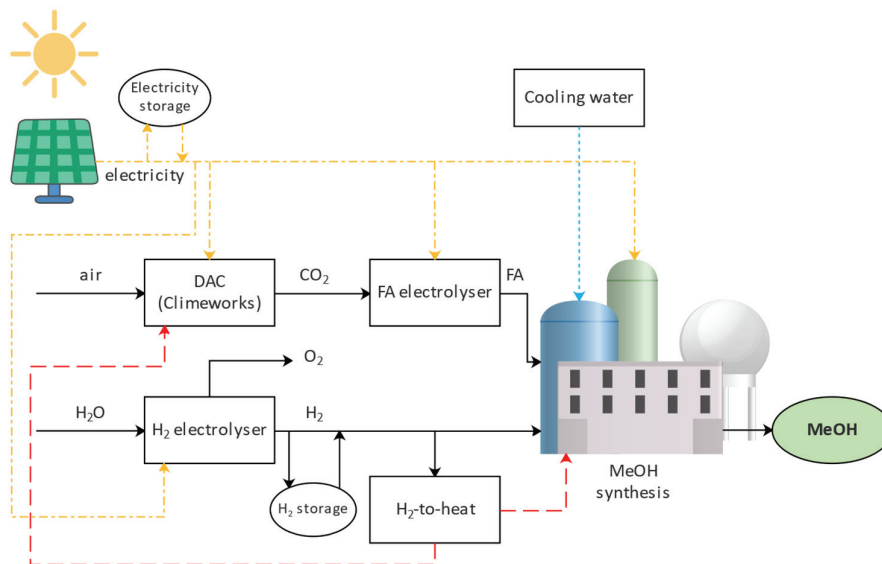


Fig. 2. MeOH production process from solar electrolytic H₂ and CO₂ from direct air capture (DAC) using the Climeworks technology. This production route includes the electrocatalytic reduction of CO₂ to formic acid (FA). A heating utility, *i.e.*, H₂-to-heat, and a cooling utility, *i.e.*, cooling water, are required after heat integration.

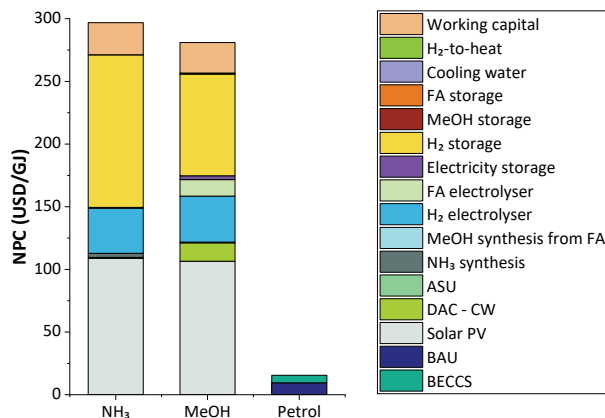


Fig. 3. NPC of the studied e-fuels, including all process and storage equipment required for the continuous operation of the chemical plant, compared with fossil petrol (BAU) + bio-energy with carbon capture and storage (BECCS) (Natural Resources Canada, 2014; Drax Group plc, 2020). Neither duty rate nor VAT is included in the given figure for fossil petrol. BECCS is regarded at the moment as the most mature carbon removal technology (IEA, 2020)

Fig. 3 shows the cost breakdown of the production routes of each e-fuel, while the required sizes of each subprocess are detailed in Table 1. It can be seen that both e-fuels have similar NPCs on an energy basis, *ca.*, 300 USD/GJ, and that the major contributors to the production costs are the solar PV and the H₂ storage processes. A large storage capacity for H₂ is required to guarantee the chemical plants' continuous operation due to the low level of solar radiation in London compared to other regions. Therefore, it is advisable to explore different energy alternatives, *e.g.*, wind power, depending on the production plants' location. It can be seen that, currently, the NPC of the studied e-fuels is approximately 30-fold higher than that of the BAU, when they are compared under the same cost basis, *i.e.*, excluding any taxes, tariffs or fees, added after the production stage. If direct offsetting of petrol CO₂ emissions via BECCS is considered, the NPC difference between the e-fuels and BAU would be about 20 fold.

Table 1. Process design required for the continuous operation of the chemical plant

Process	NH ₃	MeOH
PV [GW]	2.24	2.19
Battery storage [MWh]	40.43	204.85
H ₂ electrolyser [tonne-H ₂ /h]	17.60	17.92
H ₂ storage [tonne-H ₂]	4611.67	3077.11
FA electrolyser [tonne-FA/h]		57.61
FA storage [kt-FA]		99.79
DAC-CW [#modules]		9832.00
ASU [tonne-N ₂ /h]	16.61	
H ₂ to heat [MW]		102.59
Direct MeOH synthesis [tonne-MeOH/h]		
MeOH synthesis from FA [tonne-MeOH/h]		18.70
NH ₃ synthesis [tonne-NH ₃ /h]	20.00	
Cooling water (MW)	18.14	16.36

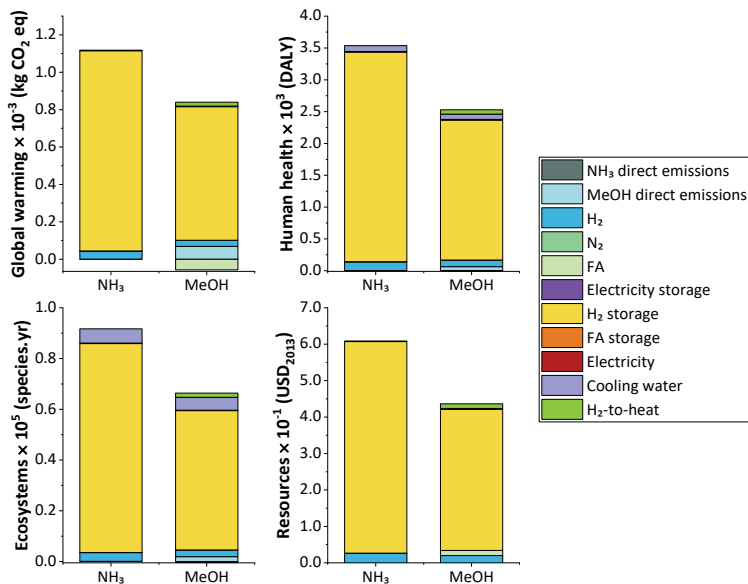


Fig. 4. ReCiPe 2016 LCIA per GJ of e-fuels: Cradle-to-grave results for GWP and endpoint impact categories, including environmental impacts and corresponding breakdowns.

As shown in **Fig. 4**, most environmental impacts come from H₂ storage. This is explained by the large storage capacity needed, and the complexity and energy-intensive nature of the processes required to manufacture carbon fibre. This fibre is the main component of the composite tanks used to store H₂ gas (Benitez *et al.*, 2021). Nevertheless, the production and utilisation of NH₃ might be a viable alternative over MeOH, if the costs and environmental impacts of H₂ storage are minimum or negligible (see **Fig. 5**), *e.g.*, if H₂ is stored in salt caverns (Sørensen and Spazzafumo, 2018).

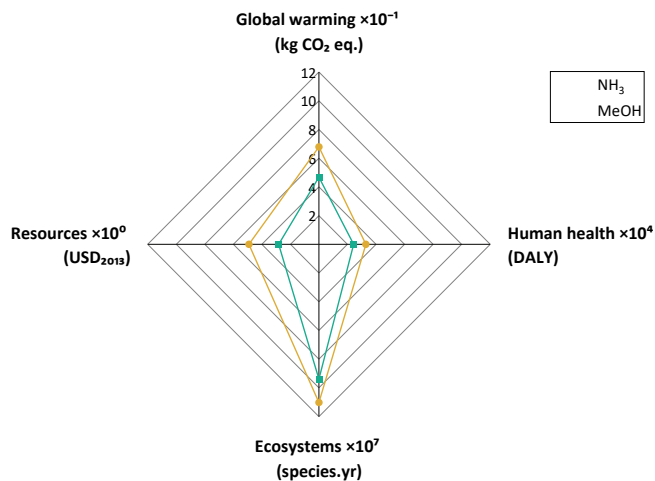


Fig. 5. ReCiPe 2016 LCIA per GJ of e-fuels: Cradle-to-grave results for GWP and endpoint impact categories, when the environmental impacts of H₂ storage in composite tanks are omitted.

4. Conclusion

In this work, we have performed a comparative techno-economic and environmental assessment of two representative carbon and nitrogen-based e-fuels, *i.e.*, MeOH and NH₃, to evaluate the hypothesis that the production and utilisation of carbon-containing e-fuels have a higher carbon footprint than their nitrogen-based analogues when delivering the same amount of energy output. This hypothesis was proven to be not always correct since the production costs and environmental burdens associated with the e-fuels strongly depend on the steady-state plant's location, which has a marked effect on the design capacity of the storage means for chemicals and electricity. Based on our results for London in 2018, it was concluded that the composite tanks required to store H₂ for the continuous operation of the chemical plant were, in aggregate terms, major economic and environmental contributors of both solar e-fuels. This fact might not hold for other locations with different energy potentials. Furthermore, it was shown that conventional petrol is still substantially cheaper than both e-fuels, even when considering BECCS to offset its fossil CO₂ emissions. It means that significant cost reductions and technology improvements are still pending for e-fuels to be widely deployed.

This analysis demonstrated the importance of a location-based approach for this kind of studies and the need to keep investing in R&D to lower equipment costs and find alternative storage options that allow not only reducing the production costs but also the associated harmful impacts of e-fuels.

5. References

- M. Al-Breiki and Y. Bicer (2021), "Comparative life cycle assessment of sustainable energy carriers including production, storage, overseas transport and utilization", *Journal of Cleaner Production*, Vol. 279, p. 123481.
- A. Benitez, C. Wulf, A. de Palmenaer, M. Lengersdorf, T. Röding, T. Grube, M. Robinius, D. Stolten and W. Kuckshinrichs (2021), "Ecological assessment of fuel cell electric vehicles with special focus on type IV carbon fiber hydrogen tank", *Journal of Cleaner Production*, Vol. 278, p. 123277.
- S. Brynolf, M. Taljegard, M. Grahn and J. Hansson (2018), "Electrofuels for the transport sector. A review of production costs", *Renewable and Sustainable Energy Reviews*, Vol. 81, pp. 1887–1905.
- D. Dominković, I. Bačković, A. Pedersen and G. Krajačić (2018), "The future of transportation in sustainable energy systems. Opportunities and barriers in a clean energy transition", *Renewable and Sustainable Energy Reviews*, Vol. 82, pp. 1823–1838.
- Drax Group plc (2020), "Likely cost for BECCS in the UK", available at: <https://www.drax.com/>.
- D. Freire Ordóñez and G. Guillén-Gosálbez (2020), "Techno-economic and Environmental Assessment of Electrofuels: a Case Study of Gasoline Production using a PEM Electrolyser", in *30th european symposium on computer aided chemical engineering, Computer Aided Chemical Engineering*, Vol. 48, Elsevier, [Place of publication not identified], pp. 595–600.
- C. Ganzer and N. Mac Dowell (2020), "A comparative assessment framework for sustainable production of fuels and chemicals explicitly accounting for intermittency", *Sustainable Energy & Fuels*.
- IEA (2020), "CCUS in Clean Energy Transitions", available at: <https://www.iea.org/reports/ccus-in-clean-energy-transitions>.
- G. Majeau-Bettez, T. Hawkins and A. Strømman (2011), "Life cycle environmental assessment of lithium-ion and nickel metal hydride batteries for plug-in hybrid and battery electric vehicles", *Environmental Science & Technology*, Vol. 45 No. 10, pp. 4548–4554.
- Natural Resources Canada (2014), "Learn the facts: Fuel consumption and CO₂", available at: https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/oe/pdf/transportation/fuel-efficient-technologies/autosmart_factsheet_6_e.pdf (accessed 14 November 2020).
- B. Sørensen and G. Spazzafumo (2018), "Social implications", in Sørensen, B. and Spazzafumo, G. (Eds.), *Hydrogen and Fuel Cells: Emerging Technologies and Applications*, Elsevier Science, San Diego, CA, USA, pp. 413–461.

Techno-economic assessment of a renewable energy-based electric vehicle fast-charging station in Qatar

Abdulla Al Wahedi*, Yusuf Bicer

Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University (HBKU), Education City, Doha 5825, Qatar

amalwahedi@hbku.edu.qa

Abstract

Electrical vehicles' (EV) deployment as an alternative eco-friendly transport solution has become a promising initiative worldwide to reduce greenhouse gas emissions and fossil fuel depletion. This fact triggers the need to roll out fast-charging stations to fulfill daily road charging demand. Securing the power requirement for those stations has become a significant challenge that would cause substantial load increase on the existing electricity generation system and distribution infrastructure if supplied from the conventional resources. Hence, grid-independent charging stations with renewable energy sources (RES) and multiple energy storage systems has become an alternative solution to overcome the raised challenge. This paper aims to assess the implementation of a stand-alone fast charging station technically and economically in the State of Qatar comprising of a wind turbine (WT), concentrated photovoltaic (CPV) system and a bio-generator as RES along with various storage systems. The proposed design is built, modelled, and simulated using Hybrid Optimization System for Electric Renewable (HOMER) software to determine the optimal techno-economic configuration to fast charge 50 EVs daily in a reliable manner. Predefined constraints such as space limitation, stochastic nature of EV demand, and site-specific metrological conditions are considered. Multiple sizing portfolios of incorporated subsystems are evaluated through simulation. Sensitivity analysis is used to evaluate the impact of selected decision variable values such as the WT height where the generated analytical results are compared from the technical and economic perspectives. The results show that a stand-alone micropower system consisting of 450 kW CPV, 250 kW WT with 60 m hub height, 100 kW bio generator, and 324 kWh batteries is the optimal configuration with minimal 2.378 million dollars net present cost (NPC), 0.284 \$/kWh cost of energy (COE) and 0.02 % unmet demand.

Keywords: Hydrogen, electrolyzer, fuel cell, clean energy, HOMER.

1. Introduction

The rising interest in clean energy and EV deployment increases the electricity demand to fulfil the associated charging power requirements causing generation fluctuations and grid tension in case it is supplied from the grid. Consequently, leading to political instability, economic variability, and environmental degradation in case conventional generation methods are used to supply the additional demand required. This made securing the needed power supply independently from the grid and based on renewable energies a global interest and motivation.

Many studies have investigated renewable energy utilization for power generation while others on electric mobility (Mazzeo, 2019). The combined research of employing stand-alone RES to supply electrical power for fast EV charging station has been insufficiently considered, and further research on grid-independent EV charging stations is required (Mehrjerdi, 2019).

In commitment to its long-term national sustainable development strategy, Qatar has embarked on the “Green Car Initiative” in 2017 (The Peninsula Qatar, 2017) to stimulate on the adoption of EVs to achieve 4 % share of its local share by 2020 (Qana, 2017). To fulfil EVs charging demand, it was declared that 400 EV charging stations would be accessible by the same year (Kumar, 2019).

Since the number of registered vehicles in Qatar in 2017 was 1.5 million and the average annual increase is 12 % (CEIC, 2019), the total number of vehicles will reach 2.7 million in 2022, out of which 100,000 shall be EVs. This fact raises the need to devise reliable nonhydrocarbon-based micropower systems to supply those charging stations with the necessary power without impacting or expanding the existing grid infrastructure.

2. Methodology

In this study, a micropower system which was designed and thermodynamically evaluated using Engineering Equation Solver (EES) program (Klein, 2018) in a previous study by the authors (Al Wahedi and Bicer, 2020a) is built using HOMER model from the National Renewable Energy Laboratory (NREL) (HOMER, 2018). The resources of the associated RES incorporated in the proposed design along with forecasted EV demand are modeled. Sensitivity analysis based on selected variables along with various parameters and sizing options for selected components are considered in the simulation process. HOMER Pro simulation and optimization results are analyzed thoroughly where multiple optimal configurations are investigated and compared to select the overall optimal design configuration and sizing based on predefined technical and economic decision criteria.

3. System description

Since Qatar is located in the Sun Belt region with plenty of annual solar irradiation, CPV solar system is selected as one of the two main RES for generating electricity. WT is selected as the second main RES since, based on the literature review the application of small to medium-size WTs is feasible to generate energy in Qatar (Moghbelli et al., 2011). Bio generator, electrolyzer, chemical storage tank, H₂ and NH₃ fuel cells (FC), DC/AC converter, and battery storage system (BSS) are the remaining components of the proposed stand-alone design.

Load-following (LF) and cycle-charging (CC) strategies are assessed for charging the incorporated BSS. In the former strategy, the battery system is charged by renewable sources but not by the incorporated bio generator. While in the latter strategy, the bio generator serves both the load and storage battery charging.

While H₂ is produced and stored in a tank on site, NH₃ is produced off site and stored in a tank on site in the proposed system.

The proposed design is built in HOMER and illustrated in Figure 1.

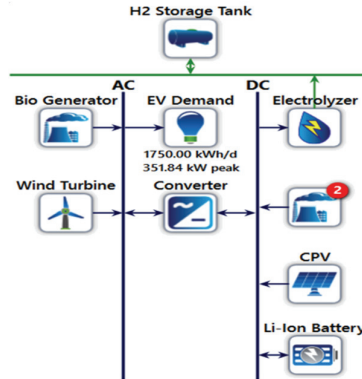


Figure 1: Proposed system configuration

Table 1 indicates the parameters to be investigated for each component in the design. Few components have predefined parameters, others have a range of parameters to be investigated, and others are determined by HOMER optimizer.

Table 1: Components and parameters investigated in the study

Component	Parameter	Remarks
CPV	450 kWp	Determined using PVSyst (PVsyst, 2018) software based on space limitation of 1,500 m ²
WT	250 kW	Determined considering space limitation and WT blade dimensions
H ₂ and NH ₃ FCs	50 kW / 100 kW	Both sizes are simulated for each FC
BSS and Converter	-	Calculated and selected by HOMER Optimizer
Bio Generator	50 kW, 100 kW and Auto size	Auto size is calculated and selected by HOMER Optimizer
Electrolyzer	50 kW / 100 kW	Both sizes are simulated
Chemical storage tank	50 kg / 100 kg	Both sizes are simulated
BSS charging strategy	LF / CC	Both strategies are simulated
WT Hub Height	40 m, 50 m, 60 m	Sensitivity analysis is carried out

4. Load profile and resources input modeling

Considering 20 % of the total EVs will need on-the-road fast charging service and the average energy required by each EV is about 35 kWh, the daily demand from a single charging station is estimated to be 1,750 kWh (Al Wahedi and Bicer, 2020b). The daily EV demand profile from the study (Bayram et al., 2016) is used here, after extrapolating the data using probability density function to represent 50 EVs daily demand. To model the demand more realistically, 10 % day-to-day and 20 % timestep randomness is added to the EV load data to represent 10 % daily variance and 20 % monthly variance. Figure 2 illustrates the daily and seasonal stochastic EV demand profile in HOMER.

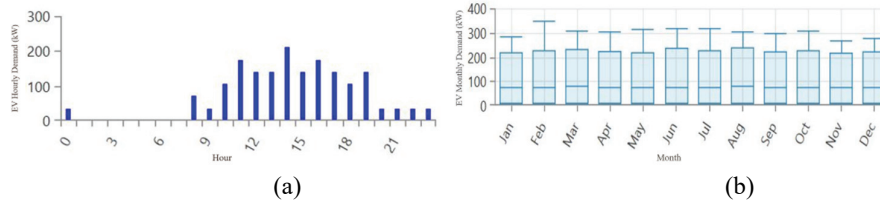


Figure 2: EV demand generated by HOMER (a) daily and (b) seasonal

For the specific site selected in Qatar, the average monthly Solar Global Horizontal Irradiance (GHI), clearness index, and wind speed data are obtained from NASA Prediction of Worldwide Energy Resources (POWER) database included in HOMER Pro software, which is illustrated in Figure 3.

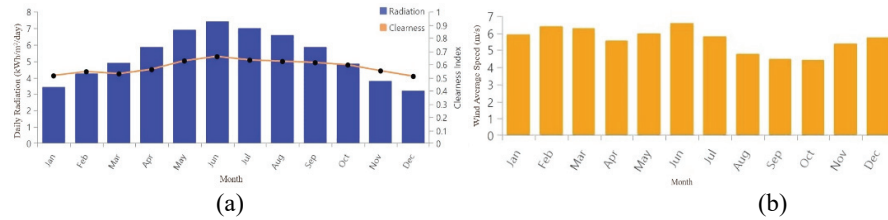


Figure 3: (a) Average monthly Solar GHI, clearness index, and (b) Wind speed at selected location in Qatar

The GHI data is for over 22 years from July 1983 to June 2005, and wind speed data is for over 30 years from January 1984 to December 2013 measured at 50 m height.

The economic analysis in HOMER is determined by assessing the life-cycle cost of a system, which is represented by calculating the NPC and COE of each configuration. Each component's costs within the project lifetime are reflected in Table 2 where electrolyzer and FCs costs are obtained from (Energy, 2015) and (Monterey Gardiner, 2014), respectively. Moreover, 5.7 % discount rate and 0.55 % inflation rate are used in this study (Bank, 2021).

Table 2: Life cycle cost of main components

Component	Capital Cost (\$)	Replacement Cost (\$)	O&M Cost (\$)	Fuel Cost (\$/L)	Lifetime
1 kW CPV	1,000	300	25/y	-	25 y
250 kW WT	250,000	250,000	1,000/y	-	25 y
1 kW NH ₃ FC	3,000	1,500	0.010/h	-	5,000 h
1 kW H ₂ FC	3,000	1,500	0.010/h	-	5,000 h
1 kWh Li-ion Battery	300	300	10/y	-	15 y
1 kW Converter	300	300	-	-	15 y
1 kW Bio Generator	200	200	0.030/h	0.50	15,000 h
1 kW Electrolyzer	380	95	20/y	-	7 y
1 kg H ₂ storage tank	700	700	-	-	25 y

5. Simulation and optimization results

The proposed stand-alone design is configured and simulated in HOMER Pro to identify the optimal configuration from the technical and economic perspectives. Based on the earlier predefined parameters such as components’ sizes, WT hub’s height, and battery charging strategy, all feasible options are simulated in HOMER Pro to determine the optimal configuration based on multiple decision criteria. Accordingly, 68,520 solutions are simulated, out of which 45,048 were found feasible while 20,400 were found infeasible due to capacity shortage constraint where the required capacity is found relatively more than the actual capacity that the system can deliver. For each WT hub height sensitivity case, nine different optimal configurations are identified with detailed associated parameters and components’ sizing.

For each WT hub height, the best configuration of each of the three suggested bio-generator capacities are listed in Table 3, along with associated components’ capacities and parameters. Techno-economic analysis is carried out by comparing NPC, COE and unmet demand among all the nine selected scenarios. As a result, 60 m WT height with 100 kW bio-generator and associated components and parameters scenario is found as the optimal configuration with minimum NPC and COE.

Table 3: Technical and cost analysis of selected optimization results

WT Hub Hight (m)	Bio Gen (kW)	H ₂ FC (kW)	NH ₃ FC (kW)	Li-Ion Battery (kWh)	Electro_lyzer (kW)	H ₂ Storage Tank (kg)	Conv_erter (kW)	Dispatch	NPC (Million \$)	COE (\$)	Unmet load (%)
40	100	50	50	344	50	50	226	CC	2.217	0.225	0.03
	500	50	50	709	50	50	281	LF	2.373	0.241	0.00
	50	50	50	817	50	50	272	CC	2.485	0.252	0.04
50	100	50	50	342	50	50	229	CC	2.206	0.224	0.03
	500	50	50	740	50	50	273	LF	2.341	0.238	0.00
	50	50	50	810	50	50	264	CC	2.464	0.250	0.04
60	100	50	50	324	50	50	253	CC	2.191	0.223	0.02
	500	50	50	729	50	50	281	LF	2.323	0.236	0.00
	50	50	50	810	50	50	253	CC	2.455	0.249	0.03

The total monthly electricity generated by the proposed configuration is illustrated in Figure 4, where WT, CPV, bio generator, and FC contributions are 41 %, 37 %, 17 %, and 5 %, respectively.

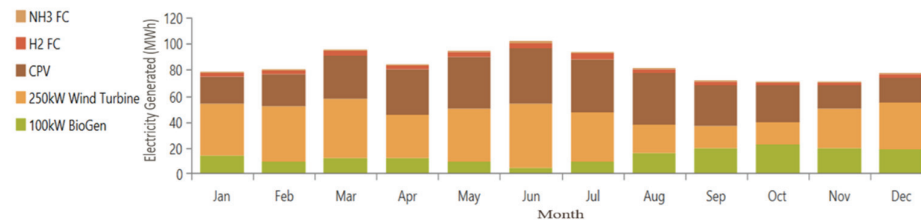


Figure 4: Monthly electricity production by the source in the optimal configuration

6. Conclusions

A techno-economic assessment is carried out in this study with HOMER Pro simulation software to model, simulate and optimize the proposed grid-dependent micropower

system consisting of multiple renewable sources and storage options to fast charging 50 EVs per day in the State of Qatar. Among many simulated feasible configurations, a system comprising of 450 kW CPV, 250 kW with 60 m height WT, 100 kW bio generator, 324 kWh batteries, 50 kg H₂ storage tank and 50 kW electrolyzer, H₂ FC and NH₃ FC is identified as the optimal configuration with 2.378 million dollars NPC, 0.284 \$/kWh COE and 0.02 % unmet demand where WT, CPV, bio generator and FC electricity contribution are found 41 %, 37 %, 17 %, and 5 %, respectively.

References

- Al Wahedi, A., Bicer, Y., 2020a. Development of an off-grid electrical vehicle charging station hybridized with renewables including battery cooling system and multiple energy storage units. *Energy Reports* 6, 2006–2021.
- Al Wahedi, A., Bicer, Y., 2020b. A Case Study in Qatar for Optimal Energy Management of an Autonomous Electric Vehicle Fast Charging Station with Multiple Renewable Energy and Storage Systems. *Energies* 13, 5095.
- Bank, C., 2021. Qatar Central Bank [WWW Document]. URL <http://www.qcb.gov.qa/English/pages/interestrates.aspx> (accessed 2.7.21).
- Bayram, I.S., Zamani, V., Hanna, R., Kleissl, J., 2016. On the evaluation of plug-in electric vehicle data of a campus charging network. 2016 IEEE Int. Energy Conf. ENERGYCON 2016. <https://doi.org/10.1109/ENERGYCON.2016.7514026>
- CEIC, C. and E.I.C., 2019. Number of Registered Motor Vehicles in Qatar [WWW Document]. URL <https://www.ceicdata.com/en/qatar/number-of-registered-motor-vehicles/number-of-registered-motor-vehicles> (accessed 7.26.19).
- Energy, U.S.D. of, 2015. DOE Technical Targets for Hydrogen Production from Electrolysis | Department of Energy [WWW Document]. U.S. Dep. Energy. URL <https://www.energy.gov/eere/fuelcells/doe-technical-targets-hydrogen-production-electrolysis> (accessed 2.7.21).
- HOMER, 2018. Hybrid Optimization System for Electric Renewable. HOMER Pro 3.14.2. <http://www.homerenergy.com>.
- Klein, 2018. EES: Engineering Equation Solver. F-Chart Software: Engineering Software. <http://fchartsoftware.com/ees/>.
- Kumar, C., 2019. 400 electric car charging stations by 2022 in Qatar - The Peninsula Qatar [WWW Document]. *Penins.* URL <https://thepeninsulaqatar.com/article/17/01/2019/400-electric-car-charging-stations-by-2022-in-Qatar>
- Mazzeo, D., 2019. Solar and wind assisted heat pump to meet the building air conditioning and electric energy demand in the presence of an electric vehicle charging station and battery storage. *J. Clean. Prod.* 213, 1228–1250. <https://doi.org/10.1016/j.jclepro.2018.12.212>
- Mehrjerdi, H., 2019. Off-grid solar powered charging station for electric and hydrogen vehicles including fuel cell and hydrogen storage. *Int. J. Hydrogen Energy* 44, 11574–11583. <https://doi.org/10.1016/j.ijhydene.2019.03.158>
- Moghbelli, H., Toliyat, H., Abu-Rub, H., 2011. A green hybrid power plant using photovoltaic and wind energy with power quality improvement in Qatar. *ASEE Annu. Conf. Expo. Conf. Proc.* <https://doi.org/10.18260/1-2--17326>
- Monterey Gardiner, C.R., 2014. DOE Fuel Cell Technologies Office Hydrogen for Energy Storage Workshop on Hydrogen Energy Storage Grid and Transportation Services at the Grand Sheraton Hotel. | *Fuel Cell Technol. Progr. Source.*
- PVsyst, 2018. Photovoltaic system software. PVsyst 6.7.8. <https://www.pvsyst.com/>.
- Qana, D., 2017. Qatar aims to reach 10% Green Cars by 2030. *Lusailnews, Online Main Page_Diverse.* <https://lusailnews.net/article/>.
- The Peninsula Qatar, 2017. Qatar aims to increase electric cars to 10%. *Penins. Qatar, Online Home_News_Qatar.* <https://thepeninsulaqatar.com/art>.

Simultaneous Synthesis and Optimization of Refrigeration Cycles and Heat Exchanger Networks

Matteo Martinelli,^a Cristina Elsidio,^a Ignacio E. Grossmann,^b Emanuele Martelli^{a,*}

^a *Politecnico di Milano, Dipartimento di Energia, Via Lambruschini 4, Milano, IT*

^b *Department of Chemical Engineering, Center for Advanced Process Decision-Making, Carnegie Mellon University, Pittsburgh, PA, USA*

**emanuele.martelli@polimi.it*

Abstract

This work proposes a simultaneous approach for the synthesis and design optimization of refrigeration cycles integrated with heat exchanger networks. The methodology includes a novel refrigeration cycle superstructure capable of reproducing a wide range of cycle architectures, and an effective solution algorithm (based on the decomposition of the problem on two levels) to tackle the challenging Mixed-Integer Nonlinear Program. In addition to optimizing the cycle and heat exchanger network structure, the methodology can optimize cycle pressures and temperatures (including superheating and subcooling degree of the working fluid). The application to a literature case study indicates that the proposed approach yields solutions that are considerably better in terms of economic than those published in the literature.

Keywords: Refrigeration Cycle Optimization, HEN Synthesis, MINLP, Heat Integration.

1. Introduction

The design of a refrigeration cycle represents a challenging task, mainly due to the wide range of designs and working fluid alternatives, resulting from the evolution of this technology during the last 100 years. All these possible alternatives have been developed to meet the needs of different applications, such as industrial refrigeration, heat pumps for industrial waste heat recovery and cryogenic applications like natural gas liquefaction. Thus, the optimal design of refrigeration cycles calls for the development of systematic optimization approaches capable of exploring all the alternative configurations and finding the best trade-off between cost and performance.

Many authors in the literature have addressed the optimization of refrigeration cycles, but to the best of the author's knowledge, only a few have considered a combination of technical, thermodynamic, and economic parameters in the process. In the study by Wallerand et al. (2018), a comprehensive literature review can be found, considering all the studies from the seventies. The main approaches adopted for optimization are Pressure and Temperature optimization approach and Synthesis approach. In the first one, a case-related structure is developed in an ad-hoc fashion, and then optimization is performed only on continuous variables since the structure is fixed. The second approach generates

the optimal structure together with the optimal operating conditions. This can be achieved by replicating multiple times a basic structure, or by selecting portions of a very general superstructure (synthesis based on reduction).

None of the approaches reported in the literature combines all the main optimization targets in one method, considering compression refrigeration cycle structure, heat exchanger network (HEN) structure, thermodynamic and economics. In contrast, the objective of this work is to develop a model that is able to consider all the alternatives available for compression refrigeration, and select the best ones for the selected purpose. In particular, the model considers:

- All the main configurations for the compression refrigeration cycles.
- The optimal configuration of the HEN: not only the general structure of the cycle, but also the heat integration is optimized, which means coupling in the best possible way hot and cold streams to guarantee a required refrigeration duty (and if necessary heating too).
- The optimization of thermodynamic properties, selecting optimal temperatures and pressures for condensations and evaporations, optimal heat exchange temperatures, and selecting the optimal values for all the technical solutions.

2. Problem Statement and Methodology

The problem addressed in this work can be summarized as follows.

“Given a set of hot and cold process streams with given mass flow rates, inlet and outlet temperatures, and a set of possible hot/cold utility systems (e.g., boilers, cooling water), and the superstructure of possible reverse Rankine cycle configurations, determine the optimal selection of utility systems, the design of the reverse Rankine cycle (selection of pressure levels, mass flow rates of each utility stream, temperatures at inlets and outlets of the evaporator/condenser/intercoolers, etc.), the HEN between process–process as well as process–utility, process–cycle and cycle–utility streams, minimizing the Total Annual Cost (TAC, sum of annualized capital cost and yearly operating costs) of the overall system (utility systems, reverse Rankine cycle and HEN)”.

The optimization of the Refrigeration Cycle and HEN is performed together with the thermodynamic and economic optimization with a synthesis approach based on a very general cycle superstructure containing all the major technological solutions from which the best configuration is selected. The superstructure is built as a series of headers representing the thermodynamic conditions (i.e., points in the p-h diagram), connected by streams and components, i.e., valves, compressors, and heat exchangers (HXs). This structure is then combined with a HEN superstructure and optimized minimizing TAC. The procedure also selects the optimal thermodynamic properties of all the streams and components. The problem is modelled and solved in GAMS.

3. Reverse Rankine Cycle and HEN Superstructures

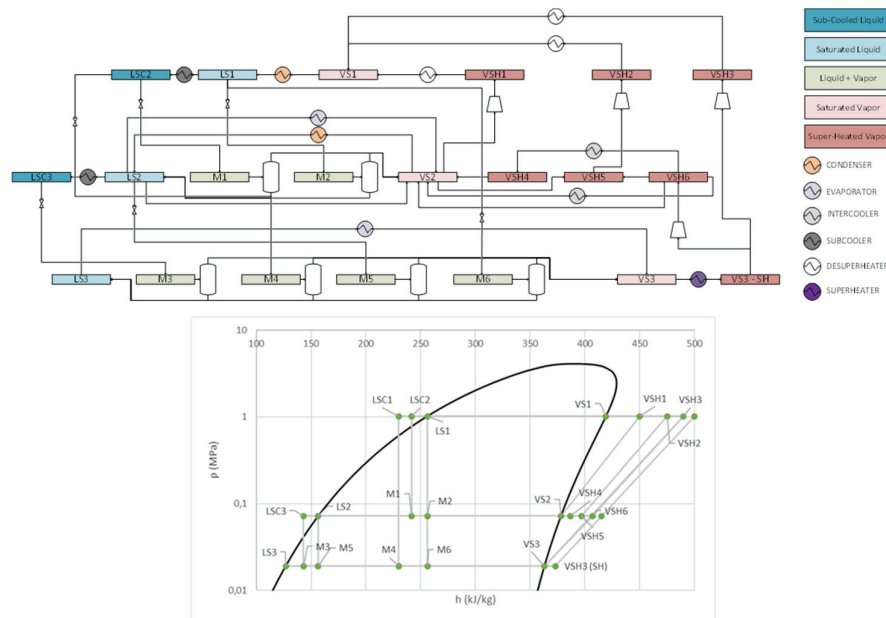


Figure 1. General Compression Refrigeration Cycle Superstructure with 3 Pressure Levels and corresponding p-h Diagram

The cycle superstructure (shown in Figure 1) is obtained by considering the most promising cycle schemes. For the sake of simplicity, Figure 1 shows a superstructure with up to 3 pressure levels, although this number can be easily increased. Indeed, in the test case, a superstructure with up to seven levels of condensation and evaporation is considered. The superstructure can model simple compression refrigeration cycles, compression refrigeration cycles with double throttling, multiple compression intercooling options (HXs, vapor mixing intercoolers and direct contact coolers), multiple condensation and evaporation levels, cascade refrigeration cycles (by replicating more than once the superstructure), liquid sub-cooling and vapor super-heating.

The corresponding p-h diagram is shown in the bottom part of Figure 1. Please notice that the p-h points reported in the diagram are optimized by the algorithm.

As for the HEN superstructure, the proposed optimization uses the SYNHEAT superstructure by Yee and Grossmann (1990) to optimize the HEN synthesis. Following the successful approach of Martelli et al. (2017) and Elsidio et al. (2019), the streams of the reverse Rankine cycle are included in the HEN in all stages and the overall structure obtained is optimized as a large scale nonconvex MixedInteger Nonlinear Programming model (MINLP).

4. Optimization Approach

The optimization problem resulting from combining the superstructure of the Reverse Rankine Cycle with the HEN synthesis is a nonconvex MINLP. It features binary variables for cycle components selection and HXs activation, and continuous variables for thermal power, areas, HEN stage temperatures, compressors power, mass flow rates of streams. The nonlinearities are bilinear products in energy balances ($m_i T_{i,j,k}$, where m_i is the mass flow rate of cycle stream i and $T_{i,j,k}$ is the HEN stage temperature), HXs areas calculation depending on the mean temperature difference (*LMTD*), economy of scale law in the capital cost of equipment units (compressors and HXs), products between mass flow rates and enthalpies of streams in energy balances, and functions $h(p, T)$ linking enthalpies h with temperature T and pressure p (or pressure and entropy for the compressor discharge stream) of each header. These functions $h(p, T)$ are evaluated using the “Extrinsic Functions” developed by Manassaldi et al. (2019) and implementing the Peng Robinson Equation of state. The isentropic efficiency definition was used to calculate the properties of the compressors outlets. Comparison with REFPROP indicates that the error incurred in the use of the Peng Robinson Equation of state in evaluating the enthalpies of the headers is lower than 1%, which is acceptable for the optimization purposes of this work.

Preliminary computational tests indicate that the nonlinear and nonconvex MINLP cannot be solved with general purpose MINLP solvers, such as BARON. Thus, a decomposition approach has been developed by adapting the bilevel decomposition approach proposed by Elsidio et al. (2019) for cases with fixed header p and T (i.e., fixed h). The adapted decomposition algorithm is shown in Figure 2. The initialization consists in guessing starting values of the headers pressures and temperatures, and then solving the problem with fixed header p and T using the same sequential initialization approach of Elsidio et al. (2019). Then, the bilevel algorithm starts with a linearized version of the problem (Mixed-Integer Linear Program, MILP) in which header enthalpies are fixed, but the capital cost functions are piecewise linearized, the area-LMTD relations are linearized using first-order Taylor approximation, and the bilinear terms $m_i T_{i,j,k}$ are linearized using the McCormick relaxation. Although of large scale, the MILP can be solved with commercially available solvers (CPLEX in our case). Once solved, the binary variables are fixed (fixing the HEN and Refrigeration cycle structure), and the lower level program optimizes the continuous variables (including headers pressures and temperatures) by solving a nonconvex Nonlinear Program (NLP). Due to the lack of the explicit algebraic expressions in the Extrinsic Functions used for $h(p, T)$, a solver with global convergence guarantee (like BARON) cannot be used for the NLP subproblem. Thus MSNLP (Multi Start Non-Linear Programming) is chosen to tackle the NLP subproblem (the risk of finding a local minimum is mitigated by considering multiple starting points). After the NLP is solved, the MINLP is linearized again around the newly found solution to derive an updated master level MILP for the subsequent iteration. “Nested integer cuts” (Elsidio et al. 2019) are added to the master level MILP to exclude already found binary solutions.

Since p and T of headers are fixed in the master level problem, its solution cannot be considered as a lower bound (useful for deriving an optimality gap and stopping criterion). Thus, as stopping criterion of the bilevel algorithm, we considered the condition of no solution improvement for 20 consecutive iterations.

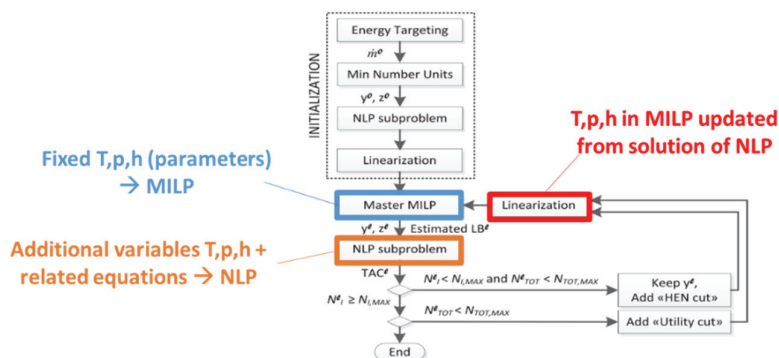


Figure 2. Schematic approach to the MINLP problem in the variable-headers optimization, adapted from Elsidó et al. (2019)

5. Case Study and Results

The proposed approach was applied to Example *E2* by Grossmann and Shelton (1986), where the synthesis and optimization of the refrigeration cycle is performed including two hot and two cold streams to be integrated in the HEN. The objective function is the Total Annual Cost (TAC) including both capital and operating costs. The bilevel decomposition algorithm converged in 33 iterations with a total computational time of 14480 s. The main results are summarized in Table 1, 2 and the optimized cycle and HEN structures are reported in Figure 3.

The results indicate major improvements in terms of economics (TAC), Coefficient of Performance (COP), and cycle simplicity (2 evaporation levels with 2 direct contact coolers and a total of 8 HXs vs. 2 evaporation levels, 4 intercoolers and 5 direct contact coolers with a total of 12 HXs of the reference solution). The key advantage of the proposed approach compared to the one adopted in Grossmann and Shelton (1986), is the possibility of optimizing the heat transfer temperature differences and the cycle pressure levels (this enables using more efficient heat integration options), the higher flexibility of the cycle superstructure (including more intercooling/throttling options) and the capability of considering simultaneously capital and operating costs.

Table 1. Main Results of Test Case Optimization

Reference TAC (k€/yr)	Optimized TAC (k€/yr)	Improvement
142,2	107,6	24,30%
Reference COP	Optimized COP	Improvement
5,73	9,59	67,36%

Table 2. Components of the TAC

	Reference Cost (k\$/yr)	Optimized Cost (k\$/yr)
Ann. CAPEX HEN	30,814	40,832
Ann. CAPEX COMP	57,135	27,011
Elect. purch. Cost	31,461	15,625
Steam import cost	22,822	24,147
TAC	142,232	107,615

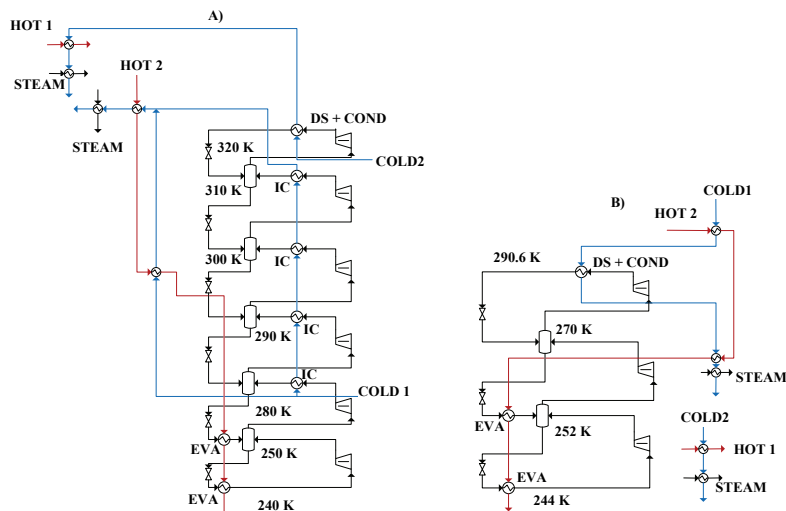


Figure 3. (A) Reference solution by Shelton and Grossmann (1986) and (B) optimized structure

6. Conclusions

The results of the test case show that the proposed approach is a very promising tool for the design of efficient and cost-effective refrigeration cycles. This is due to its capabilities of: (i) of integrating a flexible cycle superstructure with the HEN superstructure, (ii) optimizing cycle pressures and temperatures, (iii) considering both capital and operating costs in rigorous way, (iv) relying on an efficient (although without guaranteed global convergence) MINLP decomposition algorithm.

Future developments of the methodology will focus on the development of globally convergent decomposition algorithms.

References

- C. Elsido, E. Martelli, I. E. Grossmann, 2019, "A Bilevel Decomposition Method for the Simultaneous Heat Integration and Synthesis of Steam/Organic Rankine Cycles", *Computers & Chemical Engineering*, Vol. 128, 2 September 2019, pp. 228-245
- J. I. Manassaldi, M. C. Mussati, N. J. Scenna, S. Mussati, 2019, "Development of extrinsic functions for optimal synthesis and design—Application to distillation-based separation processes", *Computers & Chemical Engineering*, Vol. 125, pp. 532-544
- E. Martelli, C. Elsido, A. Mian, F. Marechal, 2017, "MINLP model and two-stage algorithm for the simultaneous synthesis of heat exchanger networks, utility systems and heat recovery cycles", *Computers & Chemical Engineering*, Vol. 106, pp. 663-689
- M. R. Shelton, I. E. Grossmann, 1986, "Optimal synthesis of integrated refrigeration systems – I. Mixed Integer Programming Model", *Computers & Chemical Engineering*, Vol. 10, No. 5, pp. 445-459
- A. S. Wallerand, M. Kemani, I. Kantor, F. Marechal, 2018, "Optimal heat pump integration in industrial processes", *Applied Energy*, Vol. 219, pp. 68-92
- T. F. Yee, I. E. Grossmann, 1990, "Simultaneous optimization models for heat integration—II. Heat exchanger network synthesis", *Computers & Chemical Engineering*, Vol. 14, Issue 10, pp. 1165-1184

Waste- to-DME: Process Synthesis and Sustainability Evaluation

Lefu A. Litheko^a, Bilal Patel^a

^aDepartment of Chemical Engineering, University of South Africa, Corner of Christiaan de Wet Road and Pioneer Avenue, Florida 1709, South Africa

Abstract

A macroscopic graphical targeting approach based on the C-H-O ternary diagram is proposed for the synthesis of waste-to-chemicals systems. Sustainability metrics are also considered to gain insight into the process efficiency and resource utilization. The process targets together with the sustainability metrics enable the screening, evaluation and comparison of various process routes. A case study for the conceptual design of a waste-to-DME facility is considered. The process includes waste gasification and DME synthesis. The DME production process has two main process targets, one based on the indirect synthesis route and the direct synthesis route, which require a syngas platform with compositions of $H_2/CO = 1$ and $H_2/CO = 2$ respectively. To achieve the syngas target required for the direct DME synthesis route, a mixture of H_2O and CO_2 or H_2O and O_2 is required as gasifying agents. A simplified input-output diagram of the overall waste-to-DME process was developed based on the process targets for the waste gasification and the direct DME synthesis process. The carbon efficiency, atom economy and global warming potential (GWP) is calculated. An additional source of hydrogen is required to achieve the syngas composition target for the indirect DME synthesis route (i.e. $H_2/CO = 2$). Introducing methane (from landfill gas) to the process results in a substantially higher carbon efficiency.

Keywords: waste biorefineries, sustainability, process synthesis, MSW, circular (bio) economy

1. Introduction

The rapidly growing consumption and commensurate depletion of the world's natural resources as well as the production of large volumes of waste has raised concerns regarding the sustainability of the current linear economy model. A shift to a circular economy model has been proposed to substantially improve resource efficiency and reduce consumption of fresh resources (Maina et al., 2017).

The valorisation of municipal solid waste (MSW) is an important aspect of a circular bioeconomy. In 2016, more than 2 billion tons of waste was generated with an increase of 70% to 3.4 billion tons in 2050 predicted (Kaza et al., 2018). Typical disposal methods like landfilling are no longer viable options due to stringent policies, limited space and environmental and health issues. The organic fraction of MSW (OFMSW) is the most abundant fraction of MSW (about 50 - 60%). The OFMSW has a relatively high carbon content (38- 51%) and calorific value (14-20 MJ/kg) (Campuzano and Gonzalez-Martinez, 2016). Waste biorefineries or waste-to-chemicals, that employ the OFMSW, play an important role in sustainably converting low-value streams to high value products (Carus & Dammer, 2018) and achieving a circular bioeconomy.

Thus, it is imperative to develop methodologies for the conceptual design of efficient and sustainable waste biorefineries that produce fuels, chemicals and other products. Insight-based approaches are useful to simplify the complexity of the waste biorefinery design problem, provide a fundamental understanding of design issues and identify holistic targets for processes (Kokossis et al., 2015). Tay et al. (2011) proposed a graphical approach, based on the C-H-O ternary diagram, for a gasification-based biorefinery to determine the syngas equilibrium composition. This study extends the use of the C-H-O ternary to waste biorefineries and presents a holistic approach to determine targets for the overall process that incorporate the downstream processes.

2. Modelling framework

Consider a waste biorefinery that converts Refuse Derived Fuel (RDF) into dimethyl ether (DME). Although the study presents an analysis for DME as the final product, any other product that contains carbon, hydrogen and oxygen can be analysed in a similar fashion. DME is used as an example to illustrate the methodology and may not represent the “best” product to produce from waste. However, the C-H-O ternary diagram (or the Van Krevelen diagram) may be applied to provide insights into which product should be produced based on carbon efficiency (Patel, 2015).

Since the system consist of inputs and outputs that consist mainly of carbon (C), hydrogen (H) and oxygen (O) atoms, a C-H-O ternary diagram is convenient graphical tool to synthesise and analyse processes at the conceptual design level (Tay et al., 2011; Litekho, 2017). For example, waste feedstock can be expressed as $C_xH_yO_z$, where x-z represent the atomic number of carbon (C), hydrogen (H), and oxygen (O) respectively, which can be obtained from the ultimate analysis of the feed material. The minor components of biomass such as nitrogen and sulphur are not considered. Possible process inputs for the biorefinery include RDF, H_2O , CO_2 , and O_2 . In this study, RDF has been specified with a chemical formula of $CH_{1.569}O_{0.541}$ (Onel et al., 2014). The thermodynamics properties of RDF are given as a function of composition according to the correlations developed for biomass (Peduzzi et al., 2016). By utilising the lever arm rule, the material balance targets for a process can be determined. By incorporating the first and second laws of thermodynamics, energy and work targets can also be obtained (Patel, 2015).

The overall material balance for the process can be written as follows:

$$a CH_{1.569}O_{0.541} + b H_2O + c O_2 + d CO_2 - 1 C_2H_6O = 0 \quad (1)$$

where a, b, c and d are the respective amounts of biomass, water, O_2 and CO_2 required (if the coefficient is positive) or produced (if the coefficient is negative) in the process for converting RDF into 1 mole of DME.

The process targets for the waste-to-DME process are assessed using sustainability metrics to determine the efficiency and/or the environmental performance of this process. Carbon efficiency, atom economy and global warming potential (GWP) in terms of CO_2 equivalence (Eq. (2)) are used to assess the process. Other sustainability metrics may also be applied to the process targets. For example, the WAR algorithm, inherent safety index and economic potential could be used to assess the process targets (Zeng et al., 2012).

$$CO_2 \text{ Equivalent} = \sum_{i=1}^n (Emissions \text{ to air})_i \times (GWP)_i \quad (2a)$$

$$\text{Atom Economy} = \frac{\text{Mass of desired product}}{\text{Total mass of feed}} \quad (2b)$$

$$\text{Carbon Efficiency} = \frac{\text{Moles of C in desired product}}{\text{Moles of C in feed}} \quad (2c)$$

The process targets developed represent an early-stage design of a waste biorefinery. These targets can be used to complement and enhance further detailed optimisation studies as well as detailed techno-economic and sustainability assessments (for example, life cycle analysis).

3. Results and Discussion

3.1. Process targets for overall system

The process target of the waste-to-DME process is represented on the C-H-O ternary diagram (Figure 1). The feed (input) line connects RDF and the H₂O/O₂ co-feed stream, while the product line joins CO₂ and DME. The reactant and product lines intersect at Point M which designates the target mixture of reactants (RDF, O₂ and H₂O) and products (DME and CO₂).

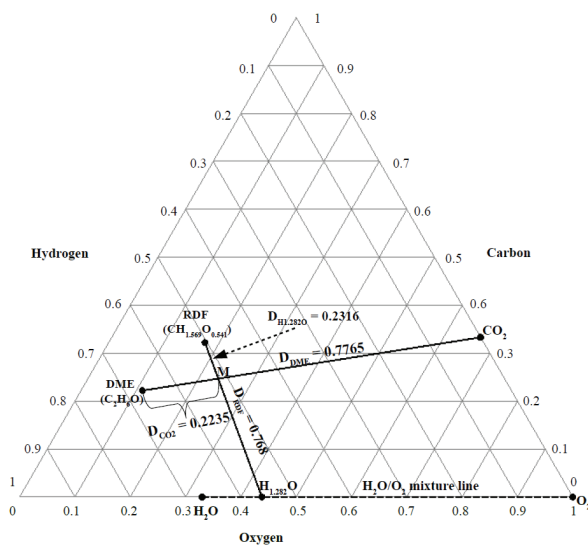
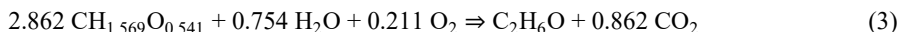


Figure 1. Material balance for the overall RDF-to-DME process

The overall material balance for the waste biorefinery process is

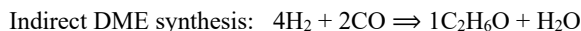
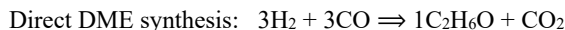


It is clear from Eq. (3) that producing 1 mole of DME from RDF requires oxygen and water as feed material. In addition, the process does not require an energy input ($\Delta H=0$) and it is thermodynamically feasible ($\Delta G < 0$). The process target show that the atom economy and carbon efficiency for the RDF-to-DME process are 55% and 70% respectively. The process also emits 0.825 tons CO₂ per ton of DME. The atom economy shows that just over half of the RDF fed to the process will be converted to DME. The

carbon efficiency indicates that of the carbon introduced to the process, 30% ends up as carbon dioxide.

3.2. Process targets for sub-systems

The overall waste-to-DME process involves two main processes - gasification and DME synthesis. DME can be produced via two process routes represented by the following material balance equations (Azizi et al., 2014):



The required syngas composition is $\text{H}_2/\text{CO} = 1$ for the direct route and $\text{H}_2/\text{CO} = 2$ for the indirect route.

Gasification of the RDF to produce syngas requires a gasifying medium, typically steam, oxygen, CO_2 or a combination of these. The C-H-O ternary diagram can be used to determine the process targets for the gasification process in which RDF is gasified using different gasifying agents to produce H_2 and CO as syngas product.

The input and output of the gasification process are plotted on the C-H-O ternary diagram (Figure 2). A product line is drawn by connecting the gasification products, i.e. H_2 and CO . The required syngas targets for the direct route ($\text{H}_2/\text{CO} = 1$) and the indirect route ($\text{H}_2/\text{CO} = 2$) are plotted based on their atomic composition as M1 and M2 respectively.

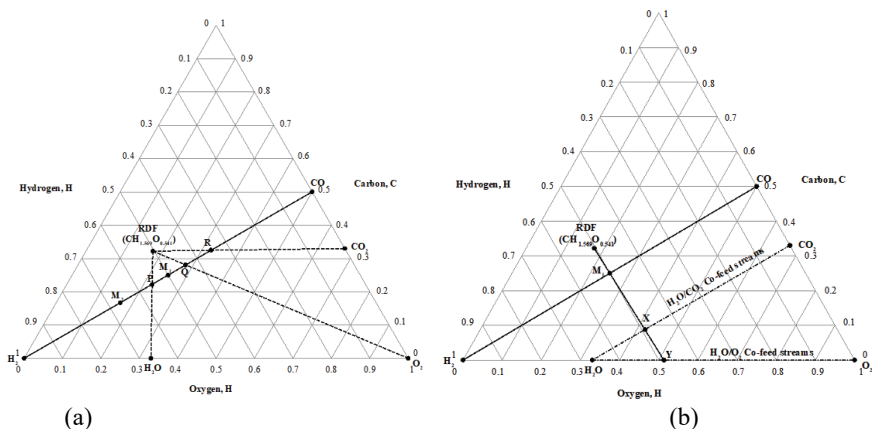
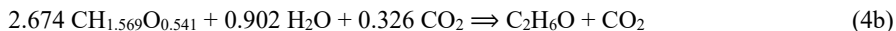
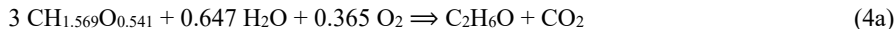


Figure 2. Evaluation of waste gasification (a) gasifying agents and (b) co-gasifying agents on a C-H-O ternary diagram

Feed lines that connect RDF with the three possible individual gasifying agents, i.e. H_2O , CO_2 and O_2 are drawn. These lines intersect the product line at points P, Q and R which represent syngas produced when RDF is gasified with H_2O , O_2 and CO_2 respectively. The gasifying agent that results in a gas composition that is on or closest to either M1 or M2 is considered optimum. From Figure 2, none of the points from the three gasifying agents is located on either M1 or M2. This means that gasifying RDF with H_2O , O_2 or CO_2 individually cannot produce syngas that is exactly $\text{H}_2/\text{CO} = 1$ or $\text{H}_2/\text{CO} = 2$. Therefore, the syngas compositions produced from the three individual gasifying agents do not meet the requirements for the intended downstream application.

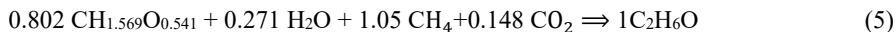
It can, however, be deduced from Figure 2 that it is feasible to produce syngas with $H_2/CO = 1$ by co-feeding H_2O with either CO_2 or O_2 . Co-feeding H_2O with either CO_2 or O_2 shifts Point P toward M1. However, producing a syngas composition of $H_2/CO = 2$ from the gasification of RDF using any combination of the gasifying agents is infeasible. All possible co-feed streams of H_2O with CO_2 (H_2O/CO_2) and H_2O with O_2 (H_2O/O_2) are represented by lines that connect H_2O and CO_2 as well as H_2O and O_2 respectively. The optimum co-feed streams for the RDF gasification are determined by drawing a feed line from RDF point through the target syngas composition, M1. The feed line intersects the two co-feed stream lines at points labelled X and Y (Figure 2b). The two points represent the optimum co-feed streams that produce the target syngas composition of $H_2/CO = 1$. Point X represents the co-feed stream of H_2O with CO_2 whereas Y represents the co-feed stream of H_2O with O_2 . By identifying the atomic species of X and Y and using the lever arm rule on the C-H-O ternary diagram, the material balance of the two process targets can be determined.

The overall RDF-to-DME process integrates the gasification process and the direct DME synthesis process. Integrating the two gasification process targets with the DME synthesis process, results in the following material balances:



The process that involves RDF gasification with H_2O and O_2 (Eq. 4a) is exothermic ($\Delta H = -70.2$ kJ/mol). The excess heat can either be used for other purposes or is lost to the environment. On the other hand, the process involving the gasification of RDF with H_2O and CO_2 (Eq. 4b) requires additional energy ($\Delta H = 95.9$ kJ/mol). For the process to be energy self-sufficient, the required energy can be supplied through combustion of additional RDF. This would result in obtaining the overall target given by Eq. 3.

As shown in Figure 2a, producing syngas with a $H_2/CO = 2$ for the downstream indirect DME synthesis is infeasible from RDF gasification with all three oxidants (i.e. H_2O , O_2 and CO_2) as well as their combinations. The syngas platform is deficient in H_2 and therefore a hydrogen source is required to produce the syngas composition target of $H_2/CO = 2$. Methane (which can be obtained from biogas or landfill gas) can be introduced into the process as a source of hydrogen. The optimum amount of methane that meets the required gasification target composition is specified from the C-H-O diagram. Thus, the overall process targets of the RDF-to-DME through the indirect DME synthesis is giving by



Eq. 5 has a carbon efficiency and atom economy of 100% and CO_2 is utilised as a feed and is not emitted from the process.

4. Conclusions

A targeting approach based on the C-H-O ternary diagram and laws of thermodynamics is presented for a waste-to-chemicals systems. The C-H-O ternary diagram is a powerful graphical tool that enable quick and insightful for representing of the overall material balance targets for waste-to-chemical processes. Various process configurations can be considered by varying process parameters and these configurations can be rapidly evaluated by applying sustainability metrics such as carbon efficiency and atom economy.

A case study for the conceptual design of a waste-to-DME facility was considered. An overall material balance target is obtained that is energy self-sufficient. The carbon efficiency, atom economy and global warming potential (GWP) is calculated. Furthermore, ways of obtaining the overall target are explored by considering a process that includes waste gasification and DME synthesis. The DME production process has two main process targets, one based on the indirect synthesis route and the direct synthesis route, which require syngas platform with compositions of $H_2/CO = 1$ and $H_2/CO = 2$ respectively. To achieve the syngas target required for the direct DME synthesis route ($H_2/CO = 1$), a mixture of H_2O/O_2 or H_2O/CO_2 can be utilised. The requirements of these gasifying agents can easily be determined from the C-H-O diagram. An additional source of hydrogen is required to achieve the syngas composition target for the indirect DME synthesis route (i.e. $H_2/CO = 2$). Introducing methane (from landfill gas) into the process results in a substantially higher carbon efficiency and atom economy

References

- Z. Azizi, M. Rezaeimaneh, T. Tohidian, and M.R. Rahimpour, 2014, Dimethyl ether: A review of technologies and production challenges. *Chemical Engineering and Processing: Process Intensification*, 82, 150-172.
- R. Campuzano, and S. González-Martínez, 2016, Characteristics of the organic fraction of municipal solid waste and methane production: A review. *Waste Management*, 54, 3-12.
- M. Carus and L. Dammer, 2018, The circular bioeconomy—concepts, opportunities, and limitations. *Industrial biotechnology*, 14(2), 83-91.
- S. Kaza, L. Yao, P. Bhada-Tata, and F. Van Woerden, 2018, What a waste 2.0: a global snapshot of solid waste management to 2050. The World Bank.
- A.C. Kokossis, M. Tsakalova, and K. Pyrgakis, 2015. Design of integrated biorefineries. *Computers & Chemical Engineering*, 81, 40-56.
- L.A. Litekho, 2017. Conceptual design of gasification-based biorefineries using the CHO ternary diagram (Dissertation)
- O. Onel, A.M. Niziolek, M.F. Hasan, and C.A. Floudas, 2014, Municipal solid waste to liquid transportation fuels—Part I: Mathematical modeling of a municipal solid waste gasifier. *Computers & Chemical Engineering*, 71, 636-647.
- B. Patel, 2015, A Thermodynamic Targeting Approach for the Synthesis of Sustainable Biorefineries, *Computer Aided Chemical Engineering*, 37, 1283-1286
- E. Peduzzi, G. Boissonnet, and F. Maréchal, 2016, Biomass modelling: Estimating thermodynamic properties from the elemental composition. *Fuel*, 181, 207-217.
- D.H.S. Tay, D.K.S Ng, H. Kheireddine, and M.M. El-Halwagi, 2011, Synthesis of an integrated biorefinery via the C–H–O ternary diagram. *Clean Technologies and Environmental Policy*, 13(4), 567-579.
- K. Zheng, H.H. Lou, P. Gangadharan, and K. Kanchi, 2012. Incorporating sustainability into the conceptual design of chemical process-reaction routes selection. *Industrial & Engineering Chemistry Research*, 51(27), 9300-9309.

Long-distance hydrogen delivery using a toluene-based liquid organic hydrogen carrier system

Mariana C. S. Carvalho^a, João Marques^b, Henrique A. Matos^a and José F.O. Granjo^{a,*}

^a*CERENA, DEQ, Instituto Superior Tecnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal*

^b*Galp, R. Tomás da Fonseca - Torre C, 1600-209 Lisboa, Portugal*

Abstract

Liquid organic hydrogen carriers (LOHCs) can be used for long-term storage and long-distance transport of hydrogen. Here, hydrogen (H₂) is loaded into the organic molecule through a hydrogenation process and then unloaded through the reverse process (dehydrogenation). These organic molecules allow hydrogen to be handled at ambient conditions and transported using the existing crude oil-based infrastructure.

In the present work, the system toluene-methylcyclohexane was identified as being one of the most promising LOHCs. Simulations for the hydrogenation and dehydrogenation processes were performed in Aspen Plus[®] for a 50 ktonne H₂/year capacity. To increase the detail and reliability of the results, both reactions were modelled using kinetic laws and by-product formation was considered. The processes were optimized through heat integration, which enabled up to 60% savings of hot utilities. An economic analysis was held, from which the break-even price of hydrogen loading and release was found to be equal to 1.9 \$/kg-H₂.

Keywords: Hydrogen storage, Liquid organic hydrogen carrier, Toluene, Process integration, Techno-economic analysis.

1. Introduction

Public pressure has increased over the years to seek alternatives for fossil-based fuels and products while mitigating the adverse effects of climate change. Renewable energies have been contributing decisively to the decarbonization of power systems, but intermittencies associated with these energy sources have hindered their deeper incorporation into the energy mix. In this context, low-carbon hydrogen (green or blue) can play a key role as a clean energy vector because it can be stored and transported in many ways (e.g., liquified, pure or blended gas, and reversibly bonded with liquid or solid compounds), and deployed as a fuel or as a chemical. However, using hydrogen as an energy vector also faces important challenges, particularly those associated with long-distance transport and long-term storage due to its rather low volumetric energy density and extreme volatility (Usman, 2010). Recently, liquid organic hydrogen carriers (LOHCs) have been screened for hydrogen transport and storage. The concept, depicted in Figure 1, involves a two-step cycle: loading (hydrogenation) and release

(dehydrogenation). The efficiency of these two steps and the inherent aspects of the LOHCs supply chains affect the final delivery costs of hydrogen decisively.

The chemical compounds receiving more attention for hydrogen transport include naphthalene, formic acid, methanol, N-ethylcarbazole, toluene, and dibenzyltoluene. The toxicity of naphthalene, decomposition reactions of N-ethylcarbazole, and the slow kinetics of hydrogenation and dehydrogenation reactions associated with formic acid and methanol still limit the adoption of these compounds as LOHCs. Contrarily, recent results of techno-economic analyzes for toluene and dibenzyltoluene cases have shown they are economically viable for the long-range transport of hydrogen (e.g., Niermann et al., 2019). However, these studies do not consider the mechanisms involved in hydrogenation and dehydrogenation reactions, neither the occurring side reactions, particularly in dehydrogenation. Furthermore, simplified models were used to design most unit operations involved, preventing optimization studies from being conducted.

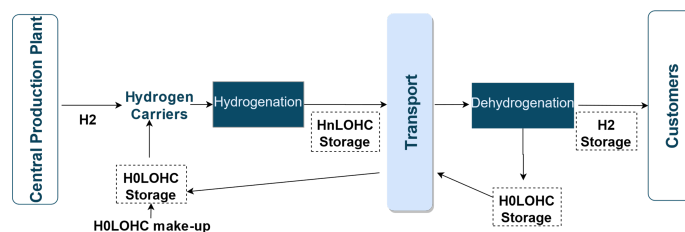


Figure 1: Two-way LOHC concept for hydrogen transportation.

This work reports ongoing research where various LOHCs are being screened for long-distance transport and storage of H₂. Here, the competitiveness of toluene as a LOHC is addressed, where hydrogenation and corresponding methylcyclohexane (MCH) dehydrogenation processes are analyzed. Kinetic models of the main reactions were used, and side reactions were considered to model the hydrogenation and dehydrogenation reactions with detail. The separation subsystems for the LOHC recovery and hydrogen purification were designed and operating conditions established using flowsheet optimization to minimize costs.

Next, the hydrogenation and dehydrogenation processes involved are described, and the modelling framework used for the techno-economic analysis presented (Section 2). In Sections 3 and 4, we show preliminary process simulation results using Aspen Plus[®] and the economic analysis. The paper closes in Section 5 with the final remarks.

2. Process overview and modelling

The toluene hydrogenation and methylcyclohexane (MCH) dehydrogenation processes, as simulated in Aspen Plus[®], are represented in Figure 2 and Figure 3, respectively. Peng-Robinson equation-of-state was used as the thermodynamic model because both processes involve hydrocarbons and are operated at moderate and high pressure. For the toluene hydrogenation reaction ($C_7H_8 + 3H_2 \rightarrow C_7H_{14}$) a non-commercial supported nickel catalyst (Ni/Al₂O₃) was selected from (Lindfors & Salmi, 1993), for which a kinetic law was derived for the range 120-200 °C:

$$r'_1 = \frac{k_1 K_{1,T} K_{1,H}^3 p_H^3}{(K_{1,T} p_T + (K_{1,H} p_H)^{0.5} + 1)^7}, \quad \text{mol}/(\text{g} - \text{cat. s}) \quad (1)$$

$$K_{1,T} = \exp K'_{1,T} \exp(-\Delta H_T/RT), \quad \text{atm}^{-1} \quad ; \quad (2)$$

$$K_{1,H} = \exp K'_{1,H} \exp(-\Delta H_H/RT), \quad \text{atm}^{-1} \quad (2)$$

$$k_1 K_{1,T} K_{1,H}^3 = A \exp K'_{1,T} K_{1,H}^3 \exp(-E_a/RT - \Delta H_T/RT - 3\Delta H_H/RT), \quad (3)$$

Where $\ln A = 69.254$, $\ln K'_{1,T} = -15.597$, $\ln K'_{1,H} = -39.666$, $E_a/R = 21288 \text{ K}$, $\Delta H_T/R = -7368.3 \text{ K}$, $\Delta H_H/R = -18097 \text{ K}$. Similarly, for the MCH dehydrogenation reaction, a single-site Langmuir-Hinshelwood kinetic model was developed for platinum supported catalyst (Pt/Al₂O₃) based on (Usman, 2010):

$$r'_2 = \frac{40.907 k p_M \left(1 - \frac{p_T p_H^3}{K_2 p_M}\right)}{1 + 40.907 p_M + 22.194 p_T + K'_2 p_T p_H^2}, \quad \text{mol}/(\text{g} - \text{cat. s}) \quad (4)$$

where

$$k = 4.064 \times 10^{-5} \exp \exp(7.652(1 - 617.2/T)), \quad \text{mol}/(\text{g} - \text{cat. s}) \quad (5)$$

$$K_2 = 3600 \exp \exp((-217650/R)(1/T - 1/650)), \quad \text{bar}^3 \quad (6)$$

$$K'_2 = 6.688 \exp \exp(-24.038(1 - 617.2/T)), \quad \text{bar}^{-3} \quad (7)$$

Here, subscript *T* represent toluene; *H*, hydrogen; and *M*, methylcyclohexane. At the conditions of MCH dehydrogenation (340-460 °C), the major side reactions reported include the formation of benzene (C₆H₆), xylene isomers (C₈H₁₀) and methane (CH₄):



For the abovementioned catalytic system, no kinetic data is available for reactions (8) and (9), but the observed reaction conversions were approximately 8.3E-4 and 2.7E-3, respectively (Usman, 2010).

Kinetic models were used to plot isobaric plots of single-pass conversion against temperature and adiabatic plots to determine the appropriate catalyst loadings and reaction conditions. Thus, the hydrogenation reaction was set to 30 bar and 170 °C, using a hydrogen molar excess equal to 6 and catalyst loading of 3065 kg, which led to a conversion equal to 99.98%. As for the dehydrogenation, the reaction occurred at ambient pressure and 450 °C, with a catalyst loading equal to 17000 kg, enabling 97.34% conversion. The fixed-bed reactors were simulated with a plug-flow reactor model (RPlug) in which the main reaction occurred and, for the dehydrogenation process, the side-reactions were modelled using the RStoic model (Aspen Technology, Inc., 2001).

Two scenarios of hydrogen compression in the toluene hydrogenation process were here considered: (i) compression from ambient conditions up to 30 bar; and (ii) hydrogen feeding at 70 °C and 30 bar, typical conditions at which it leaves the electrolyzer (David

et al., 2019). In scenario (i), the multi-compression section was designed by minimizing the energy consumption where per-stage compression ratios are adjusted, and the exit gas temperatures maintained below 204 °C to avoid exceeding equipment constraints (Luyben, 2011). Intercoolers were only used for the first two stages, as the excess heat from the last stage was used to vaporize the toluene feed. After the reactor, a distillation column, D101, is used to recover the unreacted toluene and the hydrogen excess. This equipment was initially designed using approximate methods to estimate the number of theoretical stages and feed stages and later optimized to minimize energy consumption using RadFrac rigorous distillation model (Aspen Technology, Inc., 2001).

In the MCH dehydrogenation process, H₂ is recovered by flashing the product stream and compressing it to the pressure swing adsorption (PSA) unit to obtain a hydrogen stream within fuel cell specifications (>99.97 mol%). Intercoolers and flash units are added to recover the hydrocarbon condensates and pre-concentrate H₂. The off-gas stream from the PSA containing the non-recovered hydrogen was recycled to the compression area feed, and a purge added to the accumulation of impurities. The purge stream feeds a burner unit which supplies heat to the reactor inlet stream.

Toluene, on the other hand, is recovered from the flash liquid stream through a three-distillation column scheme, which was required due to the close boiling points between toluene, MCH, benzene and xylene and also due to the high amount of MCH present in this stream. The first column, CD201, obtains the bulk of the toluene product stream and the second, which receives the distillate stream from the first column and the effluent stream from the PSA unit, separates the light contaminants (benzene and MCH) from the remaining toluene. Finally, the third column, CD203, was added to remove benzene from the MCH stream, which, in turn, is recycled to the reactor.

3. Process simulation results

The results for a delivery rate of 50 ktonne/year of H₂ and 8000 h of annual operation are now presented. In the hydrogenation process, a virtually complete reaction enables a simple purification scheme using a 7-stage distillation column, which operates at 25 bar with a reflux ratio equal to 0.175. This equipment was adjusted so that the minimum amount of MCH was contained in the recycle stream. The distillate from D101 removes most of the MCH, but still, 2.3% of MCH is recycled with hydrogen, which has a minor impact on the reaction rate, as expected from equations (1-3).

As for the dehydrogenation process, the analysis confirms the need for further purification, so that more concentrated product streams can be obtained, as both streams 209 and 216 only have 97.2 wt% hydrogen and 71.9 wt% toluene, respectively. The PSA scheme was well-suited for the hydrogen purification as the final product stream is within specifications. Regarding the toluene purification scheme, vacuum conditions were applied in all columns to facilitate the separation, while the operating pressure was estimated so that the temperature in the overhead condensers is not lower than 50 °C, to allow cooling water to be used (Turton et al., 2008). The molar reflux ratio (RR) and distillate to feed ratio (DF) parameters were determined through optimization. RR varied between 18 and 25 and DF between 0.1 and 0.15 in columns CD201-CD203. The number of theoretical stages in CD201 was 20, in CD202 was 12, and in CD203 was 10.

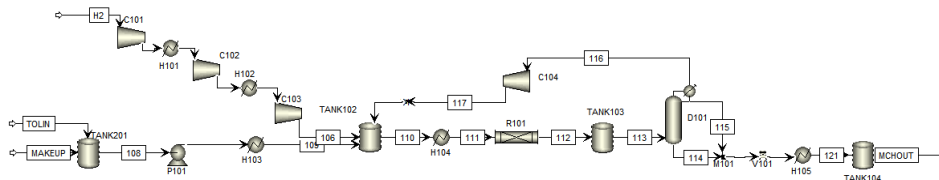


Figure 2: Toluene hydrogenation process.

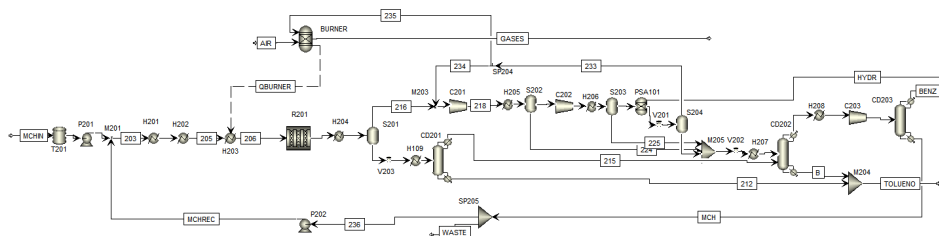


Figure 3: Methylcyclohexane dehydrogenation process.

Globally, 1.38 % of MCH in the feed and 0.03% of the toluene exiting R201 are lost in the purification process. This purification scheme can be improved using other purification methods such as extractive distillation for benzene-toluene-xylene mixtures (Kim, 2016).

4. Heat integration and economical analysis

Heat integration was conducted for both processes using Aspen Energy Analyzer[®]. The minimum temperature difference (ΔT_{\min}) assumed was 5 °C. The optimal heat exchanger network (HEN) enabled a 63% reduction of the hot utility consumption in the hydrogenation process while increasing the number of heat exchangers (HEs) from 8 to 18 and the total area from 3948 m² to 8067 m². In the dehydrogenation, a 58% reduction of hot utility consumption was achieved, and the number of HEs increased from 15 to 23, as well as the total area from 17690 m² to 26270 m². The economical analysis was performed using the CapCost tool, and it estimates the fixed capital cost and total manufacturing cost of all plants previously simulated, according to the method developed by (Turton et al., 2008). The results are summarized in Table 1.

Table 1: Summary of fixed capital cost and utilities expenses.

Consumption (kWh/kg-H ₂)	Hydrogenation		Dehydrogenation
	Scenario (i)	Scenario (ii)	
Electrical	2.24	0.119	1.40
Heat	1.05	3.36	15.9
Total	3.29	3.48	17.3
CAPEX (M\$)	60.4	4.54	78.3

As the compressors account for 80% of the capital costs in scenario (i), it is already expected that the fixed capital costs scenario (ii) is drastically lower. Nevertheless, for the same reason, the decrease in the number of heat exchangers available for heat integration leads to an increase in heat consumption, which offsets the decrease in electrical consumption. The energy consumptions reported here for the toluene hydrogenation process in both scenarios (3.29-3.48 kWh/kg-H₂) are higher than 0.47 kWh/kg-H₂ estimated from (Niermann et al., 2019). This difference can be explained by the fact that in (Niermann et al., 2019) the hydrogenation reaction is taken to occur at the liquid phase and solvent recovery was not considered. For the dehydrogenation process, the total energy consumptions here reported (17.3 kWh/kg-H₂) is slightly lower than 17.85 kWh/kg-H₂ from Niermann et al., (2019). When considering both processes, the total operating costs obtained of H₂ loading and release was 1.3 \$/kg-H₂, and the total capital investment equal to 146.3 M\$, giving an estimated break-even price of 1.9 \$/kg-H₂.

5. Conclusions

In this paper, the hydrogenation and dehydrogenation process of the toluene-MCH system was simulated using kinetic laws to model the reaction, which increased the results reliability. For the toluene hydrogenation process, two scenarios were presented, and it was found that even though the absence of a compression zone significantly reduces the fixed capital cost, the utility cost is still maintained due to the decrease in the number of heat exchangers which heat can be integrated. As for the dehydrogenation process, the complex purification scheme leads to a significant increase in utilities and fixed capital costs compared to the hydrogenation. Additionally, the heat requirements of the reactor also contribute to the heat consumption increase. When comparing to the literature results, improvement opportunities were identified towards the purification scheme.

References

- Aspen Technology, Inc. (2001). *Aspen Physical Property Methods and Models 11.1*.
- David, M., Ocampo-Martínez, C., & Sánchez-Peña, R. (2019). Advances in alkaline water electrolyzers: A review. *Journal of Energy Storage*, 23, 392–403.
- Kim, Y. H. (2016). Energy Savings in the Benzene-Toluene-Xylene Separation Process Using an Extended Divided-Wall Column. *Chemical Engineering & Technology*, 39(12), 2312–2322.
- Lindfors, L. P., & Salmi, T. (1993). Kinetics of Toluene Hydrogenation on a Supported Ni Catalyst. *Industrial & Engineering Chemistry Research*, 32(1), 34–42.
- Luyben, W. L. (2011). Compressor Heuristics for Conceptual Process Design. *Industrial & Engineering Chemistry Research*, 50(24), 13984–13989.
- Niermann, M., Drünert, S., Kaltschmitt, M., & Bonhoff, K. (2019). Liquid organic hydrogen carriers (LOHCs) – techno-economic analysis of LOHCs in a defined process chain. *Energy & Environmental Science*, 12(1), 290–307.
- Turton, R., Bailie, R. C., Whiting, W. B., & Shaeiwitz, J. A. (2008). *Analysis, Synthesis and Design of Chemical Processes, Third Edition*. Upper Saddle River, N.J.
- U.S. Department of Energy. (2016). *Hydrogen Fuel Quality Specifications for Polymer Electrolyte Fuel Cells in Road Vehicles*.
- Usman, M. (2010). *Kinetics of Methylcyclohexane Dehydrogenation and Reactor Simulation for 'On-board' Hydrogen Storage*. University of Manchester.

Assessment of energy efficiency improvement in ceramic industry through waste heat recovery modelling

Miguel Castro Oliveira ^{a,b*}, Muriel Iten ^a, Henrique A. Matos ^b

^a *Low Carbon and Resource Efficiency, R&Di, Instituto de Soldadura e Qualidade, 4415 491 Grijó, Portugal*

^b *Centro Recursos Naturais e Ambiente (CERENA), Instituto Superior Técnico, Universidade de Lisboa, Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal*

Abstract

The ceramic industry is a sector within process industry associated to a high energy intensity which is owe to the operation of thermal processes such as firing, drying, spray drying and boiling. These processes are associated to high thermal energy losses (waste heat). As such, the adoption of waste heat recovery (WHR) technologies and strategies may reveal as an excellent opportunity to improve the energy efficiency of ceramic plants by improving the operation of these thermal processes. In practice, such may be performed by applying principles of process integration through assessing the waste heat potential of all the plant material streams and furtherly elaborate a plan to use this waste heat in the several different processes. In this work, it will be assessed the energy efficiency improvement potential associated to the implementation of a WHR strategy in a case-study ceramic plant. Such is carried out by modelling and simulate the plant and the WHR strategy using the Modelica language. Through the simulation of the developed model the techno-economic assessment, it was possible to accomplish the viability of the conceptualized project, with 3508 MWh/year of energy savings corresponding to 0.2 years of payback time.

Keywords: energy efficiency, waste heat recovery, process integration, ceramic industry, Modelica.

1. Introduction

Waste heat recovery (WHR) is a set of measures with a high potential to improve energy efficiency in industry. Industrial thermal processes are associated to high energy demands and in consequence to considerable thermal energy losses (roughly representing 20 - 50 % of total energy consumption), being estimated the existence of a total 300 TWh/year waste heat potential in the European industry (Castro Oliveira et al., 2019).

The WHR strategy includes the selection of the most favourable WHR technologies, the application of process integration and energy system optimisation principles, taking into account the objective to reduce energy inputs by making use of some material streams, namely the outlet ones with considerable waste heat potential.

The ceramic industry is a high energy intensive sector. Such is due to the operation of energy intensive processes, such as firing (kilns), drying and spray drying. Some of the most implemented WHR technologies within the ceramic industry are air-gas heat exchangers to preheat the combustion air at the entrance of combustion-based processes

Organic Rankine Cycle (ORC) to produce electric energy and hot air recycling (from the cooling zone of kilns).

In this work, it will be assessed the energy efficiency improvement potential associated to the implementation of a WHR strategy in a case-study. Such will be performed by the assembling of a ceramic plant model (developed in the Modelica language) to simulate the implementation of a WHR strategy.

2. Modelling of Thermal Processes and Process Integration

The case-study is about a ceramic sanitaryware manufacturing plant, in which the most prominent thermal processes include 2 tunnel kilns, 1 intermittent kiln, 2 dryers and 4 hot water boilers. These are combustion-based processes using natural gas as fuel. It currently exists an interest to reduce the fuel consumption of the tunnel kilns and the boilers. As such, a new WHR strategy will be proposed, considering principles of process integration.

2.1. Conceptualization of the Study

This process integration study emerged within the scope of the study of industrial system retrofitting, in which most recent technologies are set to be implemented to improve the operation of the plant through the rationalization of the use of its resources such as energy and water. In this case, such is performed by integrating WHR technologies to improve the plant's overall energy efficiency. The result of such implementation is the conceptualization of a heat integration system, which encompasses the plant's thermal processes and the interdependencies which are created through the implementation of WHR technologies (such as in this case the recovery of the kiln's exhaust gas stream to be directed to two water-gas heat exchangers).

To perform a study on heat integration, several methodologies exist, such as pinch analysis, exergy analysis, use of heuristics and linear programming. This work uses a methodology based on numerical modelling: custom models for the plant's thermal processes are created and the system is created through the assembling of several component models. In practice, this methodology allows the conceptualization of new systems inserted within the operation of the plant processes (the simulation of the occurring phenomena in real time). The models for the thermal processes (tunnel kiln and boilers) and the whole system were created using the Modelica language. Two custom model were developed for the tunnel kiln and the boiler using adapting the source code present in the open-source ThermoPower Modelica library (Politecnico di Milano).

This work emerged on the scope of the objectives of the European Green Deal and the EU 2050 long-term strategy, namely on the strategy for energy system integration. In specific, it dwells with the first pillar of this strategy, which is the requirement to improve the circularity of the energy systems. The implementation of energy recovery approached in this work is in this sense a contribution to the promotion of circular economy in industry.

2.2. Case-study characterization and model

The conceptualized WHR strategy encompasses 1 tunnel kiln and 2 of the boilers of the plant. For the tunnel kiln, it is implemented the hot air recycling from the cooling zones to the combustion chambers, while for the two boilers two water-gas heat exchangers (economisers) are implemented for each one respectively, recovering the waste heat from the tunnel kiln exhaust gases to preheat the inlet water stream at both boilers. The conceptualized system is represented in Fig. 1.

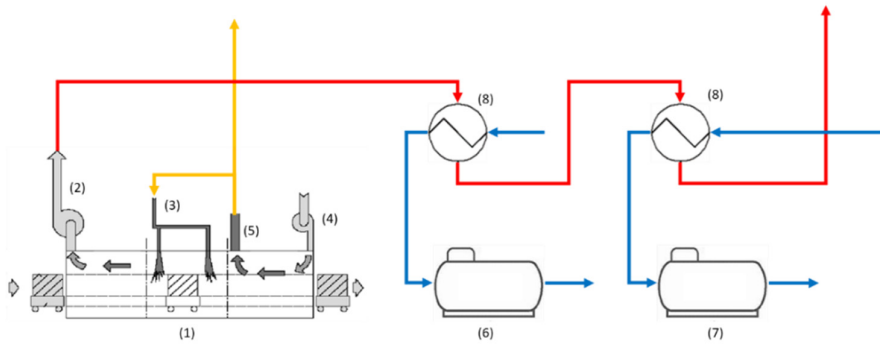


Fig. 1. Conceptualized heat integration system: (1) Tunnel kiln and its (2) exhaust gas chimney, (3) combustion chamber inlet, (4) cooling air inlet, (5) hot air chimney, (6) Boiler 1, (7) Boiler 2 and (8) Heat exchangers (Tunnel kiln scheme adapted from Dudam (2009))

The thermal process models were assembled considering the mass and enthalpy balances of a tunnel kiln and a boiler, respectively. In both cases, the total supplied heat (q_{supply} (kJ/h)) may be calculated considering the fuel's lower heating value (LHV), attending to equation (1). For the case in which more than one heat source exists to supply thermal energy to the kiln, it is necessary to consider an additional heat parcel ($q_{additional}$ (kJ/h)), as described by equation (2). In the case of the tunnel kiln, such corresponds to additional thermal energy contained in the combustion air stream ($q_{additional,kiln}$ (kJ/h)), as described by equation (3). In the case of the boiler, such corresponds to additional thermal energy contained in the inlet water stream ($q_{additional,boiler}$ (kJ/h)), as described by equation (4).

$$q_{supply} = M_{fuel,initial}LHV \quad (1)$$

$$q_{supply} = M_{fuel}LHV + q_{additional} \quad (2)$$

$$q_{additional,kiln} = M_{Comb.Air}C_{p,Comb.Air}(T_{Comb.Air} - T_{Comb.Air,initial}) \quad (3)$$

$$q_{additional,boiler} = M_{Water}C_{p,Water}(T_{Water} - T_{Water,initial}) \quad (4)$$

Where the variables above have the following meaning: $M_{fuel,initial}$ – Baseline case fuel mass flow rate (kg/h), M_{fuel} – Fuel mass flow rate (kg/h), $M_{Comb.Air}$ – Combustion air flow rate (kg/h), $C_{p,Comb.Air}$ – Air heat capacity (kJ/(°C.kg)), $T_{Comb.Air}$ – Combustion Air Inlet Temperature at the tunnel kiln (°C), $T_{Comb.Air,initial}$ – Baseline case Combustion Air Inlet Temperature (°C), M_{Water} – Water mass flow rate (kg/h), $C_{p,Water}$ – Water heat capacity (kJ/(°C.kg)), T_{Water} – Water inlet temperature at the boiler (°C), $T_{Water,initial}$ – Baseline case Water Inlet Temperature (°C).

The tunnel kiln model was assembled to allow the coupling of a PID controller, which controls the recycled air flow rate by receiving a set-point of the combustion air inlet temperature at the combustion chambers, thus simulating the recycling of hot air from the cooling zone of the kiln as a part of the total combustion air.

A set of assumptions were considered in the performance of the whole modelling:

- In respect to the tunnel kiln model, the air-to-fuel ratio is constant (not varying according to different conditions), as well as the enthalpy allocated to the ceramic product, the enthalpy allocated to the cooling air stream and heat losses;
- The ambient temperature is 25 °C and the ambient pressure is 1 bar;
- Pressure losses in the system are negligible.

The model created using OpenModelica 1.14.1, an open source distribution of Modelica, for the plant and the proposed WHR strategy is presented in Fig. 2.

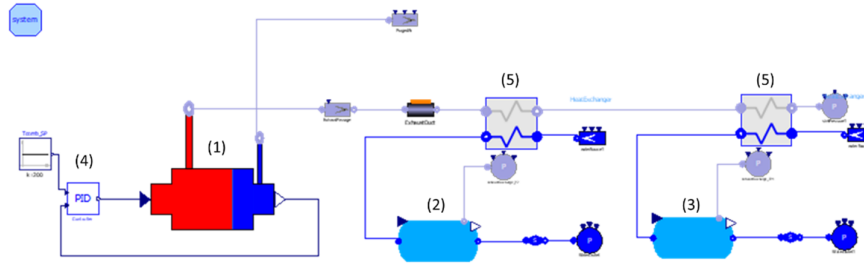


Fig. 2. Model for the ceramic plant and proposed WHR strategy: (1) Tunnel kiln, (2) Boiler 1, (3) Boiler 2, (4) PID Controllers, (5) Water-gas heat exchangers

2.3. Simulation Results and Model Validation

The model conceptualized for both components and plants are developed in a basis of dynamic simulation, although the simulation performed for the purpose of model validation was carried out in a steady state perspective – the simulation values tend to stabilize at a certain point of the simulation attaining constant values. The simulation results, real plant values and the respective deviations are presented in table 1.

Table 1. Simulation Results and Deviations from real plant values

Variable	Simulation Value	Real Plant Value	Deviation (%)
Flow rate of Exhaust Gases (Tunnel Kiln)	14455	14455	0.00
Temperature of Exhaust Gases (Tunnel Kiln)	159.4	159.3	0.09
Temperature of Hot Air (Tunnel Kiln)	93.0	93.73	0.79
Temperature of Hot Water (Boiler 1)	95.0	95.0	0.03
Temperature of Hot Water (Boiler 2)	95.0	95.0	0.03

Based on the results presented in table 1, it is possible to verify that the simulation results are overall consistent with the real plant measured data (the deviations for all the key parameters are lower than 1%).

2.4. Formulation of the Optimisation Problem

The process integration proposal may be adjusted to optimal operational conditions through the assessment of all the operational requirements needed to be considered and the energy consumption by thermal processes. This adjustment of operational conditions may be formulated as a simplified optimisation problem.

In the proposed optimisation problem, the constraints consist in limiting values for mass flow rates and temperatures of the exhaust gas and hot air streams. The hot air at the outlet of the tunnel kilns (the hot air that is purged to the environment and it is not recovered to the combustion chamber of the kiln) has associated minimum flow rate values, which is the hot air flow rate which have been recovered in the base case (5524.8 kg/h). Moreover,

as the exhaust gas must not be cooled down to the dew point temperature of the exhaust gas. Therefore, the exhaust gas stream obtain by the mixture of both kilns exhaust gas streams must be above 25 °C (a minimum temperature value). The objective-function is the minimization of fuel consumption (natural gas) in the tunnel kiln and boilers. The decision variable is the temperature of the combustion air stream at the inlet of the tunnel kiln (the set-point at the PID controller), which was set to 200°C. The decision variable is adjusted to achieve the objective while respecting the operational requirements. The mathematical formulation for the described optimization problem is presented in table 2.

Table 2. Mathematical formulation for the optimization problem

	Minimum Value	Variable	Maximum Value
Decision Variable	25°C	$T_{Comb,Air}$	200°C
Constraints	$M_{purged\ air} \leq 5524.8\ kg/h$		
	$T_{exhaust,outlet} > 25.0^{\circ}C$		
Objective-Function	$min(M_{fuel})$		

The implemented control system assisted on the implementation of the optimisation procedure (the decision variables are the PID associated controlled variables). Such allows the input of different values for the decision variable considering its variation with time, which is only possible since OpenModelica is dynamic simulation-based software. Due to the adopted steady state perspective adopted, the whole analysis performed for the techno-economic assessment is the minimum fuel consumption scenario.

3. Techno-economic assessment

The techno-economic assessment performed considers a simple payback, not considering an inflation rate of the capital costs associated to the technologies. As the system retrofitting approached in this work is essentially conceptualized in an energy efficiency improvement perspective, the overall aim is to optimise the associated plant energy costs. The capital expenditures, in its turn, correspond to the project overall investment cost (which includes the acquisition of ducts for the transportation of hot air, two water-gas heat exchangers and ducts for the transportation of the exhaust gases of the kilns). These are minimized in order to reduce the payback time associated to the implementation, although the values for the investment costs were fixed (considering the data from the catalogues of the technologies’ manufacturers).

The results for fuel savings obtained by the solving of the optimisation problem are presented in table 3. The savings were determined considering the natural gas price for Portugal of 13.28 €/GJ for industrial applications (INE-I.P./DGEG, 2018). The determination of payback time associated to the implementation of the proposed WHR strategy is presented in table 4. The calculation of this parameter was performed considering the results for the fuel savings obtained in the simulation and the total investment cost.

Table 3. Results for fuel savings

	Initial Case (kg/h)	Improved Case (kg/h)	Energy savings (MWh/year)	Savings (1000 €/year)
Kiln 1	103.5	81.4	2159	103.2
Boiler 1	10.95	4.21	659	31.5
Boiler 2	10.95	3.95	684	32.7
Total			3508	167.4

Table 4. Techno-economic assessment

Equipment	Investment Cost (€)	Savings (1000 €/year)	Payback Period (years)
Air Ducts	1754	167.4	0.2
Exhaust Gas Ducts	7700		
Two water-gas heat exchangers	25415		

Considering the payback time in which a measure is considered economically viable within the context of the European industry (2 – 3 years), the proposed WHR strategy may be considered highly economically viable.

4. Conclusions

The case-study analysed and assessed in this paper is inserted in a Portuguese ceramic plant. It is a plant with a high thermal energy consumption, with few implemented WHR strategies, currently existing an interest to reduce the fuel consumption of tunnel kilns and hot water boilers. In this sense, it was proposed and assessed the implementation of a new WHR strategy which included the hot air recycling from the cooling zone of the kilns and the installation of two water-gas heat exchangers.

Such study was performed by creating a model of the plant and WHR strategy using the Modelica language. The developed model was validated by comparing the simulation results with real plant measured data. It is verified that the simulation values are consistent with real values.

In a strategic perspective, this study revealed the significance of the use of a simulation tool for energy system-based decision making. In a technical perspective, it allowed to assess the viability associated to the project of heat integration system, having been obtained results of optimisation adding two water-gas heat exchangers, with a low payback time of 0.2 years.

The assembled References

- M. Castro Oliveira, M. Iten, P.L. Cruz, H. Monteiro, 2020, Review on Energy Efficiency Progresses, Technologies and Strategies in the Ceramic Sector Focusing on Waste Heat Recovery. *Energies*, 13 (22), 6096.
- V.P. Dudam, 2009, Numerical and Experimental Thermal Analysis of a Tunnel Kiln used in Ceramic Production, *J. of the Braz. Soc. of Mech. Sci. & Eng.*, 31 (4), 297 – 304.
- Politecnico di Milano, ThermoPower, build.openmodelica.org/Documentation/ThermoPower.html (accessed on 7 August 2020).
- EUROPA, Powering a climate-neutral economy: Commission sets out plans for the energy system of the future and clean hydrogen, https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1259 (accessed on 3 February 2021)
- INE-I.P./DGEG, 2018., PORDATA - Natural gas prices for households and industrial users (Euro), [https://www.pordata.pt/en/Europe/Natural+gas+prices+for+households+and+industrial+users+\(PPS\)-1480-313035](https://www.pordata.pt/en/Europe/Natural+gas+prices+for+households+and+industrial+users+(PPS)-1480-313035) (accessed on 7 August 2020).

Acknowledgements

The publishing procedure and conference participation have received funding by the European Union's Horizon 2020 research and innovation programmes under grant agreement “No. 810764”. This project has received funding from Portugal2020 03/SIAC/2016 under grant agreement “No. 26791”.

Targeting and Optimisation of Industrial and Urban Symbiosis for Circular Economy

Limei Gai^a, Petar Sabev Varbanov^{a,*}, Hon Huin Chin^a, Jiří Jaromír Klemeš^a, Sandro Nižetić^b

^a *Sustainable Process Integration Laboratory – SPIL, NETME Centre, Faculty of Mechanical Engineering, Brno University of Technology – VUT Brno, Technická 2896/2, 616 69 Brno, Czech Republic*

^b *LTEF – Laboratory for Thermodynamics and Energy Efficiency, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Rudjera Boskovicica 32, 21 000 Split, Croatia*

varbanov@fme.vutbr.cz

Abstract

One way of simultaneously minimising the use of fresh resources and pollution reduction is the practice of industrial and urban symbiosis, implementing Circular Economy. In the current work, it has been shown that the circularity rate alone is not a sufficient criterion to achieve resource and footprint minimisation. It is necessary to evaluate the cost and the footprints in a combined way. The formulated conceptual model allows the evaluation of the trends of the total cost, water and Greenhouse gas footprints. The provided case study clearly demonstrates the usefulness of the model in evaluating the trade-offs between the considered process characteristics, showing a 10 % difference between the optimal circularity rates for Water footprint minimisation and cost minimisation.

Keywords: Resource Integration, Resource Recovery, Industrial-Urban Symbiosis, Network Optimisation

1. Introduction

Resource recovery and reuse methods have been developed to cope with the problems of resource depletion and pollution of the environment. The main categories include Circular Economy (CE) (Hartley et al., 2020), Industrial Symbiosis (IS) (Domenech et al., 2019) and Process Integration (PI) (Klemeš, 2013). The conservation of energy and materials (Klemeš, 2013) is the most effective strategy to achieve sustainability, making the task of fresh resource supply easier. Fan et al. (2019) reviewed the methods for energy saving and pollution reduction, identifying the synergy between CE, resource, utility, and waste management as the route to make this strategy economically viable. The European Commission adopted a new Circular Economy Action Plan (EC, 2020) and policy recommendations (Hartley et al., 2020) from the Life Cycle perspective, and a set of new targets (Morsetletto, 2020) were proposed to facilitate the transition to a CE.

IS (Domenech et al., 2019) is a practical way to implement CE by sharing resources – materials, energy, water. A good example is the application of urban and industrial symbiosis for formulating a cross-industry network of multiple supply chains (Tseng et al., 2018) and fundamentally the Total Eco-Site Integration (Fan et al., 2021).

Many circularity indicators are available (Saidani et al., 2019). They can be applied to a multitude of contexts, ranging from multiple chemical species evaluation and up to regional or global system boundaries. A standardised metric for characterisation is the “Circular Material Use Rate” (CMU), officially employed by Eurostat (2018), defined as the fraction of reused materials from the total material use.

Material circularity reflects only one of the key dimensions in industrial and urban symbiosis systems. The energy dimension is also crucial, often prompting the energy-water nexus (Ahmad et al., 2020) or multi-tier interactions (Wang et al., 2019) to be considered as a whole. The importance of energy is in driving all involved processes—thermally, chemically, electrically and mechanically. In assessing the energy inputs and efficiency, it is necessary to account for the energy forms—heat and power, and the entropy generation, especially by mixing. This consideration leads to the use of exergy (Selicati et al., 2020) as the unifying criterion. Recent research has used exergy to evaluate the efficiency of material recovery from Cd-Te photovoltaic modules (Abadías Llamas et al., 2020) and evaluation of the exergy efficiency of biogas upgrading processes (Vilardi et al., 2020). A key principle is the use of embodied exergy and the linked embodied emissions for evaluating the process performance (Almeida et al., 2017). Circularity alone cannot be used as an ultimate indicator of sustainability. Environmental footprints (Holmatov et al., 2019) are also needed to quantify the impacts of both inputs and outputs.

Industrial and urban symbiosis systems have degrees of freedom for making design and operating decisions, while footprints and cost are used as performance criteria. The reviewed state of the art shows the trend to use cost, footprints or exergy alone, and in some cases cost and exergy have been used together, leaving exergy input as an auxiliary screening tool. This leaves a research gap – the use of all three criteria simultaneously for evaluating the sustainability of the symbiosis systems against the degree of circularity. The current work presents a model optimising industrial-urban symbiosis system by minimising the cost and the footprints (Čuček et al., 2012) against the circularity.

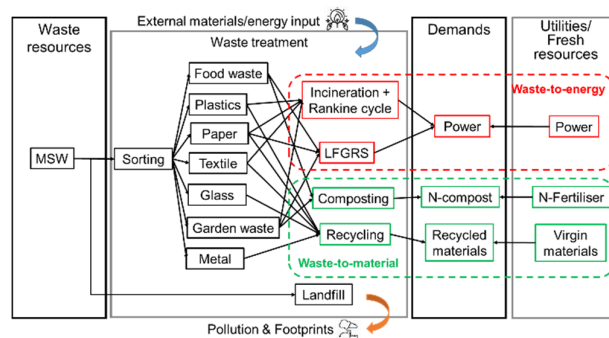


Figure 1. Concept Map of urban symbiosis waste reuse system

2. System representation, model and evaluation method

The symbiosis system can be seen as a set of paths from waste supply locations to demands for products and resources, which compete with the delivery of products from fresh resources (Figure 1). Each path includes different unit operations with varying efficiency. The processing paths also require fresh resources – energy, water, other materials, but they are often less intensive than the resource demands for new products.

The circularity is represented by the Total Circularity Index (TCI), which is the weighted sum of the CMU and a new circularity indicator – the Circular Exergy Use rate (CEU):

$$TCI = \omega_1 \times CMU + \omega_2 \times CEU; CEU = \frac{Ex_{recycle}}{Ex_{demands}} \quad (1)$$

CEU is defined (Eq.1) as the fraction of the exergy supplied from resource recycling and the overall exergy demands. The weights ω_1 and ω_2 are user specifications, modelling the importance of the material and exergy supply for each case. The model, summarised in (Eq.2), uses the material, energy and exergy balances, evaluating Total Annualised Cost (TAC) Greenhouse Gas (GHGFP) and Water Footprints (WFP). Additional constraints are stemming from the upper and the lower bounds on the stream flowrates. The optimisation variables include the flowrates of the recycle streams and those of the fresh resource supplies. The optimisations are run with three variants: minimising TAC, maximising GHGFP saving and maximising WFP saving.

$$\begin{aligned} & \text{minimise TAC} \quad \text{maximise GHGFP}_{\text{saving}} \quad \text{maximise WFP}_{\text{saving}} \quad (2) \\ \text{s.t.} \quad & TAC = f_{inv}(\text{equipment}) + f_{oper}(\text{throughput}) \\ & GHGFP_{\text{saving}} = f_{GHG1}(\text{demands}) - f_{GHG2}(\text{utilities}) - f_{GHG3}(\text{recycle}) \\ & WFP_{\text{saving}} = f_{WFP1}(\text{demands}) - f_{WFP2}(\text{utilities}) - f_{WFP3}(\text{recycle}) \\ & \text{Mass balances; Energy balances; Exergy balances} \end{aligned}$$

3. Case study

The system in Figure 1 has been evaluated on a set of sampling points of the TCI, running the optimisation model for each of the variants in Eq.2. The product demands and utilities prices are in Table 1, MSW properties in Table 2, where Degradable Organic Carbon is denoted as DOC. The maximum MSW flowrate is 137 kg/h. The unit operation characteristics and the cost data are shown in Table 3. The nitrogen content of the recycling compost is 30% (Fan et al., 2019). Table 4 provides the GHG factors, and Table 5 – the WFP factors.

Table 1. The data of the demands and utilities prices

Demands/Utilities	Demands value	Utilities Price
Power	60 kW	0.052 €/kWh
N-fertiliser	38 kg/h	0.133 €/kg
Paper	27 kg/h	0.033 €/kg
Plastic	29 kg/h	0.178 €/kg
Glass	6 kg/h	0.039 €/kg
Metal	5 kg/h	0.199 €/kg
Textile	10 kg/h	0.039 €/kg

Table 2. MSW properties

Sorting	Composition, %	LHV, MJ/kg	Carbon Content, %	DOC _F , %
Food waste	41.1	5.26	42.61	64
Plastic	22.2	5.38	60.93	0
Paper	20.9	3.08	37.37	37
Textile	7.7	2.48	60.42	0
Glass	3.6	0.00	0.00	0
Garden waste	2.5	0.48	41.47	23
Metal	2.0	0.00	0.00	0

Table 3. Efficiencies, conversion factors and Cost data of the involved processes

Processes	x_{eff} , %	UCC, €/kg	UOC, €/kg	Operating time, h/y
Incineration	30	0.020	0.026	7,008
Landfill gas recovery	8.16	0.001	0.027	8,640
Composting	60	0.005	0.010	7,008
Recycling	70	0.002	0.030	7,008
Landfill	-	0.005	0.106	8,640

Table 4. GHG emission data of Waste-to-energy and landfill

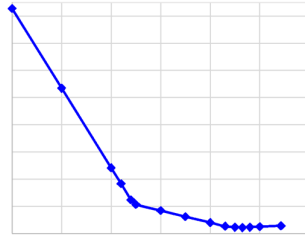
Path type	Process	Recycled Commodity GHG	Virgin-based Manufacturing GHG
Waste-to-energy	Incineration	386 kg CO ₂ eq/t	1.13 kg CO ₂ eq/kWh
	LFGRS	0.631 kg CO ₂ eq/t	1.13 kg CO ₂ eq/kWh
Landfill	Landfill	568 kg CO ₂ eq/t	-
Path type	Process	GHG mitigation	
Waste-to-material	Composting	3,600 kg/t N	
	Recycling paper	565 kg CO ₂ eq/t	
	Recycling plastic	396 kg CO ₂ eq/t	
	Recycling glass	159 kg CO ₂ eq/t	
	Recycling metal	3,205 kg CO ₂ eq/t	
	Recycling textile	230 kg CO ₂ eq/t	

Table 5. WFP emission data of Waste-to-energy

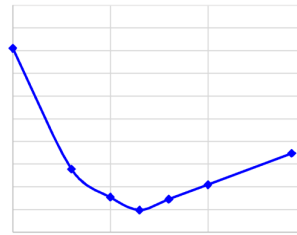
Process	Recycled Commodity WFP	Virgin-based Manufacturing WFP
Incineration	0.7 m ³ /MWh	3.2 m ³ /MWh
Landfill gas	2.4 m ³ /MWh	3.2 m ³ /MWh
Composting	0.5 m ³ /t	4,997.8 m ³ /t
Composting	0.3 m ³ /t	4,414.7 m ³ /t
Recycling paper	68.64×10 ⁻³ m ³ /t	5,272.5 m ³ /t
Recycling plastic	0.237×10 ⁻³ m ³ /t	276.8 m ³ /t
Recycling glass	65×10 ⁻⁶ m ³ /t	0.4 m ³ /t
Recycling metal	65×10 ⁻⁶ m ³ /t	52.2 m ³ /t
Recycling textile	0.237×10 ⁻³ m ³ /t	206.5 m ³ /t
Landfill	1.2×10 ⁻⁶ m ³ /t	-

The TAC evaluation results are shown in Figure 2. Figure 3 shows in part (a) the variation of the GHG and the exergy input, and part (b) plots the WFP. The TAC curve features a minimum at TCI=0.465 (1) / TAC= 84.5 k€/y. The GHG emission trend implies that higher circularity rate is, the better. The WFP curve has a maximum for TCI= 0.353 (1) and WFP saving 827 k(m³/y). The reason for this is that for TCI ≥ 0.3, for increasing the CMU values, secondary resources are redirected from energy generation because the WFP saving for material recycling is higher than for energy generation which is also more costly. These discrepancies provide information to decision-makers for the approximate

cost ($\approx 0.6 \text{ €/m}^3$) of reduction of water pollution (by reducing circularity a little) or of the environmental cost of attaining the cost optimum. However, it should be noted that the reduced circularity also reduces the GHG savings, in addition to increasing the TAC.

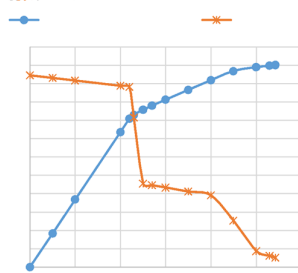


(a) The complete curve

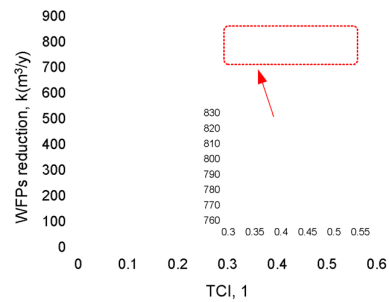


(b) TCI $\in [0.4, 0.543]$

Figure 2. TAC variation with the circularity rate



(a) GHG emission reduction (with exergy)



(b) WFPs reduction

Figure 3. GHG and WFPs emission reduction as functions of the circularity rate

4. Conclusions

The current work presents a conceptual model for evaluating the sustainability of industrial-urban symbiosis systems. The proposed model captures the key circularity rate criteria – adding the definitions of CEU and TCI to CMU, followed by a combined evaluation of the TAC, GHGFP and WFP reductions, and total exergy input. The case study demonstrates the usefulness of the proposed conceptual model for evaluating the optimal circularity rate. It also reveals that besides not coinciding with the TAC trend (WFP optimum is $\approx 10\%$ higher TAC than the cost optimum), the emission reduction trends can be adversarial to each other, pointing to the need for improved modelling in the future. Such an improvement is to add a monetary representation of the footprints and join them to the TAC, leading to the concept of Eco-Cost as the objective function.

Acknowledgement

The funding from the project 'Sustainable Process Integration Laboratory – SPIL funded by EU' CZ Operational Programme Research and Development, Education, Priority1:

Strengthening capacity for quality research (Grant No. CZ.02.1.01/0.0/0.0/15_003/000045) under the collaboration agreement with the University of Split has been gratefully acknowledged.

References

- A. Abadías Llamas, N.J. Bartie, M. Heibeck, M. Stelter, M.A. Reuter, 2020. Simulation-Based Exergy Analysis of Large Circular Economy Systems: Zinc Production Coupled to CdTe Photovoltaic Module Life Cycle. *J. Sustain. Metall.* 6, 34–67. DOI: 10.1007/s40831-019-00255-5.
- S. Ahmad, H. Jia, Z. Chen, Q. Li, C. Xu, 2020. Water-energy nexus and energy efficiency: A systematic analysis of urban water systems. *Renew Sust Energy Reviews* 134, 110381.
- S.T. de Almeida, M. Borsato, C.M. Lie Ugaya, 2017. Application of exergy-based approach for implementing design for reuse: The case of microwave oven. *J. Clean. Prod.* 168, 876–892.
- L. Čuček, J.J. Klemeš, Z. Kravanja, 2012. A Review of Footprint Analysis Tools for Monitoring Impacts on Sustainability, *Journal of Cleaner Production*, 34, 9–20.
- T. Domenech, R. Bleischwitz, A. Doranova, D. Panayotopoulos, L. Roman, 2019. Mapping Industrial Symbiosis Development in Europe_ typologies of networks, characteristics, performance and contribution to the Circular Economy. *Resour. Conserv. Recycl.* 141, 76–98.
- EC, 2020. New Circular Economy Action Plan. Interreg Eur. <<https://www.interregueurope.eu/plasteco/news/news-article/8056/new-circular-economy-action-plan/>>, accessed 08/08/2020.
- Eurostat, 2018. Circular material use rate: calculation method : 2018 edition. <<https://op.europa.eu/en/publication-detail/-/publication/cbeb5449-f2c0-11e8-9982-01aa75ed71a1/language-en>>, accessed 06/07/2020.
- Y.V. Fan, C.T. Lee, J.S. Lim, J.J. Klemeš, P.T.K. Le, 2019. Cross-disciplinary approaches towards smart, resilient and sustainable circular economy. *J. Clean. Prod.* 232, 1482–1491.
- Y.V. Fan, P.S. Varbanov, J.J. Klemeš, S.V. Romanenko, 2021. Urban and industrial symbiosis for circular economy: Total EcoSite Integration. *Journal of Environmental Management*, 279, 2021, 111829, doi: 10.1016/j.jenvman.2020.111829.
- K. Hartley, R. van Santen, J. Kirchherr, 2020. Policies for transitioning towards a circular economy: Expectations from the European Union (EU). *Res. Cons. Recycl.* 155, 104634.
- B. Holmatov, A.Y. Hoekstra, M.S. Krol, 2019. Land, water and carbon footprints of circular bioenergy production systems. *Renewable and Sustainable Energy Reviews* 111, 224–235.
- J.J. Klemeš, 2013. *Handbook of Process Integration (PI): Minimisation of Energy and Water Use, Waste and Emissions*. Woodhead Publishing/Elsevier, Cambridge, UK.
- P. Morsetto, 2020. Targets for a circular economy. *Resour. Conserv. Recycl.* 153, 104553.
- M. Saidani, B. Yannou, Y. Leroy, F. Cluzel, A. Kendall, 2019. A taxonomy of circular economy indicators. *J. Clean. Prod.* 207, 542–559.
- V. Selicati, N. Cardinale, M. Dassisi, 2020. The interoperability of exergy and Life Cycle Thinking in assessing manufacturing sustainability: A review of hybrid approaches. *Journal of Cleaner Production* 124932.
- M.-L. Tseng, R.R. Tan, A.S.F. Chiu, C.-F. Chien, T.C. Kuo, 2018. Circular economy meets industry 4.0: Can big data drive industrial symbiosis? *Resour. Conserv. Recycl.* 131, 146–147.
- G. Vilardi, C. Bassano, P. Deiana, N. Verdona, 2020. Exergy and energy analysis of three biogas upgrading processes. *Energy Convers. Manag.* 224, 113323.
- X.-C. Wang, J.J. Klemeš, X. Dong, W. Fan, Z. Xu, Y. Wang, P.S. Varbanov, 2019. Air pollution terrain nexus: A review considering energy generation and consumption. *Renewable and Sustainable Energy Reviews* 105, 71–85.

Negative CO₂ emissions in biomass gasification process with hybrid amine-deep eutectic solvents

Cristian Dinca,^{a,*} Nela Slavu,^a Calin-Cristian Cormos,^b Eliza-Gabriela Mihaila,^a

^a Politehnica University, Faculty of Power Engineering, 313 Splaiul Independentei, Bucharest, RO-060042, Romania

^b Babes-Bolyai University, Faculty of Chemistry and Chemical Engineering, Arany Janos 11, Cluj-Napoca, RO-400028, Cluj-Napoca, Romania

crisflor75@yahoo.com

Abstract

Biomass gasification is one of the most promising solution for producing clean electricity. This technology option can reduce the use of fossil fuels and limit the exhaust CO₂ emissions in atmosphere. Moreover, by integrating different CO₂ separation technologies, the biomass gasification power plant is considered to have negative CO₂ emissions due to the CO₂ absorption during its growth. In this study, the physical absorption process deep eutectic solvent (DES) was used to retain the CO₂ from syngas. Subsequently, the treated synthesis gas was used in a gas turbine (ITG) plant. The flue gases expanded in the gas turbine were introduced into the heat recovery steam generator (HRSG) for steam production and utilization in a steam turbine for mechanical work and subsequent power generation. The biomass (poplar) gasification process was analysed for different air equivalent ratio (ER). The highest lower heating value of the synthesis gas was obtained for ER = 0.2. In the case of CO₂ capture based on DES, a thermal energy consumption of 1 GJ/tCO₂ was determined, this represents an electrical energy consumption of 0.4 GJ/tCO₂. In terms of electricity production, for a syngas flow rate of 20 t/h, the maximum power obtained was of 10.8 MW, and the net cycle efficiency of 30 %.

Keywords: Biomass gasification, DES-based CO₂ capture, Energy integration

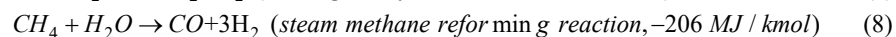
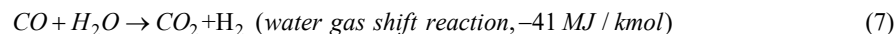
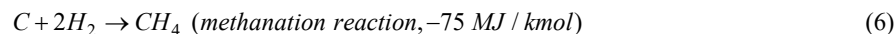
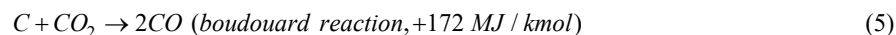
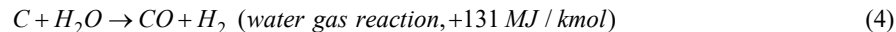
1. Introduction

The biomass gasification with CCS (Carbon Capture and Storage) can be promising solution to reduce the carbon dioxide emissions (C. Dinca et.al, 2018). The biomass is considered a energy source with neutral CO₂ emissions. The biomass gasification with CO₂ capture generates the negative emissions, therefore it could decrease the level of carbon dioxide emissions in atmosphere on the long-term (J. Kemper, 2017). The CO₂ capture technology was studied intensively in the last decade to reduce the negative impact after integration into an energy or industrial process. Thus, it is needed to develop new solvents with higher CO₂ absorption capacity, solvents that requires lower thermal energy for regeneration, and with a lower corrosivity, comparative with the conventional amines. Several researchers studied primary, secondary, tertiary amines and amine blends in order to improve the performances concerning the CO₂ absorption capacity and to reduce the thermal energy consumption (C. Dinca, 2016). Also, it is observed that primary and secondary amines have a higher rate of reaction and higher energy consumption than the tertiary ones that need a lower amount of thermal energy for regeneration due to the carbonate and bicarbonate ions form during the hydrolysis of CO₂ in the absorption

process (K. Goto et.al, 2013). Deep eutectic solvents (DES) are the promising alternative for CO₂ separation, can replace the conventional chemical solvent based on amines. DES has many advantages, like as: the thermal energy needed for regeneration is lower compared to amines, due to the physical absorption of carbon dioxide; due to the low vapour pressure, the amount lost in the absorption and desorption process is smaller; DES degradation occurs at temperatures higher than 300 °C, no reactions with other impurities from flue gases, and are not corrosive; DES preparation allows the design of absorbents with specific properties for CO₂ separation. DES still have some drawbacks to integrating at a large scale for CO₂ separation, such as the difficult production process and the expensive raw material price (R.J. Isaifan and A. Amhamed, 2018). The aim of this study consists of the biomass gasification with CO₂ capture pre-combustion based on DES and syngas utilization for energy produce.

2. Gasification process

Gasification process consists of transforming of solid fuel in gaseous fuel, rich in H₂, in the presence of an oxidizing agent (air, oxygen, steam or mixtures of these). The main reactions that occur in the gasification process are presented by Equation 1-8.



The type of biomass used in this study was wood biomass – poplar. The biomass dry composition is presented in Table 1. The initial moisture content is of 6.35 %, and the oxygen content was determined by difference. The lower heating value (LHV) of biomass is 16,760 kJ/kg.

Table 1. Biomass composition

C [%]	H [%]	N [%]	S [%]	O [%]	Ash [%]
50.94	6.34	1.81	0.1	40.562	0.248

The gasification process simulation was carried out in the specialized software Chemcad. The Peng Robinson thermodynamic model was used in the gasification and power generation process. The oxidizing agent used in the gasification process was pure air. ER represents the ratio of the actual air amount introduced into the gasification reactor and the stoichiometric air amount. The maximum volume fraction of H₂ was obtained for ER = 0.4. The concentration of methane in the syngas decreases with a larger amount of air introduced the gasification reactor, and the CO concentration increases due to the combustion reactions (Figure 1). The lower heating value of the syngas decreases with

increasing the ER ratio, due to the decrease of CH₄ concentration. The heat energy contained by the syngas was determined as the product of the lower heating value and the syngas mass flow. Thus, even if the lower heating value has a descending allure, the thermal energy contained by the syngas has a maximum value for ER ratio of 0.4 (Figures 2). After the biomass gasification, the water gas shift reaction (WGS) and the syngas treatment, the carbon dioxide from syngas is separated by physical absorption used DES. The syngas composition considered for CO₂ separation is the one for that the highest value of the H₂/CO ratio was obtained. Thus, the syngas considered for the CO₂ separation has the following volumetric composition: H₂ = 27 %, CH₄ = 1 %, N₂ = 46 %, CO = 4 %, CO₂ = 22 %.

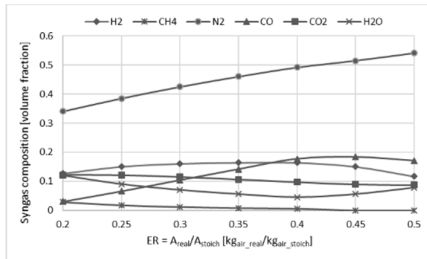


Figure 1. Syngas composition

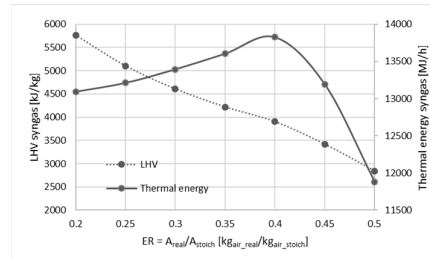


Figure 2. LHV and thermal energy content

3. CO₂ capture process

The CO₂ separation process based on DES is described in the Figure 3. It can be observed that the process is similar to the separation process based amines, with the difference that in the case of use DES-amines, the absorption process can takes place at higher pressures (>2 bar, the pressure of the absorption process for amine is the atmospheric pressure) . In this study, the thermal energy for DES regeneration rich in CO₂, was determined by calculation based the model described by Y. Zhang et.al, 2016. The absorption pressure was considered of 1.013 bar to reduce the electric energy needed for the syngas compression, and the absorption temperature of 25 °C. The desorption pressure was considered of 1.013 bar and the desorption temperature of 50 °C. Due to the desorption temperature, the thermal energy consumption for solvent regeneration was resulted of 1 GJ/tCO₂, this is equivalence of 0.4 GJ/tCO₂ electric energy consumption, for CO₂ capture efficiency of 90 %. In the Figure 4, the thermal energy consumption is presented for DES, monoethanolamine and diethanolamine. The purity of dry CO₂ is more than 98 mol %.

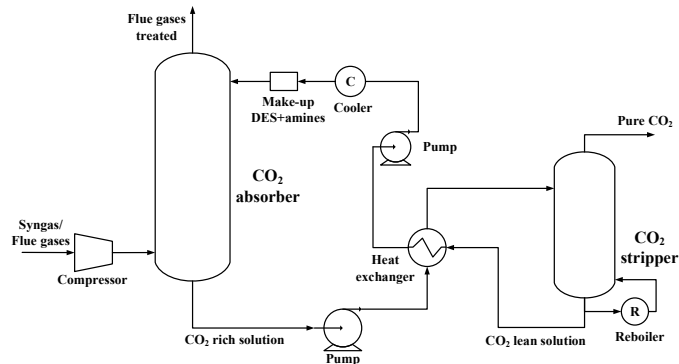


Figure 3. Schematic diagram of CO₂ separation process

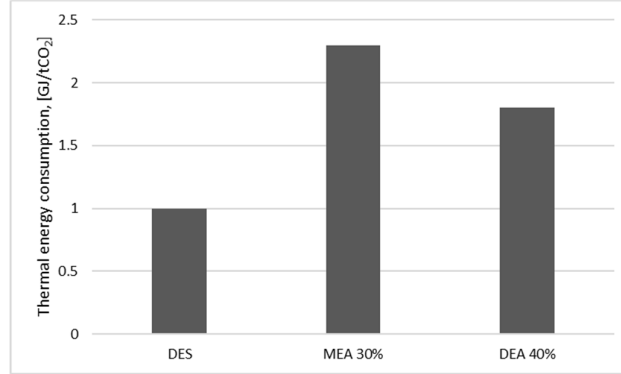


Figure 4. Thermal energy consumption for solvent regeneration

4. Energy production

The schematic diagram of the energy production chain is presented in Figure 5. The syngas composition used for energy production is presented in Table 2 before and after CO₂ capture process integration. Syngas obtained after gasification process was used to produce electric energy in 4 cases. Cases studied are:

- Case 1: Gas turbine (GT) without CO₂ capture process;
- Case 2: GT with pre-combustion CO₂ capture;
- Case 3: GT with pre-combustion CO₂ capture and recovery a part of the flue gases heat to preheat the air-syngas mixture before entering GT;
- Case 4: GT with pre-combustion CO₂ capture, recovery the heat of the flue gases to use in the steam turbine (ST), and preheat the air-syngas mixture.

The efficiency process of the energy produced was determined with the following relationship:

$$Ef_{process} = \frac{P_{GT} + P_{ST} - P_C - P_R}{\dot{m}_s \cdot LHV_s} \quad (1)$$

where: P_{TG} - gas turbine power, in kW; P_{ST} - steam turbine power, in kW; P_C - air compressor power, in kW; P_R - regeneration solvent power, in kW; \dot{m}_s - syngas flow, in kg/h; LHV_s - lower heating value of syngas, in kJ/kg.

In order to maximize the energy produced, several parameters were varied, such as the air flow introduced into the combustion chamber (CC). It was determined the optimum temperature for that the maximum power is obtained and the maximum efficiency of the combustion process. For example, in the Case 1, when the syngas is used without CO₂ capture process, for a temperature of 1,200 °C, air flow of 40,000 kg/h is needed to produce 10.8 MW. In the Case 2, when syngas is used after CO₂ capture process, for a temperature of 1,200 °C in the combustion chamber, the air flow introduced was 50,000 kg/h, and the power produced was of 11.8 MW.

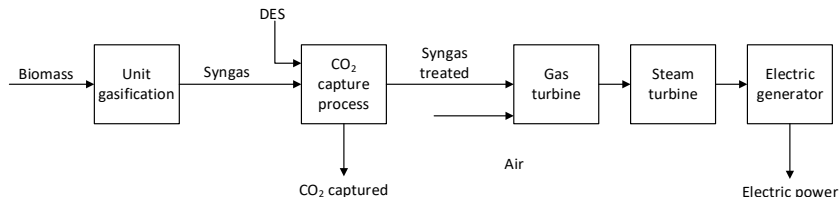


Figure 5. Biomass gasification with CO₂ capture process pre-combustion

Table 2. Syngas composition before and after CO₂ capture process

Composition (vol. %)	Before CO ₂ capture (vol. %)	After CO ₂ capture (vol. %)
H ₂	27	35
CH ₄	1	1.2
N ₂	46	57
CO	4	5
CO ₂	22	2.8
LHV (kJ/kg)	3,480	5,442

Further, it was analysed the process efficiency for a constant temperature in the combustion chamber, of 1,200 °C (Figure 6). Thus, with the heat recovery system integration of the flue gases significant improvements are known regarding the cycle efficiency. The optimal case, when it was obtained the higher process efficiency, it is Case 4 (integrated gasification combined cycle with CO₂ capture). Another important advantage of the CO₂ capture process integration consists in the amount of carbon dioxide emitted into the atmosphere with the flue gases resulting from the process. It can be observed that after the capture process integration the concentration of CO₂ in the flue gases decreases significantly, respectively from a percentage of 11 %, in the case of using of the syngas without CO₂ capture, to a maximum of 2 % for the cases when is integrated the CO₂ capture process (Figure 7). In Figures 8, there are shown the results obtained if the power is maintained constant. It can be observed, in terms of process efficiency, as in the previously hypothesis, the Case 4 has the highest efficiency, and the Case 2, the lowest efficiency. In Figure 9 is presented a comparative analysis of the process efficiency according to the type of solvent used for the CO₂ capture. The DES-based CO₂ capture is the most efficient solution due to the the lower amount of thermal energy required for solvent regeneration.

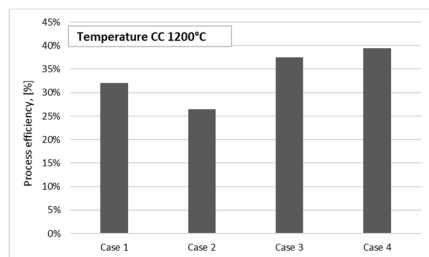


Figure 6. Process efficiency

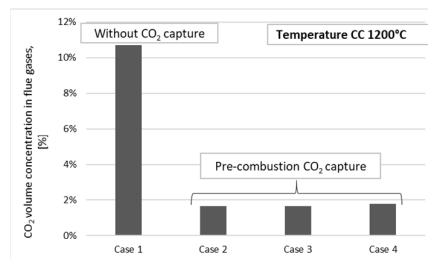


Figure 7. CO₂ concentration in flue gases

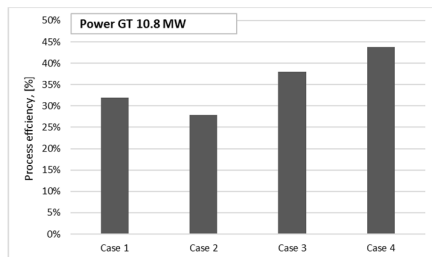


Figure 8. Process efficiency

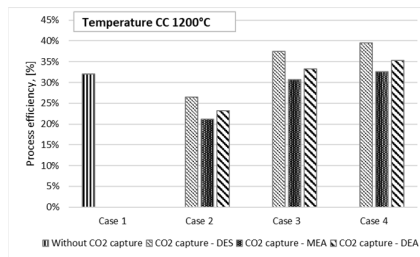


Figure 9. Comparative analysis of process efficiency

5. Conclusions

The biomass gasification with CO₂ separation process is a solution to produce thermal and electric energy with low emissions of carbon dioxide. The low heating value of the syngas obtained after air gasification process was of 3,480 kJ/kg before CO₂ capture, for an equivalent ratio of 0.4, and of 5,442 kJ/kg after CO₂ pre-combustion capture process integration. The thermal energy consumption for DES regeneration was obtained of 1 GJ/tCO₂, much lower than amines (MEA, DEA). In terms of the power produced, the combined cycle (gas turbine-steam turbine) with CO₂ pre-combustion capture and recovery heat flue gases system was optimum. In this case, it was obtained the maximum power produced.

Acknowledgments

The study has been funded by the UEFISCDI within the National Project number PED296/2020 with the title: „New Amine based DES’s SOLvent development for CO₂ capture in a demonstrative pilot Installation by Chemical Absorption Technology” – ASOCIAT. The research leading to these results has received funding from the NO Grants 2014 - 2021, under Project contract no. 13/2020.

References

- C. Dinca, 2016, Critical parametric study of circulating fluidized bed combustion with CO₂ chemical absorption process using different aqueous alkanolamines, *Journal of cleaner production*, 112, 1136-1149.
- C. Dinca, N. Slavu, C. C. Cormos and A. Badea, 2018, CO₂ capture from syngas generated by a biomass gasification power plant with chemical absorption process, *Energy*, 149, 925-936.
- K. Goto, K. Yogo and T. Higashii, 2013, A review of efficiency penalty in a coal-fired power plant with post-combustion CO₂ capture, *Applied Energy*, 111, 710-720.
- J. Kemper, Biomass with carbon capture and storage (BECCS/Bio-CCS), 2017, IEA Greenhouse Gas R&D Programme.
- R.J. Isaifan, and A. Amhamed, 2018, Review on carbon dioxide absorption by choline chloride/urea deep eutectic solvents. *Adv. Chem*, 1-6.
- Y. Zhang, X. Ji, Y. Xie and X. Lu, 2016, Screening of conventional ionic liquids for carbon dioxide capture and separation, *Applied energy*, 162, 1160-1170.

Development and optimization of an energy efficient and self-sustained integrated process for nitrogen removal for baseload LNG plant

Ajinkya Pal,^a Easa I. Al-musleh,^{b*} Iftekhar A Karimi^{a*}

^a*Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585*

^b*Department of Chemical Engineering, Qatar University, Doha 2713, Qatar*

cheiak@nus.edu.sg ; e.almusleh@qu.edu.qa

Abstract

In this study, the energy efficiency of a novel integrated process including recovery of natural gas liquids (NGL), natural gas liquefaction and nitrogen removal unit (NRU) is investigated. This study designs a prevailing single column NRU, considered as the base case and a double-column NRU integrated processes in conjugation with boil-off gas (BOG) generation in the various process steps after acid gas removal unit. The BOG generated and nitrogen rich waste stream are utilized as fuel gas (FG) for gas turbine to generate power. A simulation based optimization paradigm deploying particle swarm optimization algorithm is implemented to minimize the specific power requirement with consideration of higher heating value (HHV) of LNG product, NGL Reid vapor pressure (RVP), Wobbe index (WI) and pressure of FG. The parity between FG heating value and fuel requirements ensures efficient and self-sustained LNG plant. The optimal process parameter and specific power requirement for the single column NRU conventional and double column NRU integrated process are compared.

Keywords: LNG, Nitrogen removal, integrated process design, Boil-off gas

1. Introduction

Natural gas (NG), owing to favorable heating values, relatively low fuel costs and significantly reduced emission of carbon dioxide and a negligible amount of CO, NO_x, SO_x and particulate as compared to combustion of oil and coal makes it a lucrative option to be the bridge between renewables and non-renewables. Practical storage and transport of NG as liquefied natural gas (LNG) is viable by liquefaction at 112 K, to reduce the volume occupied by methane by more than 600 times at ambient pressure.

The literature covers the progress in designing of efficient LNG liquefaction process and sheds light on potential improvements in specific energy requirement by the integration of LNG liquefaction, NGL recovery and NRU. (Ghorbani et al., 2016) With increasing concerns of climate changes, the BOG generation minimization and handling become of paramount importance to ensure LNG supply chain has reduced carbon footprints and detrimental impact on immediate surroundings. Although many studies are addressing the two aforementioned problems exclusively, to the best of our

knowledge there is no literature painting a holistic picture of existent LNG plant by incorporating an integrated LNG liquefaction, NGL recovery and multi-column NRU process with BOG generation at various process step accounting for the fuel gas. (Lee et al., 2019) The various process constraints on the HHV of LNG product, NGL RVP, WI and pressure of FG and parity between FG heating value and fuel requirements ensuring optimal and self-sustained LNG plant, remains briefly addressed in the literature.

This study aims to analyse and improve the energy efficiency of process steps after acid gas removal unit (AGRU) in an existent baseload LNG plant. This study carries out optimization and design of a prevailing single column NRU and a double-column NRU process, both integrating LNG liquefaction, NGL recovery and NRU processes in conjugation with BOG generation in the various process steps after AGRU. Particle Swarm Optimization (PSO) algorithm was implemented to minimize specific power requirement with consideration of LNG HHV, NGL RVP, and fuel gas WI and pressure and fuel gas generation-requirement balance. (Kennedy & Eberhart, 1995) Further, the optimized results of single column NRU conventional and double column NRU integrated processes are compared and results discussed.

2. Process Description

2.1. Single Column NRU conventional process design

Figure 1 illustrates integrated liquefaction, NGL recovery and one column NRU system, with an SMR cycle, regarded as the single-column NRU conventional system and the base case for comparing the improved system proposed in this study. The top product of NGL column after liquefaction makes the feed for the NRU column operating at 2 Bar pressure. The vapour fraction of reboiler stream controls the purity and recovery of methane in LNG product stream. The bottom product of NRU column is pumped to 6 Bar and flashed in the storage tank. The BOG generated in the storage tank and the nitrogen rich waste stream is compressed and used as fuel gas. A single mixed refrigeration cycle provides the requisite cold energy for liquefaction. The retrieval of cold energy from the BOG stream and nitrogen rich waste stream in a multi stream heat exchanger (MSHE) before compression, provide additional cold energy and improve the efficiency of the process.

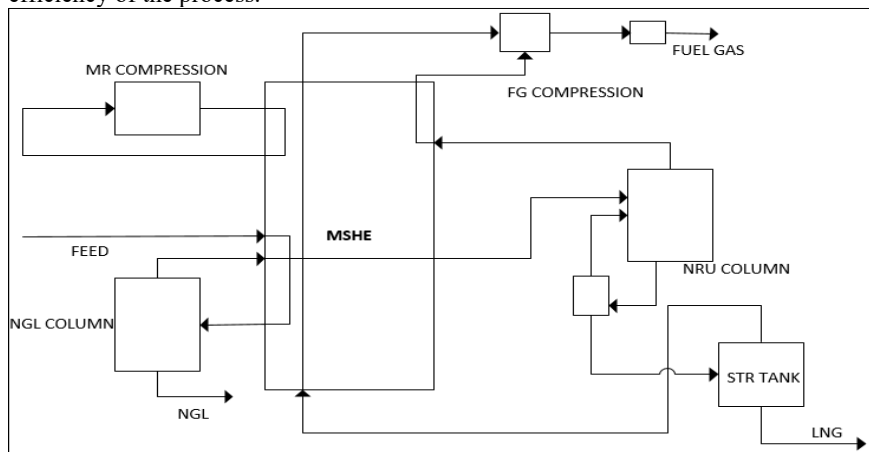


Figure 1 Schematic diagram of single column NRU conventional process

2.2 Double column NRU integrated process design

Figure 2 shows the double-column NRU process. The process along with integrated liquefaction and NGL recovery comprises of one high-pressure (HP) column and one low-pressure (LP) column in NRU. The compressed feed after AGRU is cooled and throttled before entering the NGL recovery column. The top product of HP NRU column enriched in nitrogen, divided into two streams in SPLIT-1 & 2 provide reflux to both HP and LP NRU columns after compression, splitting and cooling. A part of the bottom product of HP NRU column (SPLIT-3) contributes towards fuel gas requirement after Joule-Thomson expansion to 20 Bar and cold energy retrieval in MSHE. The feed to LP NRU column comprises of a part of the top product stream of HP NRU column. The HP and LP column operate at 26.8 and 5.3 Bar respectively. The bottom liquid stream from LP NRU is pumped to 56 Bar before sub-cooling and flashing to 2 Bar pressure. The BOG due to vapor displacement, HP flashing and heat leaks in STR-TANK and top product of LP-NRU column, after cold energy recovery in MAIN-MSHE and three-stage compression to 20 Bar, contribute to the fuel gas requirement.

3. Solution procedure

Simulation based optimization paradigm was deployed to find the optimal operating parameter of the processes. Aspen Hysys V.10 simulated each random design generated by PSO algorithm coded in MATLAB. A simulation-based optimization framework combining the accurate and fast thermodynamic calculation performed in Aspen Hysys and rigorous optimization algorithm implemented on an external platform like MATLAB can provide an accurate value for the decision variables while ensuring the constraints are met and specific power minimized. (Hamed et al 2019) The connection between MATLAB and Hysys established through component object model in ActiveX, allows direct two-way communication between Hysys and MATLAB. The process and stream data calculated in Hysys are sent to MATLAB, which performs optimization, levies constraints on variables, and sends back the values of calculated variables. This section further discusses the decision variables and constraints for the conventional and two-column NRU integrated design, which is optimized using PSO to find the minimum specific energy requirement for both cases.

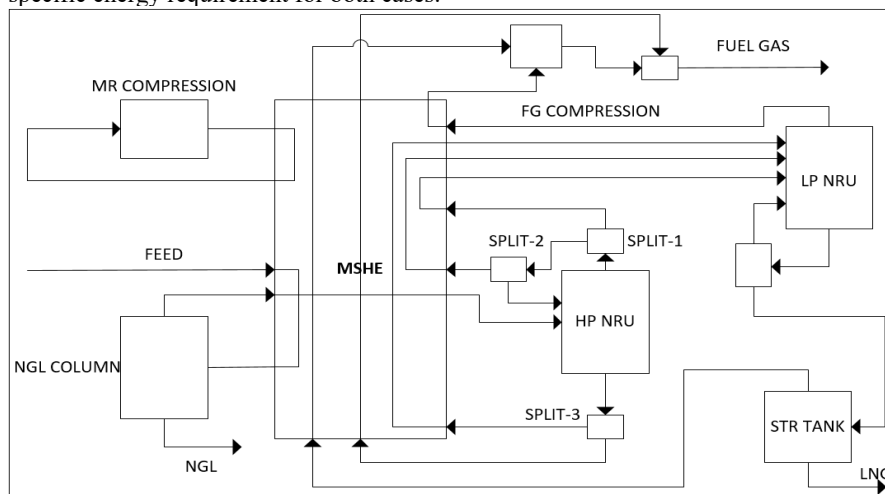


Figure 2 Schematic diagram of double column NRU integrated process

3.1 Single column NRU conventional process optimization

The major energy consumer in single-column NRU conventional process is the four-stage SMR cycle compressors. The three-stage compression of fuel gas streams generated by NRU column and BOG at the export terminal also contribute towards the energy requirements of the process. The feed compressor, SMR cycle pumps and LNG pump are other contributors towards energy consumption. The objective is to find the minimum specific energy requirement, calculated by dividing total energy consumption by LNG production in mega long tonnes per annum (Mty).

$$\text{Total power consumption by User 2} = W_{\text{FEED-COM}} + \sum_{\text{I=compressor}} (W_{\text{MR-COM-I}} + W_{\text{FG-1-COM-I}} + W_{\text{FG-2-COM-I}}) + \sum_{\text{I=pump}} (W_{\text{PMP-I}}) \quad (1)$$

Objective function, specific power consumption

$$= [\text{Total power consumption by User 2}] / [\text{PRODUCT LNG mass flow rate}] \quad (2)$$

W denotes power consumption by the compressor or pump described in the subscript. The single column NRU conventional process consists of 15 decision variables. The NGL recovery column operates at a fixed pressure of 54.7 Bar, hence the feed, reboiler and reflux temperature are the working variable to ensure the NGL flowrate and RVP specifications are met. The fuel requirement of plant before and including AGRU and post AGRU process are regarded as User 1 and User 2 fuel requirements respectively. The User 1 fuel requirement is considered same for both cases studies. The total fuel requirement for the process design takes into account the User 1 (fuel requirement before AGRU) and User 2 fuel requirements.

$$\text{Total fuel requirement} = \text{User 1 fuel requirement} + \text{User 2 fuel requirement} \quad (3)$$

$$\text{User 2 fuel requirement} = \text{Total power consumption by User 2} / 0.3 \quad (4)$$

The fuel requirement of process is satiated by the fuel gas stream available at 20 Bar.

$$\text{Total fuel generated} = (\text{Low heating value of FG}) * (\text{Molar flow of FG}) \quad (5)$$

The following specifications acting as constraints for the process:

$$\text{LNG product High heating value} \geq 39.86 \text{ MJ/m}^3 \quad (6)$$

$$\text{NGL molar flowrate} \geq 702 \text{ kmol/h} \quad (7)$$

$$\text{NGL RVP} \leq 23.9 \text{ Bar} \quad (8)$$

$$\text{Wobbe index of fuel gas} \geq 30 \text{ MJ/m}^3 \quad (9)$$

$$\text{Minimum temperature approach in MSHE} \geq 2 \text{ K} \quad (10)$$

$$\text{Total fuel requirement} / \text{Total fuel generated} \geq 0.99 \quad (11)$$

$$\text{Total fuel requirement} / \text{Total fuel generated} \leq 1 \quad (12)$$

The two inequality constraints (11) & (12), relax the equality constraint for fuel balance.

3.2 Double column NRU integrated process optimization

The double-column NRU integrated process has four compressors SMR cycle, contributing to major energy consumption and one expander at HP-NRU feed column, producing some energy. The NGL recovery column operated at a pressure similar to that of conventional design delineated in the previous section, hence the feed, reboiler and reflux temperatures of NGL recovery column are working variable to ensure NGL flowrate and RVP specification are met. The objective is to find minimum specific energy requirement containing 27 decision variables.

$$\text{Total power consumption by User 2} = W_{\text{FEED-COM}} + \sum_{\text{I=compressor}} (W_{\text{MR-COM-I}} + W_{\text{FG-1-COM-I}} + W_{\text{FG-2-COM-I}}) + \sum_{\text{I=pump}} (W_{\text{PMP-I}}) - W_{\text{EXP}} \quad (13)$$

Objective function, specific power consumption

$$= [\text{Total power consumption by User 2}] / [\text{PRODUCT LNG mass flow rate}] \quad (14)$$

The product specification and process constraints levied for the conventional process listed in section 3.1 (Equation 3-12) are active for double-column NRU integrated process optimization as well.

4. Results

The four-stage compression in SMR cycle is the major contributor towards the power requirement with 99.42 and 97.27 MW for single-column NRU conventional and double column NRU integrated process, respectively. This power requirement reduction results from lower entropy generation and lowered compression pressure. The fuel gas generated for the process design comprises of the nitrogen-rich waste stream from NRU column and BOG at low pressure and require compression. The double-column NRU integrated process has additional fuel gas stream sourced from the bottom stream of HP NRU column at high pressure. This leads to a reduction in FG compression power requirements by 56% and the absolute values are 9.64 and 4.24 MW respectively for single-column NRU conventional and double column NRU integrated process respectively.

Table 1 summarizes the values of parameters and constraints of the two process designs with fuel gas generation to satiate the process energy requirement. The NGL stream conditions are similar for the two process designs. The LNG product feed flow rate for the double column NRU integrated process increases by 4.8% as compared to the conventional case for the same feed flow rate. The total power requirement of the

Table 1
Summary of results and specific power consumption

		Single column NRU conventional design	Double column NRU integrated design
LNG	Product mass flow rate (Mty)	3.308	3.324
	Pressure (Bar)	1.2	1.2
	Temperature (K)	113.3	113.2
	Nitrogen content (mol %)	0.34	0.28
	LHV (MJ/m ³)	36.2	36.2
	HHV (MJ/m ³)	39.8	39.8
NGL	Molar flow rate (kmol/h)	703	702.5
	Pressure (Bar)	54.87	54.87
	Temperature (K)	314.8	327.8
	RVP (Bar)	23.9	23.8
Fuel Gas	Molar flow rate (kmol/h)	3815	3712
	Pressure (Bar)	20	20
	Temperature (K)	298	298
	Wobbee index (MJ/m ³)	30.78	30.36
Heat Exchanger Energy Calculation	MTA (K)	2	2
	Power requirement (MW)	109.6	104.3
	User 1 fuel requirement (MW)	210	210
	User 2 fuel requirement (MW)	365.3	347.6
	Total fuel requirement (MW)	575.3	557.6
	Total fuel generated (MW)	579.4	558.4
	Required/Generated fuel	0.993	0.999
Specific power requirement (MW/Mty)		33.13	31.3

double-column NRU integrated process witnesses a 4.8% reduction as compared to the single-column NRU conventional process. This reduction of power and fuel gas requirements lead to a reduction in required FG flow rate and higher LNG production for the former. The rigorous simulation-based optimization ensures the ratio of required and generated fuel value to be closer to unity. The objective function i.e. specific power consumptions are 33.13 and 31.13 MW/Mty for single-column NRU conventional and double-column NRU integrated process respectively. The decrease in optimal objective value can be attributed to a reduction in total power requirement and increase in LNG production.

5. Conclusions

The prevailing single column NRU conventional and double column NRU integrated processes are simulated, optimized and compared. The double column NRU integrated process shows a 5.5% reduction in specific power requirement as compared to the base case. This reduction in specific power requirement is attributed to lower power requirement leading to reduced fuel gas requirement and increased LNG production. The double column NRU integrated process witnesses a 16-kilo tonnes (long) per annum increase in production. The power required for SMR compression and fuel gas compression reduced by 2.1% and 56.1% as compared to the base case. The SPLIT-3 in double column NRU integrated process diverts 3.1% molar flow of HP-NRU bottom product to MAIN-MSHE for cold energy retrieval before utilising as fuel gas at high pressure, hence reducing the FG compression requirement. The high operating pressure of LP-NRU as compared to conventional design further contributes to the reduction in FG compression requirements.

References

- Ghorbani, B., Hamed, M.-H., & Amidpour, M. (2016). Development and optimization of an integrated process configuration for natural gas liquefaction (LNG) and natural gas liquids (NGL) recovery with a nitrogen rejection unit (NRU). *Journal of Natural Gas Science and Engineering*, 34, 590–603. <https://doi.org/10.1016/j.jngse.2016.07.037>
- Hamed, H., Karimi, I. A., & Gundersen, T. (2019). Optimization of helium extraction processes integrated with nitrogen removal units: A comparative study. *Computers & Chemical Engineering*, 121, 354–366. <https://doi.org/10.1016/j.compchemeng.2018.11.002>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, 1942–1948 vol.4. <https://doi.org/10.1109/ICNN.1995.488968>
- Lee, Y., Lim, Y., & Lee, W. B. (2019). Integrated Process Design and Optimization of Nitrogen Recovery in Natural Gas Processing. *Industrial & Engineering Chemistry Research*, 58(4), 1658–1674. <https://doi.org/10.1021/acs.iecr.8b04542>

Integration of Green Power in a Gas to Liquid Process

Mohammad Ostadi,^a Emre Gençer,^a Magne Hillestad^{b,*}

^a MIT Energy Initiative, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

^b Department of Chemical Engineering, Norwegian University of Science and Technology, NO 7491 Trondheim, Norway

* magne.hillestad@ntnu.no

Abstract

The aviation industry requires to reduce its carbon emissions. To do so, the fuel needs to be renewable-based as much as possible and have low life cycle emissions, hence be sustainable. Gas-to-Liquid (GTL) process through Fischer-Tropsch (FT) synthesis can be used to make aviation fuel. This process can be made more sustainable by lowering its carbon emissions. The addition of renewable power to the GTL process is investigated with the aim of reducing its CO₂ emissions. Two promising designs are considered which have very low CO₂ emissions: 1- Using Autothermal Reformer (ATR) to produce syngas and Solid Oxide Electrolysis Cell (SOEC) to produce H₂ and O₂; and 2- Production of syngas through electrically heated Steam Methane Reformer (E-SMR). In both designs, the addition of renewable power significantly reduces carbon emissions and increases carbon efficiency which means increased production for the same amount of natural gas feed. By assessing the two designs based on FT production, carbon efficiency, and FT catalyst volume, it is a better choice to add renewable power to the SOEC (case1) rather than using it in an E-SMR (case 2). These designs are considered in order to help us safely transit to a low-carbon society.

Keywords: Fischer-Tropsch synthesis, Renewable power integration, Low-carbon fuels, Low-emission Gas-To-Liquid process, Green Hydrogen.

1. Introduction

In order to reach the goals of the Paris Agreement (United Nations 2015) and the aims of the United Nations' Sustainable Development Goals (United Nations 2019), current industrial processes are required to move towards reducing their carbon emissions and hence becoming more sustainable. Decarbonizing the aviation industry is hard as there are few equivalent alternatives to kerosene and jet fuel, mainly because a fuel with high energy density is required. Fischer-Tropsch process produces a wide range of hydrocarbons including kerosene and jet fuel. Thus, liquid fuel production through this process is the focus of many investigations (Hillestad et al. 2018). Integration of renewable energy into chemical processes results in both increased production of chemicals and simultaneous reduction of CO₂ emissions (Agrawal et al. 2007). Hillestad et al. (2018) demonstrated that the carbon efficiency of a biomass to liquid (BTL) process can be increased by adding renewable power in the form of hydrogen and oxygen through electrolysis. The conversion of H₂ and CO₂ to liquid fuels via Power-to-Liquid

(PTL) processes is gaining attention (Dieterich et al. 2020). Van Bavel et al. (2020) proposed the hybrid “GTL-PTL” process where CO₂ and renewable H₂ are co-processed with natural gas to produce liquid fuel. Wismann et al. (2019) used electricity to drive the SMR reaction in an electrically heated wash-coated catalytic structure. As a result, the catalyst effectiveness factor is increased, production of side products is limited and a 100 fold volume reduction compared to a conventional SMR is observed (Wismann et al. 2019). In this paper, two options in reducing carbon emissions in the GTL process are investigated. Renewable power (from wind, solar, etc.) is integrated in both cases. As a result, the carbon emissions of both processes are considerably reduced. Moreover, with continued decline of cost of renewable power, these novel designs are considered to be cost competitive with the existing designs (Ostadi et al. 2015) in the future carbon-constrained world.

2. Process modelling

The specifications of natural gas feed are shown in Table 1. Aspen HYSYS V10 is used to simulate the process with Peng-Robinson as the thermodynamic model in this study.

Table 1: Specifications of the natural gas feed

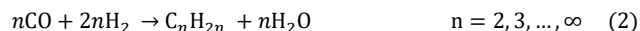
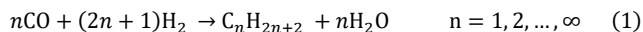
Temperature [°C]	50
Pressure [bar]	30
Flow [MMscfd]	120.2
Molar flow [kmol/h]	6000
Mass flow [t/h]	104.7
Mole fraction	
CH ₄	0.95
C ₂ H ₆	0.02
C ₃ H ₈	0.015
n-C ₄ H ₁₀	0.01
n-C ₅ H ₁₂	0.005

2.1 Syngas production

For the syngas production step, two main reformer types are considered: Autothermal Reforming (ATR) and electrically-heated Steam-Methane Reforming (E-SMR). These units are modelled as Gibbs reactors within the process simulator. To avoid coke formation in the reformer, a pre-reformer is used. Pre-reforming is done adiabatically at 420 °C. In both cases, the produced syngas is almost inert-free, due to avoiding the use of air in the reforming step. Therefore, there is no need for a costly Air Separation Unit (ASU) or a membrane unit to produce oxygen (Ostadi and Hillestad 2017).

2.2 Fischer-Tropsch Synthesis

The main products of FTS are paraffins (or alkanes) (Eq. 1) and olefins (or alkenes) (Eq. 2). The polymerization reactions taking place are hydrogenation of CO to form n-paraffins, 1-olefins, and oxygenates. Oxygenates are formed in small amounts and therefore are neglected here.



In order to account for product distribution, a variable chain growth factor, α , is applied. This factor is greatly affected by H_2/CO ratio at the reaction site on the catalyst (Ostadi, Rytter, and Hillestad 2016). It is shown that under-stoichiometric H_2/CO is desirable for selectivity to higher hydrocarbons as a result of suppression of methane and light hydrocarbon formation (Ostadi et al. 2016). The H_2/CO selected here is 2.0 which is slightly below the stoichiometric ratio. With H_2/CO being lower than the consumption ratio, it decreases as the reaction proceeds. Therefore, hydrogen addition is required to avoid an exceedingly low H_2/CO ratio. This is done by the addition of hydrogen between FT stages. The Fischer-Tropsch synthesis is staged in 3 stages, with product withdrawal and hydrogen addition between the stages. This enables a high conversion of syngas and high selectivity to heavier hydrocarbons. Once-through conversion in each FT stage is limited to 60 % to have the maximum C_{5+} selectivity and also preserve catalyst lifetime (Schanke et al. 2001). Slurry bubble-column FT reactors used in this study are modelled as a completely stirred tank reactor (CSTR). Details of the reactor and kinetic models are provided in Hillestad et. al (2018). The syngas has a pressure of about 26 bar prior to the first FT stage and a temperature of 210 °C at the feed to all FT stages.

3. Considered cases

3.1 Case 1: ATR with SOEC

Natural gas is pre-reformed and then reformed in an Autothermal Reformer (ATR) to produce syngas. The block flow diagram is shown in Figure 1. Renewable power is used in SOEC to split steam into H_2 and O_2 (Hillestad et al. 2018). Oxygen from the SOEC is used in ATR as the oxidant to keep the ATR outlet temperature at 1050 °C and therefore the produced syngas is inert-free. As a result, the majority of the tail gas (90 %) is recycled to ATR which increases the carbon efficiency of the process. Carbon efficiency is defined as the amount of the feed carbon ending up in the products. The rest of the tail gas is sent to the fired heater to provide heat for the steam to SOEC.

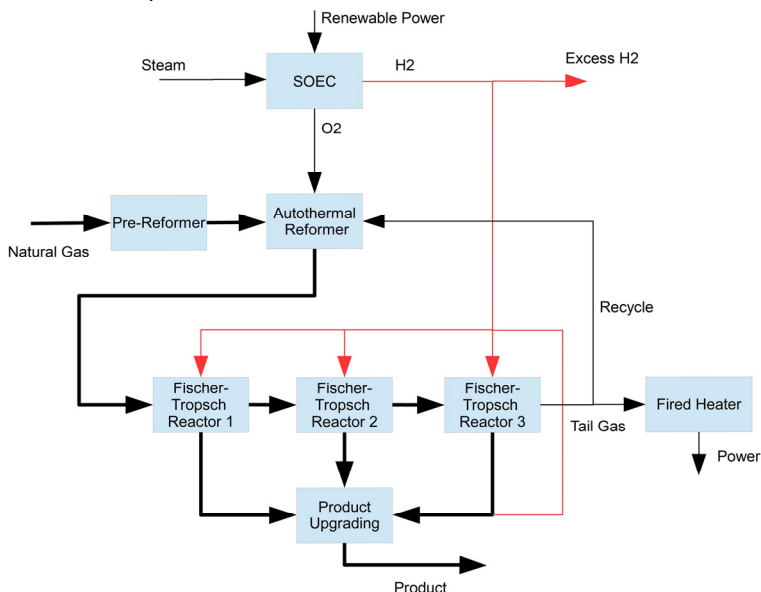


Figure 1: Block Flow Diagram of Case 1- ATR with SOEC

3.2 Case 2: Electrically heated SMR; E-SMR

In this case, renewable power derives the SMR reaction in an E-SMR. The block flow diagram is shown in Figure 2. The steam to carbon (S/C) ratio before the SMR is set to 2.0 which is a reasonable value considering the coking potential in the reformer. Natural gas is pre-reformed and heated to 500 °C prior to the SMR. The electric power input to the SMR is high enough to keep the SMR temperature at 850 °C. The H₂/CO ratio at the SMR outlet is high (> 3) which is more than the stoichiometric consumption ratio in FT process that is slightly above 2.0, as mentioned before. Therefore, part of H₂ needs to be separated to have the desired H₂/CO ratio prior to FT stages. Palladium membrane is used in this design to separate H₂. The temperature prior to the membrane needs to be at least 300 °C (Bredesen et al. 2011). There will also be excess H₂ in the process which can be used for product upgrading, provide extra heat in the fired heater or can be exported. In this design, 85% of the tail gas is recycled to the SMR.

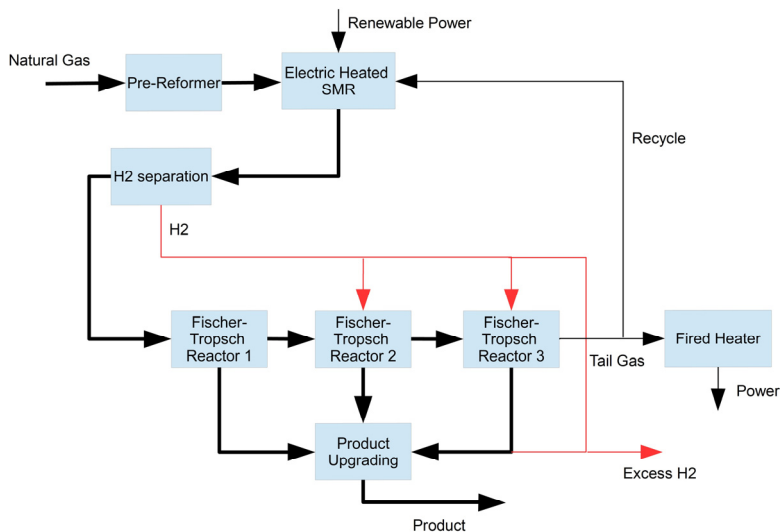


Figure 2: Block Flow Diagram of Case 2- E-SMR

3. Results

As mentioned earlier, in E-SMR case (case 2), excess H₂ is produced. In order to have a fair comparison between cases, the same excess H₂ is produced in both cases. Moreover, the same once-through conversion in three FT stages (i.e. 94 %) is used. The overall plant results in both cases are shown in Table 2. By assuming the density of syncrude (C₅₊) to be 800 kg/m³, the CO₂ emissions per liter syncrude produced are 0.20 and 0.43 kg CO₂/L_{syncrude} for case 1 and case 2, respectively. Comparing to the previous designs which had a CO₂ emission of 1.57 kg CO₂/L_{syncrude} (Ostadi et al. 2015; Ostadi and Hillestad 2016), case 1 and case 2 show 8 and 4 times less process emissions per liter syncrude produced, respectively.

Table 2. Carbon efficiency is considerably improved compared to the previous designs (Ostadi et al. 2015), from around 60% in previous designs to above 84 % and 91 %. The

power consumption in case 1 is slightly lower than case 2, while the FT production is higher. With the carbon efficiency being higher in case 1, this also indicates that the CO₂ emissions are lower in case 1. Moreover, the required FT catalyst volume is lower in case 1. The main reason for these differences is related to the higher reforming temperature used in ATR (case 1) than in the E-SMR (case 2) which are 1050 °C and 850 °C, respectively. The reforming temperature affects the CO₂ concentration out of the reformer which in case 1 and case 2 are 6% and 11%, respectively. Based on these results, it is a better choice to add renewable power to the SOEC rather than utilizing it in E-SMR. By assuming the density of syncrude (C₅₊) to be 800 kg/m³, the CO₂ emissions per liter syncrude produced are 0.20 and 0.43 kg CO₂/L_{syncrude} for case 1 and case 2, respectively. Comparing to the previous designs which had a CO₂ emission of 1.57 kg CO₂/L_{syncrude} (Ostadi et al. 2015; Ostadi and Hillestad 2016), case 1 and case 2 show 8 and 4 times less process emissions per liter syncrude produced, respectively.

Table 2: Overall plant results in both cases

	Case 1-SOEC	Case 2-ESMR
Electric power usage (MW)	633	635
FT production (t/h)	86.8	79.9
Carbon efficiency [%]	91.6	83.7
CO conversion per pass [%]	93.6	93.8
FT Catalyst volume [m ³]	1910	2680
Excess H ₂ production (kmol/h)	7908	7880
CO ₂ emission of the process (kg CO ₂ /L _{syncrude})	0.2	0.43

To get a better picture of the amount of heating and cooling required in the proposed processes, the energy composite curves for both cases are shown in Figure 3. In case 1, Figure 3a, there is a requirement for 85 MW of external heating and there is 492 MW of excess heat. In case 2, Figure 3b, the external heating and cooling requirements are 68 MW and 635 MW, respectively. The horizontal hot line at 210 °C represents the steam generated during cooling of the FT reactors. As can be observed, in both processes a large amount of excess heat is available mainly below 200 °C which can be utilized for preheating or be upgraded via heat pumps. Compared to the amount of extra heat required, there is a minor need for external heating in both designs, which can be supplied by changing important operating variables or combusting part of the excess hydrogen.

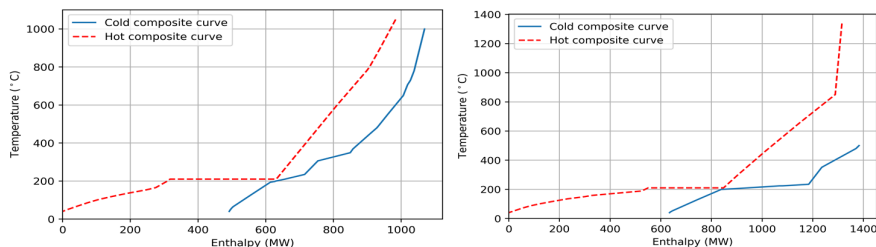


Figure 3: a) Composite curves for case 1; b) Composite curves for case 2

4. Conclusions

Two GTL designs with very low CO₂ emissions are presented and compared. In these designs, there is no need for a costly ASU or a membrane unit to produce oxygen. The generated syngas in both cases is almost inert-free because of avoiding the use of air, therefore allowing most of the FT tail gas to be recycled to the reformers. Adding renewable power increases the carbon efficiency and thus synthetic hydrocarbon production in both cases. Provided that cheap renewable power is available, it is a better choice to add power to the GTL process through SOEC (case1) rather than using it in an E-SMR (case2).

References

- Agrawal, R., N. R. Singh, F. H. Ribeiro, and W. N. Delgass. 2007. "Sustainable Fuel for the Transportation Sector." *Proceedings of the National Academy of Sciences* 104(12):4828–33. doi: 10.1073/pnas.0609921104.
- van Bavel, Svetlana, Sumit Verma, Emanuela Negro, and Maarten Bracht. 2020. "Integrating CO₂ Electrolysis into the Gas-to-Liquids–Power-to-Liquids Process." *ACS Energy Letters* 5(8):2597–2601. doi: 10.1021/acscenergylett.0c01418.
- Bredesen, Rune, Thijs A. Peters, Marit Stange, Nicla Vicinanza, and Hilde J. Venvik. 2011. "Chapter 11 Palladium-Based Membranes in Hydrogen Production." Pp. 40–86 in *Membrane Engineering for the Treatment of Gases: Volume 2: Gas-separation Problems Combined with Membrane Reactors*. Vol. 2. The Royal Society of Chemistry.
- Dieterich, Vincent, Alexander Buttler, Andreas Hanel, Hartmut Spliethoff, and Sebastian Fendt. 2020. "Power-to-Liquid via Synthesis of Methanol, DME or Fischer–Tropsch-Fuels: A Review." *Energy Environ. Sci.* 13(10):3207–52. doi: 10.1039/D0EE01187H.
- Hillestad, Magne, Mohammad Ostadi, Gonzalo Alamo Serrano, Erling Rytter, Bjorn Austbø, John Pharoah, and Odne Burheim. 2018. "Improving Carbon Efficiency and Profitability of the Biomass to Liquid Process with Hydrogen from Renewable Power." *Fuel* 234:1431–51. doi: 10.1016/j.fuel.2018.08.004.
- Ostadi, Mohammad, Kristin Dalane, Erling Rytter, and Magne Hillestad. 2015. "Conceptual Design of an Autonomous Once-through Gas-to-Liquid Process — Comparison between Fixed Bed and Microchannel Reactors." *Fuel Processing Technology* 139:186–95. doi: <https://doi.org/10.1016/j.fuproc.2015.07.022>.
- Ostadi, Mohammad, and Magne Hillestad. 2016. "Conceptual Design of an Autonomous Once-through Gas-to- Liquid Process with Microchannel Fischer-Tropsch Reactors." *Chemical Engineering Transactions* 52:523–28. doi: 10.3303/CET1652088.
- Ostadi, Mohammad, and Magne Hillestad. 2017. "Enriched Air or Pure Oxygen as Oxidant for Gas-to-Liquid Process with Microchannel Reactors." *Chemical Engineering & Technology* 40(10):1946–51. doi: <https://doi.org/10.1002/ceat.201700269>.
- Ostadi, Mohammad, Erling Rytter, and Magne Hillestad. 2016. "Evaluation of Kinetic Models for Fischer–Tropsch Cobalt Catalysts in a Plug Flow Reactor." *Chemical Engineering Research and Design* 114:236–46. doi: 10.1016/j.cherd.2016.08.026.
- Schanke, Dag, Petter Lian, Sigrid Eri, Erling Rytter, Bente Helgeland Sannæs, and Keijo J. Kinnari. 2001. "Optimisation of Fischer-Tropsch Reactor Design and Operation in GTL Plants." Pp. 239–44 in *Natural Gas Conversion VI*. Vol. 136, *Studies in Surface Science and Catalysis*, edited by E. Iglesia, J. J. Spivey, and T. H. Fleisch. Elsevier.
- United Nations. 2015. *UNFCCC, Paris Agreement, Decision 1/CP.21*.
- United Nations. 2019. *The Sustainable Development Goals*.
- Wismann, Sebastian T., Jakob S. Engbæk, Søren B. Vendelbo, Flemming B. Bendixen, Winnie L. Eriksen, Kim Aasberg-Petersen, Cathrine Frandsen, Ib Chorkendorff, and Peter M. Mortensen. 2019. "Electrified Methane Reforming: A Compact Approach to Greener Industrial Hydrogen Production." *Science* 364(6442):756–59. doi: 10.1126/science.aaw8775.

Comparative Study of Different Hybrid Life Cycle Assessment Methodologies Applied to Biomass-based Chemical Production

Yuqing Luo^a, Robert O'Dea^a, Thomas Epps^a, Marianthi Ierapetritou^{a*}

^aDepartment of Chemical and Biomolecular Engineering, University of Delaware, 150 Academy Street, Newark, Delaware, 19716, USA

mgi@udel.edu

Abstract

The environmental benefits of switching to biomass-based feedstock have been demonstrated in numerous life cycle assessment (LCA) studies. Currently, most LCA research follows the process LCA approach, which is inevitably subject to missing flows and truncation errors. Hybridization methods have been developed to reduce truncation error by incorporating the economic input-output model into the traditional process LCA. In this work, the tiered and integrated hybrid LCA, along with their double-counting elimination methods, are applied to five different biomass-based chemical productions. The results demonstrate the effectiveness of the hybrid LCA in eliminating truncation error and the significance of hybridization method selection.

Keywords: Hybrid LCA, Double counting elimination, Biomass

1. Introduction

There is a growing interest in utilizing biomass feedstocks to replace traditional fossil-based chemical production (Athaley et al. 2019). The feedstock of the process itself does not guarantee a greener production than the conventional oil-based ones. Since the utility usage and other chemical demands could differ significantly, the conclusion should only be drawn after the comparative life cycle assessment (LCA).

Two common LCA methodologies are process LCA and Economic Input-Output LCA (EIO-LCA), which suffer from the truncation error and aggregation error, respectively (Pomponi and Lenzen 2018). Many hybridization techniques are developed to take advantage of the process LCA's accuracy and the EIO-LCA's completeness (Crawford et al. 2018). Under various scenarios of data availability, different hybrid LCA methodologies use the input-output table (Yang et al. 2017) that reflects the interactions between different industries to estimate missing activities in the supply chain. These strategies will ensure a more complete system boundary than the process LCA.

The tiered hybrid and integrated hybrid LCA are the most common choices for hybridization in the literature (Crawford et al. 2018). The matrix formulation of these two methods is described in Section 2. This work assesses the current hybrid LCA methods through case studies of biomass-based p-Xylene productions and lignin-based pressure-sensitive adhesive (PSA) production with two fractionation alternatives in Section 4.

2. Process LCA and Hybrid LCA

The matrix representation of the intermediary flow exchanges within different unit processes is well suited for LCA, especially in large systems with recycling loops (Heijungs and Suh 2002). In this framework, the element a_{ij} (row i , column j) of the technological matrix, A_p , represents the physical amount of raw material i used in process j . The environmental impacts are calculated from equation (1), where the f_p represents the final demand, I is the identity matrix, and B_p is the environmental intervention matrix that connects the productions with emissions.

$$g = B_p \cdot (I - A_p)^{-1} \cdot f_p \quad (1)$$

Similarly, the environmental impacts of the EIO-LCA model are calculated by equation (2). The main difference between these two methods is that the input-output direct requirement matrix A_{IO} and its corresponding input-output environmental intervention matrix B_{IO} are used instead of A_p and B_p . Moreover, each element in A_{IO} represents the monetary flow spent on different economic sectors, rather than the physical flow in A_p .

$$g = B_{IO} \cdot (I - A_{IO})^{-1} \cdot f_{IO} \quad (2)$$

The tiered and integrated hybrid LCA are the most common choices for process-based hybridization. In equation (3), missing flows in the final demand of process LCA are supplemented by input-output elements in the tiered hybrid LCA (Crawford et al. 2018). Alvarez-Gaitan et al. defined the depth of interface between the input-output and process models and used this to include additional elements in f_{IO} . (Alvarez-Gaitan et al. 2013).

$$g = B_p \cdot (I - A_p)^{-1} \cdot f_p + B_{IO} \cdot (I - A_{IO})^{-1} \cdot f_{IO} \quad (3)$$

The integrated hybrid LCA includes the interactions between the process system and the economy (Suh 2004). These interactions are captured in equation (4) by the upstream cut-off matrix C^u and downstream cut-off matrix C^d (which is often set to null) (Perkins and Suh 2019). Many double-counting elimination methods are developed, among which the binary correction (Agez et al. 2020) and SSM correction are discussed in this paper (Strømman and Solli 2008). If a process is already covered in A_p , the binary correction method will set its corresponding economic sector in C^u to zero. The SSM correction further eliminates the elements in C^u if some flows are known to be complete. To account for uncertainty in the completeness of the background database, three scenarios (min, base, and max) on which sectors are already included in process LCA (Acquaye et al. 2011) are considered in the SSM correction.

$$g = \begin{pmatrix} B_p & B_{IO} \end{pmatrix} \cdot \begin{pmatrix} I - A_p & 0 \\ -C^u & I - A_{IO} \end{pmatrix}^{-1} \cdot \begin{pmatrix} f_p \\ 0 \end{pmatrix} \quad (4)$$

3. Methodologies

Process simulations of p-Xylene (Athaley et al. 2019) and pressure-sensitive adhesive production are simulated using Aspen Plus v11. Yields from literature and experiments are used for reactor modeling. Missing physical property data are estimated by regression of experimental data and the group contribution method. Cradle-to-gate analysis is performed including the manufacturing stage of chemical products and emissions associated with the upstream raw materials and utility production. The functional unit is

selected as 1 kg of p-Xylene or PSA produced. Raw material and utility usages are collected from the simulation and utilized in the traditional process LCA. The background data for the process LCA are obtained from the Ecoinvent v3.3 database (Wernet et al. 2016), while the USEEIO including 388 economic sectors is chosen as the economic input-output model (Yang et al. 2017) for hybridization. TRACI 2.1 method is used in the USEEIO model for impact assessment; the environmental metrics include global warming potential (GWP), ozone depletion, acidification, eutrophication, respiratory effects, smog, human carcinogenics and non carcinogenics effects. GWP is one of the most used indices for LCA and is chosen in this study. The environmental impacts of the lignin production from black liquor are calculated from the life cycle inventory of Bernier et al. (Bernier et al. 2013). Multi-functionality is treated as follows: mass allocation of the environmental impacts is chosen for p-Xylene productions when furfural is the byproduct, while the generated electricity in p-Xylene production and biofuel in PSA production are treated as avoided burdens.

4. Case Study Results: Bio-based Chemical Production

4.1. Descriptions of Bio-based p-Xylene and Pressure-Sensitive Adhesive Production

Two hydrolysis technologies for biomass-based p-Xylene production are discussed in this work: the concentrated acid (CA) (Weydahl 2014) and molten salt hydrate (MSH) hydrolysis (Sadula et al. 2017). Glucose and xylose from the hydrolysis are used to produce p-Xylene and furfural, as shown in Figure 1 (Athaley et al. 2019). The humins and lignin byproducts are incinerated to provide electricity and steam. One particular MSH hydrolysis scenario that is unable to sell the excess electricity from the power generation section to the grid is also considered and referred to as ‘MSH no grid.’

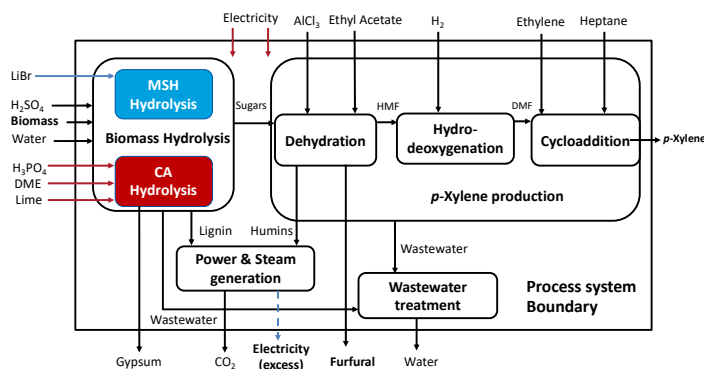


Figure 1: Process flow diagram of biomass-based p-Xylene production

The two alternatives for PSA production are illustrated in Figure 2. Using either glycerin or methanol as the solvent, lignin is depolymerized into a mixture of phenolics. Glycerin and methanol serve as the hydrogen donors for reductive catalytical fractionation, but typical processes using methanol or other low molecular weight alcohols employ hydrogen gas at high pressure (40 bar) to facilitate improved hydrogenation. Moreover, the glycerin fractionation process benefits from reactive distillation as the products are more volatile than the solvent. Unlike the pressurized methanol fractionation process, the glycerin process does not require cooling prior to separations and thus exhibits lower utility usage than the methanol fractionation process. Next, methacrylic anhydride (MAAH) functionalizes the phenolics with the help of 4-dimethylaminopyridine

(DMAP). After neutralization, monomers are polymerized with butyl acrylate (BA) by reversible addition-fragmentation chain-transfer polymerization in anisole with azobisisobutyronitrile (AIBN) as the initiator to produce the copolymer (PSA).

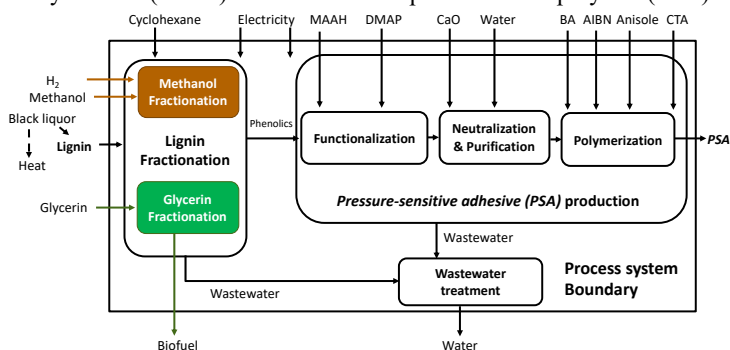


Figure 2: Process flow diagram of lignin-based pressure-sensitive adhesive (PSA) production

4.2. Truncation Errors in the Process LCA

The truncation errors in the process LCA comes from cutting off missing flows during the boundary selection. The extent of cut-off differs from case to case in process LCA studies, posing challenges to comparing different alternatives. The hybrid LCA has a more complete system boundary than the process LCA because it expands the system boundary by including previously missing upstream flows. It is often used to calibrate the truncation error in the process LCA (Mattila et al. 2010).

Table 1: Estimated truncation error (GWP) of process LCA based on hybrid LCA results

Estimated truncation Error (%)	p-Xylene production			PSA production	
	MSH	MSH (no grid)	CA	Methanol Fractionation	Glycerin Fractionation
Tiered hybrid	5.9	4.0	4.1	9.9	9.3
Binary-Integrated	14.6	13.0	12.6	26.6	21.4
SSM (min)	6.4	5.6	0.5	12.6	11.7
SSM (base)	6.7	5.9	0.9	13.7	12.1
SSM (max)	10.0	8.9	5.1	17.9	14.8

In Table 1, the truncation error estimated from missing upstreams varies based on the choices of hybridization methods and different technologies. For instance, glycerin fractionation for PSA production has a much higher estimated truncation error than the CA p-Xylene production. This is because many chemicals used in the PSA production processes are not captured well by the process LCA background database and need to be supplemented by the input-output analysis in the hybrid LCA. A high truncation error is an indicator of the original process LCA model's incompleteness, which encourages the use of hybrid LCA to replace the process LCA when comparing different alternatives.

The binary integrated hybridization estimates the highest truncation error among all methods as it removes the least amount of elements to avoid double-counting. The SSM correction is more conservative and acknowledges that irrelevant economic activities should be excluded in estimating missing flows. The minimum case of SSM only includes sectors highly probable to be neglected in the process database, while the maximum case accounts for all potential contributing sectors. These scenarios remove double-counting to different extents, reflecting the practitioner's confidence in the process LCA database.

The truncation errors vary by different magnitudes for different LCA metrics beyond carbon footprint. For instance, using the binary corrected integrated hybrid LCA as the reference, the truncation errors of PSA range from 0.11% (human non-carcinogenic effects) to 74.11% (ozone depletion).

4.3. Different Choices of LCA Methodologies in Comparative LCA

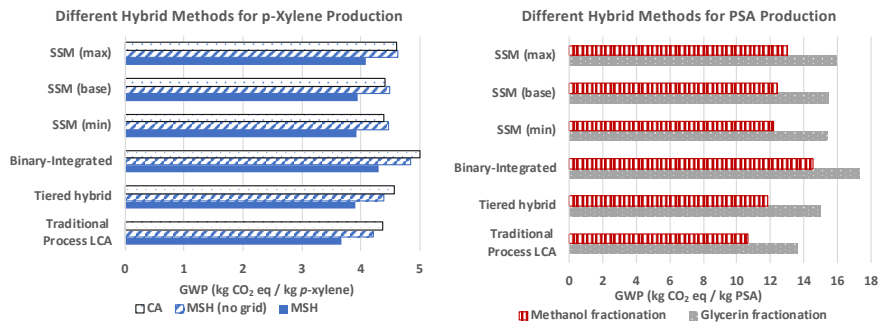


Figure 3: Effects of different hybridization methods on LCA results (Left: 1 kg biomass-based p-Xylene production; Right: 1 kg lignin-based pressure-sensitive adhesive (PSA) production)

Figure 3 shows that the MSH process of p-Xylene production without selling electricity to the grid has a lower global warming potential than the CA process according to the process LCA, tiered hybrid LCA, and binary corrected integrated hybrid LCA. After further eliminating double-counting by SSM corrections, the MSH process without selling electricity to the grid behaves similarly or even worse than the CA process. This comparison illustrates that the choice of hybridization method could lead to contradicting results. It is beneficial to choose SSM corrected integrated hybrid LCA for its flexibility. However, this method requires the practitioners to use their prior knowledge on the process and database to decide on the irrelevant economic activities. Tiered hybrid LCA should be applied if accurate technoeconomic analysis is available. If no extra information is available, the binary corrected integrated hybridization is the only option. Since the truncation errors for similar processes are very close, the binary corrected hybrid LCA and tiered hybrid LCA will still lead to reliable results in most cases. For example, glycerin fractionation outperforms methanol fractionation in PSA production, and the MSH process that sells the extra electricity to the grid behaves better than the CA p-Xylene process in all and hybrid LCA or even process LCA comparisons.

5. Conclusions

Both tiered and integrated hybrid LCA methods are applied in this work to achieve a more complete system boundary than traditional process LCA by incorporating the input-output model. In cases where the process LCA database does not sufficiently capture some inputs, the truncation error is effectively reduced through hybridization. Although different LCA methodologies generate similar results for the same process, the comparison between many technology alternatives could be affected by the hybridization method selection, as shown in Section 4.3. One limitation of the hybrid LCA, as with most of the current process LCA and EIO-LCA studies, is the model's static and linear nature, limiting its ability to assess future scenarios and significant changes (Yang and Heijungs 2019). Future studies should include more dynamic technological development features and consider the economy's complexity by nonlinear models.

Acknowledgement

The authors are grateful for financial support from the National Science Foundation (NSF GCR CMMI 1934887) and U.S. Department of Energy's RAPID Manufacturing Institute for Process Intensification (DE-EE000788-7.6).

References

- Acquaye, A. A., Wiedmann, T., Feng, K., Crawford, R. H., Barrett, J., Kuylensstierna, J., Duffy, A. P., Koh, S. C. L. and McQueen-Mason, S., 2011, Identification of 'Carbon Hot-Spots' and Quantification of GHG Intensities in the Biodiesel Supply Chain Using Hybrid LCA and Structural Path Analysis. *Environmental Science & Technology* 45(6): 2471-2478.
- Agez, M., Majeau-Bettez, G., Margni, M., Strømman, A. H. and Samson, R., 2020, Lifting the veil on the correction of double counting incidents in hybrid life cycle assessment. *Journal of Industrial Ecology* 24(3): 517-533.
- Alvarez-Gaitan, J. P., Peters, G. M., Rowley, H. V., Moore, S. and Short, M. D., 2013, A hybrid life cycle assessment of water treatment chemicals: an Australian experience. *The International Journal of Life Cycle Assessment* 18(7): 1291-1301.
- Athaley, A., Annam, P., Saha, B. and Ierapetritou, M., 2019, Techno-economic and life cycle analysis of different types of hydrolysis process for the production of p-Xylene. *Computers & Chemical Engineering* 121: 685-695.
- Athaley, A., Saha, B. and Ierapetritou, M., 2019, Biomass-based chemical production using techno-economic and life cycle analysis. *AIChE Journal* 65(9): e16660.
- Bernier, E., Lavigne, C. and Robidoux, P. Y., 2013, Life cycle assessment of kraft lignin for polymer applications. *The International Journal of Life Cycle Assessment* 18(2): 520-528.
- Crawford, R. H., Bontinck, P.-A., Stephan, A., Wiedmann, T. and Yu, M., 2018, Hybrid life cycle inventory methods – A review. *Journal of Cleaner Production* 172: 1273-1288.
- Heijungs, R. and Suh, S., 2002, The Computational Structure of Life Cycle Assessment.
- Mattila, T. J., Pakarinen, S. and Sokka, L., 2010, Quantifying the Total Environmental Impacts of an Industrial Symbiosis - a Comparison of Process-, Hybrid and Input-Output Life Cycle Assessment. *Environmental Science & Technology* 44(11): 4309-4314.
- Perkins, J. and Suh, S., 2019, Uncertainty Implications of Hybrid Approach in LCA: Precision versus Accuracy. *Environmental Science & Technology* 53(7): 3681-3688.
- Pomponi, F. and Lenzen, M., 2018, Hybrid life cycle assessment (LCA) will likely yield more accurate results than process-based LCA. *Journal of Cleaner Production* 176: 210-215.
- Sadula, S., Oesterling, O., Nardone, A., Dinkelacker, B. and Saha, B., 2017, One-pot integrated processing of biopolymers to furfurals in molten salt hydrate: understanding synergy in acidity. *Green Chemistry* 19(16): 3888-3898.
- Strømman, A. H. and Solli, C., 2008, Applying Leontief's Price Model to Estimate Missing Elements in Hybrid Life Cycle Inventories. *Journal of Industrial Ecology* 12(1): 26-33.
- Suh, S., 2004, Functions, commodities and environmental impacts in an ecological-economic model. *Ecological Economics* 48(4): 451-467.
- Wernet, G., Bauer, C., Steubing, B., Reinhard, J., Moreno-Ruiz, E. and Weidema, B., 2016, The ecoinvent database version 3 (part I): overview and methodology. *The International Journal of Life Cycle Assessment* 21(9): 1218-1230.
- Weydahl, K. R. (2014). Process for the production of alcohol. United States, WEYLAND AS.
- Yang, Y. and Heijungs, R., 2019, Moving from completing system boundaries to more realistic modeling of the economy in life cycle assessment. *The International Journal of Life Cycle Assessment* 24(2): 211-218.
- Yang, Y., Ingwersen, W. W., Hawkins, T. R., Srocka, M. and Meyer, D. E., 2017, USEEIO: A new and transparent United States environmentally-extended input-output model. *Journal of Cleaner Production* 158: 308-318.

Optimising the energy, water and food nexus node to support decision making for sustainable food security in Risky Environments

Maryam Haji,^a Rajesh Govindan,^b Tareq Al-Ansari^{a,b}

^aDivision of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha - Qatar

^bDivision of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha – Qatar

Abstract

As the population continues to rise, the need for energy, water and food (EWF) resources are increasing in parallel. Accordingly, the EWF nexus has received considerable attention in previous years as a means to manage resources given the inherent interlinkages between them. The objective of this study is to expand on geospatial nexus approaches to support decision-making for EWF resources in non-resilience environments. A simple linear optimization model is applied to a geospatial representation of decentralised EWF systems to enhance the EWF nexus node approach and analysis of resilience. The methodology is applied to hydroponic greenhouses in Qatar, where the water is the core factor that affects the overall efficiency and AHP risk level. Thus, the optimization model introduces a new EWF node that operates at minimum cost whilst reducing the AHP risk level of existing EWF nodes. The objective function of the optimisation model accounts for three cost components; the cost of groundwater transportation, pumping and desalination, that is subjected to certain constraints; one constraint is suggested for the new EWF node to select a low risk area, hence ensuring low operating costs; three constraints assigned to identify a location for the new EWF node that will minimise risks of the three existing nodes, hence, reducing high AHP value of three farms; and the last constraint used to select one node only at an optimal location. The result demonstrates that with the allocation of a new EWF nexus node at an optimal location to be the main source of water supply, the risk level for all three existing nodes will reduce tremendously.

Keywords: Decentralization, Resource Management, EWF Nexus, Optimisation.

1. Introduction

In modern society, there are concerns related to the availability and accessibility of energy, water and food (EWF) resources. Hence, the EWF nexus approach to resource management has been heavily explored in the literature in recent years since its introduction in Bonn Conference (2011), to support resource management across these sectors whilst considering the inter-linkages between them (Al-Ansari et al., 2014; 2015). Recently, numerous scholars, academics and politicians have suggested various approaches to tackle uncertainties within EWF using different types of optimisation techniques and other intelligence algorithms (Li & Singh, 2019). Numerous studies demonstrated that computational modelling is an important way of quantifying WEF interactions and fostering sustainable decision-making (Bieber et al., 2018). It has been

considered as influential tool due to the ability in addressing resource optimisation problems with various objectives and combining several management features, goals and constraints into one system. Stamou & Rutschmann (2018) optimised energy, water and food resources using the parameterization-simulation-optimisation method, which is based on the WEF Nexus approach and by considering their related existence and interrelationships. The authors aimed to promote understanding of changes by decision makers as a consequence of the altered management policies through analysing the interchanges between hydropower and competing irrigating goals and to make water supply management more effective in the Nile region. The proposed Pareto optimisation based on WEF nexus method can be used to optimise water management with similar issues and at the same time having various goals. Additionally, Niu et al. (2019) suggested a linear objective linear fractional programming to model the interactions between WEA nexus. This study provides overall insights into agriculture and the management of water resources in arid areas and provides a new approach to tackle the complicated dynamic WEA nexus through a modelling technique. In general, mathematical modelling and other optimisation techniques have been primarily used to improve the overall food systems performance. Moreover, an integrated resource modelling with an EWF Nexus framework demonstrates useful results in a multi-objective and stochastic optimisation as they can cope with the complexities of Nexus related issues. As such, Namany et al., (2018) suggested an optimum energy-water mix model to reduce environmental burdens and costs related to food production in the State of Qatar. Specifically, the research evaluated the allocation of energy and water resources using a game-theoretical modelling approach, and by focusing on the impact that natural gas prices on the stochasticity of the market and competition in the power-generation business. Lately, Haji et al. (2020) introduced the EWF Node to describe the decentralisation of interconnected EWF Nexus systems. The study utilised the EWF nexus node, where every node represented a food sector sub-system that are influenced by external risks, such as climatic, water and soil factors. Then, spatial risk factors were gathered, digitalised, and incorporated into a single geo-processing framework from various data sources focused on the food industry, allowing the visualisation as well as the further processing of risks. Thus, a composite geospatial risk maps for the different food production scenarios were then generated using an analytical hierarchy process (AHP). Consequently, this study builds on recent work reported by Haji et al. (2020), expanding the geospatial Nexus framework developed to support decision-making within EWF Nexus, especially in hazardous non resilience environments, in order to reduce the effect of various risk factors, thus improving the national adaptability of EWF systems. This is achieved by integrating the computational modelling and optimisation approaches within the EWF nexus node.

2. Linear programming optimisation framework for EWF Nexus

Supporting objectives for sustainable food security in risky environments in this study is performed by means of a simple linear programming optimisation on the composite geospatial risk map for the hydroponic greenhouse case reported by Haji et al. (2020), which is the current state of Qatar's agriculture farms. To this point, the current state of existing nodes (four hydroponic greenhouses) has been detailed and analysed. Briefly, the risks were defined based on the chosen food sector. Thus, the main risk factors that have an impact on greenhouses in general are categorised into three: climate (temperature, humidity and solar radiation), water (groundwater depth, salinity, recharge rate and pH)

and soil factor (As and Fe concentration). However, since this study considers a hydroponic greenhouse, then the soil indicators are removed from the assessment. Then, a comparison between each indicator for the particular criterion is made in pair by classifying which of these indicators were extremely critical and by how much. The preferences are expressed on a semantical scale from 1 to 9, where 1 denotes equal importance between the two indicators, and the value 9 implies that first indicator is 9 times more critical than the second indicator. Finally, using the importance percentage of each factor from AHP, a composite geospatial risk map representing is generated from GIS. Thus, the primary purpose of the optimisation in this study is to minimise the risk of existing nodes located in high-risk areas. Figure 1 illustrates that 3 out of 4 existing nodes are located in high-risk areas, and includes Al Safwa, Global and AGRICO farms with a risk level of 28.91, 22.60 and 24.29 respectively. The objective of this optimisation is to allocate a new EWF node at an optimal location, which will operate at minimum cost, whilst minimising the risk of existing nodes especially those located in high risk areas. Hence, the three main cost components are considered in this case study. The first component is the cost to transport groundwater from new to existing EWF nodes. The second component is the cost associated with pumping groundwater to the surface, and the final component is the cost of desalinating pumped water having certain groundwater salinity. The decision variables comprise of 3 parameters, including the transportation and the groundwater variables, which are elaborated in Table 1. The optimal decisions of the selected variables will minimise the operating cost of the new node locating at the optimal location and overall risk of existing nodes. The mathematical formulation for the optimisation model is presented below, where spatial (distance) and economic related data are used to run the optimisation model, along with weights associated with each risk factor. Since the optimisation process involves many decision variables and various calculation steps; then, a computer model such as Excel Solver is valuable in performing these steps, and to solve the optimisation formulation by ensuring that all functions and variables are sufficiently expressed. In terms of the spatial and temporal scales, the proposed approach will vary under various geographical characteristics and will differ over different seasons. In this study, the optimisation model is applied for hydroponic greenhouse distributed within the State of Qatar for Summer season data.

Table 1: Mathematical formulation for the optimisation model.

Objective Function:	
$\text{Min Cost} = \sum_{i=1}^{28} \sum_{j=1}^3 d_{ij} C_T x_i + \sum_{i=1}^{28} D p_i C_p x_i + \sum_{i=1}^{28} S_i C_D x_i$	Identifies the optimal distance and groundwater characteristics that minimises the total operating cost of the new EWF node that will supply the required water demands to the existing EWF nodes.
Subject to:	
$\sum_{i=1}^{28} x_i (w_s S_i + w_{ph} p H_i + w_{rr} R R_i + w_d D p_i + w_t T_i + w_h H_i + w_{sr} S R_i) \leq 17$	This constraint is suggested for the new EWF node in order to select a low-risk location with an AHP value less than 17, hence ensuring low operating cost
$\sum_{i=1}^{28} x_i (w_s S_i + w_{ph} p H_i + w_{rr} R R_i + w_d D p_i) + w_t T_1 + w_h H_1 + w_{sr} S R_1 \leq 28.9100$	This constraint is assigned in order to identify a location for the new EWF node that will minimise risks of the existing node; thus, reducing the high AHP value of Al Safwa Farm to a value less than 28.91
$\sum_{i=1}^{28} x_i (w_s S_i + w_{ph} p H_i + w_{rr} R R_i + w_d D p_i) + w_t T_2 + w_h H_2 + w_{sr} S R_2 \leq 22.6039$	This constraint is assigned in order to identify a location for the new EWF node that will minimise risks of the existing node; thus, reducing the high AHP value of Global Farm to a value less than 22.60

$\sum_{i=1}^{28} x_i (w_s S_i + w_{ph} p H_i + w_{rr} R R_i + w_d D p_i) + w_t T_3 + w_h H_3 + w_{sr} S R_3 \leq 24.2957$	<p>This constraint is assigned in order to identify a location for the new EWF node that will minimise risks of the existing node; thus, reducing the high AHP value of AGRICO Farm to a value less than 24.29</p>
$\sum_{i=1}^{28} x_i = 1$	<p>This constraint used so as to select one node only at an optimal location</p>
$d_{ij}, D p_i, S_i, x_i \geq 0$	<p>It implies that all decision variables must be strictly positive.</p>

The optimisation begins with the initialisation of the database, including the distance and cost matrix between new and existing EWF nodes, the groundwater salinity value and the cost associated to desalinate such salinity level, in addition to the groundwater depth value along with the cost associated with pumping groundwater. It is then followed by the simulation steps to calculate the operating cost of the initial data. The SAIC farm is excluded from the analysis; as it is the only farm located in a low-risk area. However, in order to calculate the capacity of the new RO desalination plant, it is assumed that the same RO characteristics installed in the SAIC farm is applied to the new EWF node. Currently, SAIC farm has two RO plants with a total capacity of 1070 m³/day that fulfilled 10,223 L/kg/year of water requirements for hydroponic farming, in addition to other water requirements within the farm. Hence, the new node should have an RO plant with an approximate unit size of 19,000 m³/day, in order to satisfy the three existing farm total water requirements, which is equivalent to 176,091.25 L/kg/year. However, by considering new EWF node water requirements satisfaction, the total capacity of the new RO plant is assumed to be 24,000 m³/day. Furthermore, the electricity consumption for RO desalination with a unit size of 24,000 m³/day is almost 2.10 kWh/m³ for brackish water with a salinity of approximately 5000 ppm (Manju & Sagar, 2017). Noting that the groundwater salinity in Qatar varies between 2206.12 and 5639.81 ppm, and the subsequent energy required in order to desalinate using RO desalination will range from 0.93 to 2.37 kWh/m³. Hence, the cost of desalinating groundwater using the RO desalination process is approximately between 1467.12 \$ and 9588 \$; where the rate associated with electricity consumption for productive farms in the State of Qatar is 1.68 QAR/kWh (KAHRAMAA, 2020). The other cost component in the objective function is the transportation cost, a previous study demonstrated that the approximate cost is 0.061 \$/m³ per 100 km to transport water, thus it is equivalent to 0.00061 \$/m³/km (Zhou & Tol, 2005). Therefore, the distance between the new node and the existing node is calculated using the distance formula. Finally, the operating cost of lifting groundwater to the surface using an electrical motor is 0.1780 \$/kWh, where 2.725 kWh is required for pumping groundwater per each meter (Robinson, 2002). Thus, the groundwater pumping costs at various locations were then calculated. Ultimately, all the previous data for the minimisation of the objective function and set of constraints required to perform optimisation are inserted into Excel Solver, where Simplex LP is used.

3. Geospatial optimisation model application

A simple linear optimisation is conducted to determine the optimal location for the new EWF node to be allocated and established. The chosen location minimised the risk of existing EWF nodes whilst operating at minimal operating costs. In this study, hydroponic greenhouses represent the food element of the nexus, where the water is the core factor that affects the overall efficiency and AHP risk level. Thus, the objective of the optimisation is to identify a location for new EWF nodes, with better water quality; in order to reduce the operational cost, such as the cost of pumping and desalinating

groundwater, and hence supply good water quality to existing EWF node to reduce their overall risk. Figure 1 illustrates the optimal node with respect to 28 potential nodes. Although the potential EWF nodes are distributed randomly, including high risk area, the four constraints assigned to the optimisation model contribute to enhancing its ability to select optimal EWF nodes located at low-risk areas with better groundwater quality. Moreover, outcomes from the linear programming optimisation validated the trade-offs between cost components. Though, the location of the new node (north-east) is slightly far from existing nodes, which implies a larger transportation cost. However, the location with better groundwater salinity and less depth is selected. This demonstrates that through the allocation of a new node in an optimal location, the risk of existing nodes decreased tremendously. Initially, Al Sawfa, Global and Agrico farms are located in high-risk areas indicating poor water qualities with an AHP risk value of 28.1, 22.6 and 24.3 respectively. However, with the allocation of a new EWF nexus node to be the main source of water supply, the risk level reduces to approximately 10 for all three existing nodes. Therefore, the optimisation result proves the necessity to use groundwater with improved qualities within EWF nodes.

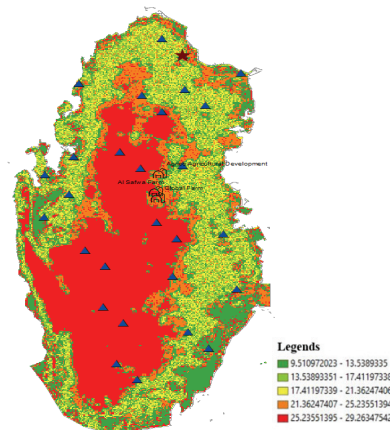


Figure 1: The location of the optimal selected EWF node corresponding to the 4 hydroponic farms and 28 potential nodes.

4. Conclusion

This study implements a simple linear optimisation model to introduce a new EWF node that will reduce the AHP risk level of existing EWF nodes. The objective function of the optimisation model accounts for three cost components; the cost of groundwater transportation, pumping and desalination. Hence, by running the optimisation model into an excel solver, a location in the north-east of Qatar is selected as an optimal location in order to allocate the new EWF node, which will operate at the least cost whilst reducing the AHP risk of three existing nodes. In conclusion, the approach of EWF nexus node deployed in this study can be utilised to reflect various farming industries that have different spatial and temporal characteristics. For instance, a dairy, fodder or combined farm are affected by various geospatial risk factor, thus it can be analysed using the same approach. Evidently, the geospatial optimisation of EWF nexus node is promising as it classifies nodes that are performing well in a methodological risk analysis, thus paving

the way for further optimization and decision-making with a view to reducing the effects of risk factors, and therefore, improving the national resilience of the EWF systems.

Acknowledgments

This research was made possible by an Award (GSRA7-1-0407-20014) and supported by proposal number NPRP11S-0107-180216 from Qatar National Research Fund (a member of Qatar Foundation). The contents herein are solely the responsibility of the authors[s].

References

- Al-Ansari, T., Korre, A., Nie, Z., & Shah, N. (2014). Development of a Life Cycle Assessment Model for the Analysis of the Energy, Water and Food Nexus. In J. J. Klemeš, P. S. Varbanov, & P. Y. Liew (Eds.), *24th European Symposium on Computer Aided Process Engineering* (pp. 1039–1044). <https://doi.org/https://doi.org/10.1016/B978-0-444-63455-9.50008-8>
- Al-Ansari, T., Korre, A., Nie, Z., & Shah, N. (2015). Development of a life cycle assessment tool for the assessment of food production systems within the energy, water and food nexus. *Sustainable Production and Consumption*, 2(July), 52–66. <https://doi.org/10.1016/j.spc.2015.07.005>
- Bieber, N., Ho, J., Wang, X., Triantafyllidis, C., Dam, K. H. Van, Koppelaar, R. H. E. M., & Shah, N. (2018). Sustainable planning of the energy-water-food nexus using decision making tools. *Energy Policy*, 113(July 2017), 584–607. <https://doi.org/10.1016/j.enpol.2017.11.037>
- Bonn2011 Conference. (2011). *The Water, Energy and Food Security Nexus – Solutions for a Green Economy*.
- Haji, M., Govindan, R., & Al-Ansari, T. (2020). Novel approaches for geospatial risk analytics in the energy–water–food nexus using an EWF nexus node. *Computers and Chemical Engineering*, 140. <https://doi.org/10.1016/j.compchemeng.2020.106936>
- KAHRAMAA. (2020). Productive Farms Tariff. Retrieved from <https://www.km.qa/Customerservice/Pages/Tariff.aspx>
- Li, M., & Singh, V. P. (2019). Sustainability of water and energy use for food production based on optimal allocation of agricultural irrigation water. *International Journal of Water Resources Development*, 0627. <https://doi.org/10.1080/07900627.2019.1649129>
- Manju, S., & Sagar, N. (2017). Renewable energy integrated desalination: a sustainable solution to overcome future fresh-water scarcity in India. *Renewable and Sustainable Energy Reviews*, (73), 594–609.
- Namany, S., Al-Ansari, T., & Govindan, R. (2018). Integrated techno-economic optimization for the design and operations of energy, water and food nexus systems constrained as non-cooperative games. In *Computer Aided Chemical Engineering* (pp. 1003–1008). <https://doi.org/10.1016/B978-0-444-64241-7.50162-2>
- Niu, G., Zheng, Y., Han, F., & Qin, H. (2019). The nexus of water, ecosystems and agriculture in arid areas: A multiobjective optimization study on system efficiencies. *Agricultural Water Management*, 223(June), 105697. <https://doi.org/10.1016/j.agwat.2019.105697>
- Robinson, D. W. (2002). Construction and operating costs of groundwater pumps for irrigation in the riverine plain. *Technical Report 20/02*, (January), 28. Retrieved from <http://www.clw.csiro.au/publications/technical2002/tr20-02.pdf>
- Stamou, A. T., & Rutschmann, P. (2018). Pareto Optimization of Water Resources Using the Nexus Approach. *Water Resources Management*, 32(15), 5053–5065. <https://doi.org/10.1007/s11269-018-2127-x>
- Zhou, Y., & Tol, R. S. J. (2005). Evaluating the costs of desalination and water transport. *Water Resources Research*, 41(3), 1–10. <https://doi.org/10.1029/2004WR003749>

Synthesis of Sunlight, Seawater and CO₂ based Industrial Parks

Elizabeth J. Abraham, Farah O. Ramadan, Dhabia M. Al-Mohannadi*, Patrick Linke

Chemical Engineering Program, Texas A&M University at Qatar, Education City, Doha, Qatar

Dhabia.al-mohannadi@qatar.tamu.edu

Abstract

In response to climate change concerns, there has been an increasing global commitment to reduce carbon dioxide emissions. Circular economy pillars have been proposed as a means to conserve resources, limit emissions and maintain sustainable revenue streams. Most CO₂ emissions stem from stationary sources, which are largely due to industrial processes (IPCC, 2005). In an attempt to reduce these emissions, there has been an increased focus on carbon capture utilization and storage solutions, e.g. integrating carbon dioxide in industrial parks to create value-added products. Ahmed et al. (2020) introduced a method that can represent any resource, material or energy, within a cluster to identify integration potential. In this work, we utilize the mixed integer linear program to synthesize waste reuse clusters at zero carbon footprint. The work mainly assesses the potential benefits of utilizing waste resources such as CO₂ emissions, free resources such as seawater (H₂O and its minerals), air (N₂, and O₂) and sunlight as an energy source to produce various value-added products to create carbon neutral industrial parks. This attempt demonstrates circular economy values on an industrial park level by conserving fresh resources, maximizing reuse of materials and creating valuable products. The system studied was able to generate profits while adhering to emission and material constraints.

Keywords: Climate Mitigation, Renewable Energy, Water Reuse, Optimization, Integration

1. Introduction

In recent years, there has been an alarming increase in greenhouse gases in the atmosphere, the most notable of these gases being carbon dioxide (CO₂). Thus, there is an urgent need to minimize these emissions and alleviate the climate change concerns that result from them, identified as one of the United Nations Sustainable Development Goals for 2030. One possible solution for CO₂ reduction is the generation of energy from renewable sources (such as solar or wind power) in place of non-renewable sources such as through burning fossil fuel with a large CO₂ footprint. Another way to reduce CO₂ emissions is through carbon capture and utilization (CCUS) systems. This involves the capture and use of CO₂ from emissions in order to produce value-added products. Both of these methods are being studied extensively to determine their applicability in industry. The concept of an eco-industrial park (EIP) where sustainable practices are incorporated into industrial processes can be used to achieve sustainable development goals. An EIP, as defined by Côté and Hall (1995) is “an industrial system which conserves natural and

economic resources; reduces production, material, energy, insurance and treatments costs and liabilities; improves operating efficiency, quality, worker health and public image; and provides opportunities for income generation from use and sale of wasted materials.” The principle governing the operation of EIPs is industrial-ecology, where they operate much like natural systems in which resources are conserved or reused (Valenzuela-Venegas, 2016), while also gaining economic, environmental, and social benefits that would be otherwise be unachievable by a stand-alone facility.

Numerous works have been applied to promote sustainability through material integration. This includes research by Lovelady and El-Halwagi (2009), whose work presented an approach to water integration in industrial parks through optimization. Al-Mohannadi and Linke (2016) uses optimization to integrate CO₂ in industrial parks. Alnouri et al. (2018) established a brine management water integration approach through a mathematical modeling program. El-Halwagi et al. (2003) developed a procedure to minimize the amount of fresh feed required for a source sink system using a set of linear algebraic equations modeling the mass balance of fresh feed and contaminants. While these works all focused on one aspect of integration, multiple resources must be integrated to achieve maximum reduction in CO₂ emissions. Ahmed et al. (2020) developed an optimization-based mixed integer linear program (MILP) for this purpose. The representation allows the exchange of any resource, material or energy within a cluster. Thus, this work uses the Ahmed et al. (2020) resource integration model to design a profitable carbon neutral industrial park from natural sources (seawater, air, sunlight, and CO₂). The aim is to use material and energy integration to have the same the amount of CO₂ emitted from the park as that entering the park.

2. Methodology

The methodology used in this work to achieve the objective involves multiple steps. First, the inputs in to the city are defined, then a database of possible reactions and processes to convert the raw materials is created. Finally, the resource integration model is applied to optimize for maximum profit while meeting the emission constraint. As seawater and air are inputs to the cluster, various desalination and air separation technologies were assessed to choose that which was best suited for the application. Next, a number of processes were chosen as potential candidates to be selected as part of the EIP. An ideal candidate is one that can utilize the raw materials or products produced by other plants, is profitable, and is feasible in the geographical location of the industrial cluster. A database was created containing the mass balances, energy requirements, capital costs, and operating costs. In addition, other relevant information for each of the specified candidate plants such as material and utility prices and market sizes, were also included. The candidate city was then optimized using the resource integration model. This gives specified optimal operational capacities for each plant and determined the connections between them. The model used was developed by Ahmed et al. (2020) and the system equations can be found in their work. The model constraints were the minimum and maximum possible capacities and the allowable inputs.

$$\begin{aligned} \text{Max} \quad & \text{Profit} = f(\text{Resources flows}) \\ & g(x_1, x_2, \dots, x_n) \leq 0 \\ & h(x_1, x_2, \dots, x_n) = 0 \end{aligned}$$

3. Case Study

The approach developed here is applied to a case study where the objective is to design an EIP that is carbon neutral and profitable as illustrated in Figure 1.

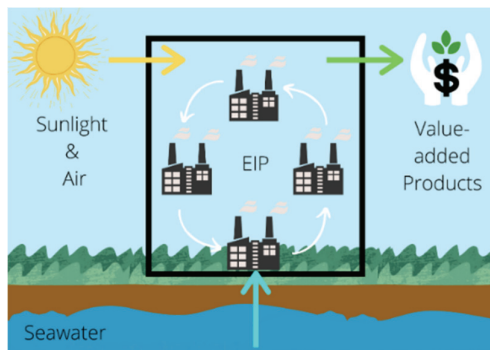


Figure 1: Illustration of the case study objective

Through the process selection phase, 12 candidate plants were determined. These include production units for hydrogen (H₂), ammonia (NH₃), sodium chloride (NaCl), formalin (CH₂O), methanol (CH₃OH), soda ash (Na₂CO₃), urea (CH₄N₂O), nitric acid (HNO₃), and sodium bicarbonate (NaHCO₃), along with a seawater reverse osmosis (SWRO) unit, an air separation unit (ASU) that gives oxygen (O₂) and nitrogen (N₂). Of these, the CH₃OH, NaHCO₃ and CH₄N₂O units have their own carbon capture units since they release emissions. A sequestration unit is also present to store any pure CO₂ that is not converted. The capital cost requirement or CAPEX parameter for each of the units was calculated by assuming a reasonable maximum capacity limit and an operational life of 20 years utilizing the CAPEX estimation techniques found in Peters et al. (2002). Table 1 lists these calculated CAPEX parameters along with the maximum allowed capacities.

Table 1: Selected processes, their reference products and CAPEX parameters

Process	Reference Product (RP)	CAPEX Parameter (\$/ton RP)	Maximum Allowed Capacity (t/y)
Sequestration	CO ₂	20.00	120,000
SWRO	H ₂ O	0.02	800,000
H ₂ Production	H ₂	779.00	100,000
Air Separation	O ₂	7.70	500,000
NH ₃ Production	NH ₃	20.00	200,000
CC Unit	CO ₂	8.00	20,000
NaCl Production	NaCl	5.65	200,000
CH ₂ O Production	CH ₂ O	52.88	200,000
CH ₃ OH Production	CH ₃ OH	14.67	200,000
Na ₂ CO ₃ Production	Na ₂ CO ₃	140.56*	200,000
CH ₄ N ₂ O Production	CH ₄ N ₂ O	7.53	200,000
HNO ₃ Production	HNO ₃	5.30	200,000
NaHCO ₃ Production	NaHCO ₃	27.46	50,000

*It should be noted that the Na₂CO₃ unit is assumed to produce sodium hydroxide (NaOH) needed for its own consumption and the capital and operating expenses for this production are reflected in the Na₂CO₃ CAPEX parameter. This NaOH production does not require any additional imports into the cluster.

The mass and utility resource parameters for these processes were obtained through mass and energy balances performed on data obtained from an extensive literature review. A negative parameter indicates that the resource is utilized in the process while a positive parameter indicates that it is an output of the process. The parameters used for the ASU, sequestration and production units for H₂, NH₃, CH₃OH and CH₄N₂O were taken from Ahmed et al. (2020), while the parameters for the capture units were calculated using the capture units in their work as a basis. Data for the SWRO and NaCl production units were obtained from Nayar et al. (2019). The data for the CH₂O unit was obtained from Millar and Collins (2017) and Bahmanpour et al. (2014). For Na₂CO₃, NaHCO₃ and HNO₃, the data was collected from Yusuf et al. (2019), Lee et al. (2019) and Wiesenberger (2001). The resource parameters for the remaining units are summarized in Table 2 where all parameters are given in t/t RP except for electricity and cooling water parameters which are given in kWh/t RP and m³/t RP respectively.

Table 2: Selected processes and their process parameters

SWRO		Sodium Chloride		Formalin	
Brine	1.41	Brine	-59.31	Air	-1.17
Seawater	-2.41	NaCl	1.00	Emissions*	0.84
H ₂ O	1.00	Wastes*	54.16	CH ₂ O	1.00
Electricity	-2.53	H ₂ O	4.15	CH ₃ OH	-0.42
		Electricity	-102.00	H ₂ O	-0.25
				CW	-42.00
				Electricity	-49.00
Soda Ash		Nitric Acid		Sodium Bicarbonate	
CO ₂	-0.66	Air	-4.48	CO ₂	-1.02
Na ₂ CO ₃	1.00	NH ₃	-0.28	Emissions*	4.49
NaOH	-0.35	Emissions*	3.96	N ₂	-3.97
Wastes*	0.02	HNO ₃	1.00	Na ₂ CO ₃	-1.59
Wastewater	0.83	H ₂ O	-0.20	NaHCO ₃	1.00
H ₂ O	-0.83	CW	-105.00	Wastes*	0.91
CW	-37.52	Electricity	-8.50	Wastewater	3.90
Electricity	-23.24	LP Steam	-0.05	H ₂ O	-3.72
		HP Steam	0.80	Electricity	-78.91
				LP Steam	-0.30
				Process Water	-0.25

*Emissions** refers to the emissions from a production unit, *Emissions*** refers to fugitive emissions from the capture units, and *Wastes** refers to process wastes from a unit.

Constraints were set to only allow seawater, air, and imported CO₂ into the cluster, while the RP material outputs can be sold. Energy requirements for the cluster are met by utilizing solar energy through a photo-voltaic system that is built outside the cluster and low-pressure (LP) steam obtained from geothermal sources. 50,000 tpy of CO₂ is imported into the system while no constraint is placed on the imported energy. The limit for the seawater and air feeds are restricted by the maximum capacity limits of the SWRO unit and the ASU. The utility prices for electricity, cooling water and LP steam were obtained from Ahmed et al. (2020) along with the prices of NH₃, O₂, and H₂O. The prices for HNO₃ and N₂ was obtained from Intratec (Intratec, 2019), while those of Na₂CO₃ and NaCl were obtained from Lee et al. (2019) and Nayar et al. (2019) respectively.

The EIPs determined for these cases are shown below where both networks activate the same candidate plants with the exception of the soda ash plant which is only activated when there is a CO₂ import. The activated units are represented by the white blocks while

those in grey are inactive. The solid lines show the material flows into and out of the cluster while the dashed lines show the exchange of resources within the cluster. Scenario 1 was able to achieve 100% utilization of all the CO₂ imported into the system at no cost making the cluster carbon-negative while achieving a profit of \$178 M/y (with selling nitrogen) and \$113 M/y (without selling nitrogen). Scenario 2 achieves a profit of \$171 M/y (with selling nitrogen) and \$109 M/y (without selling nitrogen). The capital costs of scenario 1 and 2 are \$805 M and \$716 M respectively.

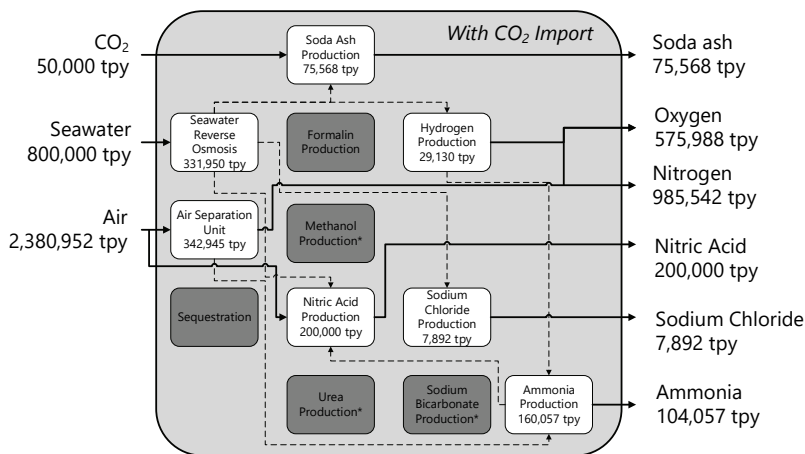


Figure 2: Scenario 1 - Industrial cluster design with CO₂ import

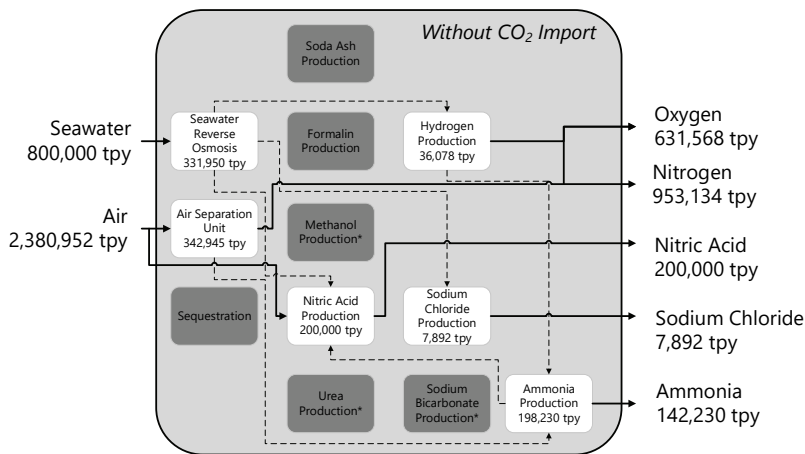


Figure 3: Scenario 2 - Industrial cluster design without CO₂ import

It was observed that a profitable carbon-neutral EIP is possible even with CO₂ as input. In fact, the importing pure CO₂ increased the profit. It should also be noted that this profit is achieved when the production capacities of the hydrogen and ammonia units were decreased. This was due to the ability to convert CO₂ with water and sodium chloride to produce soda ash. This observation encourages the expansion of the case to include more CO₂ converting processes for future work.

4. Conclusion

The profitability and environmental performance of an industrial clusters whose sole material inputs were seawater and air were investigated through the use of a MILP tool. Two scenarios were modelled, one with and the other without an import of CO₂, that activated the same units with the exception of the soda ash plant when CO₂ was allowed into the cluster. The soda ash plant serves as a sink for the CO₂, thereby generating considerable profits of \$178 M/y and \$113 M/y, with and without selling nitrogen, while achieving a carbon-negative system. Multiple scenarios can be further studied to analyze the impact that environmental restrictions have on profitability.

References

- R. Ahmed, S. Shehab, D.M. Al-Mohannadi, P. Linke, 2020, Synthesis of integrated processing clusters, *Chemical Engineering Science*, DOI:10.1016/j.ces.2020.115922
- IPCC, 2005, IPCC Special Report on Carbon Dioxide Capture and Storage, Cambridge, United Kingdom and New York, NY, USA, 442 pp
- R. Côté, J. Hall, 1995, Industrial parks as ecosystems, *Journal of Cleaner Production*, DOI:10.1016/0959-6526(95)00041-C
- G. Valenzuela-Venegas, J.C. Salgado, F.A. Díaz-Alvarado, 2016, Sustainability indicators for the assessment of eco-industrial parks: classification and criteria for selection, *Journal of Cleaner Production*, DOI:10.1016/j.jclepro.2016.05.113
- E.M. Lovelady, M.M. El-Halwagi, 2009, Design and integration of eco-industrial parks for managing water resources, *Environmental Progress & Sustainable Energy*, DOI:10.1002/ep.10326
- D.M. Al-Mohannadi, P. Linke, 2016, On the systematic carbon integration of industrial parks for climate footprint reduction, *Journal of Cleaner Production*, 112, 4053-4064, DOI:10.1016/j.jclepro.2015.05.094
- S.Y. Alnouri, P. Linke, M.M. El-Halwagi, 2018, Accounting for central and distributed zero liquid discharge options in interplant water network design, *Journal of Cleaner Production*, DOI:10.1016/j.jclepro.2017.09.236
- M.M. El-Halwagi, F. Gabriel, D. Harell, 2003, Process design and control rigorous graphical targeting for resource conservation via material recycle/reuse networks, DOI:10.1021/ie030318a
- A. Yusuf, A. Giwa, E.O. Mohammed, O. Mohammed, A. Al Hajaj, M.R. Abu-Zahra, 2019, CO₂ utilization from power plant: A comparative techno-economic assessment of soda ash production and scrubbing by monoethanolamine, *Journal of Cleaner Production*, 237, DOI: 10.1016/j.jclepro.2019.117760
- G.J. Millar, M. Collins, 2017, Industrial production of formaldehyde using polycrystalline silver catalyst, 56, 33, DOI: 10.1021/acs.iecr.7b02388
- A. Bahmanpour, A. Hoadley, A. Tanksale, 2014, Critical Review and Exergy Analysis of Formaldehyde Production Processes, *Reviews in Chemical Engineering*, 30, 6, 583-604
- J.H. Lee, D.W. Lee, C. Kwak, K. Kang, J.H. Lee, 2019, Technoeconomic and environmental evaluation of sodium bicarbonate production using CO₂ from flue gas of a coal-fired power plant, *Industrial & Engineering Chemistry Research*, 58(34), 15533–15541. DOI: 10.1021/acs.iecr.9b02253
- H. Wiesenberger, 2001, State-of-the-art for the production of nitric acid with regard to the IPPC directive
- Intratec, 2019. Fertilizers & Food Prices
- K.G. Nayar, J. Fernandes, R.K. McGovern, B.S. Al-Anzi, J.H. Lienhard, 2019, Cost and energy needs of RO-ED-crystallizer systems for zero brine discharge seawater desalination, *Desalination*, 457, 115-132 pp, DOI: 10.1016/j.desal.2019.01.015
- M.S. Peters, K.D. Timmerhaus, R.E. West, 2002, Equipment costs, Plant design and economics for chemical engineers

Planning production of solar grade silicon to yield solar panels involving behaviour of population

Esbeydi Villicaña-García,^a César Ramírez-Márquez,^b Juan Gabriel Segovia-Hernández,^b José María Ponce-Ortega^{a,*}

^a*Chemical Engineering Department, Universidad Michoacana de San Nicolás de Hidalgo, Av. Francisco J. Múgica S/N, Ciudad Universitaria, Edificio VI, Morelia, Michoacán, 58060, México*

^b*Chemical Engineering Department, Universidad de Guanajuato, Campus Guanajuato, Guanajuato, 36005, México*

jponce@umich.mx

Abstract

The need to use clean sources for producing electricity has been a great topic of discussion in recent years. Many countries have begun to take advantage of solar energy through the production of solar panels for electric power generation. In addition, knowing people's preferences allows predicting their behavior to propose better planning at macroscopic level. This work presents a mathematical programming model to address the conduct of people. The objective is to know their inclinations using the matching law, which involves the actions of users through different scenarios considering economic incentives and punishments. It includes a strategic planning of the production and distribution of solar grade silicon that is used for the construction of solar panels to meet the demand of electricity in the residential sector of Mexico as a case study. Process intensification was used to enhance different ways to obtain solar grade silicon such as an Intensified Fluidized Bed Reactor and a hybrid, which is a combination of Siemens and conventional FBR processes. Also, the Siemens process was considered the most common process to produce silicon. Results show that the difference between the analyzed scenarios lies in the behavior of people while profit maintains constant.

Keywords: Matching Law, Solar Grade Silicon, Strategic Planning, Process Intensification, Economic Incentives.

1. Introduction

In recent decades, photovoltaic systems have evolved and have been adapted to various applications of daily life. That way, large-scale manufacturing of photovoltaic cells capable of providing energy has become economically viable. It is important to produce large amounts of polysilicon due to the high demand by the solar sector. The production of polysilicon from quartz consists of two main stages: obtaining metallurgical grade silicon and purifying it to transform it into solar grade silicon (Ranjan et al., 2011). Attention has been drawn to update processes to obtain this component in order to achieve not only cost reduction but also improving safety of the process. Due to the exceeding cost of photovoltaic panels, it is necessary to look for alternatives in order to reduce the production cost of the solar grade silicon (SiSG). The photovoltaic industry relies on high-

purity silicon produced in the Siemens process; however, it requires a high power consumption. Another process is the Fluidized Bed Reactor that saves between 80-90 percent of the energy consumed compared to the Siemens process (Jiao et al., 2011) but the operating conditions are more difficult to achieve. Different alternatives have been proposed, one of them is an intensified FBR's process by substituting the conventional reactors and separation zone with a reactive distillation column to improve the chemical conversion. An additional alternative process is based on both, the Siemens and the conventional FBR attempting to reduce the use of raw material (Ramírez-Márquez et al., 2018).

The behavior of the population under various circumstances has not been addressed in a proper way. Considering the human behavior is especially important due to the high dependence of the energy system on it, this because the demand for each type of energy depends on the preference of the end user (population). Matching law is a mathematical approach that describes the relationship between the relative rate of response and the relative rate of reward in the face of a concurrent stimulus; from this law and respecting what is established, it is possible to give different interpretations and use it under diverse scenarios (Herrnstein, 1961). This way, in this paper is proposed a mathematical model that involves behavior through matching law and planning a supply chain of an important value-added compound.

The main difference with other works is that the planning of photovoltaic systems has not considered users in the system, they only included ways to evaluate costs or improve the processes but not to consider how the decision of a person could affect the whole system. Also, it has not been studied the behavior of the population through the matching law together with the strategic planning of solar grade silicon (Si_{SG}). The use of matching law is related to the economic flow of incentives and penalties that will influence the consumer's decision to select the type of solar panel to use depending on the addressed case. Matching law provides a good tool to model consumer response to incentives. This allows optimizing the supply chain and planning of poly-silicon used to yield solar panels involving producers, government and end-users.

2. Problem statement

The addressed problem seeks to meet the demands of electric power of the residential sector. One of the main limitations is the production and distribution of the raw material to produce solar grade silicon, which can be implemented through the Siemens processes, intensified FBR Union Carbide and a combination of both (Hybrid).

Matching law is adapted in a mathematical optimization model and the problem consists in determining the preference of the end user at a maximum profit. In the presented case study, the Government is the decision maker because it is responsible to pay incentives or implement economic punishments. The considered perspective is to study the effect of the government over producers and consumers. There are different studied scenarios, where it is intended to analyze those perspectives. The consumer can choose between two options: (a) use solar panels interconnected to the network, (b) use isolated solar panels. The final decision will be influenced by the incentives and punishments that the government will provide. Following those ideas, the behavior can be predicted.

3. Optimization model

The proposed mathematical model is based on the superstructure shown in Figure 1, which represents the problem to be addressed and involves all the possibilities to solve it. It considers the raw material available for using in different cities of the country. The raw material is transported to the different processes to produce solar grade silicon (Siemens, intensified FBR Union Carbide and Hybrid). This raw material must meet the compositions requirements for its processing. Once the Si_{ISG} is obtained, it is used to produce solar panels isolated or solar panels interconnected to the existing network and will subsequently be used to meet the electricity demand of the country's residential sector.

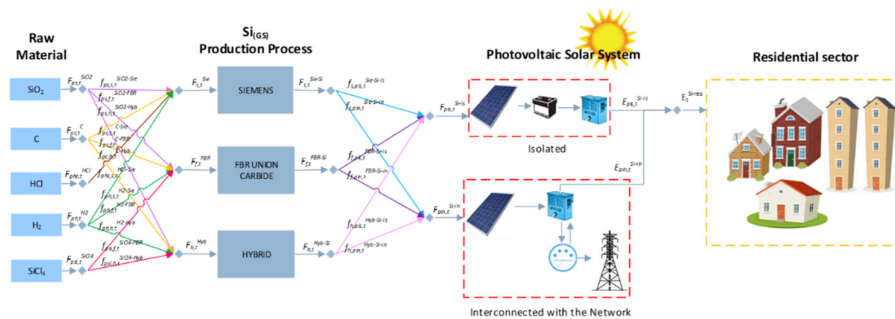


Figure 1. Proposed superstructure.

3.1. Problem behavior

Eq. (1) is used to calculate economic incentives (rewards, $R^1_{pis,t}$) and economic punishments ($P^1_{pis,t}$). The producer is asked to propose a certain amount that he considers he can meet, knowing that by meeting that amount he will obtain an economic incentive, but if he does not meet it, he will be financially punished. In this way, both the stakeholder and the producer are involved. It is necessary to involve the receiver in establishing the targets to assure that they will try their very best in order to achieve the production target and earn the incentives but also they need to be aware of their limits and do not propose more than they can do, that is why punishments are also included. $P^{is-p}_{pis,t}$ represents the number of panels proposed by the CEO of the company or the final user, according to the case. Those are the panels that CEO/user mean to achieve, it is a hypothetical situation because that number represents the number of panels that is easy to use for them but also the number must be reliable and achievable. They commit to reach the number and based on that the incentives/punishments are calculated. $P^{is-oc}_{pis,t}$ is the target quantity of solar panels established by the stakeholder. IE^{R1} is the base economic incentive that can be obtained. u^{is} , x^{is} and v^{is} are required parameters.

$Y^{R1}_{A_{pis,t}}$ will be activated if the number of produced panels is equal to the number of proposed panels. The reward is the base economic incentive and there is no punishment. $Y^{R1}_{B_{pis,t}}$ will be activated if the number of produced panels is greater than the proposed panels. The reward will be greater than the base economic incentive and there is no punishment. $Y^{R1}_{C_{pis,t}}$ will be activated if the number of produced panels is lower than the

proposed panels. There is no reward, and the punishment will be greater than the base economic incentive.

$$\left[\begin{array}{l} Y_{A_{pis,t}}^{R^1} \\ P_{pis,t}^{is} = P_{pis,t}^{is-p} \\ R_{pis,t}^1 = IE^{R^1} \cdot u^{is} \cdot \frac{P_{pis,t}^{is-p}}{P_{pis,t}^{is-oc}} \\ P_{pis,t}^1 = 0 \end{array} \right] \vee \left[\begin{array}{l} Y_{B_{pis,t}}^{R^1} \\ P_{pis,t}^{is} > P_{pis,t}^{is-p} \\ R_{pis,t}^1 = IE^{R^1} \cdot x^{is} \cdot \frac{(P_{pis,t}^{is} + P_{pis,t}^{is-p})}{2 \cdot P_{pis,t}^{is-oc}} \\ P_{pis,t}^1 = 0 \end{array} \right] \vee \left[\begin{array}{l} Y_{C_{pis,t}}^{R^1} \\ P_{pis,t}^{is} < P_{pis,t}^{is-p} \\ R_{pis,t}^1 = 0 \\ P_{pis,t}^1 = IE^{R^1} \cdot v^{is} \cdot \frac{(3 \cdot P_{pis,t}^{is} - P_{pis,t}^{is-p})}{2 \cdot P_{pis,t}^{is-oc}} \end{array} \right], \forall pis, \forall t \quad (1)$$

The equation of matching law developed by Borrero and Vollmer (2002) was adapted to the current problem statement (Eq.(2) and Eq. (3)).

$$\frac{B_{pis,t}^1}{B_{pis,t}^1 + B_{pin,t}^2} = \frac{R_{pis,t}^1 - P_{pis,t}^1}{(R_{pis,t}^1 - P_{pis,t}^1) + (R_{pin,t}^2 - P_{pin,t}^2)}, \forall pis, \forall pin, \forall t \quad (2)$$

$$\frac{B_{pin,t}^2}{B_{pis,t}^1 + B_{pin,t}^2} = \frac{R_{pin,t}^2 - P_{pin,t}^2}{(R_{pis,t}^1 - P_{pis,t}^1) + (R_{pin,t}^2 - P_{pin,t}^2)}, \forall pis, \forall pin, \forall t \quad (3)$$

The response rates B1 and B2 are reflected with the flow of money in the case of isolated and interconnected to the network solar panels:

$$B_{pis,t}^1 = IE^{R^1} \cdot P_{pis,t}^{is}, \forall pis, \forall t \quad (4)$$

$$B_{pin,t}^2 = IE^{R^2} \cdot P_{pin,t}^{in}, \forall pin, \forall t \quad (5)$$

3.2. Objective function

The considered objective function corresponds to maximize the profit, which is calculated through the sale of Si_{SG} minus all costs involved in the production and transport of the raw material and the different processes to obtain silicon, as well as rewards and punishments related to using solar panels:

$$Profit = \left[\begin{array}{l} Sale^{Si} - \\ Cost^{p-SiO_2} - Cost^{p-C} - Cost^{p-HCl} - Cost^{p-H_2} - Cost^{p-SiCl_4} - \\ Cost^{p-Sie} - Cost^{p-FBR} - Cost^{p-Hyb} - \\ Cost^{t-raw} - Cost^{t-Si} + \\ \sum_{pis,t} (R_{pis,t}^1 - P_{pis,t}^1) + \sum_{pin,t} (R_{pin,t}^2 - P_{pin,t}^2) \end{array} \right] \quad (6)$$

4. Case study

The case study is applied in Mexico. The raw material must be transported from its production point of the different plants to the silicon production plants for either the Siemens, the intensified FBR or the Hybrid process. It is necessary to meet the required

compositions for each process. For the installation of solar panels, 20 cities are selected which represent the cities where entrepreneurship is being promoted in the country. The contemplated time horizon is 30 years divided in periods of a year. According to the current consumption of electric energy, the necessary solar panels are determined if the total demand is covered, and his way the basis for subsequent calculations is obtained. Solar panels of 270 W are used. Different scenarios were proposed.

Scenario A: Government pays economic incentives and implements economic punishments to the CEO. Scenario B: CEO pays economic incentives to the final user. Scenario C: Government pays economic incentives and implements economic punishments to the CEO at the same time that CEO pays economic incentives to the final user. Under this background, it was determined the preference of population between isolated or interconnected solar panels.

5. Results

The profit obtained in scenario A is 7.9188×10^8 \$/y. Silicon production is mainly satisfied from the hybrid process (70.1%). Siemens process has the lowest cost but also the lowest silicon production rate. Intensified FBR has the highest cost but a large production. On the other hand, the hybrid process has the largest production rate with the middle production cost compared with the other two. It was shown that the Hybrid process is the most cost-effective one.

The values B1 and B2 represent the flow of money that is directed to each option considered, it means that where there is greater flow of money is the largest number of households that select that type of panel. In scenario A, results show that population is more inclined to use panels interconnected to the grid to produce electricity and to meet the demand for residential sector (Figure 2). Only Tlaxcala and Monterrey received economic punishments. Results in scenario B show a greater profit (7.9223×10^8 \$/y) than scenario A (7.9188×10^8 \$/y), because scenario B does not consider economic punishment although the incentives granted are lower than the incentives given by government in scenario A. Profit in scenario C is 7.9214×10^8 \$/y and in 40% of the cities, more than 50% of the users prefer to install isolated panels. In scenario A, 25% of the cities prefer isolated panels. In contrast, in scenario B, all cities prefer to install interconnected panels.

It should be noted that in the addressed case study there were analyzed various scenarios to see how the whole system will react under different targets proposed by different entities and the presented scenarios are only cases to show the applicability of the proposed approach. This work is the first attempt to understand the relationship with matching law and planning under the fixed parameters.

6. Conclusions

This work has presented a mathematical optimization model capable of involving the conduct of people with the planning for the construction of solar panels isolated and interconnected with the network. The proposed approach has been tested in a national case study in Mexico, taking into account the greater profit and preference towards the solar panels of the involved users.

By using the matching law to study the human behaviour, it is possible to predict preferences of population. Also, it can be created hypothetical scenarios to understand the

main factors that change the final decision of the user and to help the stakeholder to decide according to their objectives and interests.

In the presented case study, the profit remains similar between scenarios A, B and C (7.9188×10^8 , 7.9223×10^8 and 7.9214×10^8 \$/y). However, the difference lies in the behaviour of the users, the preference for which they lean. Using models that predict behavior makes possible to control or manipulate the response expected from consumers under various established situations. The proposed model is general, and it can be applied to different case studies by making the appropriate changes in the involved parameters.

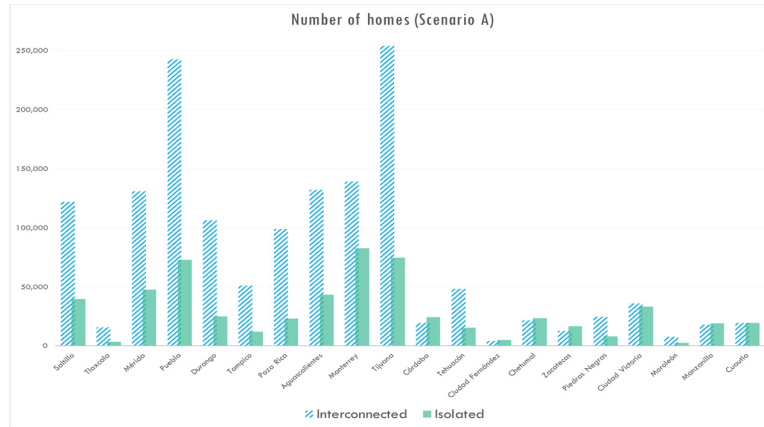


Figure 2. Selection of type of solar panel for each dwelling in each municipality considered.

References

- Borrero JC, Vollmer TR (2002) An application of the matching law to severe problem behavior. *Journal of Applied Behavior Analysis*, 35(1), 13-27. DOI: 10.1901/jaba.2002.35-13.
- Herrnstein RJ (1961) Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 4(3), 267-272. DOI: 10.1901/jeab.1961.4-267.
- Jiao Y, Salce A, Ben W, Jiang F, Ji X, Morey E, Lynch D (2011) Siemens and Siemens-like Processes for producing Photovoltaics: Energy Payback Time and Lifetime Carbon Emissions. *JOM*, 63, 28-31. DOI: 10.1007/s11837-011-0007-4.
- Ramírez-Márquez C, Otero MV, Vázquez-Castillo JA, Martín M, Segovia-Hernández JG (2018) Process design and intensification for the production of solar grade silicon. *Journal of Cleaner Production*, 170, 1579-1593. DOI: 10.1016/j.jclepro.2017.09.126.
- Ranjan S, Balaji S, Panella RA, Ydstie BE (2011) Silicon solar cell production. *Computers & Chemical Engineering*, 35(8), 1439-1453. DOI: 10.1016/j.compchemeng.2011.04.017.

Bi-level Mixed-Integer Data-Driven Optimization of Integrated Planning and Scheduling Problems

Burcu Beykal^a, Styliani Avraamidou^a, Efstratios N. Pistikopoulos^{a,b,*}

^a*Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843, USA*

^b*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, USA*
stratos@tamu.edu

Abstract

Supply chain management is an interconnected problem that requires the coordination of various decisions and elements across long-term (i.e., supply chain structure), medium-term (i.e., production planning), and short-term (i.e., production scheduling) operations. Traditionally, decision-making strategies for such problems follow a sequential approach where longer-term decisions are made first and implemented at lower levels, accordingly. However, there are shared variables across different decision layers of the supply chain that are dictating the feasibility and optimality of the overall supply chain performance. Multi-level programming offers a holistic approach that explicitly accounts for this inherent hierarchy and interconnectivity between supply chain elements, however, requires more rigorous solution strategies as they are strongly NP-hard. In this work, we use the DOMINO framework, a data-driven optimization algorithm initially developed to solve single-leader single-follower bi-level mixed-integer optimization problems, and further develop it to address integrated planning and scheduling formulations with multiple follower lower-level problems, which has not received extensive attention in the open literature. By sampling for the production targets over a pre-specified planning horizon, DOMINO deterministically solves the scheduling problem at each planning period per sample, while accounting for the total cost of planning, inventories, and demand satisfaction. This input-output data is then passed onto a data-driven optimizer to recover a guaranteed feasible, near-optimal solution to the integrated planning and scheduling problem. We show the applicability of the proposed approach for the solution of a two-product planning and scheduling case study.

Keywords: Integrated planning and scheduling, bi-level programming, data-driven optimization, mixed-integer optimization.

1. Introduction

Planning and scheduling belong to different levels of supply chain management, yet their coordination is essential for feasible decision making. The former determines the production targets by observing the market considerations (i.e., product demands), whereas the latter decides on the sequencing of tasks and their respective assignment to specific units such that the production targets set by the planning level are met.

Typically, these planning and scheduling problems are solved sequentially: initially, the planning level decides on the production targets of finished goods by observing the

demand profiles for the products; later, the production targets are set for the scheduling problem, and the respective optimal schedules are calculated. However, such a sequential approach, where the planning decisions are solely based on the product demand profiles are optimistic estimates and may lead to infeasible schedules (Grossmann, 2005). In other words, the process of converting raw materials to finished goods cannot be realized within the given capacity of the scheduling level. Henceforth, the production target set by the planning level cannot be met, and consequently, the demand is not satisfied. This interconnected network of decision making requires an integrated approach where the interdependencies among planning and scheduling levels need to be addressed simultaneously to achieve globally optimal solutions (Maravelias and Sung, 2009).

Bi-level programming enables the integration of planning and scheduling levels and creates holistic models that explicitly account for the inherent hierarchy between different levels of the supply chain. Unfortunately, many algorithmic difficulties arise when solving bi-level programming problems including, NP-hardness, nonconvexity, and discontinuity, even in the most simplistic formulations (Sinha et al., 2018). Especially the mixed-integer formulations in the scheduling level prohibit the use of the KKT transformation at the lower level, creating a necessity to tackle these problems with an algorithmic approach that can guarantee feasibility to bi-level formulations. An approach based on multi-parametric programming has been proposed to address this (Avraamidou and Pistikopoulos, 2018), but it cannot handle a relatively high number of variables. Recently, the DOMINO algorithm is introduced as a data-driven methodology for solving constrained bi-level mixed-integer nonlinear programming (B-MINLP) problems with a higher number of variables and a feasibility guarantee (Beykal et al., 2020). In this work, our goal is to (1) further advance the DOMINO framework for solving integrated planning and scheduling problems with multiple followers; and, (2) utilize this framework to provide guaranteed feasible solutions to integrated supply chain management problems.

2. The DOMINO Algorithm and its Extension to Multi-Follower Systems

DOMINO is a data-driven optimization algorithm that is tailored to solve single-leader single-follower constrained B-MINLP problems by approximating them as single-level grey-box optimization problems through series of sampling, optimization, and surrogate modeling steps (Beykal et al., 2020). It postulates candidate sampling points for the upper-level decision variables and solves the lower-level problem deterministically to global optimality at each sampling point. The collected input-output data is then passed onto a data-driven optimization step where the solver retrieves the optimal solution of the bi-level program. DOMINO ensures the feasibility of a candidate solution by evaluating the constraint violations of grey-box constraints and the optimality of the lower-level problem. If these two conditions are satisfied, the candidate solution is deemed a guaranteed feasible solution of the bi-level program. Previously, DOMINO is shown to locate near-optimal solutions to varying types of single-follower bi-level programming problems in series of benchmark problems (Beykal et al., 2020) and a case study (Avraamidou et al., 2018).

In the case of integrated planning and scheduling problems, the lower-level problem is composed of multiple scheduling problems (i.e., multiple followers) that need to be solved sequentially over the entire planning horizon. For that reason, the DOMINO

algorithm cannot be directly implemented, and a subroutine needs to be devised to collect the input-output data. The production targets for the planning horizon are the upper-level decision variables, which are used to create the input data for the sampling stage. To account for the multi-follower nature of the lower-level problem, these input production targets will be fixed at the scheduling level and the schedules for each planning period will be solved sequentially while accounting for the inventories and production cost per period. Finally, the total cost of planning will be calculated from the total inventory and production cost across the planning horizon and this value will serve as the output information that is required to be minimized in the data-driven optimization stage.

3. Case Study for Bi-level Production Planning and Scheduling

The computational case study for the bi-level production planning and scheduling problem is adapted from Li and Ierapetritou (2009) with an implementation of the continuous-time formulation of Example 2 (Ierapetritou and Floudas, 1998) at the lower-level scheduling problem. An overview of this bi-level program is provided in Eq. (1).

Total Cost of Planning

s. t. Inventory and Balance Equations (1)

Production Cost for Each Planning Period

s. t. Scheduling Constraints

3.1. The Upper-Level Problem: Planning Model

The planning level objective minimizes the total inventory and production cost over the entire planning horizon. The decision variables for the bi-level program are the production targets (Prd_s^t) of the desired products (i.e., states $s \in S_p$) per planning period t . The inventory accounting is performed at every planning period using Eq. (2), where Inv_s^t is the inventory level of the product state s at the end of planning period t , Prd_s^t is the production level of the product state s in the planning period t , and Dmd_s^t is the demand level for the product state s in the planning period t .

$$Inv_s^t = Inv_s^{t-1} + Prd_s^t - Dmd_s^t \quad \forall s \in S_p, t \quad (2)$$

If the inventory becomes negative at any period, the minimum inventory of each product is calculated, and all respective inventories are penalized by adding this minimum level of inventory. Also, the total amount of products given by the production target, and the inventory level at the beginning of each planning period should be greater than or equal to the demand to satisfy the need. The planning horizon is set to be 7 days, and the demand for the two products is assumed to be known for the entire week. Besides, the planning problem is assumed to be cyclic where the last day production target should meet the last day demand and produce the 1st day inventory of the next planning cycle. The upper-level problem contains 14 production target constraints, 14 demand constraints, and 28 positive variables (14 inventory and 14 decision variables). The upper bound on the decision variables is set to be 80-unit materials per product per planning period.

3.2. The Lower-Level Problem: Scheduling Model

In the selected case study, the scheduling problem (Figure 1) considers the production of two products through three reaction steps, a heating step, and a separation step. Three feed and four intermediate components are considered in the production schedule. The scheduling formulation considers material balances, allocation, capacity, storage, duration, time horizon, and sequence constraints (Ierapetritou and Floudas, 1998). To guarantee feasibility to the integrated multi-follower problem, a constraint on the final states of the products is added to ensure the production targets set by the upper-level problem are met at the scheduling level. The objective function of the scheduling problem minimizes the fixed cost of operating tasks in units, as well as the variable cost coming from handling raw materials, intermediates, and final products within the process. The lower-level problem contains 285 continuous variables, 160 binary variables, and 712 constraints. The detailed model equations and parameters can be provided upon request.

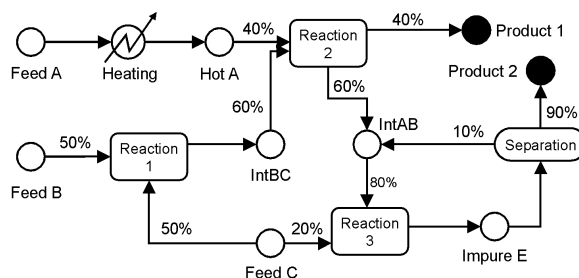


Figure 1 State-task network of the scheduling case study. Adapted from Ierapetritou and Floudas (1998).

3.3. Other Considerations for the Problem Formulation

The integrated bi-level mixed-integer program is solved for two cases: (1) assuming linear inventory and production cost, leading to an LP-MILP bi-level problem; and, (2) assuming cubic inventory cost at the planning level and quadratic production cost at the scheduling level, creating an NLP-MIQP problem. Compared to other bi-level optimization algorithms, DOMINO allows us to solve highly nonlinear systems using its data-driven optimization strategy and retrieve guaranteed feasible near-optimal solutions (Beykal et al., 2020). Specifically, this capability of DOMINO is advantageous because the nonlinear objective functions allow for a more realistic estimate of the variable cost in the integrated planning and scheduling formulation, as linear cost estimates are approximations of the nonlinear behavior. Also, the inclusion of the nonlinear terms in the bi-level formulation yields a very challenging optimization problem which the exact methods like multi-parametric programming cannot address. Data-driven evolutionary algorithms can handle nonlinearities but they fail to provide guaranteed feasible solutions.

To assess the consistency and accuracy of the DOMINO algorithm in finding the best solution for the integrated planning and scheduling problem, the algorithm is randomly executed 10 times with the NOMAD algorithm (Le Digabel, 2011) chosen as the data-driven optimizer. The results retrieved from DOMINO are presented in the next section.

4. Results

4.1. LP-MILP Solution

The best solution of the 10 random runs for the LP-MILP integrated planning and scheduling formulation is summarized in Figure 2. The results show that the demand for both products is satisfied over the 7-day planning period where the corresponding production targets are met with globally optimal schedules. We also observe that at the start of the planning period (Day 1), the system produces more than the minimum required level to sustain the inventory levels for Product 1 whereas a lower production target is set for Product 2 and the remaining demand is supplemented from the starting inventory. On Day 2, both production targets are relatively higher than the demand. This higher level of production enables Product 1 to sustain its starting inventory levels whereas Product 2 makes up for the lost inventory on Day 1. Later, when peak demand is expected on Day 4 for both products, the production target is supplemented with this accumulated inventory to satisfy the demand. On Day 6 as the demand is lower, the production can meet this demand without requiring any extra supply from the inventory. On the final day of the planning period, the production levels are increased to satisfy the 7th day demand as well as to produce the starting inventory of the 1st day of the next planning cycle.

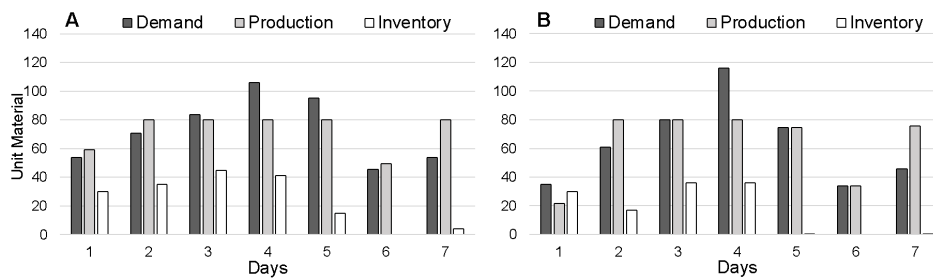


Figure 2 Demand, production, and starting inventory profiles for (A) Product 1; and (B) Product 2.

The results of the other 9 random cases also showed very consistent production and inventory levels in Days 3-7 for Product 1 and in Days 4-5, 7 for Product 2. Some variability is observed in the early days of the week where in one instance a significantly higher production is observed for Day 1 for both products which led to no production in Day 2, essentially pushing the system to spend the accumulated inventory more quickly than the best solution. In the other two instances, Day 1 production is observed to be zero where the system relied on having a higher start inventory by producing more at the end of the planning period when the demand is lower. Nonetheless, the final cost objective value for all runs was within ± 0.0386 standard error and all final solutions were guaranteed feasible with globally optimal schedules.

4.2. NLP-MIQP Solution

The best solution of the 10 random runs for the NLP-MIQP formulation is summarized in Figure 3. The nonlinear integrated planning and scheduling results show very similar production and starting inventory profiles to the LP-MILP case study where the demand is satisfied for the entire planning horizon of 7 days with a globally optimal lower-level solution. Only on Day 7, a slightly higher starting inventory level is observed for Product 1 because the production target of the prior day is higher in the NLP-MIQP

solution. Furthermore, the DOMINO solution is very consistent across all runs for the NLP-MIQP formulation. For Product 1, we observe that the same starting inventory and production levels are determined for Days 1-5 whereas a slight deviation is observed for Days 6 and 7. For Product 2, some variability is observed at the start and end of the planning period, but the production and inventory levels are consistent for Days 3 and 4. The consistency of the solutions is also reflected in the final objective values where for all runs the objective value was within ± 0.1465 standard error.

5. Conclusions and Future Work

In this work, we present a data-driven approach to solve bi-level multi-follower mixed-integer formulations of integrated planning and scheduling problems. By extending the DOMINO algorithm to solve multi-follower bi-level optimization problems and utilizing the data-driven and deterministic optimization capabilities of this framework, we solve the integrated problems to guaranteed feasibility. For all the tested cases, DOMINO identified solutions that meet the product demand and have globally optimal schedules at the lower level, which ensures meeting the production targets, DOMINO also found consistent feasible solutions for both the linear and nonlinear formulations. In the future, the results of the linear formulation will be compared to the deterministic algorithm developed by Avraamidou and Pistikopoulos, 2019. This research was funded by the U.S. National Institutes of Health (NIH) grant P42 ES027704.

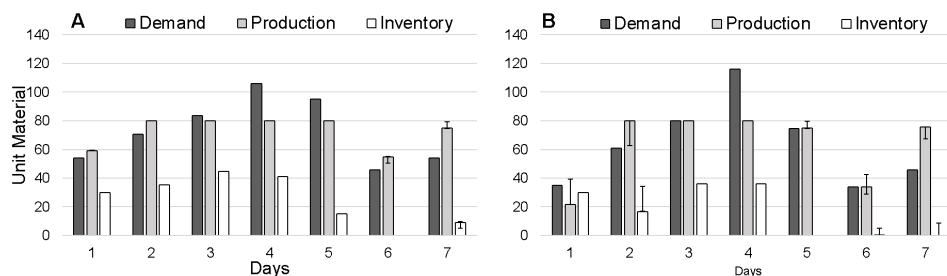


Figure 3 Demand, production, and starting inventory profiles for (A) Product 1; and (B) Product 2 in the NLP-MIQP solution. The error bars indicate the minimum and maximum deviation observed across all runs with respect to the best-found solution.

References

- S. Avraamidou, B. Beykal, IPE. Pistikopoulos, EN. Pistikopoulos, 2018, A Hierarchical Food-Energy-Water Nexus (FEW-N) Decision-Making Approach for Land Use Optimization, *Computer Aided Chemical Engineering* 44, 1885-1890.
- S. Avraamidou, E. N. Pistikopoulos, 2018, A novel algorithm for the global solution of mixed-integer bi-level multi-follower problems and its application to Planning & Scheduling integration, 2018 European Control Conference (ECC), Limassol, 2018, pp. 1056-1061.
- S. Avraamidou, E. N. Pistikopoulos, 2019, A Multi-Parametric optimization approach for bilevel mixed-integer linear and quadratic programming problems, *Computers & Chemical Engineering*, 125, 98-113.
- B. Beykal, S. Avraamidou, IPE. Pistikopoulos, M. Onel, EN. Pistikopoulos, 2020, DOMINO: Data-driven Optimization of bi-level Mixed-Integer Nonlinear Problems, *Journal of Global Optimization*, 78, 1-36.

- IE. Grossmann, 2005, Enterprise-wide Optimization: A New Frontier in Process Systems Engineering, *AIChE Journal*, 51, 7, 1846-1857.
- MG. Ierapetritou, CA. Floudas, 1998, Effective Continuous-Time Formulation for Short-Term Scheduling. 1. Multipurpose Batch Processes, *Industrial & Engineering Chemistry Research*, 37, 11, 4341-4359.
- S. Le Digabel, 2011, Algorithm 909: NOMAD: Nonlinear Optimization with the MADS Algorithm. *ACM Transactions on Mathematical Software*, 37(4), 44.
- Z. Li, MG. Ierapetritou, 2009, Integrated Production Planning and Scheduling Using a Decomposition Framework, *Chemical Engineering Science*, 64, 3585-3597.
- CT. Maravelias, C. Sung, 2009, Integration of Production Planning and Scheduling: Overview, Challenges and Opportunities, *Computers & Chemical Engineering*, 33, 1919-1930.
- A. Sinha, P. Malo, K. Deb, 2018, A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications, *IEEE Transactions on Evolutionary Computation*, 22, 2, 276-295.

A Prospective Approach in the Design of Lignin-Based Products Supply Chains

Brunelle Marche,^{a*} Javier A. Arrieta-Escobar,^a Vincent Boly,^a Juan C. Solarte-Toro,^b Carlos A. Cardona Alzate^b

^a*ERPI (Équipe de Recherche des Processus Innovatifs), Université de Lorraine, 8 rue Bastien Lepage, 54000 Nancy Cedex, France*

^b*Instituto de Biotecnología y Agroindustria, Departamento de Ingeniería Química, Universidad Nacional de Colombia, Manizales, Caldas, Zip Code: 170003, Colombia.*

brunelle.marche@univ-lorraine.fr

1. Abstract

On a global scale, the main usages of wood are energy production and materials. Wood chemistry offers a broad range of potential outlets, and the whole supply chain around it. The latter is made up of a set of actors (suppliers, producers, etc.), processes and financial, informational and material flows. Designing the supply chain is therefore a complex activity. It is therefore important to determine the different variables that will influence the design of the supply chain. From this perspective, this research focuses on the valorization of the lignin fraction, by addressing the various aspects of the development of new products and by determining the variables that will impact the design and implementation of the future supply chain.

Keywords: supply chain, prospective analysis, lignin.

2. Context

Nowadays, oil consumption accounts for 90% of the energy consumed and alternatives to replace oil are still poorly developed. In France, for example, renewable energies (hydraulic, wind, solar, geothermal or biomass energy) account for 10% of energy consumption (Arion et al. 2018). The increasing scarcity of oil could therefore require a reconsideration of the entire energy supply chains and stimulate the exploitation and development of innovative products from natural renewable resources, like lignocellulosic biomass. Nevertheless, the success of these new products and their supply chains will depend on technological progress and cost control (de Cherisey 2015). It is therefore necessary to anticipate and design the product, the process and the supply chain jointly (Hilletoft et al. 2010; Sharifi et al. 2006; Primus 2017). At any given moment, the supply chain can be characterized by its actors, processes, flows, and the value it generates described by its structure and operating dynamics (Marche et al. 2017). In general, the pathways considered to develop products derived from lignocellulosics include the following elements: biomass cultivation, storage, and treatment plants (biorefineries); intermediate product storage; final product production plants, and demand centers. Raw materials and intermediate products circulate between the elements (Panteli et al. 2017; 2018). However, this structure depends on the decisions of each actor, which

are interdependent, but also on collective decisions. Therefore, this research is a holistic and prospective approach in order to anticipate the design of future lignin supply chains.

3. Research approach

To better understand the influence of the variables on the future supply chains, as well as the effects of the decisions of the actors involved, a prospective analysis was carried out using the MICMAC (Cross Impact Matrix and Multiplication Applied to Ranking) method. This method has already been used in prospective supply chain design approaches (Attri et Grover 2017; Agarwal, Tyagi, et Garg 2021). It allows a systematic identification of the key variables determining the evolution of the supply chain and examining the combinations of hypotheses that can be excluded a priori (Elmsalmi and Hachicha 2014). Based on the influence/dependence plan, it is possible to identify 9 areas, as shown in Figure 1.

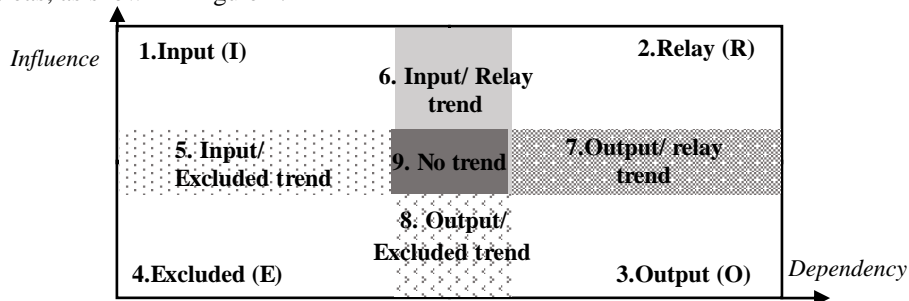


Figure 1: Zones of the influence/dependence plan

A confrontation with 8 experts made it possible to identify 29 variables, which have a higher impact on the future supply chain design. Then, these experts evaluated the impact of each variable on all other variables (0: no impact; 1: low impact; 2: medium impact; 3: high impact; 4: potential impact). According to these assessment, the MICMAC method classifies variables into four typologies (Godet and Durance 2011):

- Input variables: They are explanatory of the system under study, and condition the overall dynamics. They guide and explain the development of the system.
- Relay variables: They are inherently unstable. They play a central role in the dynamics of the system and are the main issues at stake. Any action on them will have repercussions on the other variables and a feedback effect on themselves.
- Output variables: Their behaviour can be explained by analysing (or acting on) the input and relay variables. They are the variables most dependent on others.
- Excluded variables: They have little impact on the system, either because their inertia does not alter the dynamics of the system, or because of their reduced connection with it and develop relatively autonomously.

Three types of analyses were carried out: (i) unweighted analysis and considering the pool of experts in the panel (ii) analysis with a weighting of the self-rated level of expertise (1: low expertise; 5: high expertise), and (iii) analysis with a weighting of the level of expertise (same scale) evaluated by the researchers.

4. Results

The different analyses made it possible to identify 7 variables on which there is a consensus among all the stakeholders. A strong trend may emerge for certain variables but for the majority, it seems impossible to determine the type of each variable. On the basis of the representation of Figure 1, it is possible to position the variables more accurately, as shown in Table 1.

Table 1. Variables classification

Variable	I	R	O	E	I/E	I/R	O/R	No trend
1. Minimal scale of processability			X					
2. Diversification of raw materials						X		
3. Added value of final product			X					
4. Sales volume of final products							X	
5. Transformation cost			X					
6. Source configuration	X							
7. Pre-treatment capacity			X					
8. "Purity" of the lignin				X				
9. Lignocellulosic source prices								X
10. Availability of lignin								X
11. Production capacity							X	
12. Sales price of lignin			X					
13. Cost of the final product			X					
14. Price of energy					X			
15. Geographical location	X							
16. Technological maturity	X							
17. Type of products								X
18. Type of wood	X							
19. Supply risk								X
20. Weight of suppliers								X
21. Type of processes		X						
22. Weight of biorefineries							X	
23. Potential competition								X
24. Quality of wood				X				
25. Pedological & climatic variables				X				
26. Shape and structure of the tree				X				
27. Type of solvents			X					
28. Environmental impact of products			X					
29. Presence of alternatives products	X							

Thus, Table 1 presents the classification of the 29 variables, as follows:

- Dominant trend for one of the typologies (more than 70% of identical answers from the experts): zone 1, 2, 3 and 4;
- Dominant trend for two typologies (more than 70% for the two majority typologies): zones 5, 6, 7 and 8;
- No dominant typology (total disagreement between the experts or the two majority typologies below 70%): zone 9.

4.1. Input variables

Our qualitative analysis characterizes the following variables as input variables: the source configuration; geographical location; Technological maturity; wood type; and the presence of alternative products. The source configuration, the geographic location, and the wood type are all related to lignin and its suppliers are globally dispersed, so their activities depend on the raw material available to them. The lignin source and their locations will have a direct impact on the choice of biorefinery according to their specialty or proximity. Technological maturity refers to the degree of completion of a technology with a view to making it available on the market. It is a measure of whether the technology is capable of delivering the expected services and quality of service under real operating conditions. This variable is particularly important in our study, as lignosulphonates, Kraft or soda lignins are obtained from mature and mastered technologies, but organosolv lignins are rather emerging. Thus, the technological maturity being classified as an input variable seems relevant, as the choice of technology will condition the supply chain (production capacity, supply capacity, type of suppliers, etc.). The presence of alternative products plays a key role in the supply chain if one considers supply chains competing with each other, the alternative product chains will condition the functioning of the future supply chain being created. Indeed, if the innovative product arrives in a highly competitive market, the supply chain must make its place and adopt a more aggressive behavior or develop a particular strategy, turned towards the customer's demand. In the case of the future supply chain, it should compete with similar petroleum-based product chains. These products are omnipresent on the market, the future supply chain must therefore adapt its strategy to gain market share.

4.2. Relay variables

The type of process was the only variable classified as relay. It plays an important role in the creation of the future supply chain for the product, as it characterizes the biorefineries that will be involved in the supply chain. The choice of a process for obtaining lignin will lead to other decisions in terms of the supply chain, related to the output variables.

4.3. Output variables

The output variables are the result of the decisions taken within the supply chain. In our study these are the minimum scale of processability, the added value of the final product, its transformation cost, pre-treatment capacity, selling price of lignin, cost of the final product, type of solvents and environmental impact of the product.

4.4. Excluded variables

Some variables were excluded as these are considered to be less influential or not very dependent. Variables directly related to the growing condition of the plants (wood quality, pedological and climatic variables and wood form/structure) were considered as having no impact on the creation of the future supply chain. The "purity" of lignin, on the other hand, could be considered as a strong trend. This means that it could have an impact on the supply chain in the future. Indeed, lignins exhibit different properties according to the process chosen and therefore. Probably in the future, the purity of lignin could become an important variable in the supply chain, because it would condition the processes to be favored to obtain lignin-based products with very precise and differentiating properties.

4.5. Input/Excluded variables

Energy prices and weight of suppliers are classified as input/excluded (zone 5) variables in the supply chain. The energy price variable can be very influential depending on the location one wishes to set up the supply chain in. Indeed, in order to minimize energy costs, it is in the interest of companies to deploy their supply chain in countries where energy is easily accessible and at affordable costs.

4.6. Input/Relay variables

The diversification of the raw material is found to be in the zone 6 of the Figure 1. This variable can cause the type of processes/solvents used to change, if the lignin composition goes beyond the established limits. In this case, suppliers of different types of wood can gain weight along the supply chain.

4.7. Output/Relay variables

In the zone 7 we found the volume of sales of final products, production capacity and weight of biorefineries. For example, the volume of sales of final products can be considered as a relay variable before the product launch on the market because it can be difficult to assess how customers will welcome the new product. Once the supply chain has been set up, the volume of sales becomes a known variable and is the result of decisions taken within the supply chain. Similarly, before the future supply chain is set up, it may be possible to assess the production capacity of the supply chain and the processing costs based on the individual capacities of each stakeholder and the costs generated. However, it may be difficult to estimate the evolution of demand once the product is launched, and thus to ensure that the supply chain has the capacity to keep up with changes in demand and to manage processing costs when changing the scale of production.

4.8. Dissensus

Several variables remain unclassified (zone 9) for several reasons. Firstly, when the four typologies are proposed by the experts, as it was the case for the price of lignocellulosic sources, the availability of different types of lignin, and the weight of suppliers. Secondly, when the experts' visions are antagonistic (input variable for some VS result variables for others). This is the case for the variables type of products and presence of alternative products. For the weight of biorefineries or the supply risk there was a slight trend towards relay variables, but it was not confirmed by all the experts. No trend emerged for the potential competition. Note that our qualitative analysis did not characterise any of the variables of the supply chain as outcome/excluded variables (zone 8).

The MICMAC analysis underlines that the supply chain created is influenced by relay variables, making it unstable. In order to anticipate and facilitate its creation, it is necessary for the stakeholders of the supply chain to be able to "control" the impacts of these variables. The relay variable clearly identified for this is the type of process, whose importance was illustrated in a parallel study using poplar as a basis for producing a range of products in a sustainable way thanks under the biorefinery concept (Solarte-Toro et al. 2020).

5. Conclusions

The study of the supply chain for a lignin-based product highlighted the complexity of setting up a supply chain. The mobilization of experts was an important step to describe the future supply chain scenarios that could support the industrialization of a lignin-based product. However, the analysis of the results underlines the difficulty for individuals to project themselves in a situation where the system, in this case the supply chain, does not yet exist. How can the sales volume of the future product be estimated? How can the weight of the players be assessed? Several questions arise. It is therefore necessary to develop a maximum number of scenarios to anticipate all possible situations. The choice of the lignin extraction process seems to be a relevant basis for developing different scenarios, as it has a direct impact on sales volumes, prices, production capacities and the location of biorefineries, which will give information on the type of wood or on energy prices.

References

- S. Agarwal, M. Tyagi & R. K. Garg. (2021). Commencement of Green Supply Chain Management Barriers: A Case of Rubber Industry. In *Advances in Manufacturing and Industrial Engineering* (pp. 685-699). Springer, Singapore.
- G. Arion, D. Cavaud, Y. Coltier, and F. Guggemos. 2018. « Chiffres clés des énergies renouvelables – Edition 2018 » France: Ministère de la transition écologique et solidaire.
- R. Attri, & S. Grover. (2017). Developing the weighted ISM-MICMAC framework for process design stage of production system life cycle. *International Journal of Process Management and Benchmarking*, 7(1), 94-119.
- H. de Cherisey,. 2015. « Etat de l'art sur la production de molécules chimiques issues du bois en France ». ADEME, 147.
- M. Elmsalmi and W. Hachicha. 2014. « Risk mitigation strategies according to the supply actors' objectives through MACTOR method ». In 2014 International Conference on Advanced Logistics and Transport (ICALT), 362-67. <https://doi.org/10.1109/ICAdLT.2014.6866339>.
- M. Godet and P. Durance. 2011. La prospective stratégique - 2e éd.: Pour les entreprises et les territoires. Dunod.
- P. Hilletoft, D. Ericsson, and K. Lumsden. 2010. « Coordinating new product development and supply chain management ». *International Journal of Value Chain Management* 4 (1-2): 170-92. <https://doi.org/10.1504/IJVC.2010.031808>.
- B. Marche, V. Boly, L. Morel, and J. R. Ortt. 2017. « Innovative product's supply chain: How to model it ». In 2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC), 177-88. <https://doi.org/10.1109/ICE.2017.8279887>.
- A. Panteli, S. Giarola, and N. Shah, 2017. « Biobased Supply Chain Optimisation Model under Uncertainties ». In *Computer Aided Chemical Engineering*, Eds. Antonio Espuña, Moisès Graells, et Luis Puigjaner, 40:961-66. 27 European Symposium on Computer Aided Process Engineering. Elsevier. <https://doi.org/10.1016/B978-0-444-63965-3.50162-8>.
- A. Panteli, S. Giarola, and N. Shah, 2018. « Supply Chain Mixed Integer Linear Program Model Integrating a Biorefining Technology Superstructure ». *Industrial & Engineering Chemistry Research* 57 (30): 9849-65. <https://doi.org/10.1021/acs.iecr.7b05228>.
- D. J. Primus, 2017. « A configuration and contingency analysis of the development chain ». *Technovation* 64: 1-15.
- H. Sharifi, H. S. Ismail, and I. Reid. 2006. « Achieving agility in supply chain through simultaneous “design of” and “design for” supply chain ». *Journal of Manufacturing Technology Management* 17 (8): 1078-98. <https://doi.org/10.1108/17410380610707393>.
- J. C. Solarte-Toro, J.A. Arrieta-Escobar, B. Marche, V.Boly, C. A. Cardona Alzate. « Effect of the lignin extraction process on the economics and environmental impact of a woody-based biorefinery » 2020. Accepted as an abstract for 31th European Symposium on Computer Aided Process Engineering, Istanbul, Turkey.

Flexibility Analysis in Supply Chain Management: Application to the Traveling Salesman Problem

Alessandro Di Pretoro^a, Stephane Negny^a, Ludovic Montastruc^{a*}

^a*Laboratoire de Génie Chimique, Université de Toulouse, CNRS/INP/UPS, Toulouse, France*

ludovic.montastruc@ensiacet.fr

Abstract

During the last decades, design under uncertainty has become one of the key aspects in the Process Systems Engineering with widely spread applications to technological and economic aspects. In particular, the supply chain design is seeing a constantly increasing interest and flexibility could be an added value of critical importance for the supply chain management domain. The purpose of this research work is then to define a thorough procedure able to describe flexibility in the operational research domain taking advantage of established indicators. The most commonly used deterministic and stochastic flexibility indices are then adapted to the traveling salesman problem, an NP-hard problem in combinatorial optimization, widely studied in theoretical computer science and operations research. It can be seen as the most general problem formulation in this field and as a preliminary test before extending the flexibility assessment to more specific applications. According to existing methodologies for chemical processes, the economic and flexibility analyses have been coupled over a certain flexibility range. This approach resulted in a correlation between flexibility and costs represented by the total traveling distance. Moreover, the additional costs vs. flexibility trend can be plotted both for both the indicators showing analogies and differences with respect to the results previously obtained for chemical processes. In conclusion, the methodology provided a useful tool to quantify the parameter deviations that can be withstood keeping the solution feasible. The combined costs-flexibility graphical representation allows a more conscious decision and could be further exploited in more specific applications such as the supply chain network design.

Keywords: flexibility, supply chain, operational research, travelling salesman problem.

1. Introduction

With the increasing interest toward the replacement of the conventional sources of energy by renewables and the need to deal with the fluctuating market demand, operational research and the supply management in particular belong to the research fields in the process system engineering domain for which the flexibility assessment is of critical importance.

Design under uncertainty represents a useful and reliable tool mainly used for process unit design during the last decades by means of a thoroughly defined procedure and dedicated flexibility indices (Di Pretoro et al. (2019)). In the last years, some articles focusing on supply chain design under uncertainty have been proposed in literature as well (Govindan et al. (2017)). However, this is a more recent application field and they aim at the analysis of specific case study by considering different scenarios and uncertain

parameters without providing a detailed procedure of general validity in analogy with the processes related research works.

The Traveling Salesman Problem (TSP) can be seen as the most basic formulation of operational research problem. It was first mathematically formulated by Hamilton and Kirkman in the 1800s and mathematically considered in 1930 by Flood for the solution of the bus routing problem.

It is a NP-hard problem and, in its decision version it is also NP complete, i.e. it could be solved in polynomial time on a deterministic Turing machine. The exact algorithm to solve this problem implies the calculation of all permutations with a running time lying within a polynomial factor $O(n!)$ as a consequence, where n represents the number of cities. Therefore it is evident that this approach can be employed for problems involving a small number of cities only (Applegate et al. (2006)).

For this reason, heuristic algorithms or specific cases analysis are often used to provide a reliable approximation of the optimal solution in a reasonable time. The application of several optimization methodologies such as branch-and-bound, branch-and-cut, dynamic programming, Lin-Kernighan heuristic, ant-colony optimization, genetic algorithm etc. have been proposed during the last century in order to effectively approximate the optimal solution with good results both from the accuracy and from the computational effort perspectives for case studies of practical interest.

The purpose of this research work is then to adapt the well-established flexibility assessment procedure employed for process equipment to the TSP in order to analyse the impact of input parameters uncertainties on the optimal solution and the corresponding additional distance to be travelled. Moreover, this analysis is extended to uncertainties weighted with respect to their likelihood by means of the stochastic flexibility index in order to find the probability corresponding to the different optimal pathways and provide a more reliable tool.

The obtained results, of general validity, could be then applied to more complex operational research problem by adding the desired constraints and weight functions in order to have a design procedure for each specific supply chain management case study.

Further details about flexibility analysis and the TSP under uncertainty case study as well as a thorough description of the design procedure are discussed in the following sections.

2. Flexibility assessment

Flexibility assessment is a powerful tool usually employed in process engineering in order to assess the ability of a system to accommodate a set of uncertain parameters. This property can be effectively quantified by means the several flexibility indices proposed in literature during the last decades.

In particular, two main categories of indices can be distinguished, namely deterministic and stochastic. The difference between them is that the former evaluates the magnitude of the maximum disturbance load that can be withstood by the system without becoming infeasible while the latter accounts for the deviation likelihood as well and evaluates the probability that a given set of uncertain parameters is accommodated or, equivalently, the percentage of accommodated perturbations.

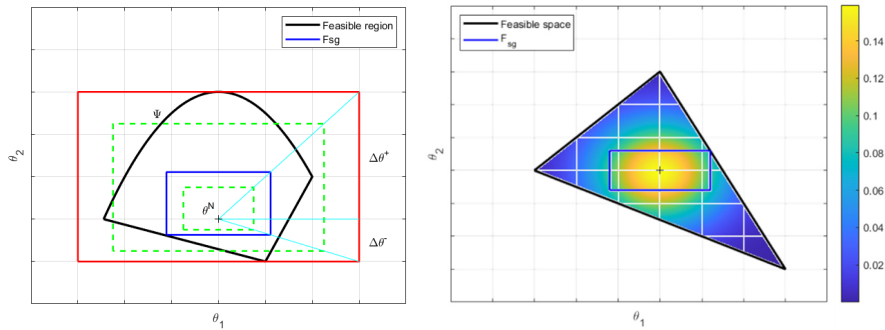


Figure 1 - Deterministic (left) vs stochastic (right) flexibility indices

The geometrical representation of both the indices typologies is reported in Figure 1.

Among the deterministic flexibility indices, the most used by far is the one proposed by Swaney and Grossmann (1985) that is the solution of the optimization problem:

$$F_{SG} = \max \delta \tag{1}$$

$$s. t. \max_{\theta \in T(\delta)} \min_z \max_{j \in J} f_j(d, z, \theta) \leq 0 \tag{2}$$

where θ refers to the uncertain variables, d to the design parameters, z to the control variables and finally δ is the scaling factor of the hyperrectangle

$$T(\delta) = \{\theta: \theta^N - \delta \cdot \Delta\theta^- < \theta < \theta^N + \delta \cdot \Delta\theta^+\} \tag{3}$$

represented in Figure 1 (left).

On the other hand, the most frequently used stochastic flexibility index was defined by Pistikopoulos and Mazzuchi (1990). Given the uncertain parameters Probability Distribution Function $P(\theta)$, the Stochastic Flexibility (SF) is evaluated as:

$$SF = \int_{\psi(d, \theta) \leq 0} P(\theta) \cdot d\theta \tag{4}$$

where ψ is the feasible domain boundary. Differently from the F_{SG} index, the value of SF is bounded between 0 and 1 and might have a non-zero value in correspondence of the nominal operating conditions.

In this research work both the indices will be used and compared in order to provide a complete overview of the TSP flexibility analysis. Although the results concerning the SF index depend on the selected PDF for the uncertain parameters, the proposed methodology has a general validity and can be applied to whatever probability function.

3. The TSP under uncertainty case study

“Given a set of cities along with the cost of travel between each pair of them, the traveling salesman problem is to find the cheapest way of visiting all the cities and returning to the starting point” is the standard formulation of the TSP.

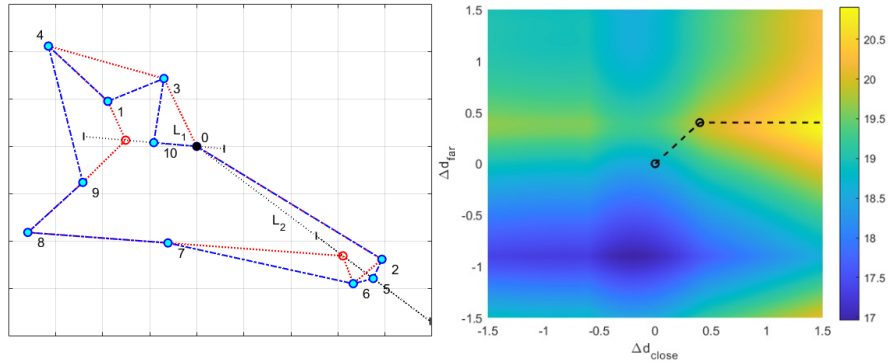


Figure 2 – The TSP case study (left) and sensitivity analysis results (right)

In this study, a set of n cities is generated randomly while the starting point is fixed at the center of the 2D space. In order to show the results with a 3D plot two uncertain parameters have been selected, namely the distances between the starting point and the closest (L_1) and furthest (L_2) cities respectively since they have been detected as the most critical points affecting the length and the shape of the optimal traveling path.

The results discussed in this paper refer to a case study with $n = 10$ whose general scheme is shown in Figure 2 (left). This value for n allows the solution of the TSP with a computational time of the order of seconds by means of the exact algorithm and does not restrain the general validity of the proposed procedure. The study of other optimization algorithms does not serve the purpose of this research.

The distances L_1 and L_2 uncertain interval for the deterministic flexibility analysis was set equal to $[L_i - 1.5, L_i + 1.5]$. On the other hand, for the stochastic flexibility analysis, a Gaussian PDF was used by setting $\mu_i = L_i$ and $\sigma_i = 0.5$ so that the 99.7 % of the deviations fall within the uncertain interval.

4. Results

The flexibility assessment was then performed by using the exact algorithm to solve the TSP case study. Hundreds of runs with ten random cities were performed to estimate an average computational time of about 6 minutes for a 31×31 discretized domain.

The sensitivity analysis results for the presented case study are shown in Figure 2 (right). As it can be noticed, the reason lying behind the relatively high computational time is the impossibility to perform the deterministic flexibility assessment by means of the vertex analysis since the traveling distance is not a quasi-convex function in the uncertain domain. Thus, all the points on the perimeter of the hyperrectangle should be evaluated to assess the flexibility index. This irregular trend can be explained by the fact that not only the length but also the shape of the optimal path can change for some values of the uncertain parameters deviations.

In fact, in Figure 2 (left) it can be noticed that, if points 5 and 10 are displaced over the uncertain domain, the shortest path (red dotted line) takes a different shape with respect to the optimal path under nominal operating conditions (the blue dashed line).

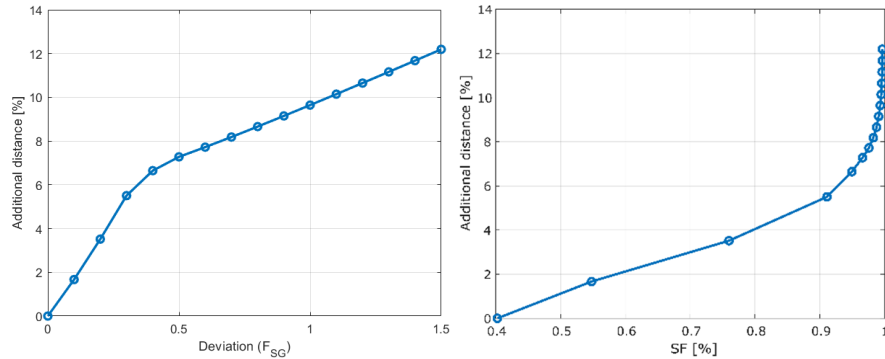


Figure 3 – Deterministic (left) and stochastic (right) flexibility assessment results

The results obtained for both the deterministic and stochastic flexibility analysis are then presented and discussed here below in the corresponding section.

4.1. Deterministic Flexibility Assessment

The flexibility analysis performed with the F_{SG} index results in an original trend worth to be analysed in detail. After an expected initial linear growth, deviations in the interval [0.3 0.5] have a lower impact on the travel path. The slope change of the trend is the indicator proving that the optimal traveling sequence between the ten cities has changed. This flexibility index behaviour can be also detected in Figure 2 – right where it corresponds to the optimal trajectory represented by the dashed line. After that, for more relevant deviations (yellow region in Figure 2 - right) the cost function restarts increasing again but with a lower slope with respect to the beginning.

In addition, from the sensitivity analysis results, it can be detected that the deviation Δd_{close} of the closest point (i.e. point 10) affects the results in a much more relevant way than the furthest one Δd_{far} and results to be the most critical parameter for the expectations concerning the travelling distance under uncertainty.

4.2. Stochastic Flexibility Assessment

The additional travel distance vs. SF index trend (Figure 3 - right) results show a couple of analogies with the main features obtained for the process units with the same PDF, namely a non-zero value in correspondence of the nominal operating conditions and an asymptotical behaviour towards $SF=1$. The first phenomenon is due to the fact that the travel path under nominal operating conditions already overestimates the value obtained for a given set of perturbed conditions, while the second one is a consequence of the residual probability and summation constraint of the PDF.

However, some relevant difference are worth to be remarked for the SF index as well. First of all a faster initial increase is observed due to the relatively limited deviation range included in an expected traveling additional distance lower than 6.65 %. On the other hand, an accumulation of points in the proximity of the same value can be observed even before approaching the vertical asymptote. In analogy with the F_{SG} results, this behaviour is somehow explained by the fact that an additional distance of 6.65 % already includes the majority of the possible deviations, with poor margins of improvement whether a higher value was taken into account.

5. Conclusions

The application of the flexibility assessment to the TSP case study was successfully achieved. The outlined procedure allows to estimate the additional travel distance with respect to nominal operating conditions as a function of the expected deviation magnitude or the withstood deviation likelihood.

Differently from flexible process equipment, where an additional flexibility usually corresponds to a proportional oversizing, in the TSP case study the change in the optimal traveling path form causes some discontinuities in the derivative of the travel distance vs flexibility plot. On the other hand, the definition of the stochastic flexibility index implies that, for a given travel distance value, the feasible domain might result by the combination of topological regions in the uncertain domain that are not connected.

However, in both cases the obtained results allow a deeper understanding of the optimal results and provide a useful decisional tool and the associated graphical representation. Nonetheless, the proposed procedure could be used in the reverse way: for a given data set, the flexibility analysis is performed by considering uncertain one city position at a time in order to detect which is the most critical one for the optimality of the solution.

In future works travel cost functions between each pair of cities and further constraining conditions could be included in order to be able to characterise each specific supply chain management problem and find the associated solution under uncertainty.

References

- Applegate, D., Bixby, R., Chvátal, V., Cook, W. (2006). *The Traveling Salesman Problem: A Computational Study*. Princeton; Oxford: Princeton University Press. doi:10.2307/j.ctt7s8xg
- Di Pretoro, A., Montastruc, L., Manenti, F., Joulia, X., 2019. Flexibility analysis of a distillation column: Indexes comparison and economic assessment. *Computers & Chemical Engineering* 124, 93–108. <https://doi.org/10.1016/j.compchemeng.2019.02.004>
- Govindan, K., Fattahi, M., Keyvanshokoh, E., 2017. Supply chain network design under uncertainty: A comprehensive review and future research directions. *European Journal of Operational Research* 263, 108–141. <https://doi.org/10.1016/j.ejor.2017.04.009>
- Pistikopoulos, E.N., Mazzuchi, T.A., 1990. A Novel Flexibility Analysis Approach for Processes with Stochastic Parameters. *Comput. Chem. Eng.* 14, 991–1000. [https://doi.org/10.1016/0098-1354\(90\)87055-T](https://doi.org/10.1016/0098-1354(90)87055-T)
- Swaney, R.E., Grossmann, I.E., 1985. An index for operational flexibility in chemical process design. Part I: Formulation and theory. *AIChE Journal* 31, 621–630. <https://doi.org/10.1002/aic.690310412>

Production scheduling of continuous make-and-pack processes with byproducts recycling

Apostolos P. Elekidis,^a Georgios P. Georgiadis,^a Michael C. Georgiadis^{a,*}

^a *Department of Chemical Engineering, Aristotle University of Thessaloniki, University Campus, Thessaloniki, 54124, Greece*

Abstract

In this work, the scheduling problem of continuous make-and-pack industries is considered. A continuous precedence-based MILP model is proposed for multistage continuous processes, considering flexible intermediate storage. Extending previously proposed precedence-based MILP models, multiple campaigns of the same recipe can be stored simultaneously in a storage tank. Explicit resource constraints related to the generation and recycling of byproduct are introduced, to achieve a better utilization of the available resources. Several case studies, inspired by a large-scale consumer goods industry have been solved, to illustrate the applicability of the proposed frameworks. It is illustrated that good quality schedules are obtained in reasonable solution times.

Keywords: scheduling optimisation, MILP, make-and-pack, continuous processes

1. Introduction

Within the current climate of business globalization, modern process industries have to produce a plethora of final products that can address the needs and demand of multiple customers. Hence, scheduling optimization is becoming a vital process and decision-makers tend to exploit recent advances in computer-aided optimization methods (Harjunkoski et al., 2014). Nowadays, several companies from various industrial sectors, such as food and beverages, pharmaceuticals, chemicals and fast-moving consumer goods (FMCGs), have adopted make-and-pack production processes. Due to variable production rates, a challenge typically met in continuous make-and-pack processes is the necessity to synchronize the production rates of consecutive stages (Klanke et al., 2020). Thus, continuous stages are often decoupled by deploying intermediate storage vessels (Méndez and Cerdá, 2002). Furthermore, product-dependent changeovers, mainly occurred by cleaning operations, have to be minimized to increase the productivity of production facilities. In cases when cleaning with water can affect the quality of products, an undesirable amount of byproduct waste is generated between two consecutive campaigns. Usually, the byproducts can be recycled into the next production campaigns. This industrial policy is typically met in liquid detergents industries. (Elekidis et al., 2019).

2. Problem statement

The scheduling problem under consideration has been inspired by a continuous make-and-pack process of a real-life, large-scale consumer goods industry (Elekidis, Corominas, and Georgiadis, 2019). Several intermediate products are produced through a continuous formulation stage (stage 1), while a plethora of final products is processed in the packing stage (stage 2), to satisfy multiple customer orders. Due to the different production rates of the two stages, a varying production bottleneck can be detected in

both stages depending on the specific product characteristics. To overcome this limitation and to achieve a better synchronization between the two stages, flexible intermediate buffer tanks are utilized. If an intermediate product is temporally stored in an intermediate vessel both stages operate at their maximum speed. Otherwise, the production rate is determined by the slowest stage. Furthermore, multiple changeovers take place between consecutive production campaigns due to the necessary cleaning operations. However, cleaning with water is not allowed, since water can affect the quality of the products, while the generation of an undesirable amount of foam also occurs. Under these circumstances, cleaning is implemented by generating an amount of byproduct waste. According to industrial policies, the waste can be temporally stored in tanks and it can then be recycled and utilized into the next intermediate products. The plant layout is also depicted in Figure 1.

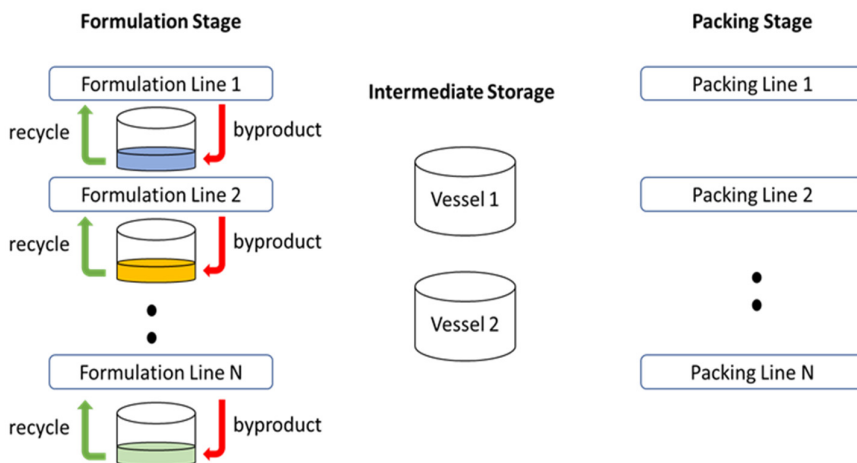


Figure 1 Plant layout

3. MILP model

A continuous-time, precedence-based, MILP model for continuous make-and-pack processes is proposed, considering intermediate storage tanks. The model consists of assignment, timing and sequencing constraints, similar to relevant frameworks (Elekidis et al. 2019). Mass balance constraints for the storage vessels are also imposed based on the framework of Méndez and Cerdá, (2002). Immediate precedence variables are used for sequencing operations in the processing units, while general precedence variables are utilized to define the sequence of storage operations. In comparison with discrete-time formulations, only a binary variable is introduced for handling the mass balance constraints. This auxiliary binary variable defines if the packing operation of a product order starts before or after the completion of its formulation process.

In the MILP model of Méndez and Cerdá (2002), it is assumed that each product can be stored in a vessel only if the packing operations of the previously stored products have been completed. However, this assumption is not in full agreement with industrial practice, since product orders produced by the same intermediate could potentially share the same storage tank. The proposed MILP model differs from the model of Méndez and Cerdá, (2002), since by integrating additional mass balance constraints and introducing a new set of auxiliary binary variables, multiple production orders, produced by the same

intermediate product type, are allowed to be stored in the same storage tank simultaneously.

The synchronisation of stages is achieved via proper timing constraints. An intermediate product can either be temporarily stored into a buffer tank or routed directly to a packing line, bypassing the storage vessels. Timing constraints guarantee that if a product bypasses the storage tank, the process of the two stages will start simultaneously.

Additionally, efficient resource-constraints, related to generation and recycling of byproduct waste are proposed to consider potential benefits by their utilization in the plant. The main objective is the minimization of the total cost, taking into account the costs of changeovers, the idle time, the processing time and the cost of the generated byproduct waste.

3.1. Mass balance constraints for product orders produced by the same intermediate product type (recipe)

Production campaigns which are made by the same intermediate product, can be stored simultaneously in the same buffer tank for some period of time. Due to the varying production rates, a production bottleneck can be posed in both stages depending on the specific product features. To prevent the overloading of storage vessels, mass balance constraints have to be enforced at two time points for each product. The first time point corresponds to the end of each formulation operation while the second one corresponds to the starting time of each packing operation (Méndez and Cerdá, 2002). Extending the work of Méndez and Cerdá (2002), a new set of auxiliary binary variables is introduced to accurately handle mass balance constraints. In case multiple intermediate campaigns are stored simultaneously at the same storage tanks, the proposed binary variables define if an operation (formulation or packing process) starts before or after the time point under consideration. The new binary variables are determined via a set of big-M constraints.

The total stored amount at a specific time point is defined as the total inserted amount minus the total amount exported from the buffer tank. The inserted and exported amounts are efficiently defined via a set of big-M constraints. It should be noted that the vast majority of scheduling approaches for make-and-pack processes rely on discrete-time formulations to handle material balances around storage equipment thus, resulting in computationally intractable models.

3.2. Modelling of byproducts

In scheduling problem under consideration, a set of product orders $i \in I$ is allocated to a set of production units $j \in J$, at each production stage $s \in S$. A subset of units j is able to process product order i , ($j \in JI_i$) at stage s ($j \in JS_s$). Additionally, a subset of processing units $j \in J$ is able to process product orders i , given by subset IJ_j . The maximum production rate of each product i , at stage s , is given by the parameter $r_{i,s}$.

The vast majority of product-dependent changeovers, usually take place among the production of consecutive production campaigns, i and i' , are related to cleaning operations. In liquid detergents industries, a significant amount of waste is generated during the changeover time, $n_{i,i'}$, which can be recycled into the next production campaigns if the quality of products is not affected. Each processing unit j of formulation stage ($j \in JS_{s=1}$), is connected with a dedicated storage vessel, in which the generated waste can be temporarily stored. Since the capacity of byproduct vessels have to be fully satisfied, a set of mass balance constraints are introduced, without utilising further binary variables or considering any discretisation of time.

$$O_i = LI_i + \sum_{i' \in I: i' \neq i} \sum_{j \in (JS_1 \cap JI_i \cap JI_{i'})} X_{i,i',j} n_{i,i'} r_{i',1} - W_i \quad \forall i \in I \quad (1)$$

$$LI_i \geq O_{i'} - dm_i \left(1 - \sum_{j \in (JS_1 \cap JI_i \cap JI_{i'})} X_{i',i,j} \right) \quad \forall i \in I, i' \in I: i' \neq i \quad (2)$$

$$LI_i \leq O_{i'} + dm_i \left(1 - \sum_{j \in (JS_1 \cap JI_i \cap JI_{i'})} X_{i',i,j} \right) \quad \forall i \in I, i' \in I: i' \neq i \quad (3)$$

$$LI_i \leq iw_j Y_{i,j} + dm_i \sum_{i' \in Ij_j: i' \neq i} X_{i,i',j} \quad \forall i \in I, j \in (JS_1 \cap JI_i) \quad (4)$$

$$LI_i \geq iw_j Y_{i,j} - dm_i \sum_{i' \in Ij_j: i' \neq i} X_{i,i',j} \quad \forall i \in I, j \in (JS_1 \cap JI_i) \quad (5)$$

$$W_i \leq dm_i a_i \quad \forall i \in I \quad (6)$$

$$W_i \leq L_i \quad \forall i \in I \quad (7)$$

$$O_i \leq cp_j Y_{i,j} \quad \forall i \in I, j \in (JS_1 \cap JI_i) \quad (8)$$

$$RW_j \geq O_i - cp_j (1 - Y_{i,j}) - cp_j \sum_{i' \in Ij_j: i' \neq i} X_{i,i',j} \quad \forall i \in I, j \in (JS_1 \cap JI_i) \quad (9)$$

The accumulated amount of byproduct O_i is calculated at the end of each changeover. According to mass balance constraints (1), the variable O_i is equal to the previously accumulated byproduct amount LI_i , plus the waste generated by the cleaning operations that take place ($n_{i,i'} r_{i',1}$), minus the recycled amount during the production of product i , W_i . In particular, it is assumed that during the changeover time $n_{i,i'}$, between two consecutive campaigns i and i' , the production rate of byproduct is equal to the maximum production rate of campaign i' ($r_{i',1}$). According to constraints (2) and (3), if product i' is produced immediately before product i , ($X_{i',i,j} = 1$), the accumulated byproduct is equal to variable $O_{i'}$. Constraints (4) and (5) ensure that the accumulated amount of byproduct at the beginning of the first campaign ($X_{i',i,j} = 0$), is either equal to zero or equal to the initial byproduct (iw_j) only if the product i is allocated at unit j ($Y_{i,j} = 1$). The demand parameter, dm_i plays the role of the big-M value in constraints (4) and (5). The maximum percentage that can be recycled is expressed by the parameter a_i . Constraints (8) ensures that the stored amount of byproduct, will not surpass the related capacity (cp_j) of each tank. The remained amount of waste RW_j at the end of the time horizon, is equal to the generated waste of the last production campaign O_i of each unit.

4. Application studies

In order to evaluate the efficiency and applicability of the proposed MILP model, 4 case studies are considered. They include 3 formulation lines, 3 packing lines and two

intermediate storage tanks. Due to the large number of final products, the decomposition algorithm of Elekidis et al. (2019), is utilized to solve the large-scale MILP model. To assess the benefits of the intermediate storage tanks, two case studies have been solved, with (decoupled layout) and without (coupled layout) the utilization of intermediate storage tanks. The minimization of total operational time is the main objective. The results are summarized in Table 1. It is illustrated that the utilization of flexible intermediate buffer tanks leads to a better synchronization of the production stages resulting in increased productivity. In particular, depending on the case study, the productivity gain ranges within 4.43% (case 2) and 17.29% (case 1).

Furthermore, the case studies have been also solved considering the minimization of total cost. Near-optimal solutions are obtained within a total CPU time of 3600s, while a zero-optimality gap is achieved at each iteration. Results are summarized in Table 2. The individual costs represent relative monetary units (rmu) and they are also presented in Table 2. It is observed that the biggest percentage of the total cost, reflects the idle time as it valued by 30 rmu/h. Although the percentage of the changeover cost is gradually decreased in larger cases, from 17% (case 1) to 8% (case 4), the percentage of idle time cost is steadily increased, from 47% (case 1) to 77% (case 4).

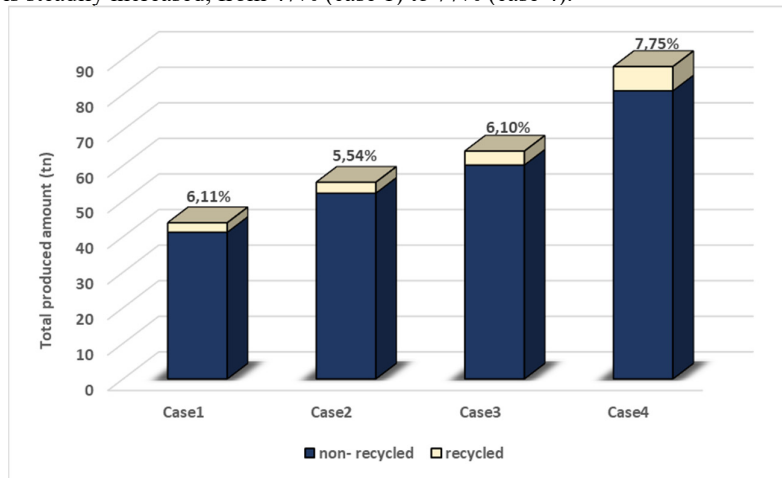


Figure 2 Total produced and recycled amount in each case

Table 1 Comparison of the total operational time of the two layouts

	Products	Coupled layout*	Decoupled layout*	Difference (hours)*	Productivity gain (%)
Case 1	20	186.63	225.63	39.00	17.29%
Case 2	50	355.34	371.82	16.48	4.43%
Case 3	70	422.14	442.70	20.55	4.64%
Case 4	100	596.23	680.88	84.64	12.43%

*the values represent the total operational time of all production units of both stages in hours

Table 2 Results for larger problem instances – Cost distribution

	Products	TC*	COC*	ITC*	PTC*	WC*
Case 1	20	1301	221 (17%)	614 (47%)	385 (30%)	79 (6%)
Case 2	50	1999	231 (12%)	1224 (61%)	470 (24%)	72 (4%)
Case 3	70	3505	299 (9%)	2592 (74%)	541 (15%)	72 (2%)
Case 4	100	5455	422 (8%)	4193 (77%)	757 (14%)	82 (2%)

*TC=Total cost, COC=Changeover cost, ITC=idle time cost, PTC=processing time cost, WC=waste cost

**COC= 10 rmu/h, ITC= 30 rmu/h, PTC= 1 rmu /h, WC= 0.5 rmu /kg

Figure 2 shows, the total produced amount and the amount of recycles as well, for each case. It is illustrated that the byproduct recycles constitute a significant percentage of the total produced amount, which ranges from 6.1% (Case 3) to 7.57% (Case 4). Hence, it is evident that the utilization of this recycling policy and the consideration of the proposed byproduct constraints lead to better utilization of raw materials and a significant reduction of material cost.

5. Conclusions

In this work a precedence-based MILP model, for the scheduling of continuous, make-and-pack industries is presented. The utilization of intermediate buffers can provide a better synchronization between the two production stages. The integration of byproduct recycling constraints leads to a better use of the available raw materials and to a significant reduction of waste. The proposed MILP model can constitute the basis of an optimization tool to assist decision-makers with rigorous scheduling solutions under a dynamic environment.

Acknowledgements

This research is co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme “Human Resources Development, Education and Lifelong Learning 2014-2020” in the context of the project “Optimization techniques for production scheduling in the process industries” (MIS 5047892).

References

- A. P. Elekidis, F. Corominas, M.C. Georgiadis, 2019, Production Scheduling of Consumer Goods Industries, *Industrial and Engineering Chemistry Research*, 58, 51, 23261-23275
- C. A. Méndez, J. Cerdá, 2002, An efficient MILP continuous-time formulation for short-term scheduling of multiproduct continuous facilities, *Computers and Chemical Engineering*, 26, 4-5, 687-695
- C. Klanke, V. Yfantis, F. Corominas, S. Engell, 2020. Scheduling of a Large-Scale Industrial Make-and- Pack Process with Finite Intermediate Buffer Using Discrete-Time and Precedence-Based Models, *Computer Aided Chemical Engineering*. 47, 1153-58
- I. Harjunkski, C. T. Maravelias, P. Bongers, P. M. Castro, S. Engell, I. E. Grossmann, J. Hooker, C. A. Méndez, G. Sand, J. Wassick, 2014, Scope for Industrial Applications of Production Scheduling Models and Solution Methods, *Computers and Chemical Engineering*, 62, 161-193

Robust Optimization for Scheduling and Lot-sizing of a Single Machine with Sequence-dependent Changeovers

Hossein Mostafaei^a, Fabricio Oliveira^{a,*}

^aDepartment of Mathematics and Systems Analysis, School of Science, Aalto University, Finland

**Email: fabricio.oliveira@aalto.fi*

Abstract

A plethora of contributions have appeared in the literature over the past decade in the area of production planning of a single machine with sequence-dependent changeovers. Most of them, however, deal with the deterministic optimization model where all the parameters are considered known, which render optimal schedules, sub-optimal, or even infeasible in practice. In this paper, we first develop a new deterministic model based on a hybrid discrete- and continuous-time mixed-integer linear programming model for the production scheduling of a single machine with sequence-dependent changeovers. The proposed model (i) considers a time interval in which the processing machine is unavailable due to some maintenance jobs, and (ii) monitors inventory levels over shorter time scales, rather than at the end of predefined long-term periods. Then, the robust counterpart of the deterministic model is devised using the Γ -robustness approach that considers uncertainty in processing and changeover times. The objective is to find robust production schedules such that the sum of holding and changeover costs is minimized. We demonstrate the performance of the proposed model with a real-world case study.

Keywords: Scheduling, Changeovers, Robust optimization, Uncertainty.

1. Introduction

Scheduling is an important planning function in the process industries and lies at the very heart of the performance of manufacturing (Stood, 1996). It is a highly dynamic process and is characterized by a high level of uncertainty, such as raw material availability, prices, machine reliability, processing times, and client demand, which vary with respect to time and are often subject to unexpected deviations (Li and Ierapetritou, 2008; Georgiadis et al. 2019). Two primary methods to address scheduling under uncertainty are stochastic programming and robust optimization (Grossmann et al. 2016). In the former (Birge and Louveaux, 2011), the uncertainty is modeled through several scenarios, using probability distributions, which is often difficult to describe in practice. Another drawback is that the number of scenarios increases exponentially with the number of uncertain parameters, which limits the application of this approach to solve practical problems. Robust optimization (Bertsimas and Sim 2004), on the other hand, defines an uncertainty set of possible realizations of the uncertain parameters and selects a solution

that can cope best with the various realizations, while keeping the problem tractable. The other advantage of robust optimization is that it only requires defining bounds for uncertain parameters, rather than their probability distribution.

There is a plethora of publications that have addressed the production scheduling of a single machine with sequence-dependent changeovers (Gupta and Magnusson, 2005; Liu et al., 2008; Mostafaei et al., 2020). Compared to its deterministic counterpart, the literature on the robust optimization of the single machine scheduling problem is still very scarce. To this end, this paper develops a novel robust optimization model for single machine scheduling with sequence-dependent changeovers (SMSSDC) to generate reliable solutions that are immunized against uncertainty. The contributions of this work are twofold: First, we develop a continuous-time mixed-integer linear programming (MILP) formulation using generalized disjunctive programming (Castro and Grossmann, 2012; Mostafaei and Harjunkoski, 2020) for SMSSDC that allows us to (i) consider the machine unavailability during the time horizon, (ii) monitor the inventory over shorter time windows, and (iii) meet customer demands over multiple intermediate due dates. Second, by means of the Γ -robustness approach, the proposed deterministic MILP model is extended to a robust counterpart propose schedules that are immunized against variability in processing and changeover times. The reason why we select the Γ -robustness approach is that it keeps the problem tractable and does not change the structure of the original problem, meaning that if the deterministic model is an MILP, its robust counterpart will also be an MILP

2. Problem statement

We focus on the production scheduling of a single machine, where a set of products p should be manufactured during a time horizon of duration H . The aim is to find the optimal sequence of production so that the costs related to backorder, changeover, and inventory are minimized. Given are the following: $[v_p^{min}, v_p^{max}]$: minimum and maximum production size during every process run; ρ_p : processing time of one unit of product p ; $\sigma_{p,p'}, c_{p,p}$: changeover time and cost; τ_t : absolute time of period t ; $d_{p,t}$: the product demand for product p at time period t ; $[S_{stop}, E_{stop}]$ idle time interval for the machine.

3. Deterministic optimization model

Here we present a deterministic MILP model for SMSSDC that is derived from generalized disjunctive programming (GDP). The model involves three sets: (i) production runs $r, r' \in R = \{1, 2, \dots, |R|\}$, in which a run r should always be performed after the completion of run $r - 1$; (ii) set of products $p, p' \in P$; (iii) set of time periods $t, t' \in T$. Note that the cardinality of set R is a tuning parameter, and to find the optimal solution, we start with an initial guess for $|R|$ and keep increasing it until no improvement is observed in the objective function.

3.1. Allocation and product size

First, let us define the binary variable $X_{p,r}$ denoting that the production run $r \in R$ processes product $p \in P$, whenever $X_{p,r} = 1$. The exclusive OR (\vee) in disjunction below states that exactly one product can be processed in the machine during each run. The size of product p processed in the machine during the run r ($V_{p,r}$) is restricted to the interval $[v_p^{min}, v_p^{max}]$. The duration of run r (L_r) processing product p is a function of $V_{p,r}$ and the

processing time of one unit of product p . The convex hull reformulation of the disjunction leads to the equations (1)-(3).

$$\bigvee_{p \in P} [X_{p,r} v_p^{\min} \leq V_{p,r} \quad \Rightarrow \quad \sum_{p \in P} X_{p,r} = 1 \quad \forall r \quad (1)$$

$$\leq v_p^{\max} L_r = \rho_p V_{p,r}] \quad \forall r \quad \Rightarrow \quad v_{p,r}^{\min} X_{p,r} \leq V_{p,r} \leq v_{p,r}^{\max} X_{p,r} \quad \forall p, r \quad (2)$$

$$L_r = \sum_{p \in P} \rho_p V_{p,r} \quad \forall r \quad (3)$$

3.2. Changeover times and costs

Changeover times and costs occur when two different products p and p' are processed in two consecutive runs $r - 1$ and r ($r \geq 2$). The changeovers are often associated with changing the operating conditions or with the cleaning of the machine. If the continuous variables CT_r and CC_r are the changeover time and cost at the start of run r , respectively, we have the following Eqs. (4)-(7). Note that the binary variable $X_{p,p',r}^{ch}$ ($= 1$ if the changeover $p - p'$ occurs at the start of run r) can be treated as a 0-1 continuous variable.

$$X_{p,r-1} \wedge X_{p',r} \Rightarrow X_{p,p',r}^{ch} \quad \forall p, p', r \quad \Rightarrow \quad X_{p,p',r}^{ch} + 1 \geq X_{p,r-1} + X_{p',r} \quad \forall p, p', r \quad (4)$$

$$\bigvee_{p \in P} \bigvee_{p' \in P} [X_{p,p',r}^{ch} CT_r \quad \Rightarrow \quad \sum_{p,p' \in P} X_{p,p',r}^{ch} = 1 \quad \forall r \geq 2 \quad (5)$$

$$= \sigma_{p,p'} CC_r \quad \Rightarrow \quad CT_r = \sum_{p,p' \in P} \sigma_{p,p'} X_{p,p',r}^{ch} \quad \forall r \geq 2 \quad (6)$$

$$= cc_{p,p'}] \quad \forall r \quad \Rightarrow \quad 2$$

$$\geq 2 \quad \Rightarrow \quad CC_r = \sum_{p,p' \in P} cc_{p,p'} X_{p,p',r}^{ch} \quad \forall r \geq 2 \quad (7)$$

2

3.3. Sequencing production runs

Let continuous variable C_r be the completion time of production run r . The beginning of the production run r should not be before the end of the completion of $r - 1$ plus the changeover time at the start of run r . Furthermore, the completion of each run is always inferior to the time horizon length (H):

$$(C_{r-1} + L_r) + CH_r \leq C_r \leq H \quad \forall r \quad (8)$$

3.4. An unavailable time interval due to maintenance jobs

Let us assume that the machine should be stopped during the time interval $[S_{stop}, E_{stop}]$ due to, e.g., maintenance. The binary variable W_r is equal to one if the production run r is completed during the time interval $[0, S_{stop}]$, i.e., earlier than time S_{stop} . If $W_r = 1$, then (i) the previous run $r - 1$ must be completed in the same interval too ($W_{r-1} = 1$), and (ii) the completion time of run r (C_r) must less than or equal to S_{stop} . Notice the inclusive OR (\vee) since multiple runs can end before the time S_{stop} .

$$W_r \Rightarrow W_{r-1} \quad \forall r \geq 2 \quad \Rightarrow \quad W_r \leq W_{r-1} \quad \forall r \geq 2 \quad (9)$$

$$\bigvee_{r \in R} [W_r C_r \leq S_{stop}] \quad \Rightarrow \quad \sum_{r \in R} W_r \geq 1 \quad (10)$$

$$C_r \leq S_{stop} + (H - S_{stop})(1 - W_r) \quad \forall r \quad (11)$$

Now let the binary variable W_r^{last} be 1 if run r is the last one completed during the time interval $[0, S_{stop}]$. If $W_r^{last} = 1$ then (i) $W_r = 1$ and $W_{r+1} = 0$, meaning that run r ends earlier than time S_{stop} ($W_r = 1$) whereas the completion of run $r + 1$ is after time S_{stop} ($W_{r+1} = 0$), and (ii) the starting time of run $r + 1$ should be after the maintenance interval $[S_{stop}, E_{stop}]$, i.e., after time E_{stop} . Notice the exclusive OR (\vee) since only one run can be the last one completed before time S_{stop} . We have thus the following conditions.

$$W_r \wedge \neg W_{r+1} \Rightarrow W_r^{last} \quad \forall r \quad \Rightarrow \quad W_r^{last} \geq W_r - W_{r+1} \quad \forall r \quad (12)$$

$$\Rightarrow \quad \sum_{r \in R} W_r^{last} = 1 \quad (13)$$

$$\bigvee_{-r \in R} [W_r^{last} C_{r+1} - L_{r+1} \geq E_{stop} W_r^{last} \quad \forall r] \quad C_{r+1} - L_{r+1} \geq E_{stop} W_r^{last} \quad \forall r \quad (14)$$

3.5. Meeting demand and inventory level

Let the binary variable $Y_{t,r}$ indicate if the production run r completes during period t . If $Y_{t,r} = 1$, it should be imposed $\tau_{t-1} \leq C_r \leq \tau_t$, where τ_t is the absolute time of period t . Notice that the disjunction in (23) is exclusive, meaning that each production run r should be completed in exactly one of the periods.

$$\bigvee_{-t \in T} [Y_{t,r} \square_{t-1} \leq C_r \leq \tau_t] \quad \forall r \quad \Rightarrow \quad \sum_{t \in T} Y_{t,r} = 1 \quad \forall r \quad (15)$$

$$\leq \tau_t] \quad \forall r \quad \Rightarrow \quad \sum_{t \in T} \tau_{t-1} Y_{t,r} \leq C_r \leq \sum_{t \in T} \tau_t Y_{t,r} \quad \forall r \quad (16)$$

Now, let us introduce the binary variable $Y_{t,r}^{last}$, indicating that run r is the last one completed in the period t . If $Y_{t,r} = 1$ and $Y_{t,r+1} = 0$, the run r is the last one that ends in period t ($Y_{t,r}^{last} = 1$). Note that $Y_{t,r}^{last}$ can take one for only one run r during each period t , as imposed in (18).

$$Y_{t,r} \wedge \neg Y_{t,r+1} \Rightarrow Y_{t,r}^{last} \quad \forall t, r \quad \Rightarrow \quad Y_{t,r}^{last} \geq Y_{t,r} - Y_{t,r+1} \quad \forall r, t \quad (17)$$

$$\bigvee_{-r \in R} Y_{t,r}^{last} \quad \forall t \quad \Rightarrow \quad \sum_{r \in R} Y_{t,r}^{last} = 1 \quad \forall t \quad (18)$$

If $Y_{t,r}^{last} = 1$ and to meet product demands, the amount of product p sent to customers during period t computed through the continuous variable $CP_{p,r}$ should be as large as $d_{p,t}$. Otherwise, the backorder demand ($BC_{p,t}$) will cause penalty costs (e.g., customer dissatisfaction). Note that $BC_{p,t-1}$ on the right-hand side of Eq. (19) indicates the shortage of product p to the customers during the period $t - 1$ to be tardily met at period t .

$$\sum_{r \in R} CP_{p,r} + BC_{p,t} - BC_{p,t-1} \geq \sum_{t' \in T} Y_{t',r}^{last} d_{p,t'} \quad \forall p, t, r \quad (19)$$

Let $IP_{p,r}$ be a continuous variable denoting the inventory level of product p at the end of run r . The stock of product p is increased by the production of the product and decreased by sending it to local customers. Note that at the start time of the planning horizon, the initial inventory of product p is known i.e., ip_p^0 . To prevent stock overloading and empty conditions, $IP_{p,r}$ should lie within the feasible range, as imposed in Eq. (21).

$$IP_{p,r} = IP_{p,r-1} + V_{p,r} - CP_{p,r} \quad \forall p, r \quad (20)$$

$$ip_p^{min} \leq IP_{p,r} \leq ip_p^{max} \quad (21)$$

3.6. Objective function

The objective is to minimize the backorder, inventory, and changeover costs.

$$z = \sum_{t \in T} \sum_{p \in P} cb_{p,t} BC_{p,t} + \sum_{r \in R} \sum_{p \in P} cip_p IP_{p,r} + \sum_{r \in R} C_r \quad (22)$$

4. Robust counterpart

4.1. Changeover time

Let us assume that the changeover times, represented by $\sigma_{p,p'}$, are subject to uncertainty, and thus are represented by the random variable $\tilde{\sigma}_{p,p'}$ with support given by the interval $\tilde{\sigma}_{p,p'} \in [\underline{\sigma}_{p,p'} - \hat{\sigma}_{p,p'}, \underline{\sigma}_{p,p'} + \hat{\sigma}_{p,p'}]$. We follow the approach in Bertsimas and Sim (2004)

in which we would like to immunize constraint (6). For that, first notice that can be equivalently rewritten as:

$$\sum_{r \in R} \sum_{r \geq 2} CT_r \geq \sum_{r \in R} \sum_{r \geq 2} \sum_{p, p' \in P} \tilde{\sigma}_{p, p'} X_{p, p', r}^{Ch} \quad (23)$$

Constraint (23) can then be restated in a robust sense, as

$$\sum_{r \in R} \sum_{r \geq 1} CT_r \geq \sum_{r \in R} \sum_{r \geq 2} \sum_{p, p' \in P} \underline{\sigma}_{p, p'} X_{p, p', r}^{Ch} + \Phi(\Gamma) \quad (24)$$

where $\Phi(\Gamma)$ is the protection term defined as

$$\Phi(\Gamma) = \sum_{r \in R} \sum_{r \geq 2} \sum_{p, p' \in P} \hat{\sigma}_{p, p'} X_{p, p', r}^{Ch} \times Z_{p, p', r} \quad (25)$$

$$\text{subject to: } \sum_{r \in R} \sum_{r \geq 2} \sum_{p, p' \in P} Z_{p, p', r} \leq \Gamma \quad (F) \quad (26)$$

$$0 \leq Z_{p, p', r} \leq 1 \quad (Q_{p, p', r}) \quad \forall p, p', r \quad (27)$$

The terms in brackets represent the associated dual variables, used to state the equivalent dual formulation (28)-(30) for (25)-(27), which is given by

$$\Phi(\Gamma) = + \sum_{r \in R} \sum_{r \geq 2} \sum_{p, p' \in P} Q_{p, p', r} \quad (28)$$

$$\text{subject to: } F + Q_{p, p', r} \geq \hat{\sigma}_{p, p'} X_{p, p', r}^{Ch} \quad \forall p, p', r \quad (29)$$

$$F, Q_{p, p', r} \geq 0 \quad (30)$$

The robust counterpart can be stated by replacing (6) with the following set of constraints

$$\sum_{r \in R} \sum_{r \geq 2} CT_r \geq \sum_{r \in R} \sum_{r \geq 2} \sum_{p, p' \in P} \underline{\sigma}_{p, p'} X_{p, p', r}^{Ch} + \Gamma F + \quad (31)$$

$$\sum_{r \in R} \sum_{r \geq 2} \sum_{p, p' \in P} Q_{p, p', r}$$

$$CT_r \geq \sum_{p, p' \in P} \underline{\sigma}_{p, p'} X_{p, p', r}^{Ch} \quad \forall r \geq 2 \quad (32)$$

$$F + Q_{p, p', r} \geq \hat{\sigma}_{p, p'} X_{p, p', r}^{Ch} \quad \forall p, p', r \quad (33)$$

4.2. Processing time

In a similar vein, let us assume that the processing time ρ_p is represented by a random variable $\tilde{\rho}_p$ with support $[\underline{\rho}_p - \hat{\rho}_p, \underline{\rho}_p + \hat{\rho}_p]$. Following the same reasoning as in §4.1, the robust counterpart considering protection in the processing time can be thus obtained by replacing (3) with the following constraints:

$$\sum_{r \in R} L_r \geq \sum_{r \in R} \sum_{p \in P} \underline{\rho}_p V_{p, r} + \Gamma' F' + \sum_{r \in R} \sum_{p \in P} Q'_{p, r} \quad (34)$$

$$L_r \geq \sum_{p \in P} \underline{\rho}_p V_{p, r} \quad \forall r \quad (35)$$

$$F' + Q'_{p, r} \geq \hat{\rho}_p V_{p, r} \quad \forall p, r \quad (36)$$

5. Real-world case study

The proposed deterministic and robust approaches for SMSSDC are applied to a real-world paper machine that was obtained by slightly modifying Example 1 previously studied by Mostafaei et al. (2020). The aim is to meet customer demands of P4₁₂₀, P6₁₇₀, P7₄₀, P8₄₀, P9₈₀, P13₂₁₈, P14₅₁, P15₄₅, P18₄₀, and P20₈₃ (with the subscripts indicating the number of paper reels to be produced during a time horizon of 11 days) at minimum backorder and changeover costs. It is assumed that: (i) the paper machine should be stopped for 12 hours during the time interval [100, 112] h, and (ii) backorder has a unit

cost of 200 \$, one hour changeover has a fixed cost of 100 \$, (iii) the processing and changeover times are uncertain parameters and have variability in $[\underline{\rho}_p - 0.02, \underline{\rho}_p + 0.02]$ and $[\underline{\sigma}_{p,p'} - 0.1, \underline{\sigma}_{p,p'} + 0.1]$, respectively.

The example was solved on a Quad 1.60 GHz Intel Core i5-8365, 16 GB RAM Laptop using GAMS/CPLEX with 8 parallel threads as the MILP solver. Figure 1 shows the Gantt chart for the deterministic model and its robust counterpart (with $\Gamma, \Gamma' = 1$). The deterministic model corresponded to a cost of \$166.47 was solved in 20.15s while the robust model with the objective function of \$6514.30 was solved in 364.12 s. As can see from the figure, the production sequence is slightly different, and some products cannot be produced when processing and changeover times are subject to uncertainty. For example, product 18 it not produced in the robust model.

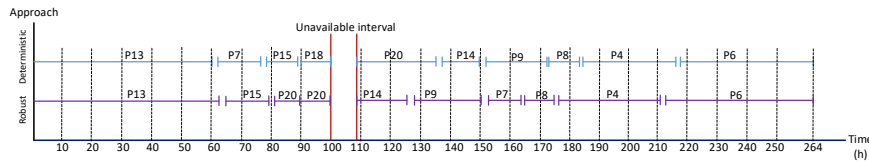


Figure 1: Solutions found for the case study

6. Conclusions

This paper has presented a continuous-time optimization formulation for the short-term scheduling of a single machine with sequence-dependent changeovers. We first developed a deterministic MILP model derived from GDP that considers the machine unavailability during the time horizon, and customer demands on multiple due dates. Then we derived the robust counterpart of the model using the Γ -robustness approach to tackle uncertain processing and changeover times. We validated the proposed model using a real-world case study from the paper-making industry, and a robust schedule for production scheduling of 11 days was successfully found in a reasonable CPU time.

References

- Stood, P. (1996), The complexity of scheduling in practice, *Int. J. Oper. Prod. Manag.* 16, 37-53.
- Li, Z., Ierapetritou, M., (2008), Process scheduling under uncertainty: Review and challenges. *Computers and Chemical Engineering* 32, 715-727.
- Georgiadis, G.P., Elekidis, A.P., and Georgiadis M.C. (2019). Optimization-based scheduling for the process industries: from theory to real-life industrial applications. *Processes*, 7(7), pp. 438.
- Grossmann, I.E., Apap, R.M., Calfa, B.A., García-Herreros, P., Zhang, Q., (2016). Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty. *Computers and Chemical Engineering*. 91, 3-14.
- Birge J.R and Louveaux F. (2011), Introduction to stochastic programming. Springer.
- Bertsimas, D., Sim, M (2004). The Price of robustness (2004), *Operations Research*. 52(1), 35-53.
- Gupta D, Magnusson T. The capacitated lot-sizing and scheduling problem with sequence-dependent setup costs and setup times. *Computers and Operations Research* 2005;32(4):727-47.
- Liu, S.; Pinto, J. M.; Papageorgiou, L. G (2008). A TSP-based MILP model for medium-term planning of single-stage continuous multiproduct plants. *Ind. Eng. Chem. Res.* 47, 7733-7743.
- Mostafaei, H., Ikonen, T. J., Krumb, J., Deneke, T., Heljanko, K. & Harjunkoski, I. (2020). Data-driven approach to grade change scheduling optimization in a paper machine. *Ind. Eng. Chem. Res.* 59 (17), 8281-8294.

- Castro, P.M., Grossmann, I.E., (2012). Generalized disjunctive programming as a systematic modeling framework to derive scheduling formulations. *Ind. Eng. Chem. Res.* 51, 5781-5792.
- Mostafaei, H., Harjunkoski, I (2020). Continuous-time scheduling formulation for multipurpose batch plants. *AIChE J.* 2019, 66. e16804.

Simultaneous minimization of minimum resource and storage requirements in batch process

Nitin Dutt Chaturvedi*, Rakesh Kumar Sinha

Process, Systems Engineering Laboratory, Department of Chemical and Biochemical Engineering, Indian Institute of Technology Patna, Bihta, Patna, 801106, Bihar, India

nitind@iitp.ac.in

Abstract

Increase in the cost of process water and stringent environmental norms result in the requirement for water reducing technologies in process industries. Batch processes have been widely used in process industry (such as pharmaceuticals, fine and specialty chemicals, etc.) as batch processing is appropriate to produce a specialized low amount of production and is flexible to adjust according to the dynamism of the market. In recent years, significant research efforts are made to minimize resources such as energy and water in batch processes. Optimizing the storage requirement is an essential aspect while considering capital investment and space availability.

This paper focuses on minimizing freshwater and storage requirement simultaneously. In this paper, a mathematical formulation is proposed to solve the proposed bi-objectives problem. The primary objective of this work is to generate a Pareto optimal front for these two objectives via using a ϵ -constraint method. The proposed linear programming formulation includes demand satisfaction, source availability and quality constraints. The trade-off between storage and resource requirements is captured through a Pareto optimal front. This Pareto optimal front facilitates decision-makers to select an appropriate operating point as per suitability.

Keywords: Batch process, Storage capacity, Resource minimization, Pareto optimal front, Water Allocation Network.

1. Introduction

The batch process is utilized for production of specialized products in several industries such as food, pharmaceutical, fine chemical, etc. Impurity concentration and time are two important constraints for resource recovery in operation of a batch process. Mathematical optimization and physical-insight based techniques are used to minimize freshwater and storage requirements in industries (Wang and Smith, 1994). Earlier, Majozi (2005) presented a continuous-time model for minimizing wastewater and freshwater with or without central water storage facilities in batch process. As water is essentially required in several operations of the process industry, so freshwater needs to be used efficiently which reduces the discharge of wastewater (Shenoy and Bandyopadhyay, 2007). Later, Gouws and Majozi (2008) presented the effect of several storages for wastewater reduction in batch process. It is reported that the optimization of the storage requirement is an essential aspect upon considering capital investment and space scarcity. Similarly, Gouws et al. (2010) reviewed and compared various methodologies of water minimization for batch processes in process industries. Mitigation of water footprint is important for enhancing profitability and reducing

environmental degradation; hence, optimization of water resources for industrial processes is essential for water management (Begatin et al., 2014). Similarly, Chaturvedi and Bandyopadhyay (2014) examined a model for optimal production scheduling and usage of freshwater. The trade-off between freshwater minimization and production maximization is presented via Pareto optimal front. Later, Adekola and Majozi (2017) presented a mathematical model for optimization of batch production and minimization of wastewater for a static mass load problem simultaneously. The minimization of wastewater is attained by exploring the order of jobs in a unit. The results depict that water utilization is minimized via suitable selection of jobs arrangements in a unit. Later, Hesran et al. (2019) studied waste minimization techniques for waste minimization. Recently, Duhbaci et al. (2020) reviewed water minimization approach via a mathematical programming in industrial process.

Earlier, Chaturvedi and Bandyopadhyay (2012) presented a pinch analysis-based approach to target minimum storage capacity and maximum freshwater requirement for a Water Allocation Network (WAN) in batch process. In this work, trade-off between storage capacity and freshwater requirement is captured by solving a mathematical programming optimization model. The proposed algorithm can be applied to any fixed flow rate and fixed scheduled batch process involving single or multiple qualities. The proposed linear programming formulation includes demand satisfaction, source availability and quality constraints.

2. Problem Statement

The general problem statement for minimizing freshwater and storage requirements for WAN in a batch process is as follows:

- For a fixed time interval (T_k), a set of internal sources (M_{sTk}) is provided. A known flow rate F_{sITk} is produced by each source which have a concentration of impurity as q_{sITk} .
- For a fixed time interval (T_k), a set of internal demands (M_{dTk}) is given. A known minimum flow rate (F_{dTk}) is required with concentration of impurity as q_{dTk} .
- The freshwater flowrate requirement is F_{wITk} with concentration of q_{wITk} .

Objective: Overall aim of this work is to minimize freshwater and storage requirements in process industry. Primary objective of the proposed model is to generate a Pareto optimal front by varying storage limit. Pareto optimal front accounts for a trade-off between freshwater and storage capacity requirement in batch process. For generating points of Pareto optimality, ϵ -constraint technique (Mavrotas, 2009) is implemented in the model. The model is evaluated for a single objective and another objective is stated as inequality constraint. The minimum storage requirement is fixed at a specified value and added as a constraint.

3. Mathematical Formulations

Let x_{ij} depicts the transfer of flow from i^{th} source to j^{th} demand. From the conservation principle, the equation for demand constraint is given by Eq. (1). The concentration of every stream and freshwater is equal to concentration of the demand which is shown in Eq. (2). The release of water from sources should be lesser than the limit of source in

any time interval. Eq. (3) presents inequality constraint for the i^{th} source in a time interval.

$$\sum_{T_{m \leq k}} \sum_i (x_{ij}) + F_{wjT_k} = F_{djT_k} \tag{1}$$

$$\sum_{T_{m \leq k}} \sum_i (q_{siT_k} * x_{ij}) + q_{wjT_k} * F_{wjT_k} = q_{djT_k} * F_{djT_k} \tag{2}$$

$$\sum_{T_{m \geq k}} \sum_i (x_{ij}) \leq F_{siT_k} \tag{3}$$

Eq. (4) represents minimizing objective function for storage requirement (Z_1):

$$Z_1 = \sum_{T_{m \neq k}} \sum_i x_{ij} \tag{4}$$

Eq. (5) represents minimizing objective function for freshwater requirement (Z_2):

$$Z_2 = \sum_k \sum_j F_{wjT_k} \tag{5}$$

These modeling equations are solved for minimizing freshwater and storage requirements in a WAN. The trade-off between storage and resource requirement are captured through the Pareto optimal front. The Pareto optimal front facilitates decision-makers to select an appropriate operating point as per its suitability.

4. Illustrative Example

For a batch process operation, considering a batch WAN. Table 1 (Chaturvedi and Bandyopadhyay, 2012) shows flow-rate and concentration of streams in batch WAN.

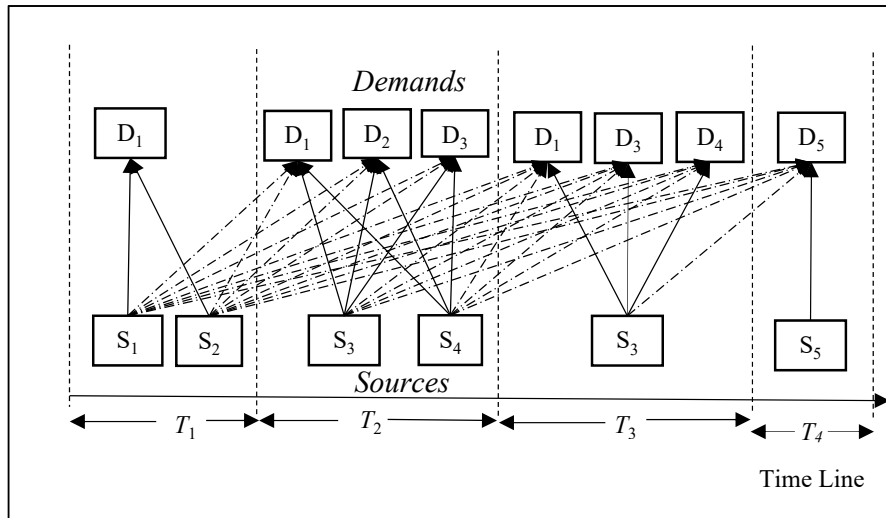


Figure 1: Schematic diagram of the Water Allocation Network.

Table 1: Demand and source availability in different time intervals
(Chaturvedi and Bandyopadhyay, 2012)

Source	Flow rate (t/h)	Conc. (ppm)	Duration (h)	Demand	Flow rate (t/h)	Conc. (ppm)	Duration (h)
S ₁	5	25	0.0-1.0	D ₁	10	0	0.0-3.0
S ₂	5	15	0.0-1.0	D ₂	40	27.5	1.0-2.0
S ₃	10	40	1.0-3.0	D ₃	5	15	1.0-3.0
S ₄	50	30	1.0-2.0	D ₄	5	25	2.0-3.0
S ₅	50	10	3.0-4.0	D ₅	20	5	3.0-4.0

Table 2: Time-interval wise flow and their concentration for the WAN problem
(Chaturvedi and Bandyopadhyay, 2012)

T ₁ (0.0-1.0) h		T ₂ (0.1-2.0) h		T ₃ (2.0-3.0) h		T ₄ (3.0-4.0) h	
Flow (t)	Conc. (ppm)	Flow (t)	Conc. (ppm)	Flow (t)	Conc. (ppm)	Flow (t)	Conc. (ppm)
25	5	10	40	10	40	50	10
15	5	50	30	-5	25	-20	5
0	-10	-40	27.5	-5	15		
		-5	15	-10	0		
		-10	0				

The streams are segregated in four time intervals as shown in Table 2. Figure 1 shows schematic diagram of sources and demands presence in several time intervals. Based on data of sources and demands of Table 2, the optimization problem is solved using GAMS with CPLEX solver for linear programming. A Pareto optimal front is generated between storage and freshwater requirement. This optimality front is generated by using ϵ -constraint method. The ϵ -constraint methodology provides the non-dominated solution for proposed mathematical model for minimizing freshwater and storage requirements in WAN.

In ϵ -constraint method, bi-objective model is solved for an objective function and another objective is expressed in form of an inequality constraint. For this inequality condition, maximum storage capacity is fixed at a certain value which is added as a constraint. For generating other optimal points, methodology is repeated with scalarly placed variations on storage requirement. The mathematical model is solved as linear programming model. The developed model is solved for minimizing freshwater and storage requirements. Thus, trade-offs between storage capacity and freshwater requirements are determined for WAN which is represented as Pareto optimal front in Figure 2.

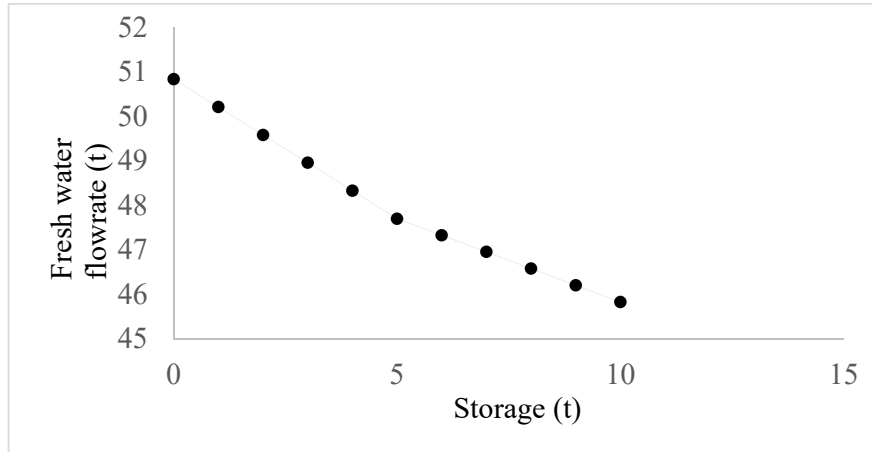


Figure 2: Pareto optimal front for bi-objective problem

As storage capacity increases from zero to 5 ton (t), fresh water requirement decreases from 50.833 t to 47.70 t. With further increase in storage capacity from 5t to 10 t, freshwater requirement decreases up to 45.833 t. A constant fresh water requirement is observed upon further increase in storage capacity. Thus, target value for minimum freshwater requirement is obtained to be 45.833 t and maximum required storage capacity is obtained to be 10 t. Note that, this targeted value equates with the problem discussed in Chaturvedi and Bandyopadhyay (2012). The developed model provides a trade-off between two objectives via Pareto-optimal front.

5. Conclusions

This paper focuses on minimizing bi-objectives problem of freshwater and storage requirements simultaneously. A mathematical model is developed to solve this bi-objectives problem and a Pareto optimal front is generated by using ϵ -constraint methodology. The proposed methodology can be applied to any fixed flow rate and fixed scheduled batch process involving single or multiple qualities. The proposed linear programming model includes demand satisfaction, source availability and quality constraints. The trade-off between storage and resource requirements are captured through a Pareto optimal front. For generating points of Pareto optimality, ϵ -constraint technique is implemented in the model. The model is evaluated for a single objective and another objective is stated as inequality constraint. This Pareto optimal front facilitates decision makers to select an appropriate operating point as per its suitability.

Nomenclatures

x_{ij}	Transfer of flow from i^{th} source to j^{th} demand
F_{djT_k}	Flow rate requirement of j^{th} internal demand in T_k time interval
F_{siT_k}	Flow rate of i^{th} internal source in T_k time interval
F_{wjT_k}	Freshwater flow rate of j^{th} internal demand in T_k time interval
M_{dT_k}	Set of internal demands in T_k time interval
M_{sT_k}	Set of internal sources in T_k time interval

q_{djT_k}	Impurity concentration of i^{th} internal source in T_k time interval
q_{siT_k}	Impurity concentration of i^{th} internal source in T_k time interval
q_{wjT_k}	Freshwater concentration of j^{th} internal demand in T_k time interval
T_k	Time interval
Z_1	Minimizing objective function for storage requirement
Z_2	Minimizing objective function for freshwater requirement

References

- C.L. Hesran, A.L. Ladier, V.B. Genoulaz, V. Laforest, 2019, Operations scheduling for waste minimization: A review, *Journal of Cleaner Production*, 206, 211-226.
- G. Mavrotas, 2009, Effective implementation of the ϵ constraint method in multi-objective mathematical programming problems, *Applied Mathematics and Computation*, 213, 2, 455-465.
- J.F. Gouws, T. Majozi, 2008, Impact of multiple storage in wastewater minimization for multi contaminant batch plants: Toward Zero Effluent, *Industrial and Engineering Chemistry Research*, 47, 2, 369-379.
- J.F. Gouws, T. Majozi, D.C.Y. Foo, C.L.Chen, J.Y.Lee, 2010, Water minimization techniques for batch processes, *Industrial and Engineering Chemistry Research*, 49,19,8877-8893.
- N.D.Chaturvedi, S. Bandyopadhyay, 2014, Simultaneously targeting for the minimum water requirement and the maximum production in a batch process, *Journal of Cleaner Production*, 77, 105-115.
- N.D.Chaturvedi, S. Bandyopadhyay, 2012, Minimization of storage requirement in a batch process using pinch analysis, *Computer Aided Chemical Engineering*, 31, 670-674.
- O. Adekola, T. Majozi, 2017, Wastewater minimization in batch plants with sequence dependent changeover, *Computers and Chemical Engineering*, 97, 85-103.
- R. Begatin, J.J. Klemes, A.P. Reverberi, D. Huisingh, 2014, Conservation and improvements in water resource management, *Journal of Cleaner Production*, 77, 1-9.
- T.B. Duhbaci, S. Ozel, S. Bulkan, Water and energy minimization in industrial processes through mathematical programming: A literature review, *Journal of Cleaner Production* (2020), doi: <https://www.sciencedirect.com/science/article/abs/pii/S095965262034796X>.
- T. Majozi, 2005, Wastewater minimisation using central reusable water storage in batch plant, *Computers and Chemical Engineering*, 29, 7, 1631-1646.
- U.V. Shenoy, S. Bandyopadhyay, 2007, Targeting for multiple resources, *Industrial and Engineering Chemistry Research*, 46, 11, 3698-3708.
- Y.P. Wang, R. Smith, 1994, Wastewater minimization, *Chemical Engineering Science*, 49,7,981-1006.

A Hybrid Solution Approach for Large-scale Batch Scheduling with Quantum Computing

Akshay Ajagekar, Fengqi You

Cornell University, Ithaca, NY 14853, USA

asa273@cornell.edu

Abstract

In this work, we demonstrate the applicability and efficiency of hybrid quantum computing (QC) based techniques for solving large-scale batch scheduling problems. We propose a novel solution approach that integrates classical and quantum techniques to overcome their individual shortcomings. The hybrid solution framework allows us to take advantage of the complementary strengths of the two methods. We demonstrate the use of this hybrid approach in developing efficient solution algorithm for an intractable batch scheduling problem formulated as mixed-integer fractional program. We then propose a novel QC-based parametric decomposition algorithm for solving large-scale batch scheduling problems. The performance and efficiency of the developed hybrid models and methods are validated by performing several computational experiments and comparing the obtained computational results with those of the state-of-the-art classical off-the-shelf optimization solvers..

Keywords: Quantum computing, batch scheduling, hybrid techniques, optimization

1. Introduction

Scheduling is a critical issue in systems operations and has a major impact on the productivity of a system. Complex scheduling problems span a wide variety of areas, especially in the chemical process industries (Kallrath, 2002). Analytical approaches based on mathematical programming are plagued by the computational complexity of scheduling problems and may demonstrate exponential time complexity (Aytug et al., 1994). While heuristic and hybrid heuristic methods (Chu et al., 2013) can be used to solve large-scale problems in a reasonable time, they tend to trade off optimality for speed. Therefore, there arises a need to develop novel solution approaches capable of overcoming limitations of current solution approaches carried out on classical computers for complex scheduling problems.

The ability of quantum computing (QC) to offer large speed advantages over classical methods (Nielsen and Chuang, 2010) has proven to be a powerful tool in solving optimization problems (Martonosi and Roetteler, 2019), as QC becomes commercially available (Mohseni et al., 2017). There are a number of QC applications in process systems engineering (Ajagekar and You, 2020). For detailed preliminaries on QC-based computational optimization, specifically adiabatic quantum optimization, we refer interested readers to theoretical background in (Kadowaki and Nishimori, 1998) and practical implementation in process systems optimization (Ajagekar and You, 2019) and power systems scheduling (Lau et al., 2009). The goal of this paper is to demonstrate

the applicability and efficiency of hybrid QC-based techniques for solving large-scale scheduling problems, building upon previous work by Ajagekar et al. (2020). To overcome the shortcomings of the purely classical and purely quantum approaches, we propose novel solution methods that integrate both approaches by decomposing the problem into smaller subproblems. The hybrid solution framework allows us to take advantage of the complementary strengths of the two methods. We demonstrate the use of this hybrid approach to developing efficient algorithms for an intractable scheduling problem, the multi-purpose batch scheduling problem.

2. Quantum Computing Background

Quantum computing, which follows the rules of quantum mechanics, has emerged as the next frontier in computation. Analogous to the bit in classical computers, the quantum bit (qubit) is the fundamental unit of information in quantum computers. This work is restricted to quantum annealing-based computers, which are purposefully built for discrete optimization applications. Quantum annealing uses quantum fluctuations to tunnel through local minima and converge to an optimal state (Kadowaki and Nishimori, 1998). The advantage of the quantum annealing model is that it uses quantum fluctuations to efficiently find the minimum-energy state of the Hamiltonian, which corresponds to the objective function in the context of an optimization problem. The annealing process begins by initializing the system to the lowest-energy state of the initial Hamiltonian, which means that all qubits are in a superposition of 0 and 1. Throughout the quantum annealing process, the influence of the initial Hamiltonian is reduced, and a user-defined problem Hamiltonian is introduced. Quantum fluctuations are used to evolve the system towards the lowest-energy eigenstate of the problem Hamiltonian as shown in Figure 1a, which represents the solution to the optimization problem (Kadowaki and Nishimori, 1998).

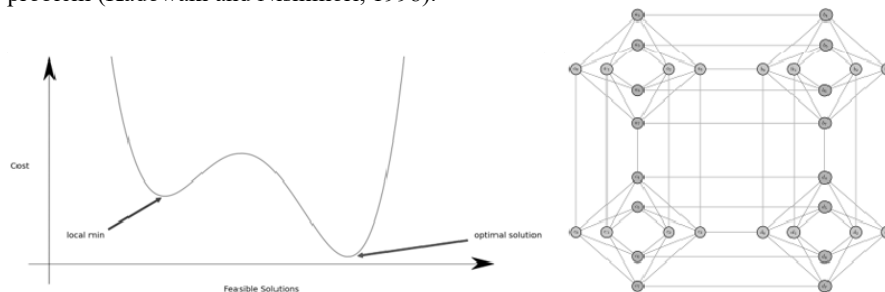


Figure 1. a) Quantum annealing and b) Chimera graph on the quantum processing unit

State-of-the-art implementations of quantum annealing are available from D-Wave Systems, which provides cloud access to users. The D-Wave system solves problems that are formulated as a quadratic unconstrained binary optimization (QUBO) model. A QUBO model can be represented by a graph, where the nodes represent the binary variables, and the edges represent the connections between these variables. On the D-Wave quantum processing unit (QPU), this graph maps to another graph where the nodes are qubits and the edges are internal couplers that connect the qubits. This lattice of interconnected qubits forms a topological structure known as the Chimera graph as shown in Figure 1b. Solving a problem on the QC system involves embedding graph of

problem QUBO on the QPU, performing quantum annealing to return the lowest-energy solutions of the problem Hamiltonian before returning the unembedded solution.

3. Multi-purpose Batch Process Scheduling

A significant scheduling problem that is typical in batch manufacturing is the multi-purpose batch scheduling problem. This operational problem pertains to determining the optimal batch sizes, the assignment of tasks to equipment resources, and the sequence of tasks on each equipment unit. In this paper, we consider a Resource Task Network (RTN)-based formulation (Castro et al., 2001) with a fractional objective, as in You et al. (2009). To solve the mixed-integer fractional programming (MIFP) model of the multi-purpose batch scheduling problem efficiently, we propose a novel approach that integrates a hybrid QC-MILP decomposition strategy (Ajagekar et al., 2020) with an inexact parametric algorithm, which was developed for general MIFP problems (Zhong and You, 2014). Consider a version of the RTN-based continuous-time batch scheduling problem modified to have a fractional objective. The binary variables of this model are $W_{S_{in}}$, $W_{f_{in}}$, and $R2_{rn}$. $W_{S_{in}}$ indicates whether task i starts at time point n . At time point n , $W_{f_{in}}$, and $R2_{rn}$ indicate whether task i finishes and whether equipment resource r is being utilized. The problem also has six continuous variables. Bs_{in} and Bf_{in} represent the batch size of task i as it starts and finishes at time point n , respectively. $R1_{rn}$ is the amount of material resource r at time point n . T_n is the time corresponding to time point n , and H is the variable scheduling horizon, or make-span, of the network. The objective (1) describes the maximization of profit per unit time, which calls for the MIFP formulation presented here.

$$\max \frac{\sum_{r \in M} price_r (R1_{rN} - init_r)}{H} \tag{1}$$

$$s.t. \quad b_i^{\min} W_{S_{in}} \leq Bs_{in} \leq b_i^{\max} W_{S_{in}} \quad \forall i \in T, \forall n \tag{2}$$

$$b_i^{\min} W_{f_{in}} \leq Bf_{in} \leq b_i^{\max} W_{f_{in}} \quad \forall i \in T, \forall n \tag{3}$$

$$b_i^{\min} \left(\sum_{m < n} W_{S_{im}} - \sum_{m' \leq n} W_{f_{im'}} \right) \leq Bp_{in} \quad \forall i \in T, \forall n \tag{4}$$

$$Bp_{in} \leq b_i^{\max} \left(\sum_{m < n} W_{S_{im}} - \sum_{m' \leq n} W_{f_{im'}} \right) \quad \forall i \in T, \forall n$$

$$Bs_{i(n-1)} + Bp_{i(n-1)} = Bp_{in} + Bf_{in} \quad \forall i \in T, \forall n \tag{5}$$

$$R1_{rN} \geq demand_r \quad \forall r \in M \tag{6}$$

$$R1_{rn} = R1_{r(n-1)} + \sum_{i \in TO_r} \rho_{ir}^O \cdot Bf_{in} - \sum_{i \in TI_r} \rho_{ir}^I \cdot Bs_{in}, \quad \forall n > 1, r \in M \tag{7}$$

$$R1_{rn} \leq storage_r^{\max} \quad \forall r \in M, \forall n \tag{8}$$

$$R2_{rn} = R2_{r(n-1)} + \sum_{i \in E_r} (W_{S_{in}} - W_{f_{in}}) \quad \forall r \in E, \forall n \tag{9}$$

$$T_m - T_n \geq \sum_{i \in E_r} (fd_i \cdot W_{S_{in}} + vd_i \cdot Bs_{in}) \quad \forall r \in E, \forall n \tag{10}$$

$$\sum_n W_{S_{in}} = \sum_n W_{f_{in}} \quad \forall i \in T \tag{11}$$

$$W_{S_{iN}} = W_{f_{i1}} = 0 \quad \forall i \in T \tag{12}$$

$$T_1 = 0, T_N = H \quad (13)$$

$$R2_{rN} = 0 \quad \forall r \in E \quad (14)$$

$$Bs_{in}, Bf_{in}, Bp_{in}, R1_{rn}, T_n, H \geq 0 \quad \forall i \in T, \forall r \in M, \forall n \quad (15)$$

$$Ws_{in}, Wf_{in}, R2_{rn} \in \{0, 1\} \quad \forall i \in T, \forall r \in E, \forall n \quad (16)$$

Constraints (2)-(4) are batch size constraints. Constraint (5) models the logical relationship between Bs_{in} , Bf_{in} , and Bp_{in} for a given task. Constraint (6) enforces required demand by the end of the scheduling horizon. Eq. (7) expresses the mass balance of chemical resources, where ρ_{tr}^O and ρ_{tr}^I represent rates of consumption and generation of resource r by task i , respectively. Eq. (8) enforces the storage capacity limitation. The equipment unit eq. (9) does not allow an equipment unit to be utilized by more than one task at any time, where I_r is the set of tasks that utilize equipment unit resource r . The timing eq. (10) ensures that the processing time of tasks on each equipment unit is not greater than the duration between two consecutive time points. Eq. (11) expresses that tasks that have started must finish. Eq. (12) prevents tasks from finishing at the first time point and does not permit tasks from starting at the end of the scheduling horizon. Eq. (13) and (14) represent the initial and final conditions at end of scheduling horizon. To solve the MIFP model of the multi-purpose batch scheduling problem efficiently, we propose a novel approach that integrates a hybrid QC-MILP decomposition strategy with an inexact parametric algorithm, which was developed for general MIFP problems (Zhong and You, 2014).

3.1. Hybrid QC-MIFP Parametric Decomposition Algorithm

We propose a novel algorithm that integrates hybrid QC-based decomposition with a computationally efficient parametric method (Zhong and You, 2014). We first discuss the QC-based decomposition and further show how to integrate it with the parametric algorithm. Consider the simplified problem that results from replacing the fractional objective in (1) by the linear objective: $\max \sum_{r \in M} price_r (R1_{rN} - init_r) - \lambda \cdot H$, where λ is a penalty parameter that is initialized to zero and is iteratively updated. We introduce a hybrid QC-MILP decomposition method that efficiently solves this MILP scheduling problem formed by replacing fractional objective with the above linear objective, in two phases. The first phase determines the start times, finish times, and batch sizes of tasks that optimize the objective function. The second phase balances equipment unit resources and searches for an optimal schedule on each unit such that the start times, finish times, and batch sizes determined in the first phase are feasible. The main idea is that if the equipment resources cannot be balanced, then the partial solution returned from the first phase is infeasible. Linear objective, constraints (2)-(8), and constraints (10)-(13) form the relaxed MILP that is optimized in phase one of this hybrid strategy.

$$\min H_r = \sum_n R2_{r(n-1)} \cdot (1 + 2K) + R2_{rn} \cdot (1 - 2K) - 2R2_{r(n-1)} R2_{rn} \left. \vphantom{\sum_n} \right\} \text{QC Step} \quad (17)$$

where $K = \sum_{i \in I_r} (Ws_{in}^* - Wf_{in}^*)$

Constraints (9) and (14) are concerned with balancing equipment unit resources and are reduced to the QC step shown in (17). During each iteration, the relaxed MILP problem is solved using the classical deterministic solver Gurobi to return partial optimal solutions. The second phase involves taking in partial optimal solutions and solving the

QC step on a quantum processor to search for a feasible schedule. If the QC step returns an infeasible solution, integer cuts are added to the relaxed MILP. This process is repeated iteratively until the QC step returns a feasible schedule or the relaxed MILP problem is proven infeasible. If the QC step returns a feasible schedule, this schedule is returned as the optimal solution for the batch scheduling problem. Each r for which the task scheduling W is infeasible, could be identified. The assignment for these specific tasks can then be excluded. This is accomplished by adding an integer cut. To tackle the intractability stemming from the fractional objective of the multi-purpose batch scheduling problem, we integrate a hybrid QC-based decomposition with an efficient parametric algorithm. To initialize the algorithm, the dynamic parameter λ should be set to zero, and a tolerance value must be defined. During each iteration, the parametric MILP problem is solved using the hybrid QC-MILP decomposition method described above. If the MILP is infeasible, the hybrid QC-MIFP parametric decomposition algorithm stops, and the original batch scheduling problem is infeasible. The algorithm updates λ to take the value of the last computed fractional objective, and it moves on to the next iteration. Convergence is achieved when the absolute difference between the fractional objective and λ is lower than the tolerance level.

Table 1. Computational results of the batch scheduling problem

Tasks	Time points	BARON v20.4		Hybrid QC-MIFP parametric decomposition method		
		Time (s)	Max obj.	Time (s)	Max obj.	Iterations
8	5	11.7	155.7	2.0	155.7	6
12	5	21.0	141.3	2.7	145.0	6
18	8	4,080	311.6	7.4	311.6	10
22	10	5,298	174.4	13.0	183.3	14
40	20	--a	--a	279	95.1	111
52	25	--a	--a	861	197.5	66

a. Run time exceeds 24 h

3.2. Computational Results

To illustrate the effectiveness of the proposed hybrid QC-MIFP parametric decomposition method, we carry out computational experiments with batch scheduling problems of various sizes, which were taken from Dow Chemical originally (Wassic et al., 2012). A RTN as reported in Chu et al. (2013) was used for modelling the batch scheduling problem. This MIFP problem can be solved using the general-purpose MINLP solvers BARON v20.4 on a classical computer. BARON is a robust global optimization solver and consistently demonstrates fast performance on benchmarks. Comparing the computational times and objective values reported in Table 1, the hybrid parametric algorithm obtains near-optimal solutions within reasonable computation times. The performance of BARON v20.4 deteriorates beyond problems with 18 tasks, and it takes more than a day without returning any feasible solution for problems with more than 40 tasks. BARON is unable to yield solutions for large-scale problems with 24 hours of computation time. The hybrid QC-MIFP parametric decomposition method, on the other hand, obtains high quality solutions for problems with 52 tasks in under 900 seconds. The proposed hybrid QC-MIFP parametric decomposition method not only performs competitively against state-of-the-art MINLP solvers for small-scale scheduling problems, but also dominates classical solvers for medium and large-scale

batch scheduling problems. Based on the computational times observed for several batch scheduling problem instances, the hybrid QC-MIFP parametric decomposition method demonstrates better scalability than BARON. Due to its computational advantage and its high-quality solutions, the hybrid QC-MIFP decomposition method is expected to yield economic advantage for multi-purpose batch scheduling problem.

4. Conclusions

In this paper, we proposed hybrid QC-based optimization methods for solving large-scale batch scheduling problems. The proposed algorithms exploited the complementary strengths of established classical computing methods and novel QC-based solution strategies. For the multi-purpose batch scheduling problem, we proposed a novel algorithm that integrates a hybrid QC-MILP decomposition method with a fractional parametric algorithm. Computational results showed that the proposed hybrid QC-based solution strategies were orders of magnitude faster than state-of-the-art MINLP solvers.

References

- A. Ajagekar, T. Humble, F. You, 2020, Quantum computing based hybrid solution strategies for large-scale discrete-continuous optimization problems. *Computers & Chemical Engineering*, 132, 106630.
- A. Ajagekar, F. You, 2019, Quantum computing for energy systems optimization: Challenges and opportunities. *Energy*, 179, 76-89.
- A. Ajagekar, F. You, 2020, Quantum computing assisted deep learning for fault detection and diagnosis in industrial process systems. *Computers & Chemical Engineering*, 143, 107119.
- H. Aytug, S. Bhattacharyya, G. J. Koehler, J. L. Snowdon, 1994. A review of machine learning in scheduling. *IEEE Transactions on Engineering Management* 41, 165-171.
- Y. Chu, J. M. Wassick, F. You, 2013, Efficient scheduling method of complex batch processes with general network structure via agent-based modeling. *AIChE Journal*, 59, 2884-2906.
- Y. Chu, F. You, 2015, Model-based integration of control and operations: Overview, challenges, advances, and opportunities. *Computers & Chemical Engineering*, 83, 2-20.
- P. Castro, A. P. F. D. Barbosa-Póvoa, H. Matos, 2001. An Improved RTN Continuous-Time Formulation for the Short-term Scheduling of Multipurpose Batch Plants. *Industrial & Engineering Chemistry Research*, 40, 2059-2068.
- T. Kadowaki, H. Nishimori, 1998. Quantum annealing in the transverse Ising model. *Physical Review E, Plasmas, Fluids, and Related Interdisciplinary Topics*, 58, 5355-5363.
- J. Kallrath, 2002. Planning and scheduling in the process industry. *OR spectrum*, 24, 219-250.
- T. Lau, C. Chung, K. Wong, et al., 2009, Quantum-inspired evolutionary algorithm approach for unit commitment. *IEEE Transactions on Power Systems*, 24, 1503-1512.
- M. Martonosi, M. Roetteler, 2019. Next steps in quantum computing: Computer science's role. *arXiv preprint arXiv:1903.10541*
- M. Mohseni, P. Read, H. Neven, et al., 2017, Commercialize quantum technologies in five years. *Nature News*, 543, 171.
- J. C. Mukherjee, A. Gupta, 2015. A Review of Charge Scheduling of Electric Vehicles in Smart Grid. *IEEE Systems Journal*, 9, 1541-1553.
- M. A. Nielsen, I. L. Chuang, 2010. *Quantum Computation and Quantum Information*, Cambridge University Press.
- J. M. Wassick, A. Agarwal, N. Akiya, et al., 2012, Addressing the operational challenges in the development, manufacture, and supply of advanced materials and performance products. *Computers & Chemical Engineering*, 47, 157-169.

- F. You, P. M. Castro, I. E. Grossmann, 2009, Dinkelbach's algorithm as an efficient method to solve a class of MINLP models for large-scale cyclic scheduling problems. *Computers & Chemical Engineering*, 33, 1879-1889.
- Z. Zhong, F. You, 2014, Globally convergent exact and inexact parametric algorithms for solving large-scale mixed-integer fractional programs and applications in process systems engineering. *Computers & Chemical Engineering*, 61, 90-101.
- J. Zhu, Z. H. Shao, C. Chen, 2019. An improved whale optimization algorithm for job-shop scheduling based on quantum computing. *International Journal of Simulation Modelling* 18, 521-530.

A Digital Twin Framework for Business Transactional Processes in Supply Chains

Hector D. Perez,^a Satyajith Amaran,^b Esra Erisen,^b John M. Wassick,^b

Ignacio E. Grossmann^{a*}

^a*Carnegie Mellon University, Pittsburgh 15213, USA*

^b*Dow Inc., Midland 48674, USA*

grossmann@cmu.edu

Abstract

An integrated framework for building a virtual replica of business transactional processes in supply chains is presented. The framework consists of two main components: a simulation module and an optimization module. Business processes are modeled as networks of queues through which requests (internal or external to the enterprise) can flow. The digital replica of the business processes creates value by providing a simulation platform upon which to: 1) test optimization strategies, 2) forecast potential delays in requests based on the current state of the real process and the historical data, 3) identify and mitigate bottlenecks, and 4) provide more accurate fulfillment dates to customers. The optimization can be performed offline or in real-time in a feed-back loop as the simulation is being executed. As an integrated simulation and optimization environment, the framework bridges and extends the literature in business process simulation and business process optimization, building upon previous work by the authors that was restricted to only business process scheduling in static deterministic environments. A case study is presented in which three mathematical programming models and a greedy heuristic are compared. A sensitivity study shows the value in using simulation to mitigate operational bottlenecks.

Keywords: Business process optimization, digital supply chain, order-to-cash, digital twin

1. Introduction

Optimization of supply chains provides significant value to enterprises that seek to reduce costs, increase customer satisfaction, and streamline operations. For this optimization to be achieved, each of the areas in the supply chain need to be improved. These areas include those governed by material flows, financial flows, and information flows. Inefficiencies in any of these areas affect all other areas in the supply chain. Traditionally, work in supply chain optimization has focused on the physical processes that are governed by material flows. This work has been led by the Operations Research community (Pourhejazy and Kwon, 2016). However, the transactional business processes that are governed by information flows have received limited attention in supply chain optimization. Previous work by the authors has shown the potential in applying scheduling models from the chemical process industry to improve the operation of

business processes (Perez et al., 2021, 2020). However, this work was limited to perfect information systems that ignore the stochastic nature of the transactional flows and future customer demands in the supply chain. Although not applied in the context of supply chain, relevant work includes that of workflow scheduling of cloud-based systems from the computer science and systems community. Both heuristic and mathematical programming approaches are presented in a handful of works in this field (Cai et al., 2016; Hoenisch et al., 2016; Li et al., 2018). However, these works are limited to static deterministic systems. To improve the applicability of scheduling models in supply chain business process optimization, the authors present a business process digital twin for supply chain. Although the digital twin is general for other business processes, it is presented in the context of the order-to-cash process.

2. Methods

The proposed digital twin uses a queuing network representation of the order-to-cash processes as shown in Figure 1. From a manufacturing perspective, the business process queuing network is represented as a multi-stage batch plant, where customer orders are analogous to batches. Discrete event simulation is used to model orders flowing through the process. Each processing stage or transaction on an order is represented as a service node with an unknown service time that is sampled from a known probability distribution, which can be determined from real process data. At these service nodes, one or more resources can be made available to perform the necessary transactions. Resources can be shared between transactions as well. Thus, each transaction has a queue, with one or more service nodes representing the resource used. As the simulation marches forward in time, optimization events can be triggered at discrete intervals, or when certain conditions occur. A snapshot of the system state is sent to an optimization algorithm that schedules active orders and assigns the resources involved in the process transactions. The resulting schedule sequences are translated into order priorities for each queue in the network. These priorities are returned to the discrete event simulation to reorganize the positions of the orders in each of the active queues. Since the optimization is an event in the simulation, the computational time spent performing the scheduling is accounted for in the duration of the optimization event. The simulation does not pause for the optimization to occur, but keeps marching forward in time, allowing for a more realistic implementation of optimization in the business process environment. Furthermore, the optimization frequency and solution time plays an important role in the simulation-optimization integration.

The digital twin framework is developed in the Julia programming language (v1.5), which brings the advantages of using several simulation and optimization libraries available within the Julia environment. As an integrated platform, the framework reduces the need for creating interfaces between different programming languages, GUIs, and databases required for simulation and optimization. The framework is designed such that transactional processes of any scale and complexity can be modeled by it. It is general in that it allows building simulation and optimization models for business processes based on four key inputs: 1) the network structure of the activities in the business process, 2) the assignment of resources to activities, 3) the resource capacities, and 4) the probabilistic distributions of the processing times for each activity or task. These

processing times can be specific to the resource type, order type, and/or customer. Making modifications to the underlying business process structure or parameters is relatively easy to make. As a unified system, any changes are automatically translated to the queueing network structure for simulation and the mixed-integer programming (MIP) models for optimization.

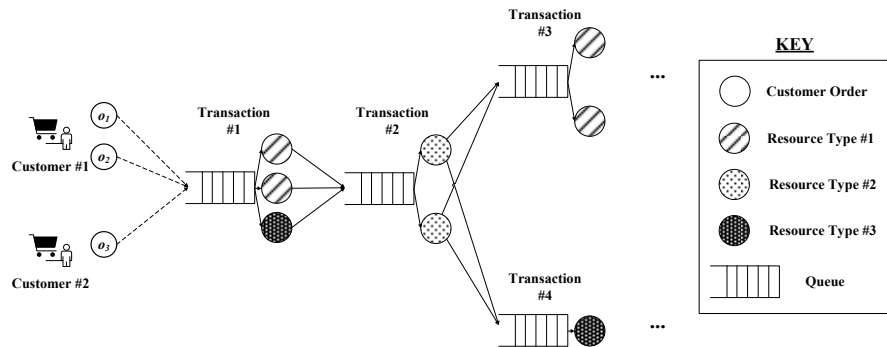


Figure 1. Queueing network representation for a business process

Two scheduling paradigms are used for optimization within the framework: a continuous-time general precedence model (Méndez and Cerdá, 2003) and discrete-time task network models. The latter include the Resource-Task Network (Pantelides, 1994) and the State-Task Network (Kondili et al., 1993). As discussed in another work by the authors, the general precedence model provides the benefits of reduced model sizes and temporal accuracy, whereas the task network models provide the benefits of reduced solution times for large systems due to tight model relaxations (Perez et al., 2021). As an alternative to mathematical optimization, a greedy heuristic is also provided, in which the order sequencing in the queueing network is determined by the order price, where orders that bring in higher revenue take priority. The benefit of using optimization over heuristics is that the mathematical models capture the different features in the system (e.g., order prices, backlog penalties, processing times, and due dates) to yield a holistic view when assigning order priorities.

3. Case Study

The digital twin is applied to an extended order-to-cash process with 11 steps as shown in Figure 2. Transactional steps (1-5, 7-9, and 11) are processed by a single agent with processing times sampled from a uniform distribution in the 15-30 min time range. The goods loading step (6) can be processed by one of two agents with processing times sampled uniformly from the 1-2 hr. range. The goods transportation step (10) has four available agents with processing times sampled uniformly from the 8-24 hr. range. 10 orders enter the system with arrival times occurring randomly within the first 28 hrs. Order due dates are set at random, ensuring that all orders are given at least the average total processing time between their arrival times and due dates. Order sales revenues vary from \$0-1K with a backlog penalty of 30% for late orders.

Queue re-prioritizing is performed at hourly intervals in the simulation. A time limit of 10 min is placed on the optimization algorithms, which are solved using CPLEX 20.1 on

a 2.11 GHz, 24 GB RAM, Intel i7 PC. The case study is run under two modes: deterministic and stochastic. The difference between the two is that the stochastic mode samples transaction times from the processing time distributions, whereas the deterministic mode uses the average values of such distributions. However, both modes are stochastic in the sense that, at any time in the simulation, future orders are unknown. Six operating models are compared: FIFO (first-in, first-out), Greedy Priority Heuristic, GP (General Precedence), D-RTN (Discrete-time Resource-Task Network), D-STN (Discrete-time State-Task Network), and PI (Perfect Information). The FIFO mode represents a simulation without any queue priority optimization. The greedy priority heuristic mode assigns order priorities in each queue according to the order value. The PI mode is an offline optimization that yields a tight upper bound on the total revenue possible when all information about present and future orders is known beforehand.

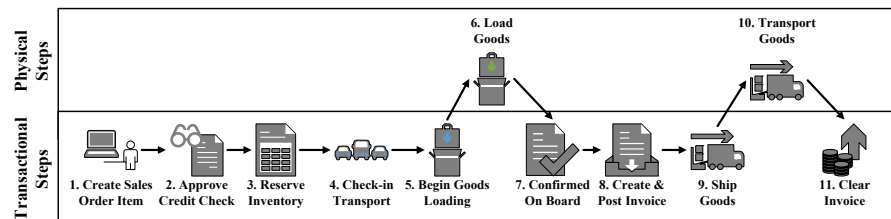


Figure 2. Order-to-cash process structure for case study

3.1. Deterministic Mode

Table 1 shows the normalized total revenues for each of the simulation modes when processing times are deterministic. The results show that performing no queue optimization yields the lowest total revenue. On the other hand, the greedy priority heuristic and the mathematical programming models produce a 57% and 105% improvement relative to the FIFO system, respectively. The superior performance of the MIP models (GP, D-RTN, and D-STN) relative to the greedy heuristic results from the fact that the assignment of order priorities in the queues is not based solely on the order prices. Instead, the MIP models provide a mathematical description of the system that accounts for processing times, backlog penalties, resource availabilities and due dates. The total revenue of the MIP models is 15% below the highest possible revenue (i.e., the perfect information model). As the optimization frequency increases, the gap in the total revenue with respect to the perfect information model is expected to decrease. The total revenue from the perfect information model serves to highlight the potential gains of having accurate demand forecasting. In terms of on-time order fulfillment, the MIP models process 5 orders on time, which is a 25% improvement with respect to the FIFO system. Despite its improved revenue relative to the FIFO mode, the heuristic does not improve the number of orders fulfilled on time. The results from the MIP models show that using mathematical optimization can improve not just the total revenue by prioritizing which orders to focus on, but can also increase the order fulfillment rate in the system.

Table 1. Results for each simulation mode with deterministic task durations

Simulation Mode	<i>FIFO</i>	<i>Heuristic</i>	<i>MIP</i>	<i>PI</i>
Cumulative Revenue	1.16	1.82	2.37	2.79
On-time Fulfillment	40%	40%	50%	80%

3.2. Stochastic Mode

Table 2 shows the results from 100 simulations in the stochastic environment. For comparison purposes, a unique random seed is used for each of the 100 runs with all simulation modes having the same seed in each run. The results show an average revenue increase on the order of 10% for the MIP models and 16% for the heuristic, relative to the FIFO system. The 95% confidence intervals correspond to a 4-6% uncertainty in the average revenues obtained. The number of simulations required is dictated by the allowed error tolerance. For a 10% tolerance, approximately 30 simulations would be required. For a 2% tolerance, the number of required simulations is in the 500-800 range. The MIP models suffer in environments with task durations that exhibit significant variance, as is the case with the uniform distributions used here. The deterministic models rely on the mean processing times, which results in a mismatch between the mathematical models and the simulations when the task durations are uncertain. However, as the variability in the processing time distributions becomes less pronounced, the MIP models should begin to outperform the Heuristic, as was observed in the deterministic case.

Table 2. Results for simulation models in a stochastic environment (95% confidence)

Sim. Mode	<i>FIFO</i>	<i>Heuristic</i>	<i>GP</i>	<i>D-RTN</i>	<i>D-STN</i>
Cum. Rev.	1.74 ± 0.10	2.01 ± 0.09	1.92 ± 0.10	1.89 ± 0.10	1.90 ± 0.11
On-time	47 ± 2%	51 ± 2%	50 ± 2%	51 ± 2%	50 ± 2%

4. Sensitivity Analysis

Besides serving as a testing ground for optimization strategies, the digital twin provides a framework upon which to identify bottlenecks and test design alternatives to mitigate these bottlenecks in a computationally inexpensive way. In the case study above, a major bottleneck is the lack of agents to perform transactional steps. A sensitivity study is performed in which the number of transactional agents is increased from 1 to 4. Table 3 shows that if the number of agents available for transactional events is doubled from the single agent in the original case study, the total revenue can be increased on average by 36% without any queue optimization. The on-time order fulfillment rate increases by 20 points. Further additions of agents do not result in increments that are statistically significant. It should be noted that performing the simulations in Julia is computationally inexpensive, with run times of less than 0.2 seconds for each simulation in this case.

Table 3. FIFO Results as the number of transactional agents increases (95% confidence)

Number of Agents	1	2	3	4
Cumulative Revenue	1.74 ± 0.10	2.37 ± 0.09	2.44 ± 0.08	2.44 ± 0.08
On-time Fulfillment	47 ± 2%	67 ± 2%	71 ± 3%	71 ± 3%

5. Conclusions

A simulation-optimization framework for supply chain transactional processes is presented. Optimization via mixed-integer linear programming is coupled in a feedback loop to a discrete event simulation of the process, which is modeled as a queueing network. The feedback loop assigns priorities to orders in each queue in the network. A greedy prioritization heuristic is compared against three mathematical optimization models in a case study. Two simulation modes are used, a deterministic and a stochastic mode. In the latter, the processing times for transactions are sampled from probability distributions. Under the deterministic scenario, the MIP models outperforms the heuristic by 30%, doubling the total revenue of the un-optimized FIFO system. Under the stochastic scenario with large uncertainty in the processing times, the heuristic slightly outperforms the MIP models, which are limited due to their deterministic representation. A sensitivity study is also presented to illustrate the value of using discrete event simulation to identify and mitigate bottlenecks in business process operations. Future work includes extending the MIP models to better respond to uncertainty in transaction times, while maintaining model tractability when executing the simulations. Reducing the uncertainty in the processing times by incorporating supply chain business rules is also expected to improve the performance of the mathematical programming models. Overall, the digital twin framework presented has promise in paving the way for industrial applications where business processes execute a very large number of transactions.

References

- Cai, Z., Li, X., Gupta, J.N.D., 2016. Heuristics for Provisioning Services to Workflows in XaaS Clouds. *IEEE Transactions on Services Computing* 9, 250–263.
- Hoensch, P., Schuller, D., Schulte, S., Hochreiner, C., Dustdar, S., 2016. Optimization of Complex Elastic Processes. *IEEE Transactions on Services Computing* 9, 700–713.
- Kondili, E., Pantelides, C.C., Sargent, R.W.H., 1993. A general algorithm for short-term scheduling of batch operations-I. MILP formulation. *Computers and Chemical Engineering* 17, 211–227.
- Li, X., Qian, L., Ruiz, R., 2018. Cloud Workflow Scheduling with Deadlines and Time Slot Availability. *IEEE Transactions on Services Computing* 11, 329–340.
- Méndez, C.A., Cerdá, J., 2003. An MILP Continuous-Time Framework for Short-Term Scheduling of Multipurpose Batch Processes Under Different Operation Strategies. *Optimization and Engineering* 4, 7–22.
- Pantelides, C.C., 1994. Unified frameworks for optimal process planning and scheduling, in: *Proceedings on the Second Conference on Foundations of Computer Aided Operations*. pp. 253–274.
- Perez, H.D., Amaran, S., Erisen, E., Wassick, J.M., Grossmann, I.E., 2021. Optimization of Extended Business Processes in Digital Supply Chains using Mathematical Programming. Submitted for publication.
- Perez, H.D., Amaran, S., Erisen, E., Wassick, J.M., Grossmann, I.E., 2020. Optimization of Business Transactional Processes in a Digital Supply Chain, in: *30th European Symposium on Computer Aided Chemical Engineering*. Elsevier, Milan, pp. 1159–1164.
- Pourhejazy, P., Kwon, O., 2016. The New Generation of Operations Research Methods in Supply Chain Optimization: A Review. *Sustainability* 8, 1033.

A Data-Driven Approach to Plan Electricity Production from Diesel Engines with Constrained Parameters

Piyush Kumar Kumawat, Nitin Dutt Chaturvedi

*Process Systems Engineering Laboratory, Department of Chemical and Biochemical
Engineering, Indian Institute of Technology Patna, Bihta, Patna, 801106, Bihar, India*

Abstract

Optimum energy utilization and constraint release of greenhouse gases are essential considerations in any production process. This challenge is generally observed in large energy extensive plants or processes involving fossil fuel combustion. Diesel engines are widely used to convert some of the chemical energy, contained by the diesel fuel, to mechanical energy through combustion. This mechanical energy can be converted into electrical energy using an alternator. Researchers claim that using alternate fuels having similar chemical properties or using foreign fuel as a fractional substance with diesel in these engines could result in a significant reduction in emissions or can alter the concentration of different pollutants. In this study, a methodology is suggested to create a feasible design space for operation with capped emission limits and energy consumption. A data-driven machine learning approach is implemented to classify feasible data points from the data set to create design space. Further, optimization is performed in the defined region to address an optimization problem associated with constrained energy and emission limits.

Keywords: Energy Consumption, Electricity Production, Machine Learning, Emissions;

1. Introduction

Energy consumption and emissions of harmful gases are of increasing concern in the world, and alternative options are continuously being explored having a reduced effect on the environment. Traditional electricity production process principally depends upon critically limited fossil fuels. The process of electricity generation by means of these fuels, several poisonous by-products adversely affect the conservation of natural eco-system.

Diesel engines are mainly used for decentralized power generation. The major drawback of these engines is the emanation of harmful gases. However, various studies claim that using other similar fuels or dual-fuel (using foreign fuel as partial substance) in these engines can significantly affect the emissions and can produce power from a few kilowatts up to several hundred kilowatts. R. Uma et al. (2004), stated that dual-fuel operation reduces NO_x and SO₂ emission without increasing particulate emission at the expense of an increase in carbon monoxide emission. Saleh (2008) studied the effect of variation in LPG composition on emissions and performance in a dual fuel (LPG-diesel) diesel engine. B.B. Sahoo et al. (2012) discussed the effect of H₂: CO ratio in syngas on the performance of a dual fuel diesel engine operation. H. Ambaria (2017) elucidated

performance and emission characteristics of a small diesel engine run in dual-fuel (diesel-biogas) mode. An appropriate trade-off between the operating diesel engine on different fuel mode will result in balanced emissions and energy consumption.

To address this challenge, support vector machines (SVM), a machine learning tool is adopted to classify feasible data from the data set in order to define a design space to operate the engine with constrained objective limits to produce electricity. The model can be used to classify the data from the embedded dataset and considered to be more reliable as it encompasses a data-driven approach. The aim of this paper is to obtain a design space for constrained energy consumption and different emission limits. The study is based on emission and energy consumption characteristics of the electricity generation system using a diesel engine. The proposed methodology is explained using literature example. The diesel engine can be operated in two different modes i.e. diesel alone and dual fuel (diesel-producer gas). In order to operate similar engines parallelly in two different modes, a design space is created using an SVM classifier. This will assist the planner to make decisions in a defined area and to optimize several objectives within the calculated predefined area using classifier algebraic equation.

2. Problem statement

Given a set of historical/simulated data with varying target load of diesel engine operated using different fuels. Each targeted load will have its respective specific energy consumption (SEC) and emission information of different gases as by-products. The energy consumption and emission of different harmful gases vary with different load conditions.

The aim of the model is to project the feasible region satisfying the capped limits of energy consumption and emissions of gases onto the space of production targets. Additionally, to generate an algebraic equation to attain objective-based optimization in the feasible region.

3. Methodology

To identify the feasible design space for the presented problem, support vector classification (SVC) technique is used. It attempts to find an optimal hyperplane or a sequence of hyperplanes in high dimension space that separates one class from those of the other class. The ideology for the classification is that the hyperplane has the largest distance from the nearest training data points in each class conducts a good separation and, in general, the larger the margin, lower the generalisation error of the classifier.

Given training vectors $x_i \in \mathbb{R}^p$, $i = 1, 2, \dots, k$ in two classes, and a vector $O \in \{-1, 1\}^k$, the goal is to find $w \in \mathbb{R}$ and $b \in \mathbb{R}$ such that the prediction given by $sign(w^T \phi(x) + b)$ is correct for most samples. Support Vector Classification (SVC) solves the following primal problem (Eqs. 1-3):

$$\frac{1}{2} w^T w + C \sum_{i=1}^k \xi_i ; s. t. \quad (1)$$

$$O_i (w^T \phi(x) + b) \geq 1 - \xi_i \quad (2)$$

$$\xi_i \geq 0, \forall i \quad (3)$$

Intuitively, to maximize the margin (by minimizing $\|w\|^2 = w^T w$), while incurring a penalty when a sample is misclassified or within the margin boundary. Ideally, the equation $O_i (w^T \phi(x) + b) \geq 1$ would be true for all data, which indicates a perfect

prediction. The problems are usually not always perfectly separable with a hyperplane, so some samples are allowed to be at a distance ' ϵ_i ' from their correct margin boundary.

The penalty term ' C ' controls the strength of this penalty, and as a result, acts as an inverse regularization parameter. The dual of the primal problem (Eqs. 4-6) is:

$$\frac{1}{2} \delta^T Q \delta - e^T \delta; \text{ s. t.} \tag{4}$$

$$O^T \delta = 0 \tag{5}$$

$$0 \leq \delta_i \leq C, \forall i \tag{6}$$

where ' e ' is the vector of all ones, and ' Q ' is a $(k \times k)$ positive semidefinite matrix $Q_{uv} = z_u z_v K(x_u, x_v)$, where $K(x_u, x_v) = \phi(x_u)^T \phi(x_v)$ is the kernel. The terms δ_i are called the dual coefficients, and they are upper-bounded by ' C '. This dual representation highlights the fact that training vectors are implicitly mapped into a higher dimensional space by the function ' ϕ '. Commonly used kernel functions are the linear kernel, polynomial kernel, radial basis function kernel (RBF) kernel with bandwidth and sigmoid kernel.

Once the optimization problem is solved, the output of decision function for a given sample ' x ' becomes:

$$\sum_{u \in SV} z_u \delta_i K(x_u, x) + b \tag{7}$$

The sum of the support vectors will be of the samples that lie within the margin because the dual coefficients are zero for the other samples. This classification using SVM is performed and practised in Python 3.0 using the library scikit-learn (Pedregosa, 2011) to process input data and to classify them on the basis of feasibility.

4. Case Study

To clarify the methodology, a case study is adapted from R. Uma et al. (2004). Biomass feedstocks are available in huge amounts in rural areas and could be an appropriate option for decentralized power generation using biomass gasifier-based systems. Producer gas from a biomass gasifier can be used either as a partial substitute for diesel in diesel engines or can be used alone in a gas engine. The mixture is known as dual fuel and considered as another fueling option to produce electricity. It is also assumed that the mixing ratio is fixed in dual fuel.

Table 1. Concentration of pollutants and SEC from diesel engine in diesel alone and dual fuel mode

Parameter	Load							
	10 kW		20 kW		30 kW		40 kW	
	Single	Dual	Single	Dual	Single	Dual	Single	Dual
CO (ppm)	175.6	656	219	669	284.3	710.6	320.6	922.3
SO ₂ (ppm)	4.234	1.167	5.4	1.234	7.367	1.667	9.834	2.167
SEC (MJ/kW)	22.8	34	15.5	18	14.0	15	13.1	16

The diesel engine system consists of a turbocharged, four-stroke I.C. engine coupled with an alternator. Table 2 describes the concentration of pollutants and SEC in diesel alone engine for four different load conditions. It is being observed that the emission of carbon monoxide (CO) in the dual fuel mode is greater than that of CO in the diesel mode alone. It indicates incomplete combustion results in the high CO content in the dual fuel exhaust, whereas, CO concentration improves at partial load condition. This suggests the need for a lower load limit for dual fuel operation. It can also be observed from the data that dual-fuel operation reduces SO₂ emissions. The SEC increases with decreasing load both in diesel alone and dual-fuel mode. This implies the considerable efficiency loss at the low-load condition. In dual fuel mode, SEC is greater than the diesel-only mode throughout the load condition tested. Increased SEC indicates a decrease in the efficiency of the dual-fuel mode, which could be caused by a decrease in the heating value of the producer gas-air mixture, a decrease in the gas pressure entering the air inlet and a decrease in the flame velocity. Due to contradictory performance of diesel engine for different parameters, it is suggested to operate diesel engine in both models parallelly with a cap over aggregate emission and energy consumption to generate a cumulative load.

A set of historical data could be useful and more accurate, however, for this study simulated data is used to train the classification model. Simulated data comprises combinations of the load to be produced by operating diesel engine in two different modes and is obtained by interpolating four load points (Table 1) using non-linear regression. A data set of 150 random production combination is generated and divided into two categories; feasible (-1) if the characteristics (emissions and energy consumption) are below the capped value for the selected combination output target, otherwise infeasible (1). Further, the input data is trained using SVC to classify the data on the basis of feasibility and to predict a feasible region in production space.

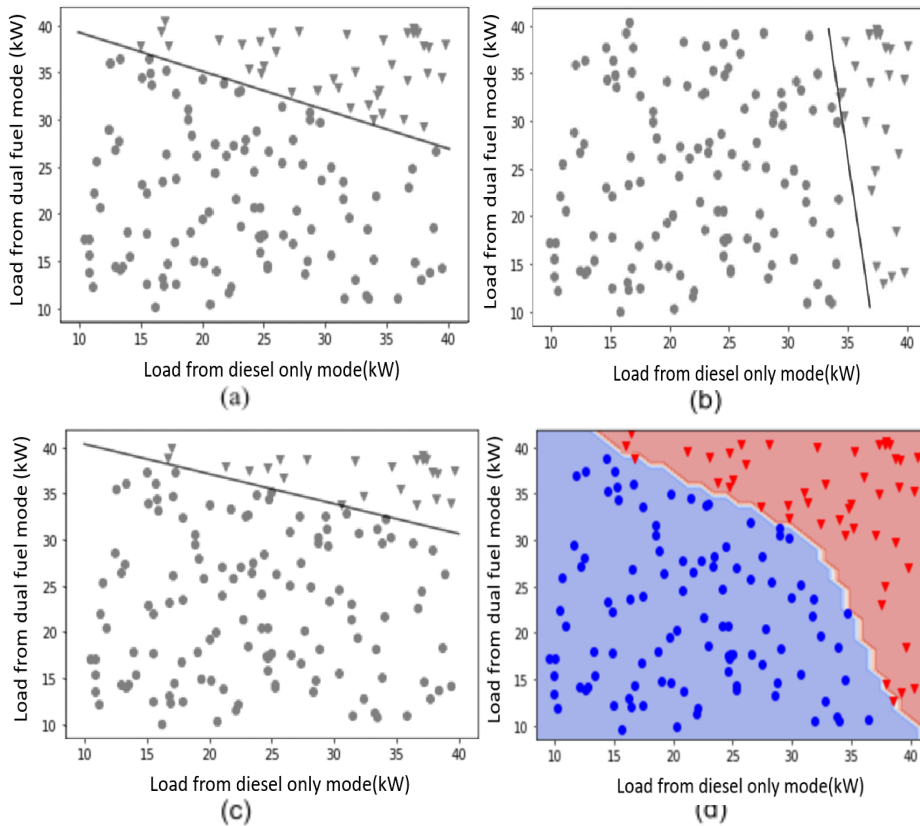


Figure 1. Performance of SVC in predicting the feasible region for (a) energy consumption, (b) CO, (c) SO₂, (d) overall (•: feasible points, ▼: infeasible points).

Figure 1 depicts the feasible region (area having feasible points) for production under specified limits. In Figure 1(a), a line is generated using SVC which separates the feasible data points. The criteria for the classification is that the total energy consumed by two engines operated on two different modes should be less than 850 MJ. Similarly, Figure 1(b) presents the feasible area lies on the left side of the generated line for SO₂ concentration to be less than 10 ppm. In the same way, data points are separated in a region with production combinations producing emission less than 1000 ppm (Figure 1(c)). Linear kernel function is used with SVC to classify input data because these data points are linearly separable in scatter plot. All the data points which resulted to be infeasible for any criteria are clustered and assigned as infeasible for overall performance. Further, a design space is obtained encompassing all the limitations over parameters, the generated curve for classifying feasible space is shown in Figure 1(d). It is generated using RBF kernel function in SVC, a hyperplane is created by mapping the data points to a higher dimension, in order to make non-linear classification possible. Table 2 summarizes the prediction result of additional 100 points generated by the model. Precision can be seen as a measure of a classifier's exactness; recall is a measure of the classifier's completeness. The F1 score is a weighted harmonic mean of precision and recalls such that the best score is 1.0 and the worst is 0, support is the number of actual occurrences of the class in the specified dataset.

Table 2. Results for the case study

	Precision	Recal	F1-score	Support
		l		
Feasible (-1)	0.90	1	0.95	65
Infeasible (1)	1	0.8	0.89	35
Accuracy			0.93	100
Macro avg	0.95	0.90	0.92	100
Weighted avg	0.94	0.93	0.93	100

This study will assist the planner to plan the generation of electricity within the specified limits or following the government norms. An algebraic equation can also be created using the fitted model parameters, i.e. support vectors, coefficients. Using this equation several objectives can be attained in the feasible design space. An example is performed to maximize diesel production can be solved using Eq. (8) as objective and Eqs. (9)-(10) as constraints.

$$P_D + P_{DF} \quad (8)$$

$$\sum_{u \in SV} z_u \delta_i \exp(-\gamma \|x_u - P\|) + b \leq 0 \quad (9)$$

$$P_D, P_{DF} \leq 40 \quad (10)$$

Here, P is the vector of production targets, $P = \{P_D, P_{DF}\}$ i.e. production using diesel-only mode and dual-fuel mode. γ be the parameter for RBF kernel function and x_u is the set of support vectors used to design classification boundary.

The resultant NLP model using support vectors from the trained model is solved using GAMS/CONOPT solver. The maximum aggregate production in the feasible region is approximately calculated to be 64.1 kW ($P_D = 32.16$ kW, $P_{DF} = 31.94$ kW). To generate this load, 893.94 MJ of energy is required with the resultant concentration of pollutants as 998 ppm CO and 9.7 ppm SO₂.

5. Conclusion

With an objective to produce electricity from diesel engine under capped emission and energy consumption, a data-driven approach is proposed. The methodology is based on classification of feasible data points using SVC and to create a design space for feasible production combinations. The proposed study is supported by a case study to use a mixture of producer gas and diesel as dual fuel. In order to operate two engines with two different fuel (diesel and dual-fuel) parallelly, a design space is created for production combination from each mode. Parameters like total CO, SO₂, and total energy consumption were considered and capped below a certain value. For individual parameters, the data was linearly separable and for the overall classification, RBF kernel was used to classify nonlinear data and are demonstrated graphically. Optimization is also performed in the feasible area to calculate maximum production satisfying capped parameters. In future work, the methodology can be implemented to other combinations for dual fuel and can be extended to scheduling and planning to satisfy the continuous electricity demand.

Acknowledgement

The authors would like to thank the Department of Science and Technology-Science and Engineering Research Board, India (DST-SERB) and Indian Institute of Technology, Patna for providing the research funding for this project under the grant no. ECR/2018/000197.

References

- B.B. Sahoo, N. Sahoo, U.K. Saha, 2012, Effect of H₂:CO ratio in syngas on the performance of a dual fuel diesel engine operation, *Applied Thermal Engineering*, 49 (2012) 139-146.
- H. Ambarita, 2017, Performance and emission characteristics of a small diesel engine run in dual-fuel (diesel-biogas) mode, *Case Studies in Thermal Engineering*, 10, 179–191.
- H.E. Saleh, 2008, Effect of variation in LPG composition on emissions and performance in a dual fuel diesel engine, *Fuel* 87, 3031–3039.
- Pedregosa F. et. al. 2011) Scikit-learn: machine learning in python. *Journal of Machine Learning Research* (12), 2825-2830.
- R.Uma, T.C. Kandpal, V.V.N. Kishore, 2004, Emission characteristics of an electricity generation system in diesel alone and dual fuel modes, *Biomass and Bioenergy*, 27, 195 – 203.

An Efficient Computational Strategy for Process Scheduling: a Case Study of the Omsk Lubricant Plant

Andrei Kostin^{**}, Yulia Mashchenkova^a, Vladimir Reflov^a, Alexander Tkachev^a

^a*LLC “Avtomatika-Service”, PJSC “Gazprom-neft, Konnogvardeyskiy Blvd. 4, Saint-Petersburg 1900000, Russia*

**kostin.am@gazprom-neft.ru*

Abstract

In this work, we propose a decomposition strategy to solve a large-scale MILP model that seeks to optimize the monthly schedule of the Omsk lubricant plant. The proposed algorithm consists of sequential solutions of two sub-problems. First, we solve a relaxed sub-problem where the integer variables representing the amount of packed lubes and their inventories are treated as positive continuous variables. The solution of the first approximate problem provides decisions on feedstock purchases and daily capacities of the continuous process units that produce oil components from vacuum residues and distillates. In the second stage, we solve the full-space sub-problem with the fixed values of the process capacities and utilization of the raw materials. By comparing with rigorous solution of the problem we demonstrate numerical efficiency of the method. The proposed algorithm provides near optimal solutions with significant reductions in CPU time as compared with the rigorous solution.

Keywords: scheduling, MILP, relaxation, refinery management.

1. Introduction

Plant-wide planning and scheduling systems play a crucial role in achieving greater profitability for oil companies. Since 2016 Gazprom-neft, a leader in the Russian oil business, has been developing own in-house optimization solutions for its refineries and lubricant plants located in Omsk and Moscow. Such planning and scheduling modules linking together will eventually compose a unified environment for supporting a decision-making process across the company’s supply chain.

In this work, we describe one of the modules -- the monthly scheduling problem of process activities within the Omsk lubricant plant. In a hierarchical planning process, the production plan is cascaded down to a more detailed production schedule (Hadidi et al., 2012). The aim of a scheduling phase is to construct the optimal sequence of production tasks, such that:

- all decisions made on the first planning stage (i.e., capacities of production units, monthly volumes of supplies, customers’ orders to deliver) should be executed with minimal deviations,
- product deliveries should be done according to the customer rating system,
- several large-capacity production units must be uniformly loaded during a month,

- blending operations must be sequenced with the minimal number of switches between incompatible oils to prevent unnecessary cleaning operations and generation of low-value oil mixtures,
- allocation of oil inventories to the tanks should be also done taking into account compatibility with preceding tank content.

Such scheduling task is formulated as a discrete-time MILP model which main equations are described in Section 2 of the paper. To reduce the computational burden associated with solution of the resulted large-scale MILP we develop a decomposition strategy, which consist of sequential solutions of two sub-problems. Details of our solution approach can be found in Section 3. Computational performance of the proposed algorithm is demonstrated by the case study of the Omsk lubricant plant in Section 4.

2. Discrete-time mathematical formulation

The scheduling model formulated as an MILP includes material balance equations, capacity constraints for every production unit and mixer, production plan limitations, and specific constraints setting preferable delivery days according to a product priority, uniform daily loads of certain production units and penalizing objective function when preparations or storing of incompatible products occur at the same mixer or tank.

The overall mass balance is determined for every stream s in a day p of a month. Thus, for every stream s its purchases $supply_{s,p}$ and initial inventories ($inventory_s$ for the first day of a month or $stock_{s,p-1}$ for other days) must be equal to the amount produced or consumed at production units ($process_{s,p}$) plus the amount delivered to customers ($demand_{s,p}$) and stored in tanks or warehouses ($stock_{s,p}$):

$$supply_{s,p} + stock_{s,p-1} + process_{s,p} - demand_{s,p} - stock_{s,p} = 0 \quad \forall s \quad (1)$$

$$supply_{s,p} + inventory_s + process_{s,p} - demand_{s,p} - stock_{s,p} = 0 \quad \forall s \quad (2)$$

The mass balance equations for the packed oils and greases (denoted as the subset of streams PS_s) are represented by Eqs. 3-4. The only difference from Eqs. 1-2 is that the amount of the packed lubes stored ($stockpack_{s,p}$) and the amount delivered to customers ($demandpack_{s,p}$) have integer nature since they are counted in pieces:

$$stockpack_{s,p-1} + process_{s,p} - demandpack_{s,p} - stockpack_{s,p} = 0 \quad (3)$$

$$inventory_s + process_{s,p} - demandpack_{s,p} - stockpack_{s,p} = 0 \quad \forall s \quad (4)$$

The amount of stream s produced or consumed in a plant ($process_{s,p}$) is calculated from the daily loads of operating modes m ($modeload_{m,p}$) and the material balance coefficients $processcoeff_{s,m}$ of a corresponding stream in an operating mode m :

$$process_{s,p} = \sum_m modeload_{m,p} processcoeff_{s,m} \quad \forall s, p \quad (5)$$

The material balance coefficients $processcoeff_{s,m}$ have a negative sign for the streams that are consumed in an operating mode m and a positive sign for the streams produced in this operating mode.

The daily loads of operating modes $modeload_{m,p}$ are limited by the minimum and maximum daily capacities ($capmin_m$ and $capmax_m$) associated with a corresponding mode m :

$$usecap_{m,p} capmin_m \leq modeload_{m,p} \leq usecap_{m,p} capmax_m \quad \forall m, p \quad (6)$$

The binary variable $usecap_{m,p}$ takes the value of 1 if a mode m is used in a day p and 0 otherwise.

The sum of all daily loads $modeload_{m,p}$ over all days p should not deviate from a monthly value $monthcap_m$ that was obtained during a planning phase:

$$\sum_p modeload_{m,p} = monthcap_m + Dcap_m^{plus} - Dcap_m^{minus} \quad \forall m \quad (7)$$

The slack variables $Dcap_m^{plus}$ and $Dcap_m^{minus}$ are incorporated in the model to guarantee finding feasible solution (Mendez at al., 2006). In Eqs. 8-10 similar slack variables are used to penalize deviations from the monthly supply and demand requirements:

$$\sum_p supply_{s,p} = monthsupply_s + Dsupply_s^{plus} - Dsupply_s^{minus} \quad \forall s \quad (8)$$

$$\sum_p demand_{s,p} = monthdemand_s + Ddemand_s^{plus} - Ddemand_s^{minus} \quad \forall s \quad (9)$$

$$\sum_p demandpack_{s,p} = monthdemand_s + Ddemand_s^{plus} - Ddemand_s^{minus} \quad (10)$$

The binary variable $usecap_{m,p}$ is used to restrict number of switches from one operating mode to another one assigned to the same process unit during the same day p at the same production unit u :

$$\sum_{m \in MU_u} usecap_{m,p} \leq switchmax_u \quad \forall u, p \quad (11)$$

where MU_u is a subset of operating modes m which can be executed at a process unit u . The amount of oil s that can be prepared in a mixer b ($VO_{b,s,p}$ in Eq. 12) is limited by the volume restrictions of a mixer b :

$$Vmin_b usemix_{b,s,p} \leq VO_{b,s,p} \leq Vmax_b usemix_{b,s,p} \quad \forall b, p, s \in BS_b \quad (12)$$

In Eq. 12 the subset BS_b represents the only oils s that can be prepared in a mixer b . The binary variable $usemix_{b,s,p}$ is 1 then an oil s is mixed in a blender b during a day p and 0 otherwise. It is used to count the number of blending operations at a mixer b during a day p and to restrict it by the maximum allowable number of blending operations NBO_b :

$$\sum_{s \in BS_b} usemix_{b,s,p} \leq NBO_b \quad \forall b, p \quad (13)$$

The binary variables $X_{b,s,s',p}$ denotes an occurrence of preparing two incompatible oils s and s' at the same blender b during period p . Another binary variable $Y_{b,s,s',p}$ is equal to 1 if a blender b prepares an oils s in a current period p and an incompatible oil s' during the next period $p+1$. Such logical variables are used to count all occurrence of blending of incompatible oils at the same blender and penalizes such schedule, because shifts to incompatible products causes additional cleaning operations and generation of low-value oils:

$$penalty^{clean} = penaltycost^{clean} \left(\sum_{b,s \neq s',p} X_{b,s,s',p} + \sum_{b,s \neq s',p < P} Y_{b,s,s',p} \right) \quad (14)$$

The aforementioned binary variables $X_{b,s,s',p}$ and $Y_{b,s,s',p}$ are derived from the binary variable $usemix_{b,s,p}$ using the procedures of disjunctive programming that translate logical expressions to inequality constraints (Grossmann, 2009). In the interest of brevity, detailed definitions of $X_{b,s,s',p}$ and $Y_{b,s,s',p}$ are not presented here.

Slack positive variables from Eqs. 7-10 are used to compute penalties for deviating from a production plan for a given month:

$$penalty^{supply} = penaltycost^{supply} \sum_s \left(Dsupply_s^{plus} + Dsupply_s^{minus} \right) \quad (15)$$

$$penalty^{cap} = penaltycost^{cap} \sum_m \left(Dcap_m^{plus} + Dcap_m^{minus} \right) \quad (16)$$

$$penalty^{demand} = penaltycost^{demand} \sum_s \left(Ddemand_s^{plus} + Ddemand_s^{mir} \right) \quad (17)$$

Every product item has its own priority according to ABC analysis (Gudehus and Kotzab, 2009). The products with a mark ‘‘A’’ should be prepared and delivered during the first days of three decads in which a month is divided. The ‘‘C’’ products should be delivered during the last days of 10-day periods, and ‘‘B’’ has no preferable days of delivery. It is modeled using the bonus coefficients $coeff_{s,p}^{bonus}$. For products s with ‘‘A’’ rating, parameter $coeff_{s,p}^{bonus}$ has the highest value for the first day p of every decad and monotonically decreases till the first day of the next 10-day interval. The bonus coefficients for the ‘‘C’’ lubes are constructed in a similar way with the only difference that they reach their maximum values at the last days of every decad. The values of $coeff_{s,p}^{bonus}$ for ‘‘B’’ items do not vary depending on a day p . Using these bonus coefficients, we compute the bonus part of the objective function:

$$bonus^{priority} = \sum_{s,p} \left(demand_{s,p} coeff_{s,p}^{bonus} + demandpack_{s,p} coeff_{s,p}^{bonus} \right) \quad (18)$$

The objective function obj to be maximized is formulated as follows:

$$obj = bonus^{priority} - penalty^{clean} - penalty^{demand} - penalty^{su} \quad (19)$$

3. Solution strategy

The MILP model includes variables of different nature. The complexity of this formulation is mainly given by the integer variable $demandpack_{s,p}$ representing amount of packed oils and greases. Our solution strategy exploits the fact that the relaxation of

these integer variables does not greatly affect the production and supply schedule. The overall MILP model can therefore be decomposed into a pair of the smaller sub-problems that are solved sequentially.

We first construct the approximate sub-model by relaxation of the integer variables $demandpack_{s,p}$ and $stockpack_{s,p}$. After solving the first sub-problem, we fix the values of the binary variables defining the optimal production schedule, i.e. the binary variable $usecap_{mp}$ denoting selection of a particular operational mode m in a day p and the binary variable $usemix_{s,b,p}$ representing a usage of a blender b for preparing a product s . Finally we solve the second sub-model in which variables $demandpack_{s,p}$ and $stockpack_{s,p}$ have their original integer nature and the values of the binary variables $usecap_{mp}$ and $usemixer_{sb}$ are fixed at the values from the solution of the first relaxed sub-model.

4. Case study

The structure of the Omsk lubricant plant is depicted in Figure 1. The plant involves a continuous processing of vacuum residues and distillates into oil components and batch operations of blending and packing. Vacuum residues from the Omsk refinery are separated from asphaltenes at three propane-deasphalting units. Two aromatics extraction units remove aromatic hydrocarbons from deasphalted oils and refinery oils cuts, thus producing dearomatized oils called “raffinates”. Finally, base oils are generated at 3 dewaxing units after removing dissolved long chain paraffines. By blending these base oils with various additives, the desired properties of lubricants are achieved. Special mixers and tanks conduct blending operations, and final products are delivered to customers by either rail cars or trucks. The Omsk lubricant plant has its own 13 packing lines that can fill oils and greases into 43 different types of packings. Most of these packings are manufactured using own extruders and injection molding machines.

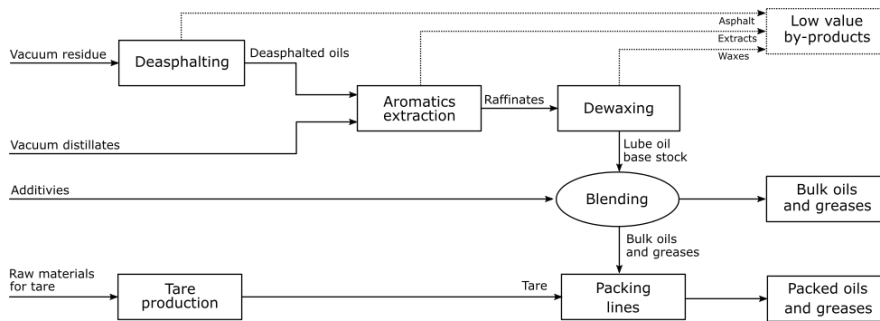


Figure 1. Structure of the Omsk lubricant plant

The mathematical model also includes several tank farms and storages. In total, there are 86 process units with 1404 different operating modes, 31 mixers, 216 tanks and 12 storages.

The model was written in GAMS 32.2 (McCarl et al., 2016) and solved with the MILP solver CPLEX 12 on a desktop PC with Intel Core I5-6500, 3.2 GHz and 8Gb of RAM. The dataset of process coefficients, supply and demand constraints used in the case study corresponds to the actual plant conditions at the beginning of August 2020. The original full-space model (i.e., without any relaxation of the integer variables) comprises 2,754,281 equations with 2,899,767 variables. 405,950 of the model variables are

discrete. The relaxed sub-problem, which we solve first according to the proposed strategy, contains 58,962 discrete variables less than the original problem.

To compare the numerical performance of our relaxation strategy and the computational results of solving the full-space model, we first attempted to solve the original scheduling MILP problem allowing the solver to use as much time, CPU and RAM resources as needed to close the optimality gap to zero. Unfortunately, after 2h and 19 minutes a memory issue interrupted the solving process at the integer solution with objective equal to -586,341,137 and the best bound of -562,994,661. The optimality gap was therefore 3.98%. The best bound we found during solving the full-space model is later used to assess the tightness of the solution of the proposed relaxation method.

Both sub-problems involved in our method were successfully solved without numerical issues in 506 seconds. Solving of the first approximate sub-problem with the relaxed integer variables took 343 seconds whereas the second sub-problem with the fixed production schedule was solved in 163 seconds provided the integer solution with the objective of -565,120,963. Thus, the proposed algorithm was able to find the better optimal solution with the optimality gap of 0.38% significantly faster than the rigorous solution of the original full-space MILP model.

5. Conclusions

In this work, we describe how the relaxation of integer variables can be exploited to effectively solve the large scheduling problems. The proposed algorithm involves sequential solution of the approximate sub-model constructed by relaxing integer variables representing delivery decisions and the second sub-problem that receives a production schedule from the solution of the first sub-model. The capabilities of the proposed method were shown through a case study of the Omsk lubricant plant. The proposed algorithm provides near optimal solutions with significant reductions in CPU time as compared with the rigorous solution.

References

- T. Gudenus and H. Kotzab. *Comprehensive logistics*. Springer, 2009.
- I. Grossmann. Overview of generalized disjunctive programming, EWO seminar, 2009.
- L. Hadidi and U. Al-Turki, Integrated models in production planning and scheduling, maintenance and quality: a review, *International Journal of Industrial and Systems Engineering*, 10(1), 2012.
- B. McCarl, A. Meeraus, P. Eijk, M. Bussieck, S. Dirkse, F. Nelissen. *McCarl Expanded GAMS User Guide, Version 24.6*. GAMS Development Corporation, 2016
- C.A. Mendez, I.E. Grossmann, I. Harjunkoski and P. Kabore. A simultaneous optimization approach for off-line blending and scheduling of oil-refinery operations, *Computers & Chemical Engineering*, 30(4), 2006

Centralised vs. decentralised production and storage: optimal design of a decarbonised hydrogen supply chain with multiple end uses

Renato Luise^{a,b}, Annabelle Brisse^a, Catherine Azzaro-Pantel^b

^a *European Institute for Energy Research, Emmy-Noether Straße 11, Karlsruhe 76131, Germany*

^b *Université Toulouse, Laboratoire de Génie Chimique, CNRS, INPT, UPS, 6 Allée Emile Monso, 31400 Toulouse, France*

Abstract

This paper addresses the design of the hydrogen supply chain (HSC) with a focus on the its decentralization degree. For this purpose, the superstructure used for HSC modelling is split into six echelons, i.e., primary energy source, H₂ production, conditioning and centralized storage, distribution, conditioning and decentralized storage, final usage. The originality of this work is also the diversification of H₂ production, conditioning and distribution methods for the H₂ supply both for the industry and mobility sectors. The problem is formulated in a mono-optimization framework, where the total cost of the HSC is minimized. The model involves a multi-period (2020-2040) mixed-integer linear programming (MILP) formulation that is implemented in the General Algebraic Modelling System (GAMS) environment, using CPLEX as a solver. The optimization results highlight some significant applications of the developed methodological framework. The assessment of the decentralization degree of the system, the diversification of the final H₂ demand, the share between liquid and compressed gaseous form are some of the Key Performance Indicators (KPIs) that can characterize a scenario. The scalability and replicability of the tool indeed allow to make forecasts on different geographical and temporal scales.

Keywords: hydrogen supply chain, multi-period optimization, mixed integer linear programming, GAMS, decentralization.

1. Introduction

The deployment of a hydrogen infrastructure constitutes one of the most challenging tasks due to the complexity and large spectrum of energy sources and technological bricks involved (e.g. water electrolysis technologies, liquid or gaseous conditioning and transport of H₂, Hydrogen Refuelling Station (HRS) options) serving a multitude of end uses. In this paper, the structure of the energy supply chain is studied with an innovative vision based on “think global, act local”, rather than “think local, act local” (Devine-Wright, 2013), emphasizing the decentralization degree of the hydrogen supply chain (HSC).

The pioneering work carried out by (Almansoori and Shah, 2006) has been followed by almost 15 years of active research in the field (Talebian et al., 2019). This work enlarges the scope of previous studies by considering the possibility for one production site to simultaneously satisfy different hydrogen demands coming from industry and mobility

via the development of a superstructure that can embed a large set of technological bricks for the construction of the HSC.

Following this introduction, the problem formulation and the key points of the methodology will be presented. A case-study will then be analysed with the most relevant results. Finally, the conclusions and critical analysis of the model will be highlighted, thus opening perspectives in terms of methodology and scenario definition.

The main text can start here. Next paragraphs should start with heading “Els-1storderheading”

2. Text

2.1. Main assumptions

A great challenge for modern day energy systems models is to capture the high degree of variability and complexity that exists in energy systems and represent the involved technologies. Several formulations have been developed including simulation vs. optimisation, top-down vs. bottom-up (e.g. MARKAL) (Pfenninger et al., 2014). The core innovation of this work lies in the modelling of the HSC with a spatial and temporal resolution adapted for addressing:

- the discretization of a territory into grids, each grid being characterized by demands targeting industry and mobility markets;
- the simultaneous satisfaction of both demands;
- various technical solutions for H₂ logistics, i.e. gaseous and liquid, corresponding to the final conditions of usage in industry or mobility (i.e., pressure level);
- interconnection between grids in order to valorise the energy produced from one grid to another grid (primary energy sources);

The skeleton of the model was taken from the works developed by (De-León Almaraz et al., 2014; Guilarte and Azzaro-Pantel, 2020). This work aims to show all the different solutions and scenarios that can be implemented in the extended tool. No investment cost has been considered for the HRS on the industrial site which suggests that the structure already exists. The advanced model focuses mainly on two KPIs, on the one hand, the diversification of the energy source, status and utilization of hydrogen, and on the other hand, the centralization degree of the production and storage facilities. The replicability and scalability of the tool make it flexible and usable for any type of scenario.

The problem is presented as a multi-period mono-optimization model in which the time slots are divided into 4 periods “t” of 5 years “n” each. The investment costs related to H₂ technologies are reduced at the beginning of each period due to a learning rate factor. Furthermore, the investment costs are involved only at the beginning of each period (i.e. end of year 0,5,10,15). The parameters of the years belonging to the same period, including the H₂ demand have been kept constant, except for the costs, which are discounted as explained in Eq. (3). The model aims finally to minimize the yearly cost of HSC.

In line with the majority of the works analysed in the literature, this formulation of the HSC design problem is based on a mixed-integer-linear-programming (MILP) approach. The model is developed in the General Algebraic Modelling System (GAMS) environment, using the CPLEX solver in a computer with Intel® Core™ i5-6200U processor with CPU @ 2.30GHz and 16GB RAM. A comprehensive explanation of all the parameters, formulas and constraints can be found in (De-León Almaraz et al., 2014; Guilarte and Azzaro-Pantel, 2020).

The design of the HSC is represented by a 6-echelon superstructure: primary energy source, hydrogen production, conditioning and centralized storage, distribution, conditioning and decentralized storage, final usage. A schematic view (Fig. 1) gives additional information on the energy sources and technologies that have been analysed in this work. Two options for H₂ production have been considered, i.e., electrolysis powered by renewable electricity sources (RES) (or by the grid if RES is not sufficient) or by Steam Methane Reforming (SMR).

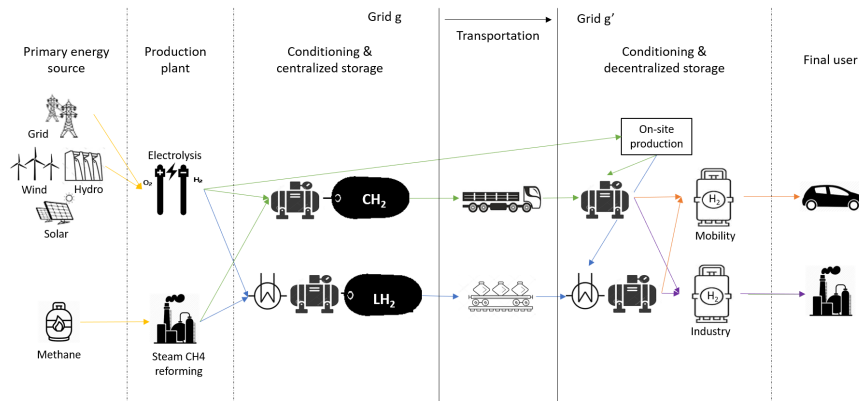


Figure 1: Definition of the H₂ supply chain superstructure

The discretization of the territory allows each area (the so-called grid) to be represented with its own characteristics (e.g. availability of RES, different H₂ demands, etc.). Furthermore, all possible exchanges between grids, both internal and external to the system have been considered (Fig. 2): primary energy sources can be imported/exported from any grid and/or bought from a source outside the system; the H₂ produced, including the one produced on-site, can be transported to any grid where it will be conditioned and made ready for end-use (i.e. mobility or industry).

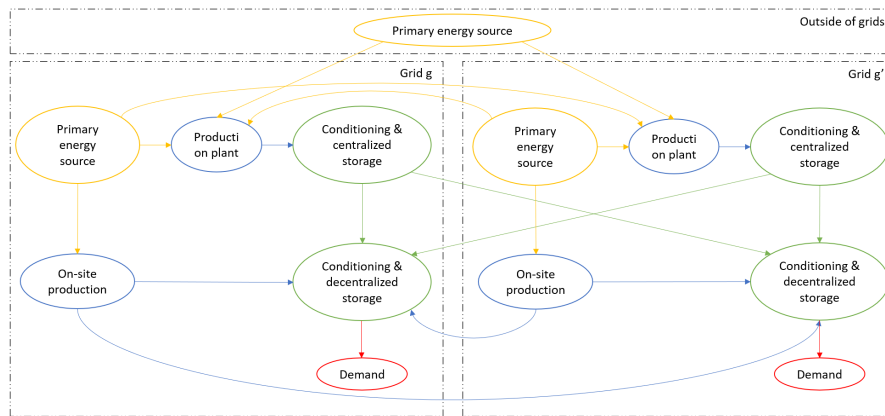


Figure 2: Interconnection between the grids of the HSC superstructure

2.2. H₂ and primary energy sources mass balance

The new mass balance constraint is shown in Eq. (1): the total H₂ production (PT) in all "g" grids, independently of its form "i" (i.e. compressed or liquid) and of the size of the plant "j" (i.e. centralized or on-site production), must satisfy the different H₂ demands (DT) "k" (i.e. mobility and industry) in each "g" grid.

$$\sum_{ijg} PT_{ijg} = \sum_{kg} DT_{kg} \quad \forall i, j, g, k \quad (1)$$

The calculation of the primary energy costs (ESC, €/day) is composed of two parts. The former considers the costs of the energy present in the territory including the transportation costs, whereas the latter considers the costs due to the purchase of energy from an off-grid source.

2.3. Economic parameters

The investment cost of the facilities (FCC) (see Eq. (2)) includes the facility capital costs (i.e. production "p" (PCC), storage "s" (SCC) and fuelling station "f" (FSCC)) respectively for the installed number of each facility (i.e. number of production plants (NP), storages (NS) and refuelling stations (NFS)). All these factors are then divided by a learning factor (LearnF) to consider cost reduction of some technologies with mass production and market penetration (e.g. electrolyzers). The facility capital costs and the learning factor are defined at the beginning of each period "t".

$$FCC_t = \frac{1}{LearnF_t} \sum_{ig} \left(\sum_{pj} PCC_{pij} NP_{pijgt} + \sum_{sj} SCC_{sij} NS_{sijgt} + \sum_{fj} FSCC_{fkj} NFS_{fkjgt} \right) \quad \forall t \quad (2)$$

2.4. Actualization factor

All the economic variables (e.g. PCC) are discounted to year "n" see Eq. (3), where "i" is the discount rate.

$$PCC_n = PCC_0 \frac{1}{(1+i)^n} \quad \forall n \quad (3)$$

2.5. Optimization variable

The total cost (TCTotal) as expressed in Eq. (4), representing the costs of the whole HSC over the timeframe analysed, is the objective function to be minimized. The TCTotal consists of the sum of investment costs for facilities and transportation means (TCC) and the sum of the total yearly operating costs (TOC) that are discounted (see Eq. (3)).

$$TCTotal = \sum_t (FCC_t + TCC_t) + \sum_n TOC_n \quad (4)$$

3. Case study

As a first case study for the validation of the model, the deployment of the HSC in the Occitania region in France was analysed considering the availability of energy sources and the H₂ demand as presented in (De-León Almaraz et al., 2014). The Occitania territory was discretized into 8 grids and in this scenario, the H₂ demand for industry and mobility has been set as an equal level corresponding to 23.5 tons per day (t/d) of H₂ at

the beginning of the study and progressively increasing up to roughly 600 t/d at the fourth period.

4. Results

Two important KPIs of the HSC design have been further analysed, i.e., the decentralization degree and the share between compressed and liquid H₂. Due to the increasing demand for H₂ in the Occitania Region, the model has to make trade-offs between centralized and decentralized H₂ production, mitigating the cost between production and distribution logistics. Starting in period 1 with 100% of H₂ produced on a decentralized basis, more than 50% of the production becomes centralized at the end of period 4 as shown in Figure 3 (a). This can be explained by the scaling effect on the investment cost of H₂ production technologies that reduces the H₂ production cost. Nevertheless, this should be in contradiction with the reduced CO₂ footprint of H₂ that was not yet considered in the first scenario.

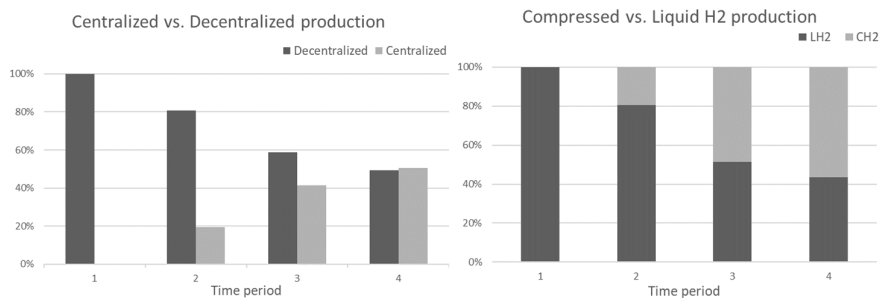


Figure 3: KPI analysis, centralized vs. decentralized (a) and CH₂ (compressed gas) vs. LH₂ (cryogenic liquid) (b)

Another result of the model concerns the mode of H₂ distribution, i.e. either as compressed gas (CH₂) or as cryogenic liquid (LH₂) (Figure 3 (b)). Despite the higher energy density of LH₂ allowing the distribution of larger volumes of H₂, the investment and operating costs for liquefaction and cryogenic tankers remain prohibitive over the fourth periods. CH₂ distribution is preferred because of the good share between centralized and decentralized production which reduces distribution distances. However, other issues have to be considered such as the availability of space for the storage of gaseous H₂ compared liquid H₂. It is interesting highlight that the solution with liquid H₂ is not considered for a centralized solution because of the currently high investment and operating costs of LH₂ vessels. Nevertheless, the bulk and footprint of the solution cannot be ignored. Road transportation is mainly dependent on the available H₂. In the first 2 periods, only tanker trucks are considered, while in the last period there is a higher number of tube trailers (14), even though the amount of H₂ that the tanker trucks (8) transport is higher (71% LH₂ vs. 29% CH₂).

HSC investment costs supporting the development of H₂ in the Occitania region integrate the contribution of H₂ production plants, centralized storage with conditioning and transportation modes, decentralized storage and refuelling stations as shown in Figure 4 (a). Around 54% of the investment costs of the HSC are related to centralized storage, conditioning and distribution due the development of centralized production plants over the fourth time periods. Concerning operating and maintenance costs, the cost of electricity and natural gas needed for the production of H₂ represents the main operating costs (76.7%) as shown in Figure 4 (b).

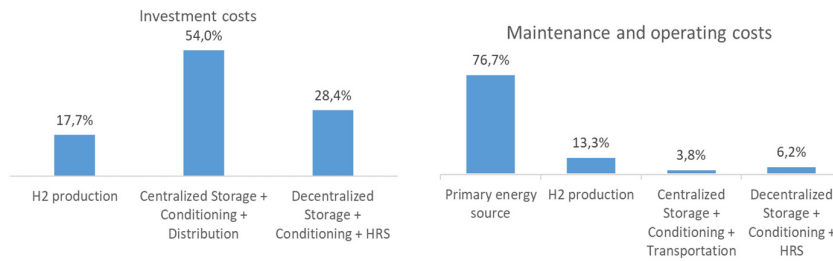


Figure 4: Share of the investment (a) and maintenance costs (b)

5. Conclusions

The model developed for the optimization of hydrogen supply chains was used for the case of the Occitania region. The optimal design of HSC was implemented on the basis of total cost minimization and some KPIs such as the decentralization degree of the production and storage plants or the share between liquid and compressed hydrogen. The model is scalable and replicable to forecast and understand where, how and in which conditions the H₂ market will evolve. The methodology can be further improved with a multi-objective analysis, including the minimization of CO₂ emissions to assess the trade-offs between low-carbon technologies (electrolysis powered by renewable electricity or SMR with CCS, Carbon Capture and Sequestration) as well as the scale-up effect of large-size electrolyzers which is a key driver for cost reduction. Furthermore, few technology solutions such as H₂ pipelines or hydrogen importing from another region, could also be implemented.

References

- Almansoori, A., Shah, N., 2006. Design and Operation of a Future Hydrogen Supply Chain. *Chemical Engineering Research and Design* 84. <https://doi.org/10.1205/cherd.05193>
- De-León Almaraz, S., Azzaro-Pantel, C., Montastruc, L., Domenech, S., 2014. Hydrogen supply chain optimization for deployment scenarios in the Midi-Pyrénées region, France. *International Journal of Hydrogen Energy* 39. <https://doi.org/10.1016/j.ijhydene.2014.05.165>
- Devine-Wright, P., 2013. Think global, act local? The relevance of place attachments and place identities in a climate changed world. *Global Environmental Change* 23. <https://doi.org/10.1016/j.gloenvcha.2012.08.003>
- Guilarte, E.C., Azzaro-Pantel, C., 2020. A Methodological Design Framework for Hybrid “Power-to-Methane” and “Power-to-Hydrogen” Supply Chains: application to Occitania Region, France. <https://doi.org/10.1016/B978-0-12-823377-1.50114-2>
- Pfenninger, S., Hawkes, A., Keirstead, J., 2014. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*. <https://doi.org/10.1016/j.rser.2014.02.003>
- Taleblian, H., Herrera, O.E., Mérida, W., 2019. Spatial and temporal optimization of hydrogen fuel supply chain for light duty passenger vehicles in British Columbia. *International Journal of Hydrogen Energy* 44. <https://doi.org/10.1016/j.ijhydene.2019.07.218>

Optimal Design of the Reverse Logistics Network of Empty Agrochemical Containers

Glenda N. Yossen^{a,b}, Gabriela P. Henning^{a,b,*}

^a *INTEC (UNL, CONICET), Ruta Nacional N° 168, Km. 0, Santa Fe, 3000, Argentina*

^b *FIQ (UNL), Santiago del Estero 2829, Santa Fe, 3000, Argentina*

ghenning@intec.unl.edu.ar

Abstract

This contribution addresses the optimal design and operation of a reverse supply chain of empty agrochemical containers. The problem is tackled by means of a multi-period MILP model, which relies on a superstructure that includes all the potential/existing collection centers and possible/existing plastic treatment plants, as well as the farms. Its decisions involve selecting the new facilities, their sizes, and installation periods, along with the operation of the facilities and the material flows among the various nodes in each planning period. A realistic case study has been solved under several scenarios in order to analyze the model scalability and to test its response to changing conditions.

Keywords: Reverse Logistics, Supply Chain Design, Empty Agrochemical Containers.

1. Introduction

In recent years, reverse logistics has become a field of importance for many organizations due to growing environmental concerns, legislation, corporate social responsibility and sustainable competitiveness (Agrawal et al., 2015). The importance increases significantly when the products handled are harmful to the health of people, animals, and the environment, and must be treated in a special way prior to disposal or recovery. Reverse logistics refers to the process of planning, implementing and controlling backward flows of raw materials, in-process inventory, packaging and finished goods, from a manufacturing, distribution or use point, to a point of recovery or point of proper disposal (Dekker et al., 2004). The definition of reverse logistics has been changing over time and widening its scope with the interest of researchers. The number of scientific publications related to the field has steadily grown during last decade (Govindan et al., 2014), showing the attention that the subject entails. However, regarding the reverse supply chains of empty agrochemical containers, few contributions have been found. Recently, Sorichetti et al. (2018) reported a preliminary approach to address the design of the reverse network of empty agrochemical containers in a region of Buenos Aires, Argentina. Various aspects of relevance are not taken into account, such as the operation and installation costs of facilities and the transportation from farms to collection centers. The optimal configuration of the reverse supply chain network of empty agrochemical containers is a challenging problem that deserves attention and will be addressed in the remaining sections of this contribution, which is organized as follows. The mathematical model is presented in Section 2. In Section 3, a realistic case-study, which allowed testing the formulation, is described. Computational

results associated with various scenarios are also presented. The contribution ends with concluding remarks and a summary of future work.

2. Mathematical Programming Formulation

The multi-period mixed-integer linear programming (MILP) model presented in this section allows making the following decisions: (i) the number and location of collection centers – intermediate points for collecting containers –, as well as plastic treatment plants to be built, (ii) the time period in which each facility will be installed and the ones in which it will operate, (iii) the capacities of the various storage and processing facilities, (iv) for each time period, which are the flows of materials between farms, which generate the containers, and the collection/treatment points, as well as the flows between collection centers and plastic treatment plants, with the goal of minimizing the total actual discounted cost of the network.

2.1. Sets/indices

F/f	farms
J/j	collection centers
P/p	plastic treatment plants
J^i/P^i	subset of collection centers/treatment plants that are already installed at the beginning of the planning horizon
M/m	set of available sizes of collection centers
N/n	set of available sizes of treatment plants
$T/t, t'$	time periods

2.2. Parameters

g_{ft}	number of containers to be generated by farm f during period t
$d_{ff}/d_{fp}/d_{jp}$	distance between the different types of nodes
k_m^{min}/k_n^{min}	minimum capacity of a collection center/treatment plant of size m/n
k_m^{max}/k_n^{max}	maximum capacity of a collection center/treatment plant of size m/n
nt^j / nt^p	number of time periods required by a new collection center/treatment plant to start operating from the period it is installed
$c^{fj}/c^{fp}/c^{jp}$	transportation cost per distance unit between the different nodes
ir	interest rate
fc_m/fc_n	fixed cost of operation of a collection center/treatment plant of size m/n
vc	variable production cost of plastic treatment plants
cc_m^j/cc_n^p	construction cost of a collection center/treatment plant of size m/n , including the cost of infrastructure and in the case of treatment plants, the equipment cost.
cap	loading capacity of the trucks that travel between collection centers and treatment plants
nf_{ft}	number of trips required to transport the total generation of containers of farm f during period t
$nmax$	$\max(nt^j ; nt^p)$ number of periods that must be considered previous to the beginning of the planning horizon for the installation of facilities

2.3. Binary variables

Y_{jt} / Y_{pt}	equal to one if collection center j /treatment plant p operates in period t
I_{jmt} / I_{pnt}	equal to one if collection center j of size m /treatment plant p of size n is installed in period t

2.4. Positive variables

$S_{fjt}/S_{fpt}/S_{jpt}$ number of containers sent between nodes of different type in period t

2.5. Equations

$$g_{ft} = \sum_{\forall j \in J} S_{fjt} + \sum_{\forall p \in P} S_{fpt} \quad \forall f \in F, \forall t \in T, t > n_{\max} \quad (1)$$

$$\sum_{\forall f \in F} S_{fjt} = \sum_{\forall p \in P} S_{jpt} \quad \forall j \in J, \forall t \in T, t > n_{\max} \quad (2)$$

$$\sum_{\forall m \in M} \sum_{\forall t' \in T, t'+nt' \leq t} I_{jmt'} \geq Y_{jt} \quad \forall j \in J, j \notin J^i, \forall t \in T \quad (3)$$

$$\sum_{\forall n \in N} \sum_{\forall t' \in T, t'+nt' \leq t} I_{pnt'} \geq Y_{pt} \quad \forall p \in P, p \notin P^i, \forall t \in T \quad (4)$$

$$\sum_{\forall m \in M} \sum_{\forall t \in T} I_{jmt} \leq 1 \quad \forall j \in J \quad (5)$$

$$\sum_{\forall n \in N} \sum_{\forall t \in T} I_{pnt} \leq 1 \quad \forall p \in P \quad (6)$$

$$\sum_{\forall f \in F} S_{fjt} \leq k_{m=\max}^{\max} Y_{jt} \quad \forall j \in J, \forall t \in T, t > n_{\max} \quad (7)$$

$$\sum_{\forall f \in F} S_{fpt} + \sum_{\forall j \in J} S_{jpt} \leq k_{n=\max}^{\max} Y_{pt} \quad \forall p \in P, \forall t \in T, t > n_{\max} \quad (8)$$

$$\sum_{\forall f \in F} S_{fjt} \leq \sum_{\forall m \in M} \sum_{\forall t' \in T, t'+nt' \leq t} k_m^{\max} I_{jmt'} \quad \forall j \in J, \forall t \in T, t > n_{\max} \quad (9)$$

$$\sum_{\forall f \in F} S_{fpt} + \sum_{\forall j \in J} S_{jpt} \leq \sum_{\forall n \in N} \sum_{\forall t' \in T, t'+nt' \leq t} k_n^{\max} I_{pnt'} \quad \forall p \in P, \forall t \in T, t > n_{\max} \quad (10)$$

$$\sum_{\forall f \in F} S_{fjt} \geq k_{m=1}^{\min} Y_{jt} + \sum_{\forall m \in M} \sum_{\forall t' \in T, t'+nt' \leq t} (k_m^{\min} - k_{m=1}^{\min})(Y_{jt} + I_{jmt'} - 1) \quad \forall j \in J, \forall t \in T, t > n_{\max} \quad (11)$$

$$\sum_{\forall f \in F} S_{fpt} + \sum_{\forall j \in J} S_{jpt} \geq k_{n=1}^{\min} Y_{pt} + \sum_{\forall n \in N} \sum_{\forall t' \in T, t'+nt' \leq t} (k_n^{\min} - k_{n=1}^{\min})(Y_{pt} + I_{pnt'} - 1) \quad \forall p \in P, \forall t \in T, t > n_{\max} \quad (12)$$

$$\begin{aligned}
TC_t = & \sum_{\forall m \in M} \sum_{\forall t' \in T, t'+nt^j \leq t} fc_m(Y_{jt} + I_{jmt'} - 1) + \sum_{\forall j \in J} \sum_{\forall m \in M} cc_m^j I_{jmt} + \\
& \sum_{\forall p \in P} \sum_{\forall n \in N} cc_n^p I_{pnt} + \sum_{\forall n \in N} \sum_{\forall t' \in T, t'+nt^p \leq t} fc_n(Y_{pt} + I_{pnt'} - 1) + \\
& \sum_{\forall p \in P} vc \left(\sum_{\forall f \in F} S_{fp} + \sum_{\forall j \in J} S_{jpt} \right) + \sum_{\forall f \in F} \sum_{\forall j \in J} c^{fj} 2d_{fj} \frac{S_{fjt}}{g_{ft}} nf_{ft} + \\
& \sum_{\forall f \in F} \sum_{\forall p \in P} c^{fp} 2d_{fp} \frac{S_{fpt}}{g_{ft}} nf_{ft} + \sum_{\forall j \in J} \sum_{\forall p \in P} c^{jp} 2d_{jp} 1.2 \frac{S_{jpt}}{cap} \quad \forall t \in T
\end{aligned} \tag{13}$$

$$\min \sum_t \frac{TC_t}{(1+ir)^{t-1}} \tag{14}$$

Eq. (1) prescribes that all the containers generated by a farm in a certain period, must be sent to a collection center and/or a treatment plant in the same planning period. In turn, Eq. (2) states that all the containers that are received by a collection center in a given period must be sent to a plastic treatment plant in the same period. Expression (3) assures that a collection center can operate in a certain period only if it has been installed in a previous one. Expression (4) forces treatment plants to obey the same restriction. Expressions (5) and (6) force every facility to be installed at most once with a unique size. According to Expressions (7) to (10), the number of containers sent to a given collection center/treatment plant in each period should not overpass the maximum storage capacity corresponding to the installed size. Following analogous ideas, the minimum capacity of each collection center or plastic treatment plant to be installed is given by expressions (11) and (12), respectively. Eq. (13) comprises the different costs involved in the design and operation of the logistics network in each period of the planning horizon. In order to determine how much future projected cash flows are worth at present, these cost components are affected by an annual interest rate, according to the multi-period total discounted cost formula given by Expression (14).

3. Computational Results

The model has been implemented in GAMS 23.6, running on an Intel Core i8 computer, with 3.4 GHz, 8 GB of RAM, and using CPLEX as a solver. As shown in Table 1, a base case (Case 1) has been initially tackled. It takes into account a super-structure composed of 300 farms, 20 potential collection centers, 8 potential treatment plants and no existing facilities that are already installed. Two different capacity modules for collection centers and treatment plants have been allowed for. The planning horizon considers 5 years of operation, plus an initial year for the installation of the first facilities. Thus, during the initial year the system does not operate, but the required facilities are being built. This base case has been extended to test the scalability of the model, analyzing different dimension scenarios presented in Table 1. In a second stage, the sensitivity of the formulation to parameter disturbances has been assessed, analyzing the scenarios presented in Table 2, which consider the same super-structure as the base case. All the model instances were run to optimality and the obtained results are presented in Table 3.

Table 1. Dimensionality of some of the analyzed scenarios.

	Farms	Collection centers/ available sizes	Treatment Plants/available sizes	Periods
Case 1	300	20/2	8/2	5
Case 2	450	20/2	8/2	5
Case 3	600	20/2	8/2	5
Case 4	300	20/2	8/2	10
Case 5	300	20/3	8/3	5

Table 2. Scenarios associated with perturbations of the base case parameters.

	Disturbed parameter	Disturbance type
Case 6	Containers generation	Gradual decrease from period 3 onwards
Case 7	Transportation costs	10% increase
Case 8	Transportation costs	20% increase
Case 9	Installation costs	10% increase
Case 10	Installation costs	20% increase

Table 3. Computational results corresponding to the analyzed scenarios.

	Total variables	Binary variables	Total constraints	Objective Function	CPUs	Centers/ plants
Case 1	54,447	504	2,799	1,513,663.31	717	7/2
Case 2	77,647	504	3,549	2,288,839.52	1,670	3/4
Case 3	102,847	504	4,299	2,833,478.15	3,459	3/4
Case 4	96,152	870	5,444	2,307,106.48	487*	9/4
Case 5	52,615	672	2,799	1,238,574.16	1,880	7/1
Case 6	54,447	504	2,799	1,478,809.46	1,301	7/2
Case 7	54,447	504	2,799	1,542,416.85	608	7/2
Case 8	54,447	504	2,799	1,571,170.39	410	7/2
Case 9	54,447	504	2,799	1,618,496.74	883	7/2
Case 10	54,447	504	2,799	1,723,330.16	1,208	7/2

*Base Case solution taken for the five initial periods

As shown in Table 3, the optimal solution of Case 1 comprises seven collection centers and two treatment plants, which are all installed before the start of the planning horizon and operate during all the periods. Two collection centers are of the smallest size, allowing a maximum capacity of 5,000 containers per period, while the remaining five double this capacity. Regarding treatment plants, one small and one large are chosen, with a maximum capacity of 60,000 and 100,000 empty containers, respectively. The solutions of Cases 2 and 3, which correspond to an increase in the number of farms, comprise 3 collection centers and 4 treatment plants. All the treatment plants associated with the solution of Case 2 have the smallest size, while two small and two large correspond to Case 3. In both situations, two collection centers are installed with the highest capacity. A comparison of these results reveals that the effect of managing a greater number of farms is to increase the investment in treatment plants, allowing more direct shipments to them. Case 4 takes into account the same superstructure as Case 1, but doubles the length of the planning horizon. This scenario has been tackled with two different solution approaches. The first one considers the solution obtained in Case 1 as given, and extends the horizon to evaluate potential new facilities. In this case, the optimal solution has been found in only 487 s. The second approach does not resort to any initial network and the 10 periods are taken altogether. A solution with a gap of

10% has been reached in 2,530 s. The solution obtained is the same in both cases; it comprises 12 facilities: 9 collection centers and 3 treatment plants. A comparison of the solutions corresponding to Cases 1 and 4 shows that the effect of having a longer planning horizon is to increase the investment in the installation of facilities. The aim is to reduce transportation costs along a longer planning period, that otherwise would have been much higher. Whereas Case 1 solution has installation (59.15%) and transportation (19%) costs as the main components, the solution of Case 4 has the following cost participation: installation (61.42%) and transportation (11%). Case 5 also considers the same superstructure as Case 1, but allows three different capacity modules for collection centers and treatment plants, instead of two. The obtained solution includes 7 collection centers and only one treatment plant of the largest size, with a capacity of 200,000 containers per period. With respect to the sizes of collection centers, only one has the smallest capacity and four are installed with the largest size (20,000 containers per period). These results indicate that, when larger facility sizes are allowed, economies of scale are taken advantage of. Case 6 considers a reduction in the number of empty containers starting from period 3. The solution reports the same initial network as Case 1, but starting periods 4 and 5 two collection centers stop operating in order to cope with the reduction of empty containers to be handled. As seen in Table 2, Cases 7 to 10 take into account different increments in the installation and operation costs with respect to Case 1. Since the increases are not significant, the obtained networks remain the same as in Case 1 and only augment the value of their objective function. Finally, Table 3 reveals that the proposed model always exhibits a satisfactory computational behavior.

4. Conclusions and future work

An MILP model for a multi-period reverse logistics network design problem has been presented, taking into account several realistic features. The model has been applied to the case of empty agrochemical containers and has been tested under various scenarios, exhibiting a very good computational behavior. The proposed approach can be easily adapted to address other recyclable waste collection systems (electronic and electric devices, household plastic waste, cardboard and paper, etc.) having two tiers: collection centers and disposal/recycling facilities. The incorporation of environmental and social objectives, in addition to the economic ones, will be considered in the future. Further testing will be done to better study scalability and sensitivity issues.

Acknowledgements. The authors acknowledge the support given by CONICET under project PIP 11220170101131CO.

References

- Agrawal S., Singh R., Murtaza Q., 2015, A literature review and perspectives in reverse logistics. *Resources, Conservation and Recycling*, 97, 76-92.
- Dekker R, Fleischmann M., Inderfurth K., Van Wassenhove L., 2004, *Reverse Logistics Quantitative Models for Closed-Loop Supply Chains*, Springer Verlag, Berlin.
- Govindan K., Soleimani H., Kannan D., 2015, Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operations Research*, 240 (3), 603-626.
- Sorichetti A., Mammini L., Savoretti A., Bandoni A., 2018, Gestión de envases vacíos de agroquímicos, dos propuestas para el Sudoeste Bonaerense. *Simposio Argentino de Informática Industrial e Investigación Operativa*, pp. 55-70. Buenos Aires, Argentina.

Designing a Rectification Strategy for Managing Disruptions in LNG Supply Chain

S.V.Ramanan,^a and Rajagopalan Srinivasan^{a,*}

^a*Indian Institute of Technology Madras, Chennai 600036, India*

raj@iitm.ac.in

Abstract

Global LNG trade continues to grow owing to increasing energy demand, especially from the Asia-Pacific region. However, the trade of LNG has inherent vulnerabilities that can eventually manifest as disruptions. Typical consequences of disruptions include excess or shortfall of inventory, and plant shutdown, all of which can significantly impact profit. In this paper, we seek to develop a systematic methodology that can be used by an LNG receiving terminal to mitigate the effects of disruptions. We develop an agent-based dynamic model of the LNG supply chain and use it to systematically design operational interventions that can ameliorate the adverse consequences of disruptions. As a case study, we focus on a sudden decrease in demand. The disruption is managed through fire sales of a product. We demonstrate that a key parameter in this rectification strategy, the fire sale price, can be systematically determined using the proposed agent-based model.

Keywords: LNG supply chain, Agent-based model, disruptions management, rectification strategies.

1. Introduction

Supply Chains are highly interlinked networks that involve a set of companies collaborating with one another for the timely delivery of products to end customers. Given the complexity of interactions, supply chains are susceptible to various disruptions that can inhibit the routine and systematic operations, leading to overall inefficient performance. There is always a trade-off to consider while designing for the reliable and unhindered operation of any supply chain (SC). This trade-off can either choose the right location of sourcing raw-materials, the best mode of transport for safe and faster delivery to the buyer, and the understanding of the internal operational processes to deliver the right products at the right place and at the right time. While nearby sourcing can always be a better option considering shorter lead times, resource constraints largely inhibit the development of proximate SC networks. Most present-day supply chains are globalized mainly due to increased customer demands, high competitiveness, improved communication, and reduced transit times leading to a highly interlinked and complex SC. With the increased dependence on imports, the SC becomes vulnerable to disruptions.

The LNG SC involves NG extraction and liquefaction at an export terminal, transportation through LNG carriers, and receiving and regasification at an import terminal. The receiving terminal is attached to a regasification plant that sells natural

gas (called Regasified LNG) in large quantities through conventional gas pipelines or in small quantities (called Retail-LNG) through road or rail-based virtual pipelines to various customers. Some receiving terminals also cater to maritime customers like FSRUs and LNG cooling services; this product is called Reload-LNG. The material flow diagram in a typical LNG supply chain is depicted in Figure 1.

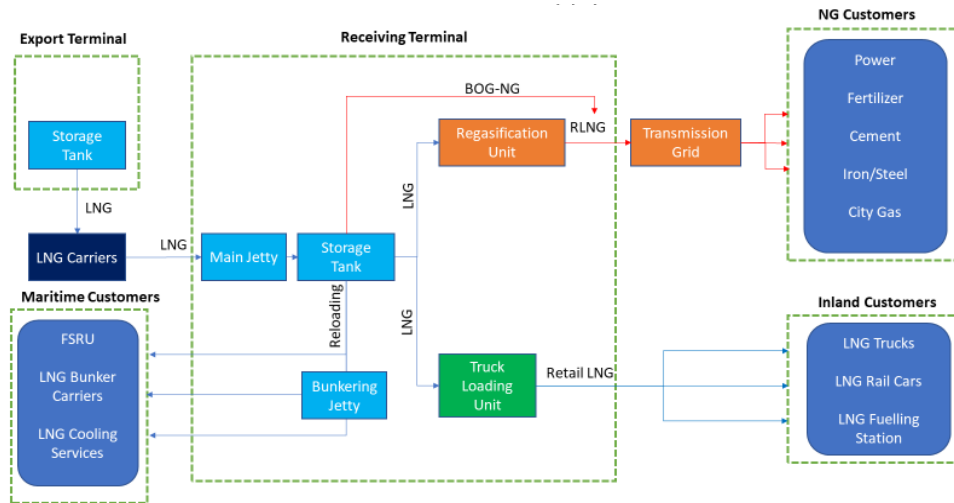


Figure 1. Overview of LNG Supply Chain

1.1. LNG Supply Chain Disruptions

Each section of the supply chain is exposed to various disruptions. Disruptions in LNG supply can happen due to political/economic situations, unplanned maintenance, or strikes at the liquefaction terminal. Marine transport of LNG can be affected by strait closure, marine accidents, piracy, and other exogenous factors like hurricanes, tsunami, etc., leading to either delay or loss of shipment. Unexpected orders or order cancellations by customers can also perturb the normal operation of the supply chain.

While an exhaustive description of real-world instances of LNG supply chain disruptions is beyond the scope of this work, the present glut in the global supply of LNG and the fall in demand due to the covid-19 pandemic is a good instance that illustrates the potential effects of supply-demand disruptions. As an example, Indian LNG import terminals were forced to declare force majeure as their customers shut down operations owing to a nation-wide lockdown (Jaganathan, 2020). Such abnormal situations can lead to major financial implications and motivate a systematic approach to disruption management.

1.2. Literature review

Significant research has been conducted on techniques for managing disruptions in supply chains. Most of them involve qualitative reasoning, relying on expert opinions to manage disruptions (Urciuoli et al., 2014). Only a few quantitative studies have reported quantitative methods such as simulation and modeling. An integrated procedure to handle disruptions in a supply chain was developed by (Behdani et al., 2012), consisting of separate pre-disruption and post-disruption considerations. The pre-disruption steps involve risk identification, risk quantification, risk evaluation & treatment, and risk

monitoring. The post disruption steps involve disruption detection, reaction, recovery, and learning. A survey of the literature suggests that there is very limited research on post-disruption recovery procedures. Also, managing the ripple effect arising from low frequency-high impact disruptions has attracted limited attention. Based on the modeling paradigm, approaches for modeling disruptions can be classified into optimization-based, simulation-based, control-theory based (Dolgui et al., 2018) schemes. Given the limited research towards managing disruptions in LNG supply chains, in this paper, we propose an agent-based dynamic simulation for disruption management in the LNG SC.

We consider a major LNG receiving terminal in India to test the effectiveness of our approach. We take up a case study of an unexpected decrease in RLNG demand and study its impact on the receiving terminal. We devise rectification strategies to counteract the effect of the disruption and as well provide insights on a key decision parameter involving the recovery action that will lead to economic benefits for the focal company. The rest of the paper is organized as follows. Section 2 gives an overview of the agent-based modeling technique. An application of the developed model is analyzed with an illustrative case study in section 3. Finally, section 4 provides some concluding remarks and future areas of research.

2. Agent-Based Model of LNG Supply Chain

Agent-based modeling is a widely used technique used to capture the various entities and their interactions in the supply chain. An agent is described as any entity that can make decisions, initiate or perform activities, manage resources, and interact with other agents in the system. Each agent has a set of behavioral attributes. The heterogeneous nature of the supply chain makes agent-based modeling a suitable simulation paradigm to capture the various dynamics inherent in the system.

We have developed a dynamic simulation model of the LNG SC. Eight distinct agents interact, exchange information and make decisions leading to the overall operations of the supply chain. The various entities in the supply chain, including LNG supplier, LNG carrier, receiving terminal, and customers, are modeled as agents that interact with one another. The various internal departments of the LNG receiving terminal, which is the entity in focus in this study, including the sales, operations, procurement, and jetty, are also modeled as rational, self-interested decision-making agents. The sales agent keeps track of customer demands, the operations agent oversees the regasification process, the procurement agent makes decisions regarding when to procure LNG and the requisite quantity, and the jetty agent oversees the cargo unloading. We implemented the agent-based model using AnyLogic® Simulation Modelling Tool, version 8.6.0 (AnyLogic, 2020)

We performed the simulations for the normal operation of the LNG supply chain using the parameters listed in Table 1. The overall simulation horizon is 365 days with each time tick corresponding to an hour of operations. The demands for the various products are stochastic and drawn from normal distributions with a standard deviation of 5% of the nominal value. Various KPIs, including profit, customer satisfaction index, capacity utilization, number of unloaded cargos, and demurrage period, are calculated for each run of the simulation. Due to the stochastic nature of the demand and disruptions, we use Monte Carlo simulation to arrive at statistically significant results. All results are

reported as average over 1000 runs of the simulation. The computational time for conducting a set of Monte Carlo simulations is 60 minutes. For the normal operation, the receiving terminal makes a profit of \$798 million, with 242 cargos procured annually.

Table 1. LNG SC Simulation Parameters

Parameter	Value
Simulation Horizon (days)	365
Demand Cycle (days)	7
LNG Procurement Cost (\$/MMBtu)	4.16
Nominal Product Demand (10^3 m ³) [RLNG Retail-LNG Reload-LNG]	[647 6 70]
Product Sales Price (\$/MMBtu) [RLNG Retail-LNG Reload-LNG]	[5.0 10.0 5.43]

The LNG supply chain is subjected to a number of disruptions. We used the HAZOP analysis based approach proposed by (Adhitya et al., 2009) to identify the possible sources of deviation from normal operations, their causes, and consequences. The study requires the development of a comprehensive supply chain flow diagram and analyzing the activities of each agent. Each activity involves performing a sequence of tasks that can be subject to deviations. Each deviation is generated using a set of guidewords. For example, the sales department collates product demand from various customers. Any deviation from the nominal demand towards a "high" or "low" (guidewords) value can have consequences on the receiving terminal leading to "low" or "high" storage tank inventory. This can further propagate both upstream and downstream of the supply chain, as illustrated below. We then analyze the existing safeguards available that can help the SC recover from the deviation (for example, safety stocks). If the safeguards are inadequate, we propose additional mitigative actions to manage the disruption, as illustrated next using a demand disruption case study.

3. Demand Disruption Case Study

Consider the case of a significant decrease in demand of RLNG by 75% due to exogenous factors affecting the customers' businesses for a long duration (average of 6 months). This leads to challenges to the supply chain, specifically related to LNG inventory. The existing inventory management policy of the terminal uses the traditional fixed reorder level (ROL) based replenishment strategy. This policy works well during normal conditions where the demand is close to the nominal values. Here, we seek to evaluate its suitability during demand decrease disruptions. Various disruption scenarios are generated stochastically with both the time of disruption initiation and the total duration of disruption sampled from a uniform distribution (Behdani et al., 2019). During each disrupted case, when the RLNG demand reduces drastically, there is a supply-demand mismatch due to the underlying ROL-based LNG procurement policy. As the quantity of LNG procured is aligned with the nominal situation, it is much greater than the actual customer demand during the period of the disruption.

Consequently, there is an issue of inadequate ullage space in the receiving terminal. This causes a ripple effect when LNG carriers arrive at the terminal for unloading and are forced to wait until the inventory in the LNG storage tanks falls to a level sufficient for the unloading of the cargo. This leads to demurrage charges and a continuous domino effect on all subsequent LNG carriers.

Our simulation studies of this scenario reveal that the decrease in demand would lead to a reduction of profit for the terminal to \$536 million (a decrease of \$260 million) due to the demand decrease. However, the loss is exacerbated by \$2.2 million due to demurrage charges. While little can be done about the exogeneous demand disruption, the impact due to demurrage can be tackled by triggering a suitable rectification strategy when the terminal becomes aware of the demand reduction. Among the various options, we chose the RLNG fire sale as a possible rectification strategy. In this rectification strategy, the receiving terminal sells the excess LNG inventory at a very low price (called the fire-sale price) to facilitate smooth unloading of arriving cargo. To evaluate the economic implications of this, we performed a simulation-based study by varying the key decision parameter, i.e., the RLNG fire-sale price, to identify the point of economic break-even. If the fire-sale of the RLNG is possible at any price above this break-even price, the terminal will make a higher profit compared to the unrectified case. By evaluating the entire supply chain performance for various RLNG fire-sale prices, we arrive at the break-even plot depicted in Figure 2. It can be concluded that at any RLNG fire sale price above \$3.52/MMBtu, the terminal will be economically better-off. For instance, at an RLNG fire-sale price of \$3.75/MMBtu, the demurrage dropped drastically by 98%, as shown in Table 2. The mean profit with the rectification increases by \$4 million as compared to the unmanaged case.

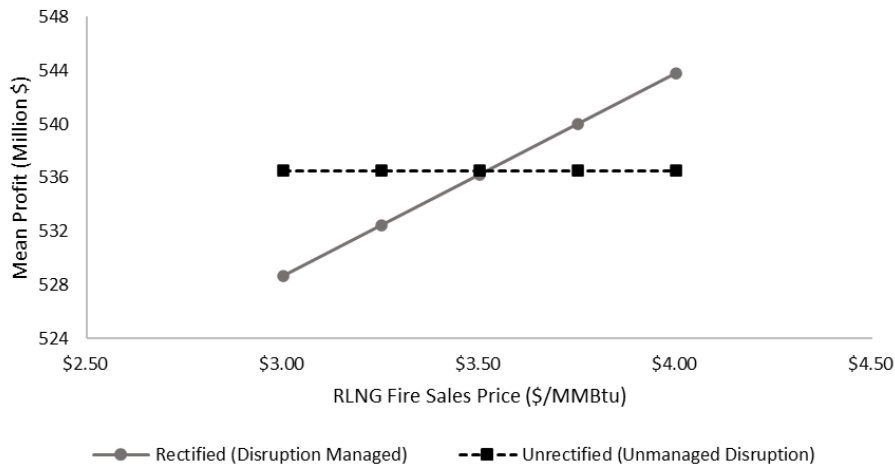


Figure 2. Effect of RLNG Fire Sales price on mean profit

Table 2. Effect of rectification strategy on performance

KPIs	Nominal Operation	Disrupted operation	
		Unrectified	Rectified
Overall Profit (Million \$)	798 (± 7.27)	536 (± 13.76)	540 (± 10.7)
Demurrage (h)	-	658 (± 209.05)	11 (± 20.1)
RLNG Fire Sales Quantity (10^3 m^3)	-	-	628 (± 45)

4. Concluding Remarks

This paper addresses the management of disruptions in LNG supply chains using an agent-based dynamic simulation technique. The model can be used to design rectification strategies that can be triggered when a disruption occurs in the supply chain. Various disruptions that can occur in the LNG supply chain are demand decrease, demand increase, reduction in production, and transportation delay. We have focused on an RLNG demand decrease case study in this paper. The consequences of this disruption are increased waiting time of LNG carriers, which can be ameliorated by RLNG fire sales to prevent a tank ullage shortfall. This recovery strategy requires the determination of a break-even price for the fire sales of RLNG. If the LNG terminal is able to find new customers willing to buy RLNG above this break-even price, the resulting profit would be higher compared to the unrectified case. Our simulations show that, against a normal RLNG sales price of \$5/MMBtu, fire-sales at any price higher than \$ 3.52/MMBtu would be advantageous to the LNG terminal. Similar insights for designing rectification strategies for other disruptions can be obtained using the proposed agent-based model. This work did not consider various operational aspects of the LNG regasification plant tanks, such as boil-off gas generation, which can play an important role in terminal operations. These would be incorporated in the future. Further, we plan to include strategic mitigative actions like additional buffer stock, capacity expansion, and multiple sourcing in our future development.

References

- Adhitya, A., Srinivasan, R., Karimi, I.A., 2009. Supply chain risk identification using a HAZOP-based approach. *AIChE J.* 55, 1447–1463. <https://doi.org/10.1002/aic.11764>
- AnyLogic: Simulation Modeling Software Tools & Solutions for Business <https://www.anylogic.com/>
- Behdani, B., Adhitya, A., Lukszo, Z., Srinivasan, R., 2012. How to Handle Disruptions in Supply Chains – An Integrated Framework and a Review of Literature. *SSRN Journal*. <https://doi.org/10.2139/ssrn.2114201>
- Behdani, B., Lukszo, Z., Srinivasan, R., 2019. Agent-oriented simulation framework for handling disruptions in chemical supply chains. *Computers & Chemical Engineering*, 2017 Edition of the European Symposium on Computer-Aided Process Engineering (ESCAPE-27) 122, 306–325. <https://doi.org/10.1016/j.compchemeng.2018.09.027>
- Berle, Ø., Norstad, I., Asbjørnslett, B.E., 2013. Optimization, risk assessment and resilience in LNG transportation systems. *Supp Chain Mngmnt* 18, 253–264. <https://doi.org/10.1108/SCM-03-2012-0109>

- Dolgui, A., Ivanov, D., Sokolov, B., 2018. Ripple effect in the supply chain: an analysis and recent literature. *International Journal of Production Research* 56, 414–430. <https://doi.org/10.1080/00207543.2017.1387680>
- Jaganathan, N.V.& J., 2020. Indian LNG importers issue force majeure notices as gas demand slumps. *Business Standard India*.
- Urciuoli, L., Mohanty, S., Hintsa, J., Gerine Boekesteijn, E., 2014. The resilience of energy supply chains: a multiple case study approach on oil and gas supply chains to Europe. *Supply Chain Management* 19, 46–63. <https://doi.org/10.1108/SCM-09-2012-0307>

Statistical Decision-Theoretic Risk Management for Planning Renewable Energy Pathways

Houd Al-Obaidli,^a Rajesh Govindan,^b Tareq Al-Ansari^{a,b,*}

^a*Division of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

^b*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

talansari@hbku.edu.qa

Abstract

The world has witnessed an increase in the level of investments for renewable energy technologies in the past decade mainly owing to the pressure for mitigating greenhouse gas (GHG) emissions, whilst addressing the issue of climate change. According to recent statistics, global investments edged up by 2 % in 2019 to \$301.7 billion, taking the value of cumulative investments since 2004 to \$3.5 trillion, which has been attributed to the falling costs of solar and wind technologies. With the commissioning of additional capacities from renewable energy sources each year, there is a growing need for managing the associated risks and uncertainties from the perspective of different stakeholders throughout the planning, development and operational phases. Renewable energy sources entail considerable technological and financial risks exposure, depending on the location where the technologies are implemented, and thus needs to be managed using techniques that would provide both the quantification of risks and optimal decisions that lead to risk mitigation. The objective of the proposed research is to develop a probabilistic framework which broadly includes: (a) statistical modelling of financial risks - such as variability of revenue due to electricity price, demand fluctuations, generation costs, or other market conditions; and (b) evaluating options that maximise the stakeholders' utility/reward functions, or minimise risks, for a given technology mix. This research demonstrates the implementation of binomial lattice model in real options analysis (ROA) for the valuation of investments on diversified energy portfolios. The framework is applied to analyse the impact of risks and uncertainties on capital budgeting decisions relating to project size (expand or contract); project life and timing (initiation, deferment or abandon); and project operation (flexibility in the technology mix) for scenarios involving large-scale deployment of renewable energy sources.

Keywords: Binomial lattice model, Cogeneration, Real options analysis, Renewable energy, Risk management.

1. Introduction

Investment in the renewable energy sector continues to increase globally for the fulfilment of the Paris Agreement obligation of sustaining global temperature rise to well below 2 degrees Celsius. The amount of renewable energy provision in 2019 was the highest recorded at 184 GW of which 118 GW was through solar energy and 61 GW through wind turbines. This can primarily be attributed to the ever decreasing levelised

cost of energy (LCOE) for both solar and wind technologies due to continuous technology improvements and economies of scale. The current LCOE levels for solar and wind energies are lower by 83 % and 50 %, when compared to the last decade (FS-UNEP, 2020). Prior to committing to a project, investors need to evaluate all the associated risks in terms of risk probability and impact. Some risks are intolerable leading to a no investment decision and some to the other extreme are very low impact and are ignored entirely. However, the majority of risks fall in the middle where a formal evaluation is required to determine what is acceptable and what is not. Risk mitigation measures are usually implemented to reduce the risk level or to eliminate it altogether at a premium that need to be factored into the project overall cost. In the energy domain, fluctuations of fossil fuel, carbon and electricity market prices cause undesirable uncertainty that impacts future investment predictions and, in general, a portfolio of production technologies are mixed together to mitigate the risk and lower the overall impact. For large power projects, project economics and investment decisions are typically based on metrics such as the net present value (NPV) and the discounted cash flow (DCF). These metrics are well suited and have been long established for the conventional power projects. Recently, real options methods gained wide spread and are considered more appropriate for renewable energy projects where project flexibilities such as flexible project designs and flexible investment timings are possible during the planning phase.

Several studies were conducted previously to either address the weaknesses of traditional project assessment tools using basic ROA methodology or propose improved ROA techniques. Yang and Blyth (2007) developed a computational model to address the impacts of climate change policy including energy and carbon price uncertainties on power projects investment using ROA. Stochastic variables were used to map volatile energy and carbon prices and optimisation techniques were developed to search for the best time to invest. The model was applied on twelve case studies all with satisfactory results. The aim of the model is to assist policy stakeholders in the process of decision-making and critical comparison of the effects of different policy designs, and the impacts of climate change policy and market uncertainty on power project investments. Martinez-Cesena and Mutale (2011) employed an advanced real options methodology for the renewable energy generation projects. The scheme was implemented on a simplified hydropower plant as a case study in which the results demonstrated higher expected profits compared to conventional investment assessment tools. Santos *et al.* (2014) proposed a case study for a small hydropower plant where a comparison between traditional project evaluation methods and ROA was conducted. The analysis involved the development of real options and investment decision flexibilities based on ROA using the binomial decision tree. The results show that when the traditional methods support an immediate investment decision, ROA provides more flexibility to the investor in terms of project re-evaluation or redefinition of the project strategy. Pringles *et al.* (2020) applied ROA to determine the value of investment deferral while awaiting favourable market conditions in terms of system acquisition costs. Some of the conducive flexibilities that were valued are site location/relocation, network accessibility and regulatory policies. A stochastic valuation framework based on the least square Monte Carlo method was developed and applied to a proposed photovoltaics (PV) facility where site relocation and alternative tariff schemes were the more attracting options.

Based on the review above, the majority of the analyses focus on policy changes or uncertainties in fuel and carbon prices. Also, integrated power and water production facilities, co- and polygeneration in general, don't receive the same attention as single output facilities which are usually less complex with a lower risk profile. The aim of this paper is to develop a statistical based risk framework for the economic evaluation of integrated electricity and water cogeneration facilities to identify potential future expansion options where electricity market prices are uncertain in an unregulated electricity market.

2. Methodology

The integrated facility configurations are based on proposals previously made by Al-Obaidli *et al.* (2019, 2020) for a utility-scale production of electric power and freshwater. A base configuration and 4 expansion options are selected and presented in Figure 1. The base configuration is composed of: natural gas combined-cycle gas turbine (CCGT) for power generation and seawater desalination based on the multistage flash (MSF) process. Two additional technologies are considered: PV and seawater reverse-osmosis (SWRO). The base plant capacities are 500 MW for power and 100 km³d⁻¹ for water and the expansion capacities are the same except for the PV farm which has a higher capacity (1385 MW) due to the lower plant efficiency and the requirement to maintain the same level of energy throughput for an impartial assessment of the 4 options.

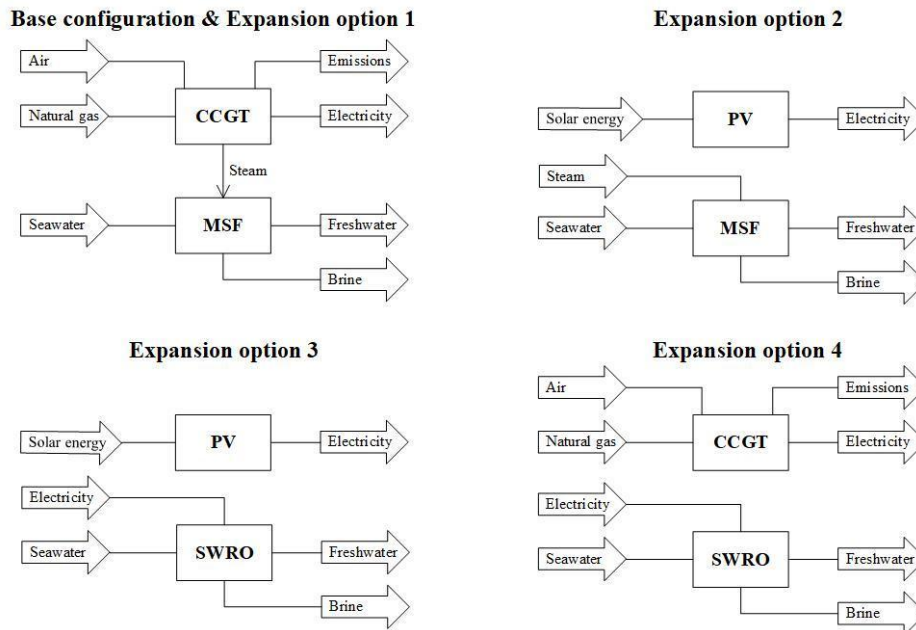


Figure 1. Facility block diagram of base configuration and expansion options.

The NPV is determined using:

$$NPV = -Capex + \sum \frac{Rev-Opex}{(1+d)^T} \quad (1)$$

where *Capex* is the total capital expenditure for each expansion option, *Rev* is the annual revenue, *Opex* is the total operating expenditure, *d* is the discount rate, and *T* is the facility lifetime. For the power block, the gas section is composed of multiple H-class combined-cycle turbines in a typical 1×1×1 configuration. The PV farm is fitted with a single-axis tracking system. In this case, battery storage was not considered due to the nature of the power mix and the operating philosophy since CCGT and PV are mixed to supply base and peak demands based on weather conditions and seasonality. For the water block, typical MSF and SWRO configurations suitable for the Middle-East region are used. The financial parameters for each of the production technologies are listed in Table 1. The discount factor and facility lifetime are estimated at 10 % and 25 years and the capacity factors for CCGT and PV are 67.4 % and 24.3 %, respectively.

Table 1. Financial parameters for the production technologies.

	CCGT	PV	MSF	SWRO
Capex (\$/kW, \$/m ³ d ⁻¹)	1,084	1,313	2,100	1,500
Fixed Opex (\$/kW, \$/m ³ d ⁻¹)	14.10	15.25	79.20	140.30
Variable Opex (\$/kWh, \$/m ³)	2.55	-	30.80	89.70

Source: EIA (2020) and World Bank Group (2019).

A binomial lattice model is developed for each of the options and is based on the exponential Brownian motion using the electricity market price as the stochastic variable:

$$\frac{\delta S}{S} = e^{\mu(\delta t) + \sigma \varepsilon \sqrt{\delta t}} \quad (2)$$

where μ is the price escalation rate and σ is the market price volatility. The asset value based on the deterministic NPV of each option is escalated across 6-time steps where each of the steps is equivalent to 2.08 years. At the end of the 6th step, equivalent to 12.5 years or half facility lifetime, the option is to be exercised in a European-like options system. The payoff value is compared across different market price evolutions and the decision to expand or not to expand is based on the difference between the escalated asset value and the payoff. The payoff values are then discounted across all evolution branches from step 6 back to the starting point resulting in the extended NPV for each option. The extended NPVs and the option values are then compared to identify the rank of the options from most attractive to least.

3. Results and Discussion

The binomial lattice model evolution was first applied to expansion option 1 which resulted in the binary tree depicted in Figure 2. At time step 0, deterministic NPV of the option is escalated through 6-time steps using different price escalation rates based on Equation 2. The upper branches represent higher escalation rates compared to the lower branches. As shown in Figure 2, expansion option 1 offers more positive results at multiple branches of electricity market price evolutions at half facility lifetime. This provides an incentive for a more likely “Expand” decision with a predicted payoff value of around \$45.6 billion at the highest anticipated price escalation branch. Also, the extended NPV at \$5.1 billion provides another indication regarding the feasibility of this particular option.

0	1	2	3	4	5	6	Time step
0	2.08	4.17	6.25	8.33	10.42	12.5	Years
						23,198,217,269.32	
						45,644,434,538.65	Expand
					16,171,194,070.00		
					31,731,812,007.69		
				11,272,741,978.99		11,272,741,978.99	
				22,049,735,004.06		21,793,483,957.97	Expand
			7,858,090,823.39		7,858,090,823.39		
			15,313,665,052.54		15,105,605,514.48		
		5,477,779,186.63		5,477,779,186.63		5,477,779,186.63	
		10,628,740,520.46		10,459,809,419.35		10,203,558,373.26	Expand
	3,818,493,001.90		3,818,493,001.90		3,818,493,001.90		
	7,371,699,427.92		7,234,469,409.56		7,026,409,871.50		
2,661,824,858.00		2,661,824,858.00		2,661,824,858.00		2,661,824,858.00	
5,108,496,289.42		4,997,139,446.00		4,827,900,762.09		4,571,649,716.00	Expand
	1,855,525,614.72		1,855,525,614.72		1,855,525,614.72		
	3,446,849,589.36		3,309,911,393.17		3,100,475,097.13		
		1,293,464,254.99		1,293,464,254.99		1,293,464,254.99	
		2,264,464,216.94		2,097,342,002.50		1,834,928,509.99	Expand
			901,658,142.40		901,658,142.40		
			1,416,659,713.52		1,220,323,610.61		
				628,534,884.22		628,534,884.22	Don't
				816,077,606.93		628,534,884.22	Expand
					438,143,995.06		
					438,143,995.06		
						305,424,830.38	Don't
						305,424,830.38	Expand

Figure 2. Binomial lattice model evolution and valuation (expansion option 1).

Evolution and valuation of the other 3 expansion options is also applied and similar to expansion option 1, the results in general are positive with the exception of PV based expansion options (2 and 3) where price escalation evolution must lean towards the higher escalation rate branches for the options to become feasible. A positive option value does not necessarily mean that the option is feasible as both sides of the tree should be investigated together with current price escalation rates at the facility half lifetime, when a decision must take place. Furthermore, the extended NPV for all of the options together with the option values are presented in Figure 3. An informed decision is possible once all the options are properly ranked from the most feasible to the least based on current market data.

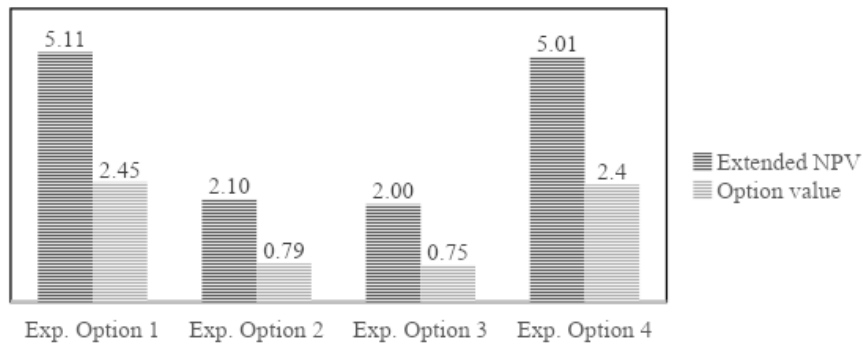


Figure 3. Comparison of extended NPV and option value for all expansion options.

The poor financial performance of renewable energy technologies can be attributed to the higher initial capital costs compared to conventional technologies. Moreover, the lower plant efficiencies that are characteristic to renewable energy technologies also amplify the financial gap especially when a persistent energy throughput is in demand. From a pure financial perspective, CCGT technology is privileged when compared to PV and for water MSF process is slightly better than SWRO. In order to overcome these shortcomings with renewable energy technologies, additional criteria must be included in the assessment such as natural gas and carbon market price fluctuations.

4. Conclusions

Future expansion pathways for existing CCGT-MSF electricity and water production facility were evaluated using ROA. A discretised binomial lattice model was developed for 4 possible configurations composed of CCGT, PV, MSF and SWRO technologies. The results show that all of the options were feasible at half the existing facility lifetime with expansion option 1 providing more attractive financial results across various future electricity market price scenarios. Since only a single risk factor was considered in this study, renewable energy-based systems might have been at a disadvantage due to the higher capital costs and the lower throughput owing to the inherent lower energy efficiencies. Future work will focus on the introduction of variability in other risk factors including but not limited to natural gas and carbon market prices and the application of Bayesian based methods.

References

- H. Al-Obaidli, A. AlNouss, Y. Bicer, T. Al-Ansari, 2020, Portfolio Optimisation of Integrated Renewable Energy Cogeneration Systems, *Comput. Aided Chem. Eng.*, 48, 1435–1440
- H. Al-Obaidli, S. Namany, R. Govindan, T. Al-Ansari, 2019, System-Level Optimisation of Combined Power and Desalting Plants, *Comput. Aided Chem. Eng.*, 46, 1699–1704
- EIA, 2020, Capital Cost and Performance Characteristic Estimates for Utility Scale Electric Power Generating Technologies, <http://www.eia.gov>
- FS-UNEP, 2020, Global Trends in Renewable Energy Investment 2020, <http://www.fs-unep-centre.org>

- E.A. Martinez-Cesena, J. Mutale, 2011, Application of an advanced real options approach for renewable energy generation projects planning, *Renew. Sustain. Energy Rev.*, 15, 2087–2094
- R. Pringles, F. Olsina, F. Penizzotto, 2020, Valuation of defer and relocation options in photovoltaic generation investments by a stochastic simulation-based method, *Renew. Energy*, 151, 846–864
- L. Santos, I. Soares, C. Mendes, P. Ferreira, 2014, Real Options versus Traditional Methods to assess Renewable Energy Projects, *Renew. Energy*, 68, 588–594
- World Bank Group, 2019, The Role of Desalination in an Increasingly Water-Scarce World, <http://www.worldbank.org>
- M. Yang, W. Blyth, 2007, Modeling Investment Risks and Uncertainties with Real Options Approach, *Climate Policy Uncertainty and Investment Risk*, IEA

Supply Chain Planning with Vehicle Allocation for Gas Industry

Yena Lee^a, Sivaraman Ramaswamy^b, Jose M. Pinto^b and Lazaros G. Papageorgiou^{a,*}

^a*Centre for Process Systems Engineering, Department of Chemical Engineering, UCL (University College London), Torrington Place, London, WC1E 7JE, UK*

^b*Linde Digital, Linde plc., 10 Riverview Drive, Danbury CT 06810, United States*
l.papageorgiou@ucl.ac.uk

Abstract

In this work, we address a problem of optimal production and distribution planning of industrial gas supply chains and the associated truck allocation. Two different optimisation models, a multi-objective optimisation (MOO) model and a mixed integer linear fractional programming (MILFP) model, are developed. The proposed models simultaneously determine the purchasing schedule for raw material, transportation schedule, network structure of supply chains, and optimal allocation of truck resources. To solve the problem, we employ the ϵ -constraint method for the MOO model, while a literature approach based on the Dinkelbach algorithm is adopted for the MILFP model as a solution method. Finally, an industrial case study is presented to demonstrate the applicability of the proposed optimisation models.

Keywords: Mixed-integer linear fractional programming, multi-objective optimisation, integrated supply chain planning, vehicle allocation

1. Introduction

Industrial gas supply chains involve domestic and/or international transportation systems between multi-site production plants, intermediate storage locations and several customers (Chima, 2007). Consequently, a large portion of supply chain costs come from transportation, and improving transportation efficiency has become increasingly important to improve overall profitability. Transportation efficiency can be improved by optimally allocating vehicle resources. Furthermore, there is a need for considering not only the transportation efficiency but also the economic performance, such as total cost or profit, in an integrated manner. There are existing works in literature, which address the production and distribution planning for the gas industry. Campbell and Savelsbergh (2004) addressed combined vehicle routing with scheduling problems for industrial gas distribution, and more recently by Dong et al. (2014). Concerning the integrated production and distribution planning, Marchetti et al. (2014) developed a mixed integer linear programming (MILP) model which coordinates production and distribution decisions. The production model considers plants operating at various production modes, while the distribution model accounts for a combined vehicle routing and inventory management. You et al. (2011) developed an MILP model accounts for planning decision of tank sizing at customer locations, inventory and truck routing decisions of industrial gas supply chain simultaneously. Zhang et al. (2017) proposed an MILP model and an iterative heuristic approach for the multi-scale production routing problem which combine production, distribution, and inventory decisions. Additionally, Lee et al. (2020) developed an MILP model which integrates supply contracts, production and transportation scheduling, and inventory management. They also proposed a relaxation approach to tackle computational limitation when dealing with large-size problems.

However, to the best our knowledge, none of works has considered the issue on improving transportation efficiency. In this work, we investigate a problem for the production and

distribution planning involving truck allocation of industrial gas supply chains, with total cost and total number of allocated truck resources as objectives. The aim is to develop optimisation frameworks which integrate the decision on the truck allocation into our previous model (Lee et al., 2020), and to adopt two solution approaches for the developed models; ϵ -constraint method and Dinkelbach algorithm.

2. Problem Statement

This work addresses the integrated optimisation of the industrial gas supply chains, including the production and distribution planning, and truck allocation. Figure 1 shows the network of the industrial gas supply chains considered in this work. On the production side, each plant purchases raw material from an external supplier and transform into liquid product at high purity. This liquid product can also be purchased from third-party suppliers. On the distribution side, the customer demand is satisfied through two primary transportation modes. The trucks deliver the product from plants, depots, and third parties to customers, whereas the rail-cars transport the product from plants and third parties to depots.

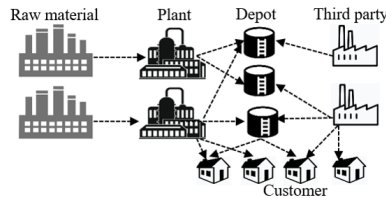


Figure 1: Network of industrial gas supply chains

Overall, the optimisation problem is described as follows:

Given (a) maximum available supply of raw material and purchasing cost; (b) plant location, production capacity, and production cost; (c) third party location, maximum purchasing amount from third party, and related cost; (d) unit outsourcing cost (e) initial, minimum, and maximum inventories; (f) loading capacities, available quantities, and transfer unit costs for both rail-car and truck; (g) total number of available truck resources.

Determine (a) production schedule; (b) purchasing schedule of raw material and product from external suppliers; (c) product allocation; (d) rail-car schedule; (e) allocation of truck resources to plants, depots, and third parties.

so as to (a) minimise total cost; (b) minimise total number of allocated truck resources.

3. Mathematical Frameworks

In this section, brief outlines of two different optimisation models are presented based on our previous work (Lee et al., 2020).

3.1. MOO model

Firstly, we formulate the problem as an MOO model, which comes with two objective functions. The first objective, total cost which is comprised of raw material cost, production cost, plant start-up cost, third party cost, truck cost, rail-car cost, and outsourcing cost, is minimised:

$$\min TC^{total} \quad (1)$$

$$TC^{total} = TC^{raw} + TC^{prod} + TC^{st} + TC^{third} + TC^{rail} + TC^{truck} + TC^{os} \quad (2)$$

The second objective considers the minimisation of total number of truck resources allocated to plants, depots, and third parties:

$$\min NT^{total} \quad (3)$$

$$NT^{total} = \sum_i NT_i^{new} + \sum_j NT_j^{new} + \sum_m NT_m^{new} \quad (4)$$

where NT_i^{new} , NT_j^{new} and NT_m^{new} are the number of allocated trucks to each plant i , depot j , and third party m , respectively.

Inventory mass balance. Inventory of each customer k in time period t is given by:

$$I_{kt} = I_{k,t-1} |_{t>1} + I_k^{ini} |_{t=1} + \sum_i Q_{ikt}^{truck} + \sum_j Q_{jkt}^{truck} + \sum_m Q_{mkt}^{truck} + Q_{kt}^{os} - D_{kt} \quad \forall k, t \quad (5)$$

The inventory at the end of time period t (I_{kt}) is equal to the inventory at previous time period ($I_{k,t-1}$) plus incoming product delivered by trucks from any plants, depots, and third parties ($\sum_i Q_{ikt}^{truck}$, $\sum_j Q_{jkt}^{truck}$, $\sum_m Q_{mkt}^{truck}$) plus outsourcing product (Q_{kt}^{os}), which can be used when existing plants and third parties can not satisfy customer demand, minus customer demand (D_{kt}). Here, the initial inventory is defined by I_k^{ini} . The similar equation returns to plant i and depot j for inventory mass balance of each plant and each depot (I_{it} and I_{jt}).

Truck constraint. The formulation to limit both the total product amount delivered and total travel time performed by trucks during each period follows our previous work (Lee et al., 2020):

$$\sum_k \frac{\gamma Q_{ikt}^{truck}}{CAP_i} \theta_{ik} \leq NT_i^{new} \Delta_t \quad \forall i, t \quad (6)$$

Here, the summation of the delivery amount/capacity ratio (Q_{ikt}^T/CAP_i) multiplied by the duration of the round-trip between the plant and customer (θ_{ik}), estimates the total travel time executed by trucks during the period. The γ multiplied to the delivery amount/capacity ratio is the tuning parameter to avoid underestimating the total travel time. The total travel time is restricted by the length of each time period (Δ_t) and the number of allocated trucks at each location (NT_i^{new}).

3.2. MILFP model

In MILFP model, the objectives are formulated as a fractional objective function:

$$\min \frac{TC^{total}}{AUR^{total}} \quad (7)$$

As we minimise both objectives, total cost and total number of trucks, we can not directly adopt one of them as a denominator. Here, we introduce the overall truck utilisation rate (AUR^{total}) as a denominator, which is maximised as the total number trucks is minimised.

Overall truck utilisation rate. The overall truck utilisation rate is calculated by:

$$AUR^{total} = \frac{\sum_i UR_i + \sum_j UR_j + \sum_m UR_m}{\sum_i X_i + \sum_j X_j + \sum_m X_m} \quad (8)$$

where UR_i , UR_j and UR_m are the truck utilisation rates for truck resources which are allocated at each plant, depot and third party. The binary variables, X_i , X_j and X_m , are 1 when at least one truck is allocated to each plant, depot and third party.

Eq. (8) includes the bilinear term; thus, it is linearised with an auxiliary variable, a big-M value, and additional constraints:

$$AUR^{total} = \sum_{l=1}^q \frac{TUR_l}{l} \quad (9)$$

$$\sum_l^q \overline{TUR}_l = \sum_i UR_i + \sum_j UR_j + \sum_m UR_m \quad (10)$$

$$\sum_i X_i + \sum_j X_j + \sum_m X_m = \sum_l^q l \cdot Y_l \quad (11)$$

$$\sum_l^q Y_l = 1 \quad (12)$$

$$\overline{TUR}_l \leq M \cdot Y_l \quad \forall l = 1, \dots, q \quad (13)$$

where Y_l is the binary variable to express the integer variable in the denominator of eq. (8) by its decimal representation. q is the maximum value of the number of plants, depots, and third parties which have at least one truck resource. i.e., q is the total number of plants, depots, and third parties considered in the problem.

Truck utilisation rate. The truck utilisation rate at each plant i during a given planning horizon is calculated by:

$$UR_i = 100 \cdot \left(\frac{\sum_k \sum_t \gamma Q_{ikt}^{truck}}{CAP_i NT_i^{new}} \theta_{ik} \right) / T \quad \forall i \quad (14)$$

where T is the number of time periods. As eq.(14) also includes the bilinear term, it is linearised by the linearisation scheme applied in eqs.(9)-(13). For the truck utilisation rates at each depot and third party (UR_j and UR_m), the similar equation and linearisation scheme are adopted.

Finally, the total number of allocated trucks is limited by the maximum number of available truck resources:

$$\sum_i NT_i^{new} + \sum_j NT_j^{new} + \sum_m NT_m^{new} \leq NT^{max} \quad (15)$$

4. Solution Approaches

4.1. ϵ -constraint method

To solve the developed MOO model, we employ the ϵ -constraint method (Haimes, 1971), where one of objectives is kept as an objective function and the other objective is considered as a constraint. The MOO model is transformed as follows:

$$\begin{aligned} & \min TC^{total} \\ & s.t. NT^{total} \leq \epsilon; \\ & \quad Eqs. (2), (4) - (6); \\ & \quad \text{Supply chain production and distribution planning constraints.} \end{aligned}$$

Here, ϵ is the user-defined upper bound for the total number of allocated truck resources.

4.2. Dinkelbach algorithm

For the MILFP model, we implement the Dinkelbach algorithm (Liu et al., 2014), which iteratively solve an MILP model given below:

$$\begin{aligned} & \min TC^{total} - f \cdot AUR^{total} \quad (16) \\ & s.t. Eqs. (2), (5) - (6), (9) - (13), (15); \\ & \quad \text{Linearisation constraints for the truck utilisation rates;} \\ & \quad \text{Supply chain production and distribution planning constraints.} \end{aligned}$$

The procedure of the Dinkelbach algorithm implemented for the problem is as follows:

- Step 1. initialise f ;
- Step 2. solve the MILP model (16), and denote the optimal solution as TC^{total*} and AUR^{total*} ;
- Step 3. If $|f - \frac{TC^{total*}}{AUR^{total*}}|/|f| \leq \delta$ (stopping criteria), the optimal solution for the MILFP model is TC^{total*}/AUR^{total*} ; otherwise, update $f = TC^{total*}/AUR^{total*}$, and go to Step 2.

5. Case study and Results

To demonstrate the applicability of the developed models, an industrial case study is conducted. The supply chain network considered for the case study consists of total 30+ of plants, depots, and third parties, 700+ customers, 300+ rail-cars, and 100+ trucks. A planning horizon is one month and it is divided into 30 days.

The developed models are implemented using Gurobi 9.0.0 in GAMS 30.3 on Intel 3.60 GHz, 32.0 GB RAM computer. For the MOO model, the optimality gap is set to 1%, while 98% of optimality gap and 0.01 of stopping criteria (δ) are set for the MILFP model.

Firstly, we solve the problem with the developed MOO model by fixing ϵ to 7 values between 50 and 80 of the total number of allocated truck resources. The resulting MOO model includes 137,268 equations, 160,146 continuous variables, and 5,415 discrete variables. The entire process to get Pareto optimal solutions with different values of ϵ takes 3,575 s CPU time.

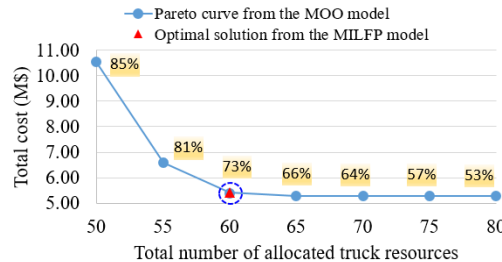


Figure 2: Pareto optimal solutions obtained from the MOO model

Figure 2 shows the optimal solutions and the piecewise linear approximation of the Pareto curve, together with the overall truck utilisation rate calculated by eq.(4) using obtained optimal values. The total cost remains relatively stable when the total number of trucks is decreased from 80 to 65, while the total cost is increased significantly as the total number of trucks is reduced from 55 to 50. We can also see that the overall truck utilisation rate is improved as the total number of truck resources is decreased. Furthermore, we can find a “good choice” solution marked with a blue circle, which dramatically reduces the total number of trucks and improves the overall truck utilisation rate with only a small increase in total cost.

Secondly, we solve the problem with the proposed MILFP model using the Dinkelbach algorithm. The model consists of 141,389 equations, 160,511 continuous variables, and 5,777 discrete variables, and it takes 1,914 s CPU time to gain an optimal solution. The red point shown in figure 2 is the optimal solution, and the point lies on the “good choice” solution which was identified by comparing the Pareto optimal solutions obtained from the MOO model. This result reveals that the proposed MILFP model can find an optimal solution, which is evaluated as the “good choice” solution by the MOO model. Figure 3 presents the number of allocated trucks at each plant and the cost breakdown gained from the MILFP model. As observed, the number of allocated trucks to each plant varies,

and the number lies on the values between 1 to 7. The result of cost breakdown shows that truck cost accounts for the highest percentage of total cost (41%), followed by raw material cost (26%) and production cost. The outsourcing cost and plant start-up cost show the lowest percentages of the total cost (2% and 1%, respectively).

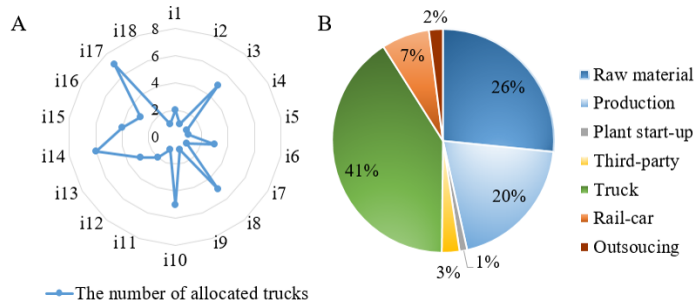


Figure 3: (A) The number of allocated truck resources and (B) the breakdown of the total cost obtained from the MILFP model

6. Conclusions

This work has addressed a problem of optimal production and distribution planning involving truck allocation for industrial gas supply chains, through the development of two different optimisation models; the MOO model and the MILFP model. The models have considered both economic performance (i.e. total cost) and truck efficiency simultaneously. To solve the MOO model, we applied the ϵ -constraint method, whereas the Dinkelbach algorithm was employed as a solution method for the MILFP model. The proposed models successfully solved an industrial case study. The results from the MOO model showed the trade-off between total cost and the total number of trucks. The decision-makers can select their preferred planning from the obtained Pareto optimal solutions. Finally, the MILFP model found a "good choice" optimal solution, which was identified by examining optimal solutions gained from the MOO model.

References

- Campbell, A. M., Savelsbergh, M. W., 2004. A decomposition approach for the inventory-routing problem. *Transportation Science* 38 (4), 488–502.
- Chima, C. M., 2007. Supply-chain management issues in the oil and gas industry. *Journal of Business & Economics Research (JBER)* 5 (6).
- Dong, Y., Pinto, J. M., Sundaramoorthy, A., Maravelias, C. T., 2014. Mip model for inventory routing in industrial gases supply chain. *Industrial & Engineering Chemistry Research* 53 (44), 17214–17225.
- Haimes, Y., 1971. On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on systems, man, and cybernetics* 1 (3), 296–297.
- Lee, Y., Carrero-Parreño, A., Papageorgiou, L. G., Ramaswamy, S., Jose, M., 2020. Integrated planning of industrial gas supply chains. In: *Computer Aided Chemical Engineering*. Vol. 48. Elsevier, pp. 1207–1212.
- Liu, S., Simaria, A. S., Farid, S. S., Papageorgiou, L. G., 2014. Optimising chromatography strategies of antibody purification processes by mixed integer fractional programming techniques. *Computers & Chemical Engineering* 68, 151–164.
- Marchetti, P. A., Gupta, V., Grossmann, I. E., Cook, L., Valton, P.-M., Singh, T., Li, T., André, J., 2014. Simultaneous production and distribution of industrial gas supply-chains. *Computers & Chemical Engineering* 69, 39–58.
- You, F., Pinto, J. M., Capón, E., Grossmann, I. E., Arora, N., Megan, L., 2011. Optimal distribution-inventory planning of industrial gases. i. fast computational strategies for large-scale problems. *Industrial & engineering chemistry research* 50 (5), 2910–2927.
- Zhang, Q., Sundaramoorthy, A., Grossmann, I. E., Pinto, J. M., 2017. Multiscale production routing in multicommodity supply chains with complex production facilities. *Computers & Operations Research* 79, 207–222.

Optimal Day-Ahead Load Curtailment for Central Chiller Plants

Gustavo Campos,^a Yu Liu,^a Nael H. El-Farra,^a Ahmet Palazoglu^a

^aDepartment of Chemical Engineering, University of California, Davis, One Shields Ave, Davis, CA 95616, USA

Abstract

Demand Response (DR) programs with the goal of maintaining the electricity grid's generation-load balance and ensuring transmission reliability have been on the rise in the last few decades. Services such as load shifting and curtailment have become economically attractive alternatives for large end-users with flexible operation. Different ways exist for providing these services, one of them being the Day-Ahead (DA) bidding as a generator in the energy or ancillary services market. In this work, we propose an optimization-based framework for performing DA bidding of load curtailment services, specifically in the context of large-scale central chiller plants with Thermal Energy Storage (TES) used for District Cooling. The framework integrates the two problems of real-time load shifting and day-ahead load curtailment bidding. Using the proposed framework, we evaluate the techno-economic feasibility of providing load curtailment services. The case-study is performed with real data from the central chilled water plant located on the campus of the University of California, Davis. Simulation results of the closed-loop implementation provide insights into the potential benefits from providing load curtailment.

Keywords: Demand Response, Load Curtailment, Process Scheduling, Energy Systems, HVAC.

1. Introduction

With the advent of deregulated electricity markets in the last few decades, there has been a push for increasing competitiveness in the generation and consumption of electrical energy. Unlike traditional commodities, electricity exhibits a few unique characteristics; among the most important are its continuous generation and delivery as opposed to batch-wise, and the need to balance generation and demand on a second-by-second basis. In the case that there is such an imbalance, or the transmission capacity exceeds its limit, there is a high risk of blackout followed by a restoring process that could take up to days (Kirschen and Strbac, 2019). To help mitigate these issues, system operators have been intensifying the use of Demand Response (DR) services, which have become especially useful to offset the impact of growing renewable penetration. From 2017 to 2018, participation of DR services in wholesale markets in the US increased by approximately 8 percent, while the California Independent System Operator (CAISO) had the greatest increase on a regional basis (FERC, 2019). The most common types of DR services include load shifting, load curtailment and the dispatching of controllable distributed generation systems.

The topic of load curtailment has been addressed in the literature mostly from the power generator or provider's perspective (e.g., Tsimopoulos and Georgiadis, 2019), which frequently makes use of oversimplified models for the consumer side (Zhang and

Grossmann, 2016). However, more recently the user's perspective has been addressed more heavily in the literature. Vujanic *et al.* (2012) considered the flexible schedule of cement milling plants using robust optimization. The provision of load curtailment was treated as a preliminary step in which the jobs having flexible starting times were selected arbitrarily. Zhang and Hug (2015) proposed a day-ahead bidding strategy using stochastic programming for aluminum smelters participating in both energy and spinning reserve markets. The continuous scheduling problem was solved daily and generated bidding curves for both markets for the following day. Zhang *et al.* (2015) assessed the benefits of adding cryogenic energy storage to air separation plants, including the provision of spinning reserve in ancillary services (AS) markets. The scheduling problem was solved weekly and uncertainty was assumed to arise mainly from the spinning reserve demand, which was addressed using robust optimization. Zhang *et al.* (2016) extended the work by considering recourse decisions using adjustable robust optimization. In these works, it was assumed that the reserve provider is rewarded based on the accepted bids, even if no actual dispatch is required. The authors highlighted the existing trade-off between load shifting and curtailment, favored by a low and high level of plant utilization, respectively. More recently, Chalendar *et al.* (2019) proposed a planning approach for bidding load reduction in the form of long-term capacity for HVAC plants. The bids are submitted 25 days before the start of each month, which results in a planning problem with long-term horizon and uncertain forecasts.

Most works in the literature address the open-loop schedule and consider bids that are either submitted a long time in advance or are valid for a long time into the future (week or month). The shorter-time interactions between load shifting and curtailment, specially under closed-loop operation, thus, remain mostly unexplored. In this work, we develop a framework for assessing the provision of load-curtailment services through bidding in a DA market, either energy or AS, specifically for large-scale central chilled water plants with TES used for District Cooling. The optimal load curtailment bidding problem is solved once every day and calculates the magnitude and duration of the load curtailment for the following day. In contrast, load-shifting is performed hourly and generates the production profiles that minimize the operating cost and implements the accepted curtailment bids. Emphasis is given on the following features: (i) the integration of the load curtailment bidding with real-time load shifting; (ii) the derivation of results from the closed-loop implementation of the framework, as opposed to the open-loop; (iii) the modeling of a large-scale chilled water plant using real operational data; and (iv) the consideration of DR programs that allow demand resources to bid as generators in the DA market.

2. Methodology

This paper considers the central chiller plant located on the campus of the University of California, Davis. The university has two separate plants at different locations, referred to here as plant 1 (P1) and 2 (P2). Plant 2 contains a TES tank with a maximum capacity of 5 million gallons of water. Each plant has four centrifugal chillers, all of them in parallel except for chillers 7 and 8 in Plant 2. The chillers have different maximum cooling capacities varying between 2,000 and 2,500 ton, adding up to a total of 18,000 ton maximum cooling capacity. The water inside the TES tank is distributed into two layers, a bottom layer of cooler water (supply) and a top layer with warmer water (return), separated by a thermocline. In addition, the tank operates full of water at all times. An important constraint particular to this system is that only the chillers at Plant 2 are allowed

to charge the TES tank, due to the configuration of the distribution system. A simplified representation of this system is presented in Figure 1.

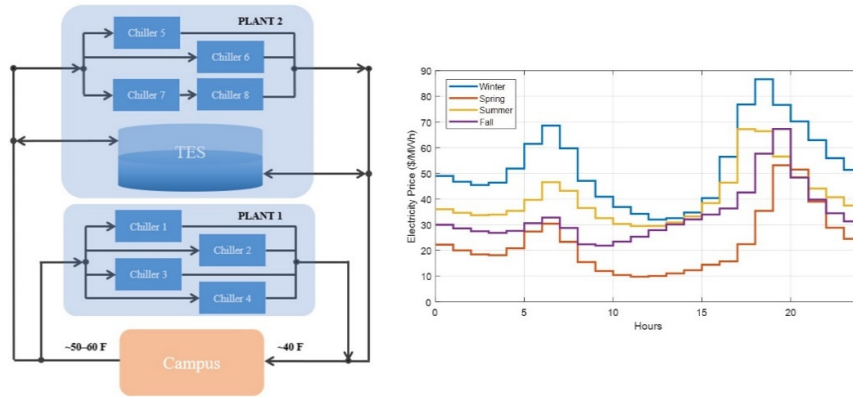


Figure 1. A schematic of UC Davis central chilled water plant (left); CAISO seasonal average day-ahead energy prices in 2019 (right).

The campus participates in the wholesale energy market from CAISO under a DA program, in which hourly electricity prices for a period of 24 h of the next day (from 00:00 to 23:00) are supplied to the customer one day ahead at 13:00. It is assumed that the real-time rate is equivalent to the DA price. The seasonal-average DA prices from CAISO in 2019 are also shown in Figure 1. In addition to the participation in the wholesale energy market for purchasing electricity, in this work we consider that the resource can also offer load curtailment capability as a generator in a DA market, either as energy or AS (e.g., spinning or non-spinning reserve). Bidding in the energy DA market as a generator provides an economic incentive to the user when the prices for purchasing and providing energy are settled using different node regions. Since the market prices vary considerably according to the location and system operator, here we focus on the unifying feature regarding the timing of the DA program. In the case of CAISO, bids for load curtailment are submitted before 10:00 on the day before the operating day. The market results are known at 13:00, along with the DA energy purchasing prices for the following day. This means that at the moment the bid is made, the prices for both services (load shifting and curtailment) are not known. Here we assume that the baseline (normal energy consumption profile without curtailment) is calculated by solving the optimal scheduling problem without load curtailment, instead of statistical historical averages. Real examples of economic-based DR programs include CAISO's Proxy Demand Resource (PDR), NYISO's Day-Ahead Demand Response Program (DADRP), and ISONE's Price-Responsive Demand (PRD).

2.1. Optimization Framework

The optimal load-shifting (LS) problem is formulated as follows:

$$\min \sum_{t \in T} (\psi_t^{opr} + \psi_t^{swi} + \psi_t^{rlx}) \quad (1)$$

$$\text{s. t. } \psi_t^{opr} = \sum_{c \in C} P_{c,t} C_t \quad (2)$$

$$P_{c,t} = (\beta_0 + \beta_1 T_t^S + \beta_2 T_t^{WB}) y_{c,t} + \beta_3 Q_{c,t} \quad (3)$$

$$ST_{t+1} = \left(ST_t + \sum_{c \in C_{P_2}} Q_{c,t} - D_t^{P_2} \right) \frac{(T_{t+1}^R - T_{t+1}^S)}{(T_t^R - T_t^S)} \quad (4)$$

$$\sum_{c \in C_{P_1}} Q_{c,t} = D_t^{P_1} \quad (5)$$

$$D_t^{P_1} + D_t^{P_2} = D_t \quad (6)$$

$$y_{c,t} Q_{c,t}^{min} \leq Q_{c,t} \leq y_{c,t} Q_{c,t}^{max}, ST_t^{min} \leq ST_t \leq ST_t^{max}, \sum_c P_{c,t} \leq P_t^{max} \quad (7)$$

where the parameters (both known and forecasted) include the electricity price C_t , the chilled water supply and return temperatures T_t^S and T_t^R , the wetbulb temperature T_t^{WB} , the campus cooling demand D_t , and the upper and lower bounds. Decision variables include the cooling load of chiller c at time t , $Q_{c,t}$, the on-off status $y_{c,t}$, the power consumption $P_{c,t}$, the storage level in cooling load equivalent ST_t , the portion of total demand supplied by plants P_1 and P_2 , $D_t^{P_1}$ and $D_t^{P_2}$, the operating cost ψ_t^{opr} and the penalties for switching and constraint relaxation ψ_t^{swi} and ψ_t^{rlx} (omitted in the model). A time horizon of 48 hours was employed.

The load-curtaiment (LC) problem is formulated as follows, in addition to Eqs. (2)-(7):

$$\min \sum_{t \in T} (\psi_t^{LS} - \psi_t^{LC}) \quad (8)$$

$$\text{s. t. } \psi_t^{LS} = P_t^{eff} C_t \quad (9)$$

$$\psi_t^{LC} = P_t^- C_t^{LC} \quad (10)$$

$$P_t^{eff} = P_t^{BL} - P_t^- + P_t^+ \quad (11)$$

$$P_t^{eff} = \sum_{c \in C} P_{c,t} \quad (12)$$

$$y_t^+ P_t^{+,min} \leq P_t^+ \leq y_t^+ P_t^{+,max}, y_t^- P_t^{-,min} \leq P_t^- \leq y_t^- P_t^{-,max} \quad (13)$$

$$y_t^+ + y_t^- \leq 1 \quad (14)$$

where C_t^{LC} is the price/reward for load curtailment, P_t^{BL} is an auxiliary variable that is fixed as the baseline during the bidding period, P_t^- and P_t^+ denote the negative and positive deviations from the baseline (only negative deviations are rewarded as load curtailment), P_t^{eff} is the effective power consumption after applying the negative and positive deviations, and ψ_t^{LS} and ψ_t^{LC} denote the economic cost and gain from load shifting and curtailment. Binary variables y_t^+ and y_t^- and Eq. (14) are employed to avoid increasing both P_t^+ and P_t^- by the same amount.

2.2. Integration

To effectively integrate the solution of the two problems, we can divide the time horizon of problem (LC) into three periods. In the first period, from 09:00 to 23:00 of the current day (first 15 hours), the problem reduces to a regular (LS), in which the bids from the previous day are implemented (setting an upper limit for power consumption) and load shifting is performed with the remaining load. The load deviation variables P_t^- and P_t^+ are set to zero by adjusting their lower and upper bounds in Eq. (13), making the effective power consumption P_t^{eff} equal to P_t^{BL} , both free decision variables. This step is important because it will determine the available storage at the beginning of the next day, which in turn determines the available load curtailment capacity for bidding. In the second period, from 00:00 to 23:00 of the following day (next 24 hours), the time and magnitude of the load curtailment bids for the following day are calculated. Prices are unknown and must be forecasted, and the base power consumption P_t^{BL} is fixed and set equal to the baseline

previously calculated by solving the (LS) problem. The load deviation variables P_t^- and P_t^+ are now allowed to vary freely by relaxing their upper and lower limits. The third period consists of the consecutive time steps. If a time horizon of 48 h is used, this period corresponds to the last 9 hours. This additional step is important to diminish the effect of the finite horizon approximation or the associated terminal constraint on the generated schedule or bids, e.g., to avoid storage depletion by the end of the following day, which could incur in demand and curtailment constraint violations. For this period, one can either solve a pure (LS) problem, in which case the problem becomes analogous to the first period, or also consider load curtailment bids, in which case the problem becomes analogous to the second period. The integration of the solution of problem (LC) into (LS) is easily done by enforcing upper bounds on the power consumption P_t^{max} in Eq. (7).

3. Case-Study Results

As a case-study, we generate and implement load curtailment bids for an average week from summer 2019. Seasonal average profiles were generated for DA electricity cost and building cooling demand, as shown in Figure 2. We assume that the load curtailment incentive C_t^{LC} has a magnitude proportional to the DA energy purchase cost (e.g., 10%), and that it only assumes a nonzero value during peak hours (e.g., 12 highest values in each day), as is typically the case for reserve prices. Furthermore, the price profiles are assumed to be known ahead of time for this analysis.

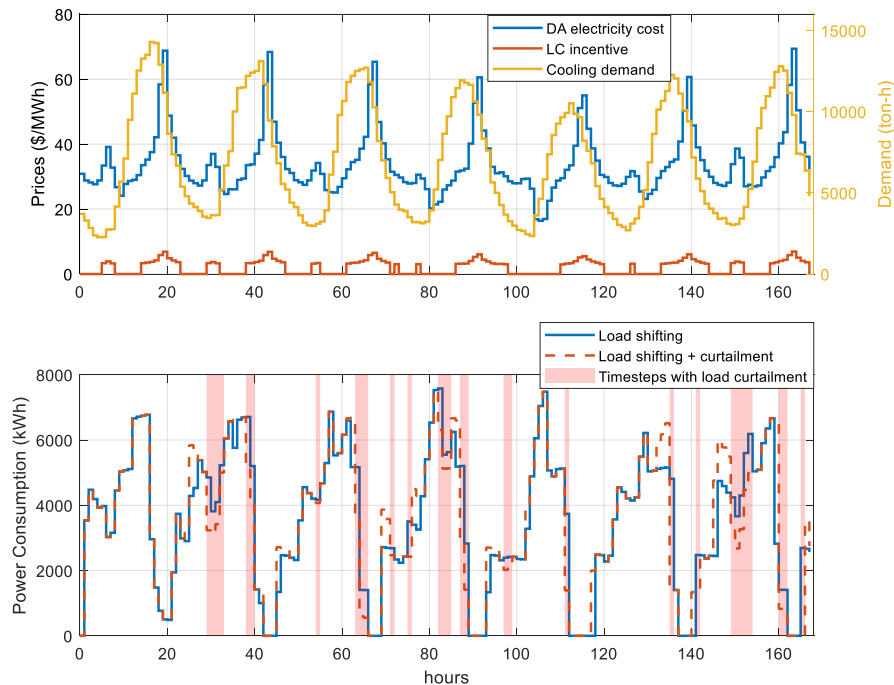


Figure 2. Energy prices and cooling demand (top), and power consumption profile with and without load curtailment (bottom).

Power utilization profiles are shown in Figure 2 for two cases; the first considering only load shifting and the second integrating load shifting and curtailment. It can be observed that most of the economic gains derive from the load shifting, while the curtailment

algorithm implements finer corrections to maximize the curtailment rewards. Providing load curtailment reduced the operating cost by approximately 2% considering a reward profile with a magnitude of 10% of the DA energy price. Naturally, the extent of the achievable economic gains will depend on the local cost profile and may or may not suggest the provision of load curtailment services. It can also be observed that periods with a higher level of plant utilization (i.e., higher power consumption) do not necessarily correspond to higher curtailment when performed together with load shifting. This can be attributed to the fact that the load shifting algorithm acts first by assigning the bulk of power consumption to low-cost regions, which in turn provide low to zero rewards for curtailing.

4. Conclusions

In this work, we presented a framework for the optimization of load-curtailment services in the DA market for large-scale central chilled water plants with TES. The framework integrates the DA bidding of load curtailment with real-time load shifting, and addresses closed-loop implementation. The results of a case-study demonstrated the potential benefits of providing load curtailment services, and provided insights into the interactions between load shifting and curtailment when implemented in a closed-loop fashion. Future research will focus on using real prices from DR programs, analyzing how price and demand forecasts and storage size affect the potential load curtailment benefits, and assessing the impact of aggregating loads.

References

- J. A. Chalendar, P. W. Glynn, S. M. Benson, 2019. Experimental Investigation of a Capacity-Based Demand Response Mechanism for District-Scale Applications. Proceedings of the 52nd Hawaii International Conference on System Sciences.
- FERC, 2019. 2019 Assessment of Demand Response and Advanced Metering, Technical Report.
- D. S. Kirschen, G. Strbac, 2019. Fundamentals of Power System Economics, 2nd Edition. John Wiley & Sons.
- E. G. Tsimopoulos, M. C. Georgiadis, 2019. Strategic offers in day-ahead market co-optimizing energy and reserve under high penetration of wind power production: An MPEC approach. *AIChE J.* 65.
- R. Vujanic, S. Mariéthos, P. Goulart, M. Morari, 2012. Robust integer optimization and scheduling problems for large electricity consumers. Proceedings of the 2012 American Control Conference, 3108–13.
- Q. Zhang, I. E. Grossmann, 2016. Enterprise-wide optimization for industrial demand side management: Fundamentals, advances, and perspectives. *Chemical Engineering Research and Design* 116, 114–131.
- Q. Zhang, I. E. Grossmann, C. F. Heuberger, A. Sundaramoorthy, J. M. Pinto, 2015. Air separation with cryogenic energy storage: optimal scheduling considering electric energy and reserve markets. *AIChE Journal* 61 (5), 1547–58.
- Q. Zhang, M. F. Morari, I. E. Grossmann, A. J. Sundaramoorthy, J. M. Pinto, 2016. An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load. *Computers and Chemical Engineering* 86, 106–119.
- X. Zhang, G. Hug, 2015. Bidding Strategy in Energy and Spinning Reserve Markets for Aluminum Smelters Demand Response. In: Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE PowerEnergy Society, 1–5.

Interplaying of food supply chain resilience, industry 4.0 and sustainability in the poultry market

Ahmed H. A. Mohamed,^{a,b} Brenno C. Menezes,^{b,*} Tareq Al-Ansari^{b,c}

^aDivision of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar Foundation, Qatar

^bDivision of Logistics and Supply Chain, Brazilian Foods, Doha, Qatar

^cDivision of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar Foundation, Qatar

bmenezes@hbku.edu.qa

Abstract

With today's rapid growth of human populations along with environmental concerns worldwide, it becomes increasingly difficult to meet future requirements for food, specifically in meat production systems (fish, poultry, cattle, etc.), to satisfy protein dietary demands of humans. The necessity of increasing the production of food, in both quantity and quality aspects, pressures the utilisation of natural resources (fresh water, land, etc.) and other inputs (as fertilisers) to fulfill demands of higher yields and improved properties of crops and livestock. Therefore, it is important to develop a sustainable model or framework integrated to the state-of-the-art technologies that helps to satisfy efficient use of natural resources in a consistent path by addressing food security issues without compromising food quality and the environment. To achieve sustainable processes within such limited resource world, the agri-food sector must move towards zero waste generation or at least minimise environmental impacts by (1) creating value in the agri-food manufacturing and supply chains and (2) recovering resources from by-products if techno-economically viable. Furthermore, considering novel technologies from the industry 4.0 (I40) age, it can be reached by efficient utilisation of resources via optimisation and integration of processes of production management among agri-food producers, traders, retailers, etc. Towards sustainable development in agri-food systems and supply chains, every operation must focus on closing loops to reduce, reuse, and recycle of material, energy, etc., without losing quality, within the concepts of the so-called circular economy (CE). These food supply chains, and agriculture productions can be improved using I40 solutions, yielding optimal operations, higher amounts of food, better quality of products, among others. An example in the poultry production and market is given considering interconnections of the triple of supply chain resilience, industry 4.0 and sustainability.

Keywords: Food supply chain, poultry market, industry 4.0, sustainability, resilience.

1. Introduction

Industry 4.0, supply chain resilience and sustainability have rapidly become strategic milestones for various industries, as they play into the competitive advantage of the chains and their overall efficiency. The three mentioned frameworks address the topics of sustainable processes in excellence and exchange value transformation between the two ends of the chain. Supply chain resilience (SCR) is an assessment of how prepared the chain would be for sudden disruptions. Supply chains must be designed to incorporate event readiness, provide an efficient and effective response, and be capable of recovering to their original state or even better post the disruptive event (Ponomarov and Holcomb, 2009). Industry 4.0 (I40) is identified as the framework that allows artificial autonomous machines to inspect, model, and visualise critical decisions based on mass data, wirelessly fed through electronic collectors that are distributed across the value chain. Moreover, this I40 element permits intelligent data-based decision-making to be done rigorously while considering the impact of the decision on the totality of the supply chain. This is achieved by using internet-of-things (IoT), radio frequency identification (RFID) technologies, big data and data analytics, advanced robotics, cloud computing, etc.

On the other hand, sustainability (SUS) focusses on the utilisation of current resources and assets to meet the demands of the present while accommodating the future generations' ability to do the same (Brandenburg and Rebs, 2015). Additionally, sustainability is the accountability measure of the negative externalities caused by the chain on the environment, society, and economy in the present and the future under the so-called circular economy fundamentals. Luthra and Mangla (2018) discuss the criticality of incorporating sustainability factors in the decision-making environment. They foresee that the framework's exclusion of those variables will intensify the damage of ecological imbalances like resource over-consumption, global warming, climate change problems and energy requirements for supply chains as their demand grows. Therefore, the design of a model that balances the efforts spent on the three frameworks would be the best approach for the business, consumer, and environment.

This work evaluates one major food supply chain that is found in the poultry industry. The significance of such study lies in the inevitable increasing demand of food due to the ever-growing human population. This continuously increasing demand pressures governments to ensure the availability of sufficient supplies through imports and local production, without damaging the source of their natural biological ecosystems. Assessing the relationship between supply chain resilience (SCR), industry 4.0 (I40), and sustainability (SUS) throughout the value chain of the poultry products will allow for a holistic analysis of the critical parameters in this specific market (Wang et al., 2018). As an example of the optimisation element of the I40, Menezes et al. (2020) consider a poultry production planning for both feed (ration, vitamins, etc.) and number of animals (from the hatchery) to be placed into appropriate cages for their proper growth. The animals are assigned to unit-places modelled as batch-processes with limited capacity and variate time-of-growing of the livestock batches. How the I40 elements can assist in the reduction of bottlenecks throughout a chain will define how such novel states of technologies may impact SCR and SUS in the chains. Various processes in the food supply chains are interlinked, allowing for some multi-dimensional surpluses with minimal improvements.

2. Supply chain resilience and sustainability connections

Many researchers have addressed the connection between sustainability and supply chain resilience, as they seek to understand how external and internal disruptions would impact a sustainable supply chain and how quickly can it recover to normal state. Derissen et al. (2011) affirmed that there are four possible relationships between the two frameworks: *a) resilience of the system is necessary, but not sufficient, for sustainability; b) resilience of the system is sufficient, but not necessary, for sustainability; c) resilience of the system is neither necessary nor sufficient for sustainability; and d) resilience of the system is both necessary and sufficient for sustainability.* Fahimnia et al. (2015) conducted an in-depth trade-off analysis between the supply chain costs incurred while minimizing environmental degradation. They found that when it comes to lean versus green directions, not all lean interventions at the planning level will result in a greener chain and that a flexible chain (resilient chain) is the most efficient alternative to balancing the focus on greenness and leanness.

Mari et al. (2014) tested the relationship between sustainability and resilience through a network design and optimisation. Differently to Azevedo et al. (2013) who assigned an index for the *ecosillient* relationship and Fahimnia et al. (2015) with their trade-off analysis approach, they created a multi-objective programming model that optimises sustainability (SUS) through carbon emission reduction and resilience through location-specific risks at each node in the network. Taking into consideration other conflicting goals like supply chain costs, carbon emissions and disruption costs, this research will attempt to use the different previously addressed models for well-rounded relationship testing.

This study will shed the light on various links and opportunities on the ability to retain mutual utility from the two far ends of the chain as the chain converges towards the equilibrium balance of prioritizing each framework. If the suppliers' ability to transfer value through their chains becomes highly responsive and visible through industry 4.0 components, then it would enhance the chains' absorption ability to external disruptions. This in turn would increase the resilience of the chain. The strategic strength of a more resilient chain will allow for better matching of supply and demand. In an ideal theoretical space, where a perfect match between supply and demand occurs, the reduction of food waste while meeting the nutritional demands of the consumers, can be reached. The sustainability value could be reflected when industry 4.0 systems consider the present and future tensions caused on the external environment of a supply chain. With a valid link between the frameworks, firms can focus on the optimization of resources, which will be travel in a sustainable efficient network. Eventually this would accommodate for the present and the future, without any decrease in retained utility of the firm and society. In other words, the ideal match between supply and demand could be more attainable as the supply chain management moves towards the equilibrium between all frameworks.

3. Interplaying of food supply chain resilience, industry 4.0 and sustainability

This research targets on using the previously developed concepts, metrics, indices, and models to calculate the impact factor of industry 4.0 (I40) integration on supply chain resilience (SCR) and sustainability (SUS). Most of the studies discuss thoroughly each framework exclusively as a major strategic focus, then propose one or multiple directions

for supply chain managements to consider. Some of the researchers suggest the link between resilience and sustainability is due to the implicit benefits between the two terms. However, a thorough comparative study that addresses the concept of a resilient sustainable supply chain with industry 4.0 integrated components (in short, RSSC4.0) has not been conducted in the food sector. The goal of this research is to (1) validate the triangular connection between the frameworks, (2) conceptualise the benefits and limitations of adopting RSSC4.0 in the supply chain operations. Firstly, contributing to the introduction of the concept of interplay between the three studied frameworks, and the business value of taking the strategic directions to reach the equilibrium. Secondly, contributing to an in-depth study of how industry 4.0 integration can solve or ease various bottlenecks in food supply chains. This would allow for enhanced, sustainable, and resilient networks that satisfy their fiduciary duties to all their stakeholders. The points to be raised in the SCR-I40-SUS triple as show in Figure 1 to be applied in the food industry or more precisely in the poultry market are:

- Visualize the supply network design of a *farm-to-fork* poultry firm.
- Identify and extract the critical parameters that cause bottlenecks for each process across the chain.
- Identify industry 4.0 technologies that can resolve or ease the bottlenecks.
- Compare the previously address models like the *ecosillient* index of the supply chains before and after the industry 4.0 integration.
- Divide the impacts of the industry 4.0 integration to two categories:
 - Direct or indirect impact demonstrated on the resilience score (R)
 - Direct or indirect impact demonstrated on the sustainability score (S)
- Express through the findings the future implications and trends of the study of RSSC4.0.

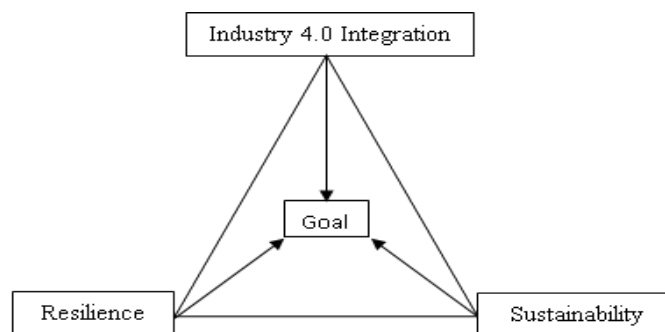


Figure 1. Supply chain resilience (SCR), industry 4.0 (I40) and sustainability (SUS) triple.

4. Poultry market value chain

As an example, we consider in the analyses of segment of the poultry supply network, where farms regularly place orders to suppliers of feed (like corn) to satisfy chickens' nutritional needs. The selection of the supplier is based on the ideal balance between pricing, quantity, and quality of the feed to ensure that the standard food conversion rate

(FCR) is met the livestock production in the farmyards. Variables that allow farms to decide on the quantity and quality of the feed would be mass inputted in cloud-based machine, which would be analysed and the resulted information would be shared to:

- (1) allow cameras or other optical-verification technologies to flag the rating of the quality of the feed when an order is received for quality assurance.
- (2) optimise the order point thresholds by balancing the increase of FCR obtained with a higher quality feed, allowing for a reduced quantity ordered and variable cost per chicken.
- (3) allow for a swift well-rounded decision on the procurement selection that is most ideal for the farm's sort and long-term needs.

This cloud-based system would be integrated with the remainder of the information flowing through the network possibly assisting in an optimised decision elsewhere in the network. If all farms aim to optimise the balance of variables in their procurement selection, it will allow the feed suppliers to focus on the metrics that they require if made public to them, which enhances the suppliers' ability to categorise farms based on their quality needs and direct existent and future products to the farms with a high likelihood of acceptance.

With the implementation of some of I40 components in the poultry market for a sustainable supply chain in Figure 2, the total SCR of the feed procurement process will increase through helping the two-ends effectively predict the versatility of each other's demand and supply patterns. Furthermore, by enhancing their cooperation and data visibility, an efficient production schedule will be made, reducing the amount of waste and pollution caused by unnecessary oversupply of poultry feed, enhancing the supply's network SUS and circularity. Similar practices could be implemented on agribusiness value chain yielding optimal operations, higher amounts of food, better quality of products, among others.

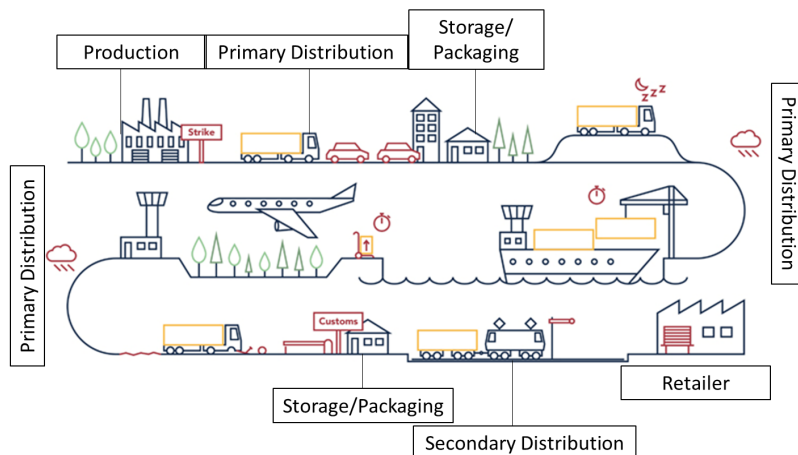


Figure 2. Poultry market components for a sustainable supply chain.

5. Conclusion

The improvements and changes that could be done to the current processes in the poultry production industry using the triple SCR-I40-SUS framework could enhance the value added from farm-to-fork to the businesses, governments, people, environment, and economy. The matching of supply and demand is the fundamental goal for each player in the supply network as it enhances their short-term profitability and long-term financial security. I40 would support in attaining that ideal matched point of supply and demand, while also allowing the firms in the network to consider other environmental, social, and economic impacts that could be easily overlooked when decisions are focused on short-term outcomes. The future objectives of this study would be to identify an ideal equilibrium point or optimisation model to balance the benefit and cost of a RSSC4.0 focused supply network.

Acknowledgment

The authors are grateful to Hamad Bin Khalifa University, Doha, Qatar, for supporting this research.

References

- S.G. Azevedo, K. Govindan, H. Carvalho, V. Cruz-Machado, 2013, Ecosilient index to assess the greenness and resilience of the upstream automotive supply chain, *Journal of Cleaner Production*, 56, 131-146.
- M. Brandenburg, T. Rebs, 2015, Sustainable supply chain management: a modeling perspective, *Annals of Operations Research*, 229(1), 213-252.
- S. Derissen, M.F. Quaas, S. Baumgärtner, 2011, The relationship between resilience and sustainability of ecological-economic systems, *Ecological Economics*, 70(6), 1121-1128.
- B. Fahimnia, J. Sarkis, A. Eshragh, 2015, A tradeoff model for green supply chain planning: A leanness-versus-greenness analysis, *Omega*, 54, 173-190.
- S. Luthra, S.K. Mangla, 2018, Evaluating challenges to industry 4.0 initiatives for supply chain sustainability in emerging economies, *Process Safety and Environmental Protection*, 117, 168-179.
- S. Mari, Y. Lee, M. Memon, 2014, Sustainable and resilient supply chain network design under disruption risks, *Sustainability*, 6(10), 6666-6686.
- B.C. Menezes, J.D. Kelly, T. Al-Ansari, 2020, Livestock production planning with batch-lines in the agriculture industry, *Computer Aided Chemical Engineering*, 48, 465-470.
- S.Y. Ponomarov, M.C. Holcomb, 2009, Understanding the concept of supply chain resilience, *The International Journal of Logistics Management*, 20(1), 124-143.
- J. Wang, R. Dou, R.R. Muddada, W. Zhang, 2018, Management of a holistic supply chain network for proactive resilience: theory and case study, *Computers and Industrial Engineering*, 125, 668-677.

Integrated Spatially Explicit Landscape and Biofuel Supply Chain Network Design

Eric G. O'Neill^{a,b}, Christos T. Maravelias^{a,b,c,*}

^a*Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08540, USA*

^b*DOE Great Lakes Bioenergy Research Center*

^c*Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ 08540, USA*

maravelias@princeton.edu

Abstract

An integrated biofuel supply chain (SC) system, from field to product, has the potential to help fuels from second generation bioenergy crops become competitive with fossil fuels and grain-based fuels. Landscape design, deciding where to plant bioenergy crops and how to manage that land, has been shown to improve the environmental impact of farm-scale biomass production (including soil carbon sequestration) but has been studied largely separately from biofuel supply chain network design (SCND). This paper proposes methods for the spatially explicit and simultaneous optimization of landscape design and SCND to explore their economic and environmental tradeoffs. Field scale data methods are described, key constraints for a two-stage stochastic mixed-integer linear model are introduced, and results from a realistic case study in southern MI, USA are presented that show economic and environmental benefits from the proposed approach.

Keywords: Mixed-integer Stochastic Programming, Biofuels, Landscape Design

1. Introduction

Efficient design and operation of cellulosic biofuel production systems are crucial factors for the future scale up and competitiveness of renewable fuels (Puigjaner et al., 2015). In the context of biofuel production, mathematical programming methods such as mixed-integer linear programming (MILP) and stochastic programming can simultaneously account for tradeoffs between decisions needing to be made under uncertainty that span a range of both time and spatial scales. Researchers have applied a variety of techniques to the biofuel supply chain network design (SCND) problem. Osmani and Zhang (2013) developed an optimization model that considers supply chain (SC) strategic and tactical decisions at the county level and the minimization of total cost. Ng et al. (2018) presented a high-resolution SC model for corn stover with preprocessing depots (referred to subsequently as depots) and demonstrated that the spatial distribution of biomass influences the optimal SC configuration and operation. Gao and You (2017) presented a modelling framework that optimized the SC environmental impact in addition to the economic impact through a functional-unit-based life cycle approach under uncertainty.

One of the primary ways cellulosic crops achieve sustainability is through soil organic carbon (SOC) sequestration (Gelfand et al., 2013). Both biomass yield and SOC

sequestration potential depend on where in the landscape a specific crop is established. The term 'landscape design' refers to the process of choosing where in the landscape to establish crops and how to manage that land to achieve a specific goal. Some recent studies applied optimization techniques to landscape design, usually with environmental objectives that consider SOC sequestration. Cobuloglu and Büyüktaktakin (2017) developed a two-stage stochastic MILP model for landscape design around a fixed biorefinery that makes crop establishment decisions, considers environmental objectives, and analyses the competition between food and fuel. Field et al. (2018) simulated crop growth at different management conditions on lands around a fixed biorefinery and 'optimized' the landscape design by aggregating the lowest cost and greenhouse gas (GHG) parcels to meet a specific demand. Nguyen et al. (2019) developed a high-resolution MIP model that includes crop establishment and management decisions to optimize a set of environmental objectives around a pre-existing biorefinery.

The SCND problem and the landscape design problem have been studied largely separately in the literature. Landscape design studies typically do not make strategic or tactical SCND decisions (e.g., facility location, inventory planning) and SCND studies assume fixed crop locations and availability. Integrating these two aspects of biofuel production by including crop location and management as decision variables, rather than parameters as in most SC studies, offers an opportunity to optimize their spatially explicit tradeoffs and provide a tool for decision makers to design holistically sustainable systems.

In this paper we present methods for integrating the biofuel SCND and landscape design problems. In section 2 we present methods for gathering necessary high-resolution data, flexible spatially explicit model features, and selected model equations for landscape design. In section 3 we report results from a case study in southern MI, USA over a range of model instances at varying levels of detail. Finally, in section 4 we present conclusions and discuss application opportunities for the integrated model.

2. Spatially explicit two-stage stochastic model

2.1. Model Parameters and Data Methodology

Spatially explicit environmental and economic interactions among many of the SCND and landscape design decisions motivate the integration of high-resolution data sources with optimization models that can capture those tradeoffs. For example, the landscape design decision of where to plant crops will affect, among other things, both the SOC sequestration and the transportation cost.

Because yield and SOC sequestration potential is highly field specific, we model the available land for planting biomass on the field level and choose lands with United States Department of Agriculture Land Capability Classifications (LCC) V-VIII in southern Michigan, USA as our study area; although any set of geographically identifiable fields are compatible with the model equations presented in section 2.2.

Field-specific SALUS biogeochemical crop model simulations are used to obtain yield and SOC sequestration data for switchgrass grown on the available land (Basso et al., 2006). SALUS uses a daily time step approach to simulate crop growth and harvest including the flows of carbon and nitrogen. Simulations were performed over thirty years (1989 – 2018) with the stand re-planted every ten years at two levels of nitrogen fertilization (0, 50 kgN/ha) for each of the ~40,000 fields within our study area. SALUS

accounts for local weather at a 4 x 4 km resolution and, with thirty years of simulation, provides an estimate of the spatially explicit year-to-year uncertainty in biomass yield.

The two-stage stochastic model is solved for a representative year. Establishment years for which crop yields are artificially low are removed and uncertainty scenarios are sampled from the remaining 24 simulation years. The sampled yield years are taken as uncertainty scenarios $s \in S$ with an equal probability of occurring.

As the number fields within the study area grows, large model instances are unable to be solved in a reasonable amount of time. To maintain tractability of the integrated model while still accounting for the uneven distribution of biomass yield and the spatially explicit nature of landscape design, we introduce a gridded approximation. A grid of user defined resolution is overlaid on the study area. Grid cells containing no fields are discarded. The yield and SOC sequestration of fields within a cell are aggregated and the location of the new ‘harvesting site’ is defined as the yield-weighted average location of the fields it contains. Finer grid resolutions better approximate the uneven distribution of yield and SOC sequestration but lead to instances that are computationally difficult.

2.2. Integrated Landscape Design and SCND Model

To integrate the SCND and landscape design problems, we propose a spatially explicit two-stage stochastic model. The SC structure largely follows the deterministic approach of Ng et al. (2018) with extensions for the stochastic biomass yield uncertainty. Compounds $i \in I$ move through the SC according to material balance and capacity constraints defined over the harvesting sites, depots, and biorefineries with second-stage operational decisions made for a representative year divided into time periods $t \in T$.

Extending the work of Ng et al. (2018), we introduce a set of landscape design constraints to model crop establishment at high-resolution harvesting sites (from the gridded approximation) in addition to land management decisions that affect biomass yield and SOC outcomes. The binary variable $W_{i,j,r'}$ is equal to one if crop $i \in I^F$ is planted at harvesting site $j \in J$ using discrete management (e.g., harvest method, tillage method) $r' \in R^D$. Eq. (1) enforces no more than one crop may be established at each harvesting site and only one discrete management option may be chosen. A combination of discrete management options can be considered if each element of R^D is taken as a unique combination of discrete management options.

$$\sum_{i,r'} W_{i,j,r'} \leq 1 \quad \forall j \tag{1}$$

Landscape design decisions also involve continuous land management considerations (e.g., N-level, residue removal rate). Because the yield and environmental responses to land management are often non-linear, we introduce non-negative Special Ordered Set II (SOS2) variables to approximate the response to management as linear combinations of the available simulated data points for each field described in section 2.1. SOS2 variables are an ordered set of variables such that no more than two may be non-zero and they must be adjacent. SOS2 variables $Z_{i,j,r',r}$ are defined for each feedstock, harvesting site, discrete management option, and continuous management data point $r \in R$ (SOS2 on index r). Eq. (2) enforces the constraint that the SOS2 variables must sum to one if the harvesting site is selected for crop establishment and zero otherwise.

$$\sum_r Z_{i,j,r',r} = W_{i,j,r'} \quad \forall i \in I^F, j, r' \tag{2}$$

Eq. (3) constrains the yield $Y_{i,j,t,s}$ of crop i at each harvesting site j at each time period t and uncertainty scenario s to the linear combination of the SOS2 variables, the simulated yield potential $\alpha_{i,j,t,s,r,r'}$ (Mg/ha) from SALUS, and the area for establishment σ_j (ha).

$$Y_{i,j,t,s} = \sum_{r,r'} Z_{i,j,r,r'} \sigma_j \alpha_{i,j,t,s,r,r'} \quad \forall i \in I^F, j, t, s \quad (3)$$

$$D_j = \sum_{i,r,r'} \chi_r Z_{i,j,r,r'} \quad \forall j, r' \quad (4)$$

The level of continuous management D_j at each harvesting site is given by Eq. (4), where χ_r are the levels of the continuous management option at each data point $r \in R$. A two-stage stochastic formulation is implemented to account for the year-to-year uncertainty in biomass yields. Because the SCND decisions involving facility locations, technology selection, and capacity are capital intensive and difficult to change, model solutions should behave well in expectation over the lifetime of the SC in the face of the uncertain biomass yields which are influenced by variations in local weather, soil quality, and farming practices. The SC specific first stage decisions are biorefinery and depot location, technology, and capacity. The landscape design specific first stage decisions are crop establishment location and management level. Second stage decisions made after the realization of uncertainty include the SC operational decisions (inventory, shipment, and production management) and the landscape design decision, harvest amount. The optimization objective is the minimization of the expected value of the total annualized cost, TAC , of biofuel production defined by Eq. (5).

$$TAC = C^{CAP} + \sum_s p_s (C_s^F + C_s^{PR} + C_s^{INV} + C_s^{TR} + C_s^{SFALL} - C_s^{BY} + C_s^{ENV}) \quad (5)$$

where C^{CAP} , C_s^F , C_s^{PR} , C_s^{INV} , C_s^{TR} , C_s^{SFALL} , C_s^{BY} are the capital cost, feedstock cost, production cost, inventory cost, transportation cost, shortfall from unmet demand, and revenue from by-products respectively; p_s represents the probability for uncertainty scenario s to occur; and C_s^F defines the cost of landscape design (Eq. (6))

$$C_s^F = \sum_{i,j,t} \lambda_i H_{i,j,t,s} + \sum_j D_j \rho \sigma_j + \sum_{i,j,r'} \phi_{i,r'} W_{i,j,r'} \sigma_j \quad \forall s \quad (6)$$

which includes the per-ton cost λ_i for harvesting biomass $H_{i,j,t,s}$, the per-unit cost of continuous land management ρ , and the annualized per-area cost $\phi_{i,r'}$ of establishing and managing a crop. Finally, C_s^{ENV} defines the environmental impact of the landscape design and SCND on a cost basis by using the social cost of carbon (SCC) (\$/MgCO₂e). We model the emissions from transportation, production, and landscape activities (GT_s^{LAND}). The spatially explicit modelling approach described in section 2.1 allows us to include field-level SOC sequestration as negative landscape emissions (Eq. (7)).

$$GT_s^{LAND} = \sum_{i,j,r'} W_{i,j,r'} \sigma_j GHG_r^{HARV} + \sum_{i,j,t} H_{i,j,t,s} GHG_i^{MG} + \sum_j \sigma_j D_j GHG^{MAN} - \sum_{i,j,r,r'} \sigma_j GHG_{i,j,r,r'}^{SOC} Z_{i,j,r,r'} \quad \forall s \quad (7)$$

GHG_r^{HARV} is the per-area emissions from establishing and harvesting crops, GHG_i^{MG} is the per-ton emissions from harvest operations, GHG^{MAN} is the per-unit emissions from continuous land management, and $GHG_{i,j,r,r'}^{SOC}$ is the potential for sequestering SOC at a harvesting site under management options r and r' . A baseline value of 50 \$/MgCO₂e is used for the SCC (ToI, 2011), but by adjusting from the base value, a decision maker can determine how much emphasis to put on reducing emissions versus SC specific costs.

3. Case study results

To demonstrate the capabilities of the combined SCND and landscape design model and the methods used to integrate the two, we present the results of a case study located in southern MI, USA. Figure 1. displays the configuration of a detailed instance considering 5 uncertainty scenarios randomly sampled from the simulated years described in section 2.1, a spatial resolution of 4 x 4 km, and considering nitrogen fertilization 0 – 50 kgN/ha.

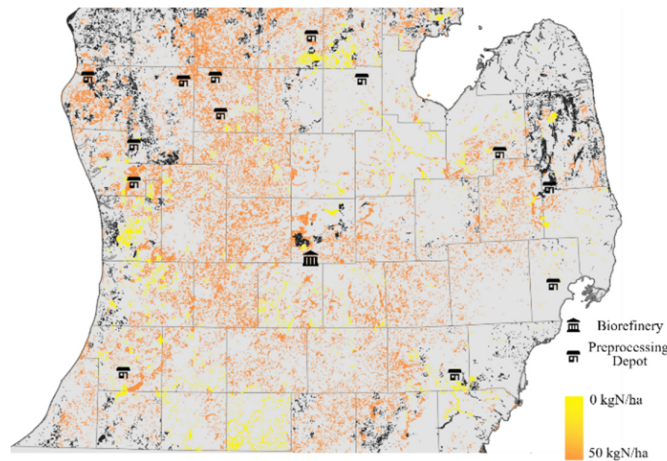


Figure 1. Optimal first-stage decisions for an instance considering 5 uncertainty scenarios and a 4 x 4 km grid resolution. Large centralized biorefinery and 13 depots located far from the refinery.

The model solution finds a large centralized biorefinery with 13 depots located far from the refinery. Landscape design, including crop establishment locations and fertilization, is driven by both the proximity to the refinery and depots and the SOC sequestration potential at each harvesting site. Computational results for instances at different spatial resolutions and number of uncertainty scenarios are detailed in Table 1. Implementation is with GAMS 32.2 solved via CPLEX 12.10 on 2.4 GHz Linux cluster machines with a time limit of 24 hours. Each row represents the average of 10 runs with different uncertainty scenario samples. High-resolution instances with many scenarios are difficult to solve to optimality. The value of the stochastic solution (VSS) is reported in Table 1. for instances able to be solved to a 1% optimality gap. The VSS represents the benefit to the objective value re-evaluated over the course of 10 years with first-stage decisions from solving the stochastic solution as opposed to a solution assuming the biomass yield at each field is that field's average yield. Table 1. also shows the average percent decrease of total annualized emissions (MgCO_2e) from the corresponding county level instance.

4. Conclusions

The spatially explicit and stochastic methods described in section 2 contribute to capturing the effect of landscape scale decisions on the SC outcomes to find solutions that perform well in expectation compared to the average information case. Instances with a higher spatial resolution lead to solutions with much better environmental performance largely due to better control of crop establishment, leading to higher SOC sequestration. The high-resolution instances rely on the availability of high-resolution and realistic data. The data methodology combining field-level crop model simulations with a gridded approximation enables tractable simultaneous landscape design and SCND optimization.

Table 1. Computational results for model instances at resolutions: county – 16 km² and uncertainty scenarios 1 – 12. Each row is the average 10 model runs for that number of scenarios.

# of Scenarios	Resolution	Soln. time (s)	Gap	VSS (\$)	GHG Benefit
1	County	64	1%	N/A	0%
	400	94	1%		-103%
	100	316	1%		-125%
	16	6,281	1%		-151%
3	County	406	1%	1.05 E7	0%
	400	779	1%	1.02 E7	-138%
	100	5,232	1%	1.67 E7	-151%
	16	64,336	1%	1.48 E7	-155%
5	County	1,204	1%	9.88 E6	0%
	400	2,447	1%	1.32 E7	-136%
	100	25,071	1%	2.04 E7	-128%
	16	86,400	15%	N/A	-142%
12	County	13,333	1%	1.46 E7	0%
	400	39,838	1%	1.58 E7	-180%
	100	68,517	9%	2.07 E7	-182%
	16	86,400	40%	N/A	N/A

An integrated landscape and SCND model paves the way for impactful system level analyses. The model generality, features, and data methodology allow for future analysis of SCND outcomes resulting from inputs such as alternative definitions of marginal land, uncertainty associated with future climate, and coarse regional study areas as well as small field-level systems through the flexible spatial approach. Insights will aid decision makers in designing efficient, cohesive, and sustainable field-to-product systems.

References

- B. Basso, J.T. Ritchie, P.R. Grace, L. Sartori, 2006. Simulation of Tillage Systems Impact on Soil Biophysical Properties Using the SALUS Model. *Ital. J. Agron.* 1, 677.
- H.I. Cobuloglu, I.E. Büyüktaktin, 2017. A two-stage stochastic mixed-integer programming approach to the competition of biofuel and food production. *Comput. Ind. Eng.* 107, 251–263.
- J.L. Field, S.G. Evans, E. Marx, M. Easter, P.R. Adler, T. Dinh, B. Willson, K. Paustian, 2018. High-resolution techno-ecological modelling of a bioenergy landscape to identify climate mitigation opportunities in cellulosic ethanol production. *Nat. Energy* 3, 211–219.
- J. Gao, F. You, 2017. Modeling framework and computational algorithm for hedging against uncertainty in sustainable supply chain design using functional-unit-based life cycle optimization. *Comput. Chem. Eng.* 107, 221–236.
- I. Gelfand, R. Sahajpal, X. Zhang, R.C. Izaurralde, K.L. Gross, G.P. Robertson, 2013. Sustainable bioenergy production from marginal lands in the US Midwest. *Nature* 493, 514–517.
- R.T.L. Ng, D. Kurniawan, H. Wang, B. Mariska, W. Wu, C.T. Maravelias, 2018. Integrated framework for designing spatially explicit biofuel supply chains. *Appl. Energy* 216, 116–131.
- T.H. Nguyen, J. Granger, D. Pandya, K. Paustian, 2019. High-resolution multi-objective optimization of feedstock landscape design for hybrid first and second generation biorefineries. *Appl. Energy* 238, 1484–1496.
- A. Osmani, J. Zhang, 2013. Stochastic optimization of a multi-feedstock lignocellulosic-based bioethanol supply chain under multiple uncertainties. *Energy* 59, 157–172.
- L. Puigjaner, M. Pérez-Fortes, J.M. Laínez-Aguirre, 2015. Towards a carbon-neutral energy sector: Opportunities and challenges of coordinated bioenergy supply Chains-A PSE approach. *Energies* 8, 5613–5660.
- R. Tol, 2011. The social cost of carbon. *Environ. Forum* 28, 38–41.

Techno-economic and environmental assessment of Gasoline produced from GTL and MTG processes

Saad A. Al-Sobhi,* Ahmed AlNouss, Mohammad Alhamad

Department of Chemical Engineering, College of Engineering, Qatar University, Doha, Qatar

saad.al-sobhi@qu.edu.qa

Abstract

Natural gas is expected to grow faster than oil or coal with predictions that it will be the leading primary fuel between 2020 and 2030. Technically, there are a number of ways to monetize the natural gas resource depending for example on the resource location, quantity, and quality. Beside the conventional way of transferring natural gas through pipelines, other options considered promising include: liquefied natural gas (LNG), compressed natural gas (CNG), gas to solid (GTS), gas to wire (GTW), gas to liquids (GTL) and gas to commodity (GtC). The utilization ways vary as being physical compression and mixing process like the case of LNG, CNG, and GTS or chemical conversion process like the case of GTL. Some of these utilization options are described as mature, while others are developing or prospective options that need more investment. In this study, the gas to methanol option is explored considering the methanol to gasoline (MTG) production route and compared to GTL process and products. The objective of this work is to conduct rigorous simulation models for two NG utilization routes namely, GTL and MTG in order to help the decision-maker in gas processing industry. Given a natural gas feedstock, it is desired to assess economically and environmentally the two production routes of gasoline product. In addition, the study aims to investigate the comparative and sensitivity analyses to incorporate the sustainability perspectives between GTL and MTG production routes. The results indicate a higher overall yield for the MTG process with a net economic value of around \$568 per tonne of gasoline produced and 0.2 tonne of CO₂-e emitted per tonne of product. Furthermore, increasing the percentage of methanol being processed to gasoline indicates to have more net economic profitability but with increased amounts of CO₂ emissions. These results indicate the significance of MTG process and can be further studied to identify the optimal percentage of methanol being processed to gasoline as a trade-off between economic and environmental benefits.

Keywords: MTG, GTL, Process simulation, Techno-economic-enviro, Natural gas

1. Introduction

Fossil fuels are dominating the major energy sources worldwide and serve as the prime feedstock in the production of a wide variety of chemicals. Reserves of natural gas are larger than other fossil-based fuels; coal and oil. However, with the significant depletion of fossil-based sources and the increasing demand for energy, especially liquid fuels, there is a critical need to establish alternative routes that archive higher sustainability in the generation of liquid fuels. Conversion of natural gas to liquid fuels through what is known as gas to liquid (GTL) process is technically mature and considered economically promising for large-scale production. Alternatively, liquid fuels generation through

methanol production route known as methanol to gasoline (MTG) is considered a feasible pathway. Methanol can be generated through reforming of natural gas to produce syngas intermediate. The first discovery of MTG technology was in 1970's by ExxonMobil in which zeolite catalysts were used due to their well-defined structure, acidity and surface areas (Li et al., 2015). After that, MTG process has received significant interest from industry and research. Most of the previous studies in literature focused on analysing a single natural gas conversion pathway and less attention was given to comparative studies that present the trade-offs between the different routes to support decision making. Aspen Plus process simulation along with published data were used to establish a basis for the techno-economic analysis of shale gas-to-methanol production. The results of the techno-economic analysis demonstrate that production of methanol from shale gas would be profitable for a broad range of methanol selling prices and shale gas costs. For example, a desirable 31% ROI is achieved for a methanol selling price of \$2.00/gal and shale gas price of \$3.50/kSCF (Ehlinger et al., 2013). Sequential process simulation and optimization techniques were implemented through a developed framework to natural gas processing and production networks, consisting of LNG, GTL, and methanol facilities. Initially, a steady state simulation using Aspen Plus software was used to determine accurately mass and energy balances, operating conditions, and equipment specification. Several benefits obtained by simulating the network, for example, determining accurate product yields and feed-gas requirements, capital and operating cost estimates linked to plant capacities, environmental impacts in terms of quantified emissions. The yields and costs obtained from the simulation are then used as inputs for the optimization step. The formulated linear programming (LP) optimization model with the objective function set to maximize profitability, evaluates the processing and production network over a wide range of forecasted economic changes (i.e. product prices) (Al-Sobhi and Elkamel, 2015). The same has been also applied in other studies to support the techno-economic decision making in process synthesis such as the different LNG liquefaction process routes (AlNouss et al., 2018) and to assess sustainability of production pathways (Al-Sobhi and AlNouss, 2018). The objective of this research proposal is to address and identify ways to increase the value of NG resource in Qatar by considering more utilization option such as methanol to gasoline (MTG) process. The MTG option is explored considering the methanol to gasoline (MTG) production route and compared to GTL process and products. The novelty of this approach is the economic and environmental comparison of two different routes for generating the same final product that is originating from the same energy source. This can be achieved by considering rigorous simulation and advanced mathematical optimization models while taking into account market information (i.e., supply and demand curves, product prices) for liquid fuels and methanol products. This presents a comparative analysis framework which benefits addressing the grand challenges currently facing Qatar and any NG-rich county worldwide through identifying the potential downstream value-added products to be generated and the right investment decisions. It is worth mentioning that the developed framework can be applied to different processing sectors such as oil refining, petrochemical, etc. One of the main expected outcomes of this study is to present a state-of-the-art NG processing and production planning tool to help the decision maker to assess the current monetization and identify any emerging/potential technologies to be considered to improve the economic portfolio of the country.

2. Model development

The application and implementation of integrated simulation-optimization approach proved to be helpful in utilizing the data from process simulation to deliver decision outcomes through optimization models. This study focuses on the assessment and the comparative analysis of synthesis gasoline produced from GTL and MTG processes. GTL is a process that convert natural gas into more valuable liquid products such as gasoline. It has three major steps; reforming of natural gas to produce syngas, Fischer-Tropsch (FT) synthesis to generate liquid fuels, and lastly, product upgrading to refine the generated products into different cuts. Natural gas reforming is reaction of the feed with oxygen and steam to produce H_2 and CO . Auto-thermal reforming (ATR) is selected in this study to be the method of syngas production since it has the ability to control large-scale production rates. ATR is endothermic reaction utilizing both oxygen and steam, and has the most suitable hydrogen to carbon monoxide ratio for cobalt-based catalyst operation (Iandoli and Kjelstrup, 2007). Air separation unit is utilized in the process to provide the required oxygen for syngas production unit. Hydrogen to carbon monoxide ratio ($H_2:CO$) is a significant parameter in GTL process, and it can be adjusted by water gas shift reaction to obtain the optimum ratio for the process. For cobalt-based catalyst process the ratio should be between 1.8:2.1 and in this study its equal to 2.1 (Steynberg et al., 1999). The produced syngas is then fed into FT reactor to convert it into long-chain hydrocarbons which are upgraded using hydrocracked into more valuable products such as naphtha, gasoline, kerosene, and wax. The final step of the GTL process is to refine the generated liquid fuels through series of distillation columns into the different cuts by means of difference in boiling points. Figure 1 illustrates the block flow diagram of the process.

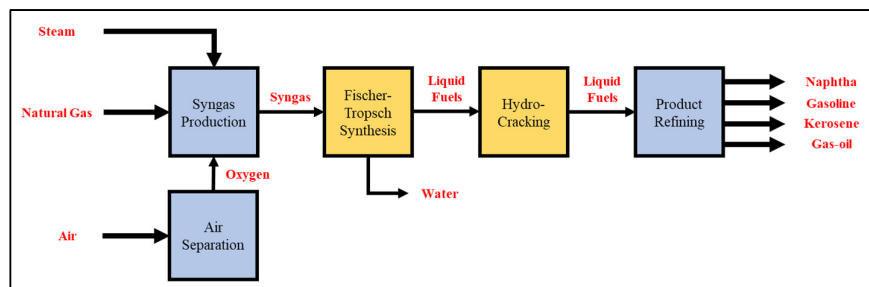


Figure 1: Block flow diagram of the GTL process.

Methanol to gasoline conversion is a modern process for generating gasoline from natural gas and it is considered as an alternative route to GTL process. The process starts by converting natural gas to syngas (CO and H_2) in the presence of steam, then converting this syngas into methanol followed by MTG. MTG process normally take place into two steps, first methanol is transformed into a mixture of dimethyl-ether (DME) and water with some unreacted methanol over a non-zeolitic catalyst usually. Secondly, this mixture is introduced into a zeolite catalyst to be converted to gasoline. This process was developed in the early 1970 by Mobile and it was commercialized later, and now other technologies are introduced in the area of MTG such as Topsoe integrated gasoline synthesis (TIGS) (Galadima and Muraza, 2015). Figure 2 illustrates the block flow diagram of the MTG process.

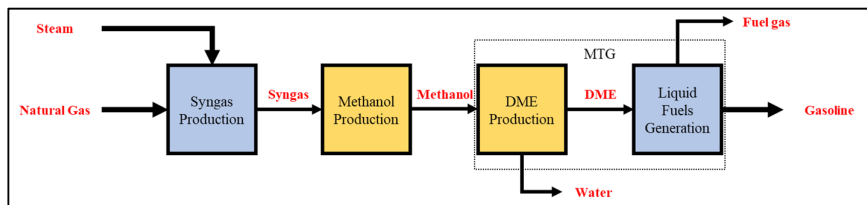


Figure 2: Block flow diagram of the MTG process.

The steady state simulation of the GTL and MTG process flowsheets is carried out using Aspen HYSYS software and taking into account the feed specifications presented in Table 1. The simulated flowsheets have been then analysed from technical, environmental and economic perspectives. Aspen Process Economic Analyzer and Aspen Process Energy Analyzer have been used to estimate the capital and operating costs of the two models along with the environmental emissions in terms of CO₂ equivalent. Moreover, a sensitivity analysis has been applied on MTG process to study the effect of changing the NG feed flowrate on the generated gasoline product and the associated CO₂ emissions. Another sensitivity analysis has been applied on GTL process to the effect of chain growth probability (α) on the products distribution according to Anderson-Schulz-Flory model.

Table 1: Feed specifications for the GTL and MTG models.

Parameter	GTL	MTG
NG molar flowrate (kmol/h)	37540	37540
NG Composition (mol%)		
Methane	92%	92%
Ethane	3%	3%
Nitrogen	5%	5%
NG feed Temperature (°C)	500	345
NG feed Pressure (bar)	25	30
Air to NG molar ratio	0.60	
Air Temperature (°C)	144	
Air Pressure (bar)	25	
Air Composition (mol%)		
Nitrogen	6%	
Oxygen	94%	
Steam to NG molar ratio	0.20	3.00
Steam Temperature (°C)	500	345
Steam Pressure (bar)	25	30

3. Results and discussion

The results of the techno-economic and environmental study demonstrate a higher overall product yield for MTG compared to GTL process with a higher net profit per product and CO₂ emissions. The detailed results of the two models are presented in Table 2. The raw material costs are calculated based on a NG feed price of \$137.2 per and the total product sales are calculated following a liquid fuels price of \$775.4 per tonne. The product distribution in GTL process is generated at an α value of 0.75 following Anderson-Schulz-Flory equation. The product distribution indicate a high share for gasoline with more than 50% followed by Naphtha 25%, Kerosene 22.5% and wax 1%. Whereas, the product in MTG process is totally gasoline with an overall yield of around 11% compared to 6% in the case of GTL process. The capital and operating costs are higher in the MTG process compared to the GTL process which was compensated by the higher production yield and sale. The raw material cost is the same for both cases as the same amount of NG is utilized. As a result, the net profit per product is higher for MTG process with \$568 per tonne of

product compared to \$558 for the GTL case. Similarity, the CO₂ emissions per product for MTG process is lower with 0.2 tonne of CO₂-e emitted per product compared to 1 for the GTL process.

Table 2: Summary of the main results for GTL and MTG processes.

Parameter	GTL	MTG
Liquid fuels product (kmol/h)	2151.43	4121.50
Liquid fuels product (tonne/y)	3.77x10 ⁶	4.20 x10 ⁶
Product distribution (wt%)	25.1% Naphtha	100% gasoline
	51.1% Gasoline	
	22.5% Kerosene	
	1.3% Wax	
Product yield	6%	11%
Total capital cost (\$)	2.20 x10 ⁸	4.62 x10 ⁸
Total operating cost (\$/y)	4.48 x10 ⁶	7.26 x10 ⁶
Total raw materials cost (\$/y)	7.70 x10 ⁸	7.70 x10 ⁸
Total product sales (\$/y)	2.92 x10 ⁹	3.25 x10 ⁹
Total annualized cost (\$/y)	8.20 x10 ⁸	8.72 x10 ⁸
Net profit (\$/y)	2.10 x10 ⁹	2.38 x10 ⁹
Net profit per product (\$/T product)	558.10	567.55
CO ₂ emissions (T CO ₂ -e/h)	433.07	93.45
CO ₂ emissions per product (T CO ₂ -e/T product)	1.0	0.2

The distribution of the liquid fuels generated from the GTL process are highly affected by the characteristics of the FT reactor and α probability following Anderson-Schulz-Flory equation (Al-Yaeeshi et al., 2019). The effect of α on the products distribution has been studied to determine the trend of gasoline production. Figure 3 demonstrate that the gasoline production increases with the increase in α until a certain peak point at around 0.75 which beyond the trend decreases. The curves of kerosene and wax indicate an increasing trends in general while the trend of Naphtha production decreases with the increase in α .

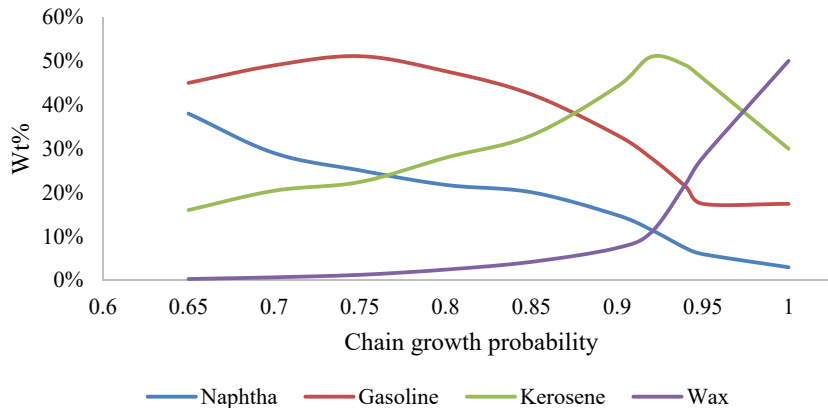


Figure 3: Effect of chain growth probability (α) on the products distribution of the GTL process.

Similarly, the effect of changing the NG feed flowrate on the generated gasoline product and the associated CO₂ emissions has been studied for MTG and GTL processes. The trends presentenced in Figure 4 indicate an increasing gasoline production and CO₂ emissions with the increase in NG feed. This increasing trend become less signification after 60,000 kmol/h of NG feed.

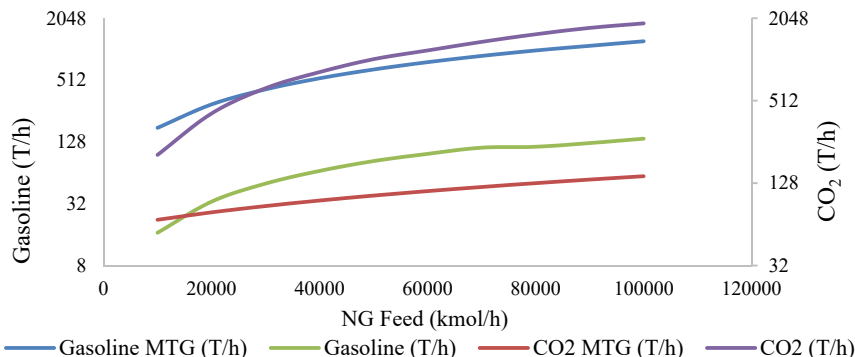


Figure 4: Effect of NG feed flowrate on the gasoline product and the associated CO₂ emissions.

4. Conclusions

The importance of transportation fuels is growing rapidly while the downstream investment in its production indicates critical operational costs. Hence, there is a need to establish alternative routes that archive higher sustainability in the generation of liquid fuels from fossil fuels. This study investigates the methanol to gasoline (MTG) production route as a GTL process alternative. The objective is to conduct rigorous simulation models for two NG utilization routes namely, GTL and MTG in order to help the decision-maker in gas processing industry. The results indicate a higher overall yield for the MTG process with a net economic value of around \$568 per tonne of gasoline produced and 0.2 tonne of CO₂ -e emitted per tonne of product. As a future work, the optimal percentage of methanol being processed to gasoline can be investigated as a trade-off between economic and environmental benefits.

References

- S.A. Al-Sobhi, and A. AlNouss, 2018, Applying New Sustainability Metric in Different Natural Gas Liquid (NGL) Recovery Configurations to Extend Investment Decision and Incorporate Sustainability Analysis, *Computer Aided Chemical Engineering*, 43, 145-50.
- S.A. Al-Sobhi, and A. Elkamel, 2015, Simulation and optimization of natural gas processing and production network consisting of LNG, GTL, and methanol facilities, *Journal of Natural Gas Science and Engineering*, 23, 500-08.
- A.A. Al-Yaeshi, A. AlNouss, G. McKay, and T. Al-Ansari, 2019, A Model based analysis in applying Anderson-Schulz-Flory (ASF) equation with CO₂ Utilisation on the Fischer Tropsch Gas-to-liquid Process, *Computer Aided Chemical Engineering*, 46, 397-402.
- A. AlNouss, M. Ibrahim, and S.A. Al-Sobhi, 2018, Potential energy savings and greenhouse gases (GHGs) emissions reduction strategy for natural gas liquid (NGL) recovery: Process simulation and economic evaluation, *Journal of Cleaner Production*, 194, 525-39.
- V. Ehlinger, K. Gabriel, M. Noureldin, and M. El-Halwagi, 2013, Process Design and Integration of Shale Gas to Methanol, *ACS Sustainable Chemistry & Engineering*, 2, 30-37.
- A. Galadima, and O. Muraza, 2015, From synthesis gas production to methanol synthesis and potential upgrade to gasoline range hydrocarbons: A review, *Journal of Natural Gas Science and Engineering*, 25, 303-16.
- C.L. Iandoli, and S. Kjelstrup, 2007, Exergy Analysis of a GTL Process Based on Low-Temperature Slurry F-T Reactor Technology with a Cobalt Catalyst, *Energy & Fuels*, 21, 4, 2317-24.
- J. Li, P. Miao, Z. Li, T. He, D. Han, J. Wu, Z. Wang, and J. Wu, 2015, Hydrothermal synthesis of nanocrystalline H[Fe, Al]ZSM-5 zeolites for conversion of methanol to gasoline, *Energy Conversion and Management*, 93, 259-66.
- A.P. Steynberg, R.L. Espinoza, B. Jager, and A.C. Vosloo, 1999, High temperature Fischer-Tropsch synthesis in commercial practice, *Applied Catalysis A: General*, 186, 1, 41-54.

Simplified targeting models for Sustainable Supply Chains retrofitting in process industries

Shabnam Morakabatchiankar^a, Fernando D. Mele^b, Moisés Graells^a, Antonio España^a

^a*Chemical Engineering Department, Universitat Politècnica de Catalunya, EEBE. Av. Eduard Maristany, 10-14, Edifici I, Planta 6, 08019 Barcelona, Spain*

^b*Departamento de Ingeniería de Procesos, FACET, Universidad Nacional de Tucumán (UNT), Avenida Independencia 1800, S. M. de Tucumán T4002BLR, Argentina
shabnam.morakabatchiankar@upc.edu*

Abstract

The main goals of retrofitting strategies are focused on addressing fundamental improvements in energy, environmental, and cost performance. This paper presents a novel formulation model aimed at determining optimal configurations for integrating a multi-energy type system to the process industries. The design and operation tasks are posed in mathematical terms as a bi-criteria mixed-integer linear programming problem (bi-MILP) that seeks to maximize the net present value (NPV) of the process supply chain while minimizing its environmental impact. Hence, the major contribution is the adoption of a mathematical approach to capture the cause-effect relationship between material/energy consumption/demand and the associated environmental impacts. The model capability is validated through a case study based on the Argentinean sugar cane industry and the benefits of integration are demonstrated.

Keywords: Retrofitting, MO optimization, sustainability, energy integration.

1. Introduction

Optimization strategies should be improved within the framework of industrial symbiosis systems to meet sustainability goals. Besides, the development of sustainable design and planning models for the process industries has strongly stimulated academia during the last three decades. Hence, the main goals of retrofitting strategies are focused on addressing fundamental improvements in energy, environmental, and cost performance. These strategies can be implemented in different levels as a unit, process segment, and overall system. For the case of process industries, the following challenges must be addressed (i) The integration of material/energy strategies in process industries; (ii) The integration of economic and environmental aspects within the framework of supply chain management; (iii) The development of efficient retrofit techniques to production processes. These challenges can be tackled by adopting a holistic systems-based analysis. Such an approach is supposed to propose an integrated solution by minimizing the global impact while considering feasibility constraints introduced by universal physical laws and current regulations. Particularly, multi-objective optimization is applied to the sustainable design and planning of a wide variety of industrial systems (Grossmann and Guillén-Gosálbez, 2010; Arora, 2012).

Subsequently, the incorporation of environmental concerns in SCs optimization models leads to Green Supply Chain Management. Hence, an approach has gained wider interest in recent years about combining mathematical programming and life cycle assessment and economic performance of the network. For instance, (Sabio et al., 2014) proposed an integrated LCA model via multi-objective mathematical programming to demonstrate that the combination of optimization techniques and LCA provides a powerful tool for balancing the environmental and economic performance of the production system.

However, very few types of approaches have been addressed multi-objective optimization of an integrated material/energy supply chain with considering economic and environmental impact factors. Hence, a general model that simultaneously optimizes the economic and environmental benefits of single-energy resource integration in process industries, particularly, exploiting bioenergy to run the production process, is the novelty proposed through this work. The proposed MILP model extends the traditional multi-objective approaches by considering the energy cogeneration, as a particular case, and its impact on the conflicting objectives of the whole system.

2. Problem statement

The production network retrofitting considers both its economic performance and the environmental impact (see Fig. 1). The network contains a set of raw materials and energy suppliers, production system utilizing various technologies, storage facilities, and final markets. For a more sustainable option, a penalty cost is assigned to the landfill. A standard high-level network is considered to determine the mathematical formulation, encompassing material resources that play the role of suppliers.

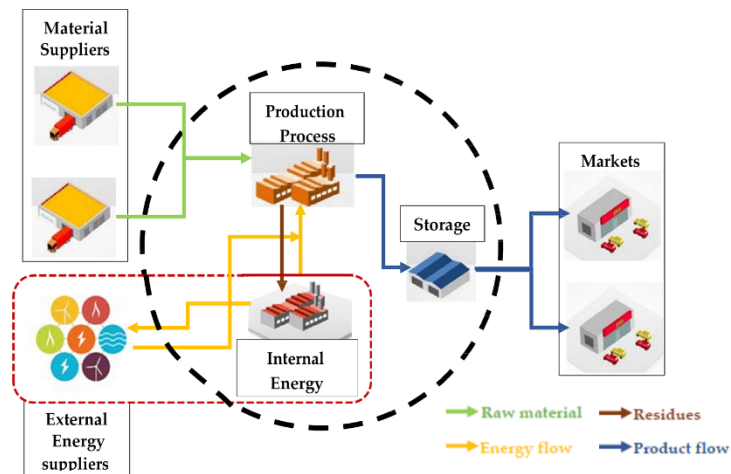


Fig. 1. Integrated energy/material SC Network

The network also includes process industries with different production technologies, acting as raw material consumers along with potential sites for storage technologies. The goal is to identify the best design and planning decisions about the system configuration and operation in terms of economic performance along with the environmental impacts.

3. Mathematical Model

The following equations formulate the network shown in Fig.1. In particular, mass and energy balances, as well as capacity constraints for each part of the network, are next described. A mixed-integer linear programming model is used to optimize sustainable supply chains. The proposed MILP is extended from the model introduced by (Mele et al., 2011). The emissions and wastes generated during the production tasks are considered in the model formulation. Any excess of energy can also be marketed.

3.1. Material balances and constraints

The mass balance for each potential site is defined by Eq.1. Hence, for each material type i , the initial inventory $ST_{i,s,g,t-1}$ maintains in region g from the previous period plus purchased raw material $PU_{i,g,t}$, the total produced material $PT_{i,g,t}$ and input material flow rate $Q_{i,l,g,g',t}$ is equal to the current inventory $ST_{i,s,g,t}$ added to delivered product $DT_{i,g,t}$ plus the output material flow $Q_{i,l,g,g',t}$, and generated waste $W_{i,g,t}$.

$$\sum_{s \in IS(i,s)} ST_{i,s,g,t-1} + PT_{i,g,t} + PU_{i,g,t} + \sum_{l \in (i,l)} \sum_{g' \neq g} Q_{i,l,g,g',t} = \sum_{s \in IS(i,s)} ST_{i,s,g,t} + DT_{i,g,t} + \sum_{l \in (i,l)} \sum_{g' \neq g} Q_{i,l,g,g',t} + W_{i,g,t} \quad \forall i, g, t \quad (1)$$

3.2. Energy balances and constraints

Due to single source energy generation, total energy demand is defined as follows:

$$\sum_{ei} [EnIJ_{ei,g,t} \times EfIJ_{ei}] + \sum_{ex} [EnXJ_{ex,g,t} \times EfXJ_{ex}] = TotalDemand_{g,t} \quad \forall g, t \quad (2)$$

Here, it is assumed that different types of energy are generated internally and purchased from external resources, reflected by ei and ex indices, respectively. The total amount of energy demand $TotalDemand_{g,t}$ of a process plant located in region g is equal to the summation of all different types of energy flows $EnIJ_{ei,g,t}$, generated inside the system, plus the total energy flow that comes from the external resources $EnXJ_{ex,g,t}$. The conversion efficiencies between energy resources and the process plant are represented by $EfXJ_{ex}$ and $EfIJ_{ei}$.

3.3. Objective functions

The model includes two objective functions, being the Net present Value (NPV) as the economic objective function whereas the environmental impact quantified regarding the Life Cycle Assessment (LCA) principles. The NPV can be determined from the discounted cash flows (CF_t) generated in each of the time intervals in which the total time horizon is divided, updated according to the interest rate (ir)

$$NPV = \sum_t \frac{CF_t}{(1+ir)^{t-1}} \quad (3)$$

In the current application, the total environmental damage is simplified to only consider the Global Warming Potential, through the GWP100 metric and calculated through the summation of damages could be caused in each stage of SC: the production of the main

feedstock GWPPU, the manufacturing and storage tasks GWPPE, and the transportation of materials between regions GWPQ. Hence,

$$DAM = GWPPU + GWPPE + GWPQ \quad (4)$$

$$GWPPU = GWP100PU \times \sum_i \sum_g \sum_t PU_{i,g,t} \quad (5)$$

$$GWPPE = \sum_i \sum_p \sum_g \sum_t GWP100PE_p \times PE_{i,p,g,t} \quad (6)$$

$$GWPQ = \sum_i \sum_l \sum_g \sum_{g'} \sum_t GWP100Q_l \times EL_{g,g'} \times Q_{i,l,g,g',t} \quad (7)$$

Here, $GWP100PU$, $GWP100PE_p$, and $GWP100Q_l$ represent the heat absorbed by any greenhouse gas (as a multiple of the heat that would be absorbed by the same mass of carbon dioxide) in the atmosphere by the feedstock production, main products operations and material transportation process, respectively. The environmental impact, as an objective function, is defined through the variable DAM as an environmental metric to be minimized.

$$DAM = GWPPU + GWPPE + GWPQ \quad (8)$$

3.4. Multi-objective equations

The mathematical model presented herein capitalizes on the mixed-integer linear programming (MILP) formulation and seeks to optimize simultaneously the NPV and DAM objectives described in the bi-dimensional objective function as presented in model M . The overall bi-MILP formulation can be expressed briefly as follow:

$$(M) \quad (9)$$

s.t. constraints 1-7

$$x \in R, X \in \{0, 1\}, N \in Z^+ \quad (10)$$

Note that, despite being the work by Mele et al. (2011) the basic model, particularly in those equations regarding production, it is extended to consider the energy section to amplify solutions borders and create more opportunities to make more flexible decisions.

4. Case Study: Retrofitting of integrated Sugar-bioethanol SCs

The proposed formulation is illustrated through its application to a case study based on the sugar-bioethanol industry of Argentina. The country is divided into 24 regions with associated ethanol, raw and white sugar demands. We use the data in Mele et al. (2011).

4.1. Cogeneration power plant

Cogeneration power plant is considered as a single source of energy generation to produce two types of energy (heat and electricity). It is assumed that the energy output per ton of sugarcane is 66 kW/t and the minimum energy generation coefficient is equal to 1. It is also assumed that the capacity of the power plant is 50 GW , and power

generation is available continuously for 7800 hours annually. On the basis of IEA (2010), the estimated installation costs of the cogeneration plant is $800 \frac{\text{€}}{\text{kW}}$.

5. Results

Regarding the aim of proposing a quantitative tool, we target at determining the optimal integration of renewable technologies to meet the energy (in this particular case electricity) demand in order to produce the main products (particularly sugar and ethanol) and cover the market demand over 12 months and during 4 years of planning horizon. The bi-criteria model was written in GAMS and solved with the MILP solver CPLEX 12.9 on an Intel® Core™ i7-3770 Octa-core Processor 3.40 GHz and 7.88 GB of RAM. The results include the environmental sustainability of the network to illustrate the environmental impacts minimization simultaneously with maximization of NPV. It takes approximately 1200 seconds to identify the global optimum in every instance. 12 Pareto points were generated, as shown in Fig. 1, in which nadir and utopia points have been also included.

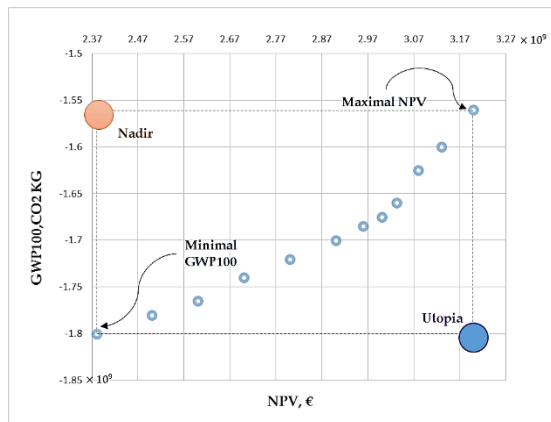


Fig. 1. Pareto set of solutions NPV vs GWP100

The decision on the optimum configuration of the system mainly depends on its local resource (energy/raw material) availability and the installation and operational costs. In this, the SC consists of three sugar mills and five distilleries. In the maximal NPV solution, all the production plants are implemented in 10 regions, 4 of them having sugarcane plantations. This optimality comes from the raw material availability and stability regarding climatic conditions during a year. There are cogeneration power plants in 4 regions that have the most sugarcane plantation capacity. The installed production technologies and the number of the power plants can be seen in Table 1 for both cases of maximal NPV and the minimal GWP100.

Table.1 Output data for the two extreme solutions

	NPV, €	GWP100, CO ₂ KG	Retrofitted Regions	Non Retrofitted Regions	Distillery Production Technology	Sugar Mill Production Technology
Maximal NPV	3.20×10^9	-1.56×10^0	4	6	T3, T4	T1, T2
Minimal GWP100	2.380×10^9	-1.8×10^0	10	0	T3	T1

Table 2. Regional Installed power capacity (kW) of power plant

Region	G13 (Salta)	G16 (Tucumán)	G17 (Jujuy)	G18 (Santa Fe)
Cogeneration power plants	2.5	2.5	58.6	378.8

Hence, they are autonomously able to provide their energy demand and Tucumán (G16) sells its exceeded energy to the grid. The results assure that for all demand levels, an energy surplus of 10% to 40% especially in lower demands create possibilities to market it. Hence, regions with constant climatic condition such G16, G17 and G18 may implement cogeneration plants and so advantage results in a high contribution of residue to energy in the energy/material SC and as it has the most effective share of 60% in the integrated energy/material SC and the NPV improves to 43% than stand-alone system.

6. Conclusions

In this work, we introduced a general model to optimize an integrated energy/material SC with economic and environmental concerns. The model includes four segments of equations: material production, energy generation and integrated energy/material equations and objective function. It is aimed to show the quantitative profitability of integration for both economic and environmental issues.

The model has been validated in a country size case study of Argentinean sugar cane industry that is considered 24 regions of Argentina as potential sites. For this case, several system configurations proposed and analyzed the exploitation benefits of renewable energies to generate different energy types required in a production process system.

Compared to Mele et al. (2011) work as an initial model, by considering a single type energy provider as an internal resource, the economic criteria represented by NPV, is increased 20% in a deterministic condition (Morakabatchiankar et al., 2017). Furthermore, by developing the model to a multi type energy generation, results show an increase of 43% in NPV. In the particular case study, not only we can satisfy the internal energy demand to produce main products that can satisfy 97%, 37% and 40% of the whole country demands for ethanol, white sugar and raw sugar respectively, but also there is generated 64 GWh of exceeded energy to sell.

Acknowledgments

Financial support received from the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (both funding the research Project AIMS, DPI2017-87435-R), and PCIN-2015-001/ELAC2014/ESE0034 is fully acknowledged.

References

- Arora, J. S. (2012) 'Multi-objective Optimum Design Concepts and Methods', in *Introduction to Optimum Design*. Elsevier, pp. 657–679. doi: 10.1016/b978-0-12-381375-6.00017-6.
- Grossmann, I. E. and Guillén-Gosálbez, G. (2010) 'Scope for the application of mathematical programming techniques in the synthesis and planning of sustainable processes', *Comput. Chem. Eng.* Elsevier, 34(9), pp. 1365–1376. doi: 10.1016/j.compchemeng.2009.11.012.

- IEA (2010) 'Energy Technology system analysis programme: Combined Heat and Power', (May), pp. 1–6. Available at: www.etsap.org.
- Mele, F. D. *et al.* (2011) 'Multiobjective Model for More Sustainable Fuel Supply Chains. A Case Study of the Sugar Cane Industry in Argentina', *Industrial & Engineering Chemistry Research*, 50(9), pp. 4939–4958. doi: 10.1021/ie101400g.
- Morakabatchiankar, S. *et al.* (2017) 'Developing a Multi-Objective Strategic-Tactical Optimization Model for Sustainable Production Supply Chains Considering Electricity Cogeneration: Sugar Cane Bioenergy Industry', in *Computer Aided Chemical Engineering*. Elsevier, pp. 2179–2184.
- Sabio, N. *et al.* (2014) 'Multiobjective optimization under uncertainty of the economic and life-cycle environmental performance of industrial processes', *AIChE Journal*. Wiley Online Library, 60(6), pp. 2098–2121. doi: 10.1002/aic.14385/full.

Flexibility-expansion planning for enhanced balancing-power market participation of decentralized energy systems

Niklas Nolzen^{a,b}, Ludger Leenders^{a,b}, André Bardow^{a,b,c,*}

^a*Institute for Technical Thermodynamics, RWTH Aachen University, 52062 Aachen, Germany*

^b*Energy & Process Systems Engineering, Department of Mechanical and Process Engineering, ETH Zürich, 8092 Zürich, Switzerland*

^c*Institute of Energy and Climate Research – Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany*
abardow@ethz.ch

Abstract

The rising share of renewable energy sources in power supply and the shut-down of conventional power plants lead to a need for new providers of balancing power. Balancing-power could be newly provided by the flexibilization of decentralized energy systems. In this contribution, we propose flexibility-expansion planning for decentralized energy systems to account for investments in additional units for flexibilization. Flexibility-expansion planning leads to a stochastic optimization model for optimal investment decisions towards increased operational flexibility. The stochastic optimization model consists of a design optimization and a two-stage stochastic program. In a case study, the method is applied to a decentralized energy system participating in the German tertiary balancing-power market. Savings of up to 5.9 % can be achieved with additional investments in heat storage. Thus, flexibility-expansion planning allows operators of decentralized energy systems to optimally invest in flexible technologies for improved balancing-power market participation.

Keywords: Ancillary service, utility system, stochastic optimization, control reserve

1. Introduction

Balancing power settles short-term imbalances between supply and demand in the electricity grid. Balancing power is an ancillary service offered by electricity providers or consumers. Positive balancing power is provided by increasing the demand or decreasing the supply; negative balancing power works vice versa. Today, conventional power plants supply most balancing power. However, many conventional power plants are shut down due to political, ecological, and economic reasons. These power plants no longer provide balancing power, while studies show an increasing demand for balancing power (DENA, 2014). Hence, new balancing-power providers are needed.

Deciding to provide balancing power is challenging due to the inherent uncertainty. Zhang et al. (2015) therefore use robust optimization to model the balancing-power market participation of an air-separation unit with cryogenic energy storage. The number of balancing-power requests is specified exogenously and the system is scheduled to the worst case of balancing-power requests. Kumbartzky et al. (2017) demonstrate savings for a combined-heat-and-power plant with heat storage participating in the German electricity and balancing-power market. Stochastic optimization is used to account for market uncertainties. However, the model excludes the actual request of balancing power. For a production system with an on-site utility system, Leenders et al. (2019) model the participation in the German tertiary balancing-power market. The operation of the overall

system is explicitly modeled if balancing power is requested. Schäfer et al. (2020) and Bohlayer et al. (2020) demonstrate savings from balancing-power market participation for an energy-intense aluminum mill and cement plant, respectively. However, only a few studies investigate investments in additional units to increase the flexibility of decentralized energy and production systems. Muche et al. (2016) optimize a combined-heat-and-power plant participating in the tertiary balancing-power market in Germany. The investment in heat storage is studied to increase flexibility. The investment decision itself is not part of the optimization problem and is only considered *a posteriori*. Hence, the sizing of storage is suboptimal and turns out to be non-profitable. Teichgräber and Brandt (2020) design a chlor-alkali electrolyzer with real-time electricity market data using stochastic optimization. Large differences in system design are shown depending on the shape and magnitude of the used market price data. The authors point out that other short-term markets should be considered in design optimization besides the real-time electricity market. Commonly, decentralized energy systems are designed only to cover exogenous demands of, e.g., heating, cooling, electricity. As reviewed, the balancing-power market participation is often studied for a given energy system design. Thus, the energy system is not designed for participation in balancing-power markets. In this work, a method is proposed for flexibility expansion of decentralized energy systems to increase balancing-power provision. The method employs a stochastic optimization model that combines long-term investment decisions with short-term operational decisions from participation in balancing-power markets. The operational model is based on Leenders et al. (2020). We extend this model with investment decisions and time-coupling storage equations.

2. Flexibility-expansion planning for balancing-power markets

The goal of flexibility-expansion planning is the optimal long-term investment for participation in the balancing-power market (Figure 1). Long-term investment decisions are considered by the design optimization. Therein, the investment in flexibility options is modeled to increase the flexibility of the decentralized energy system. The design optimization chooses the optimal capacity CAP_u of units $u \in U$, e.g., heat storage or electrode boiler. Short-term operational decisions are considered by a two-stage stochastic mixed-integer linear program. The stochastic program models the course of the information revelation in the balancing-power market and covers the uncertainty from the request of balancing power. Bids in the balancing-power market need to be made one day before operation. The 1st stage of the stochastic program models the pay-as-bid balancing-power market auction for the decentralized energy system. Bids in the balancing-power market consist of an amount of positive and negative balancing power $BP_{s,t}$, the capacity price $cp_{s,t}$ and the energy price $ep_{s,t}$ for each request scenario $s \in S$ and time step $t \in T$. The capacity price is paid for the provision of capacity in the balancing-power market. The energy price compensates for the respective request of balancing power. The energy price bid is discretized as in Leenders et al. (2019) with a Big-M formulation to avoid nonlinearities from the bidding decision. The 2nd stage of the stochastic program is the scenario-dependent operation $OP_{s,t,u}$ of the decentralized energy system. Since the request of balancing power is uncertain in each time step $t \in T$, three request scenarios $s \in S = \{ 'no', 'pos', 'neg' \}$ can occur with their probabilities $\omega_{s,t}$: no request of balancing power ('no'), and request of positive or respectively negative balancing power ('pos' / 'neg'). In each time step, the operation of the decentralized energy system depends on the request scenario for balancing power. In the scenarios 'pos' / 'neg', the produced electricity increases/decreases by the offered amount of balancing power $BP_{pos,t}$ and $BP_{neg,t}$, respectively. Hence, the operation of the decentralized energy system $OP_{s,t,u}$ is different for each request scenario and in each time step.

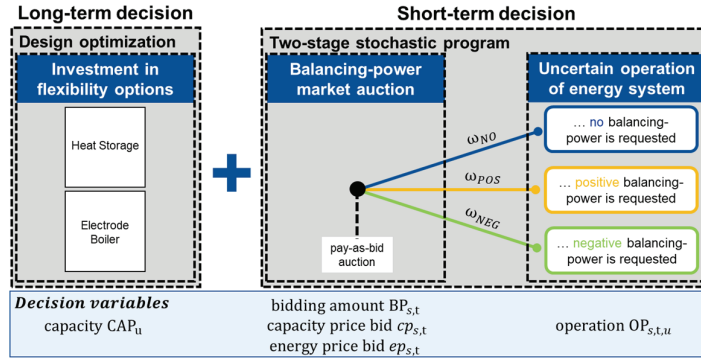


Figure 1: Flexibility-expansion planning combines a design optimization determining long-term investment decisions with a two-stage stochastic mixed-integer linear program for optimal operational decisions in the balancing-power market. Long-term and short-term decision variables are listed below the figure.

The long-term investment decisions cannot be taken independently without considering the short-term decisions. Hence, we combine both optimization models for long- and short-term decisions in the flexibility-expansion planning problem. The flexibility-expansion planning minimizes the total annual costs TAC subject to energy balances, storage equations, and other operational constraints of the energy system:

$$\begin{aligned} \min TAC &= CAPEX + OPEX & (1) \\ s.t. \text{ energy balances } &\forall p \in P, t \in T, s \in S & (2) \\ \text{ storage equations } &\forall p \in P, t \in T, s \in S & (3) \\ &\vdots & \end{aligned}$$

The total annualized costs TAC consider annualized capital expenditures $CAPEX$ in newly installed units $u \in U$ with capacity CAP_u and the yearly operational expenditures $OPEX$. The operational expenditures $OPEX$ arise from the scenario-based operation of the decentralized energy system $OP_{u,s,t}$. At each time step t , the operational expenditures $OPEX$ consider the costs for all three scenarios $s \in S$ weighted with the respective request probability $\omega_{s,t}$. The operational expenditures $OPEX$ consist of costs for gas and electricity purchase and revenues from the sale of electricity and the participation in the balancing-power market.

The energy balance needs to be fulfilled for all products $p \in P$, e.g., heating, cooling, electricity, in each time step $t \in T$ and request scenario $s \in S$. Details about the formulation of the operational expenditures and the energy balance formulation are given by Leenders et al. (2020).

2.1. Modeling of time-coupling energy storage

Energy storage is used to store a product in a specific time step and withdraw it at a later time step. Hence, energy storage couples the time steps in an optimization problem. Modeling energy storage in stochastic optimization increases complexity. In each time step, storage can operate in 3 modes, since the three request scenarios lead to different storage levels in general. In result, the number of possible storage levels grows with $3^{|T|}$. Hence, the problem gets computationally intractable with already a few time steps.

Here, we propose a method to restrict the possible storage level in each time step. To keep the problem tractable, we consider only one distinct storage level in the next time step. For this purpose, we propose two formulations based on different assumptions (Figure 2).

Same-storage-level formulation

In the same-storage-level formulation, the storage is operated the same regardless of the request of balancing power. This formulation leads to an operational schedule of the decentralized energy system, which is always feasible but restricts operational flexibility. Storage level $LEV_{p,t+1,s}^{st}$ of the next time step $t + 1$ is derived by withdrawing product $OUT_{p,t,s}^{st}$ or adding product $IN_{p,t,s}^{st}$ to the storage level $LEV_{p,t,s}^{st}$ in time step t

$$LEV_{p,t+1,s}^{st} = LEV_{p,t,s}^{st} \cdot \eta_p^l + \Delta_t \cdot \left(\eta_p^{st} \cdot IN_{p,t,s}^{st} - \frac{1}{\eta_p^{st}} \cdot OUT_{p,t,s}^{st} \right) \forall t \in T, s \in S, p \in P. \quad (4)$$

The storage losses η_p^l in each time step t are assumed to be a constant share of the storage level $LEV_{p,t,s}^{st}$. Furthermore, constant charge and discharge efficiencies η_p^{st} are assumed. In the same-storage-level formulation, the storage level $LEV_{p,t,s}^{st}$ is the same in each scenario $s \in S$ and, thus, the operational variables for charging $IN_{p,t,s}^{st}$ and discharging $OUT_{p,t,s}^{st}$ are the same in each scenario $s \in S$. Hence, the storage operation has no recourse. However, the problem remains a two-stage stochastic optimization as the other units $u \in U$ adapt to the respective scenario that materializes.

Lowest-storage-level formulation

In the lowest-storage-level formulation, the storage operates independently in each request scenario, but only the lowest storage level is considered in the next time step $t + 1$. Thereby, we avoid the exponential growth of scenarios but have to allow for overproduction if the storage is fully charged. The storage-level equation Eq. (4) is changed to an inequality:

$$LEV_{p,t,s}^{st} \cdot \eta_p^l + \Delta_t \cdot \left(\eta_p^{st} \cdot IN_{p,t,s}^{st} - \frac{1}{\eta_p^{st}} \cdot OUT_{p,t,s}^{st} \right) \geq LEV_{p,t+1,s}^{st} \forall s \in S, t \in T, p \in P. \quad (5)$$

In this formulation, no non-anticipativity constraints are needed for the operational variables $IN_{p,t,s}^{st}$ and $OUT_{p,t,s}^{st}$, but the non-anticipativity constraints for the storage level $LEV_{p,t,s}^{st}$ are kept. Each time step, the storage operates differently for each request scenario, but the storage level is equal to the lowest storage level. This formulation offers more operational flexibility of energy storage, but the energy storage might be sized too small in some unfavorable cases as overproduction is allowed.

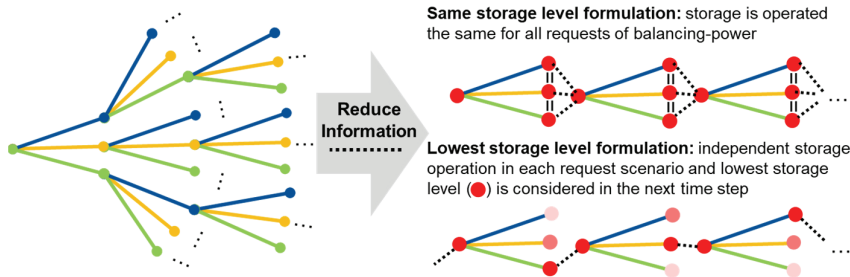


Figure 2: The scenario tree for an energy system with energy storage in stochastic optimization (left). The scenario tree grows exponentially due to time-coupling constraints. To keep the problem tractable, we propose 2 formulations (right) to reduce the information, i.e. storage level, considered in the next time step.

3. Case study

3.1. Description

Flexibility-expansion planning is applied to a real-world case study of a decentralized energy system based on Baumgärtner et al. (2019) with 3 combined-heat-and-power engines (2, 2.25, 3 MW), 2 boilers (5, 2 MW), 2 absorption chillers (1.4, 0.3 MW), and 3 compression chillers (2.2, 2.2, 1.3 MW) to cover a time-varying demand of heating, cooling, and electricity. Investments in heat storage and electrode boiler are considered. The energy system participates in the German tertiary balancing-power market while considering hourly fluctuating electricity and gas prices. We use market price data in the current balancing-power market design from 01.08.2019 to 31.07.2020. Capacity prices, discrete energy prices, and request probabilities are derived similarly to Leenders et al. (2019). The maximum amount of balancing power to be offered is set to 5 MW. Bids in the balancing-power market need to be the same for four hours to comply with the current market rules. The time-sensitive data are derived as time series of one year with an hourly resolution. We aggregate the time series with the TSA-module from Kotzur et al. (2018) using hierarchical clustering to receive one typical day with 24 time steps. Four cases are compared: No balancing-power market participation and no possibility to invest (*no market*), balancing-power market participation and no possibility to invest (*no invest*), balancing-power market participation and possibility to invest in an electrode boiler and heat storage using the same storage level formulation (*same*) or the lowest storage level formulation (*flexible*). All optimization problems are solved with the solver Gurobi on an Intel Xeon W-2155 @3.30 GHz with 128 GB RAM with an optimality gap of 1 % and a time limit of 7200 s. The 4 cases needed the following time to solve the problem to the predefined gap: 3 s (*no-market*), 13 s (*no invest*), 7200 s with a gap of 1.4 % (*same*) and 7200 s with a gap of 2.0 % (*flexible*).

Table 1: Expected costs (positive) and revenues (negative) for the four cases *no market*, *no invest*, *same* and *flexible* (upper part). Average offered amount of positive and negative balancing power at the typical day for the four cases *no market*, *no invest*, *same* and *flexible* (lower part).

<i>In 1000 €</i>	<i>No market</i>	<i>No invest</i>	<i>Same</i>	<i>Flexible</i>
Investment costs	0	0	12	11
Expected gas costs	2653	2687	2790	2831
Electricity costs	274	232	96	60
Balancing-power market revenues	0	-81	-97	-146
Total annual costs	2927	2838	2802	2756
Savings	0.0 %	3.0 %	4.3 %	5.9 %
Average positive balancing power [in MW]	0.0	0.33	0.67	2.0
Average negative balancing power [in MW]	0.0	4.5	4.5	4.5

3.2. Results

In the case study, flexibility-expansion planning saves 4.3 % (*same*) and 5.9 % (*flexible*) compared to the case *no market* (Table 1). Participation in the balancing-power market without investment saves 3.0 % compared to the case *no market*. Balancing-power market participation decreases electricity costs while gas costs increase. More electricity is produced to reduce electricity production upon request of negative balancing power.

In the case *same*, the optimal heat storage capacity is 2.9 MWh and in the case *flexible* 2.6 MWh. The smaller heat storage capacity in the case *flexible* than in the case *same* indicates that additional heat is not necessarily profitable to store when overproduction is possible. Instead of using the excess heat in other time steps, the heat storage is sized smaller in the case *flexible* than in the case *same* to reduce capital expenditures. Neither the case *same* nor the case *flexible* regard an investment in electrode boilers as beneficial.

The decentralized energy system offers an average amount of 4.5 MW of negative balancing power for the cases *no invest*, *same* and *flexible* (Table 1). Hence, the decentralized energy system has sufficient flexibility to provide negative balancing power. The average amount of offered positive balancing power increases with heat storage installation from 0.33 MW in the case *no invest* to 0.67 MW in the case *same* (Table 1). In the case *same*, heat storage gives heat-controlled combined-heat-and-power engines more flexibility since heat production is shifted to time steps with higher electricity demand. In the case *flexible*, the average offered amount of positive balancing power increases to 2.0 MW, as heat storage enables the combined-heat-and-power engines to be even more flexible.

4. Conclusions

The proposed flexibility-expansion planning combines long-term investment decisions with short-term operational decisions in one stochastic optimization model. For the flexibility-expansion planning, we proposed two methods to model energy storage in a stochastic optimization model for balancing-power market participation. In a case study of a decentralized energy system, flexibility-expansion planning achieves cost savings of up to 5.7 %. Thereby, flexibility-expansion planning supports participation in the balancing-power market.

Acknowledgment

This study is funded by the German Federal Ministry of Economic Affairs and Energy (ref. no.: 03EI1015A). The support is gratefully acknowledged.

References

- N. Baumgärtner, R. Delorme, M. Hennen, A. Bardow, 2019, Design of low-carbon utility systems: Exploiting time-dependent grid emissions for climate-friendly demand-side management, *Applied Energy*, 247, 755-765.
- M. Bohlayer, M. Fleschutz, M. Braun, G. Zöttl, 2020, Energy-intense production-inventory planning with participation in sequential energy markets, *Applied Energy*, 258, 113954
- DENA, 2014, dena-Studie Systemdienstleistungen 2030., Sicherheit und Zuverlässigkeit einer Stromversorgung mit hohem Anteil erneuerbarer Energien, Endbericht
- L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, 2018, Impact of different time series aggregation methods on optimal energy system design, *Renewable Energy*, 117, 474-484.
- N. Kumbartzky, M. Schacht, K. Schulz, B. Werners, 2017, Optimal operation of a CHP plant participating in the German electricity balancing and day-ahead spot market, *European Journal of Operations Research*, 261, 1, 390-404.
- L. Leenders, A. Starosta, N. Baumgärtner, A. Bardow, 2020, Integrated scheduling of batch production and utility systems for provision of control reserve, *Proceedings of ECOS 2020*, 699-710.
- T. Muche, C. Höge, O. Renner, R. Pohl, 2016, Profitability of participation in control reserve market for biomass-fueled combined heat and power plants, *Renewable Energy*, 90, 62-76.
- P. Schäfer, H. G. Westerholt, A. M. Schweidtmann, S. Ilieva, A. Mitsos, 2019, Model-based bidding strategies on the primary balancing market for energy-intense processes, *Computers and Chemical Engineering*, 120, 4-14.
- H. Teichgraeber, A. R. Brandt, 2020, Optimal design of an electricity-intensive industrial facility subject to electricity price uncertainty: Stochastic optimization and scenario reduction, *Chemical Engineering Research and Design*, 163, 204-216.
- Q. Zhang, I. E. Grossmann, C. F. Heuberger, A. Sundaramoorthy, J. M. Pinto, 2015, Air Separation with Cryogenic Energy Storage: Optimal Scheduling Considering Electric Energy and Reserve Markets, *AIChE Journal*, 61, 5, 1547-1558.

A fair profit allocation model for the distribution plan optimisation of refined products supply chains

Bohong Wang,^{a,b} Jiří Jaromír Klemeš,^{a,*} Taicheng Zheng^b, Yongtu Liang^b

^a*Sustainable Process Integration Laboratory – SPIL, NETME Centre, Faculty of Mechanical Engineering, Brno University of Technology – VUT Brno, Technická 2896/2, 616 69 Brno, Czech Republic*

^b*National Engineering Laboratory for Pipeline Safety/Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, Fuxue Road No.18, Changping District, Beijing, 102249, China*

Abstract

Refined products supply chain optimisation is a crucial issue in the oil industry. The task is to transport refined products from refineries to storage depots and retail markets through multiple transport modes. The optimisation of the refined products supply chain has been mostly based on maximising the total profit. However, each company, along with the supply chain, pursues their own maximum profit. The solution obtained by the total chain method may not satisfy all the participants in the refined products supply chain. Bargains have existed between these companies, and steps are required to adjust the transport plan, which may affect the optimality of the plan. Also, the bargain abilities of these companies are different. A mixed-integer nonlinear programming (MINLP) model considering both cooperative game theory and transfer prices to simulate the behaviours of stakeholders is proposed to solve this problem of the optimal refined products distribution planning. Material balance, production rate, transport modes and capacities are developed as constraints. Nash bargaining solutions for the multi-player bargaining game are obtained via maximising a tailored objective function. In this way, a refined products distribution plan with more fair profit allocation can be obtained, and the bargaining ability of each participant can be considered. The model is applied to a refined products supply chain in China as a case study to show its validity and applicability. This method can also be used in the optimal planning of other refined products supply chain and help to enhance their performance.

Keywords: Refined Products Supply Chain (RPSC), fair profit allocation, mixed-integer nonlinear programming (MINLP), Nash bargaining.

1. Introduction

As a sort of primary energy, the consumption of the refined products achieved 325.14 Mt in 2018, with 6% year-on-year growth in China (National Development and Reform Commission, 2019). Refined products supply chain (RPSC) plays an important role in linking refineries and retail markets. Optimisation of the RPSC is widely studied from many aspects to reduce the operating cost and improve efficiency (Sule et al., 2011). Wang et al. (2019) developed a mixed-integer linear programming (MILP) model that considered both RPSC distribution and new pipeline route planning to achieve the goal of minimising annual transport and pipeline investment costs. Lima et al. (2018) developed a multistage stochastic programming model for the optimal distribution of

refined products. Yuan et al. (2019) developed a framework that coupled a mathematical programming technique with the energy-economy-environment (3E) assessment method to quantify the impacts of pipeline network reform on the downstream oil supply chain of China. Zhang et al. (2019) proposed a stochastic linear programming model for the reliable design of the oil products supply chain. The stochastic hub disruption and the demand uncertainty were considered, and the effect of uncertainties on the supply chain design was analysed. Wang et al. (2020a) proposed a framework for reliability assessment of a regional RPSC under several scenarios. Yuan et al. (2020) evaluated the province- and region-level supply security of refined products in China by an integrated assessment framework. Wang et al. (2020b) developed a framework for the optimisation of downstream oil supply chains with emission-cost nexus using P-graph model (Friedler et al., 1992).

However, the majority of companies in the supply chain pursue their maximum profit. The solution obtained by the current methods may not satisfy all the participants in the RPSC. Bargains have existed between these companies, and steps are required to adjust the transport plan. In this case, it is practical to optimise the operation of RPSCs, considering the profit allocation. Gjerdrum et al. (2001) proposed a MILP model for fair and optimised profit distribution between members in a supply chain. Liu and Papageorgiou (2018) developed a MILP model framework with the consideration of fair profit distribution to optimise an active ingredient supply chain. Zheng et al. (2020) studied the fair profit allocation problem in third-party take-back waste cooking oil-to-biodiesel supply chain, and they also compared different policy impacts on integrated and opening markets. In this paper, a MILP model is developed for the fair profit allocation of an RPSC. A Nash-bargaining solution is solved to obtain the distribution plan. Results of total profit maximisation and Nash-bargaining solution are compared to show how the proposed model can lead to a fair profit allocation between participants in the RPSC.

2. Problem description

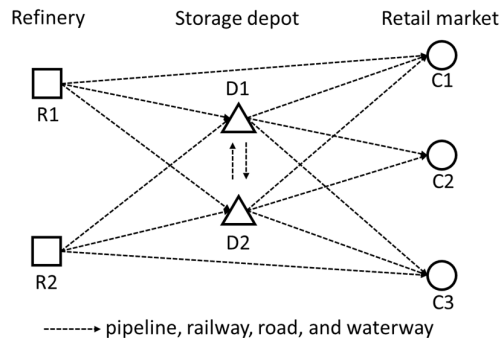


Figure 1 Schematic diagram of an RPSC

The RPSC in China has 3 echelons: refinery, storage depot, and the retail market. Refineries produce refined products and sell them to storage depots and retail markets. Storage depots purchase refined products from refineries and sell them to retail markets. Storage depots usually have advantages of lower transport cost. Some have pipelines link from the upstream and link to the downstream. Some have access to water transport. The unit transport prices of these two ways are lower than transport by rail and road. Retail markets purchase refined products from refineries and storage depots to satisfy the

demand. If the demand is not satisfied, a lost sale cost would be counted. The schematic diagram of the RPSC is shown in Figure 1.

3. Mathematical model

Two models are developed, one is to maximise the overall profit of the whole RPSC, and the other is to maximise the profit based on the fair profit allocation. In this study, Nash-bargaining is applied to get a fair profit solution. The objective function of the Nash-bargaining is shown as follows:

$$\text{maximize } F = \sum_{i \in I} \ln Pr_i^I + \sum_{j \in J} \ln Pr_j^J + \sum_{k \in K} \ln Pr_k^K \quad (1)$$

Where Pr_i^I , Pr_j^J , and Pr_k^K are the profit of refinery i , storage depot j , and retail market k .

The basic constraints for supply chain optimisation include material balance and capacity constraints. These constraints can be referred to Wang et al. (2019).

In this model, there are three parts of profit constraints. The first is for refineries, the second for storage depots, and the third for retail markets. For the refineries, their profits are obtained according to sales revenue, depreciated investment, construction cost, purchasing cost of crude oil, and inventory cost. For the storage depots, their profit equations consist of sales revenue, refined products purchasing cost, inventory cost, and transport cost. For the retail markets, sales revenue, refined products purchasing cost, inventory cost, transport cost, and cost for lost sales make up the profit equations.

For the cooperative bargaining mode, Nash bargaining is adopted in this model. Except for the basic constraints for the RPSC optimisation, following transfer price constraints should be modelled.

$$TP_{i,p}^I = \sum_{e \in E} TPL_{p,e}^I \cdot X_{i,p,e}^I, \forall i, p \quad (2)$$

$$TP_{j,p}^J = \sum_{e \in E} TPL_{p,e}^J \cdot X_{j,p,e}^J, \forall j, p \quad (3)$$

$$\sum_{e \in E} X_{i,p,e}^I = 1, \forall i, p \quad (4)$$

$$\sum_{e \in E} X_{j,p,e}^J = 1, \forall j, p \quad (5)$$

Where $TP_{i,p}^I$ is the transfer price for refined product p from refinery i , $TP_{j,p}^J$ is the transfer price for refined product p from storage depot j , $TPL_{p,e}^I$ and $TPL_{p,e}^J$ are parameters that indicate the transfer prices when price level e is selected for refineries and storage depots, $X_{i,p,e}^I$ is a binary variable which indicates if the price level e is chosen for refined product p from refinery i , and $X_{j,p,e}^J$ is a binary variable which indicates if the price level e is chosen for refined product p from storage depot j . Constraints (2) and (3) obtain the transfer prices, and constraints (4) and (5) ensure that only one price level should be chosen for each refinery and storage depot.

4. Case study

In this section, an RPSC in East China is studied as an example. As shown in Figure 2, the supply chain includes refineries (represented by dot i), storage depots (represented by triangle j) and retail markets (represented by diamond k). Refined products can be transported between each individual by pipeline, railway, road, and waterway. The production capacities of refineries and storage capacities of storage depots are shown in Table 1 and Table 2. In this case, a total of 10 periods are selected as the study period; each period is one month long.

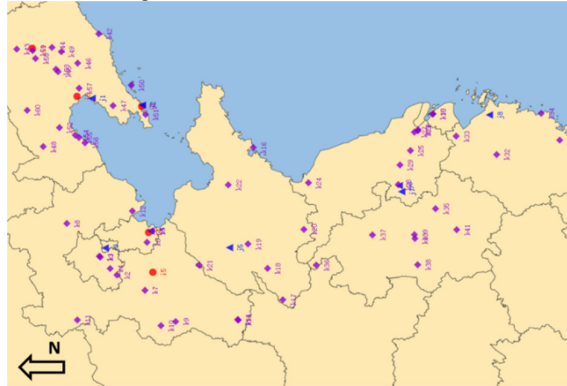


Figure 2 Locations of facilities in the studied RPSC

Table 1 Production capacities of refineries

Refineries	Petrol (kt/month)	Diesel oil (kt/month)
i1	6,250	9,500
i2	8,500	16,500
i3	2,600	4,900
i4	6,600	7,850
i5	9,000	5,700

Table 2 Inventory capacities of storage depots

Storage depots	Inventory capacities of petrol (kt)	Inventory capacities of diesel oil (kt)
j1	66.5	90.25
j2	95	114
j3	95	171
j4	38	58.9
j5	3.8	53.2
j6	47.5	57.95
j7	76	86.45
j8	142.5	209

The profit allocations of different individuals are shown in Figures 3, 4 and 5. It can be seen from Figures 3 and 4 that in the profit allocation of refineries and storage depots, the effect of the Nash bargaining solution to make a fairer profit allocation is not obvious. This is because refineries and storage depots handle a relatively large quantity of refined products in the RPSC than retail markets. To satisfy the demand for retail markets, these

upstream facilities has to be used. Also, it can be observed that the profit allocation of refineries and storage depots are more related to their handling capacities.

As shown in Figure 5, in the maximising total profit model, the profits of some individuals will be sacrificed to achieve the goal of maximising the profit of the whole supply chain. Taking retail market k17 as an example, it obtains a negative profit around $-4,800 \times 10^6$ CNY, but under this circumstance, the entire supply chain achieves the largest total profit. The distance between k17 and refineries is quite far. The solved model suggests a shortage in this retail market to increase total profit. In the Nash bargaining solution, the profit of k17 is closer to other retail markets.

The maximising total profit model can also lead to some individuals to make more profit, such as k52, k53, and k54. Their profits are significantly higher than in other retail markets. In the Nash bargaining model, the profit allocation tends to be fair.

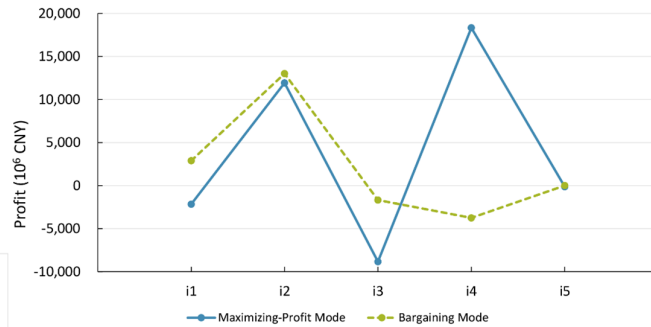


Figure 3 Profit distribution among refineries

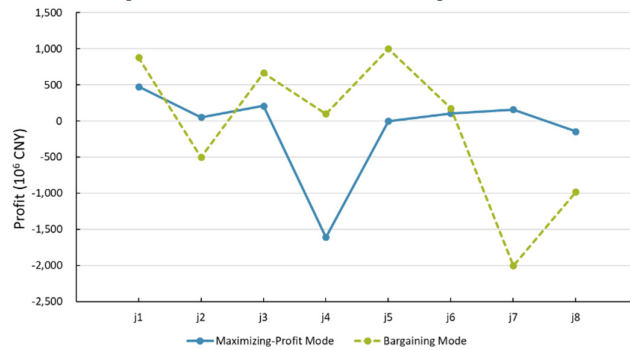


Figure 4 Profit distribution among storage depots

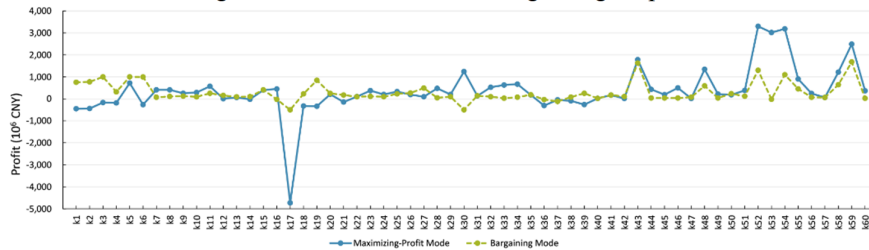


Figure 5 Profit distribution among retail markets

5. Conclusions

This paper studies the fair profit allocation in the RPSC, and a mathematical model is developed. A case is studied to test the proposed method. The results of the Nash-bargaining solution are illustrated. Results show that the Nash-bargaining solution will reduce the profit of markets by 19.3% compared to the integration, while a fairer profit allocation can be achieved. Future studies will focus on the performance evaluation of RPSCs, indicators including the supply rate of refined products and operating costs would be proposed. The sensitivity of alternative fuel prices, as well as the products purchasing, processing, and sale prices, should be further analysed.

6. Acknowledgements

The EU project "Sustainable Process Integration Laboratory – SPIL", project No. CZ.02.1.01/0.0/0.0/15_003/0000456 funded by EU "CZ Operational Programme Research, Development and Education", Priority I: Strengthening capacity for quality research and the National Natural Science Foundation of China (51874325).

References

- National Development and Reform Commission, 2019. Overview of refined products operation in 2018. <https://www.ndrc.gov.cn/fgsj/tjsj/jjyx/mdyqy/201901/t20190131_1024891.html> accessed 30 Nov 2020.
- F. Friedler, K. Tarjan, Y.W. Huang, L.T. Fan, 1992, Combinatorial algorithms for process synthesis. *Computers & Chemical Engineering*, 16, S313–S320.
- J. Gjerdrum, N. Shah, L.G. Papageorgiou, 2001, Transfer Prices for Multienterprise Supply Chain Optimisation, *Industrial & Engineering Chemistry Research*, 40(7), 1650-1660.
- C. Lima, S. Relvas, A. Barbosa-Póvoa, 2018, Stochastic programming approach for the optimal tactical planning of the downstream oil supply chain, *Computers & Chemical Engineering*, 108, 314-336.
- S. Liu, L.G. Papageorgiou, 2018, Fair profit distribution in multi-echelon supply chains via transfer prices, *Omega*, 80, 77-94.
- Z. Sule, B. Bertok, F. Friedler, L. Fan, 2011, Optimal Design of Supply Chains by P-graph Framework Under Uncertainties, *Chemical Engineering Transactions*, 25, 453-458.
- B. Wang, Y. Liang, T. Zheng, M. Yuan, H. Zhang, 2019, Optimisation of a downstream oil supply chain with new pipeline route planning, *Chemical Engineering Research and Design*, 145, 300-313.
- B. Wang, H. Zhang, M. Yuan, Z. Guo, Y. Liang, 2020a, Sustainable refined products supply chain: A reliability assessment for demand - side management in primary distribution processes. *Energy Science & Engineering*, 8, 1029-1049.
- B. Wang, Y.V. Fan, H.H. Chin, J.J. Klemeš, Y. Liang, 2020b, Emission-cost nexus optimisation and performance analysis of downstream oil supply chains, *Journal of Cleaner Production*, 266, 121831.
- M. Yuan, H. Zhang, B. Wang, R. Shen, Y. Long, Y. Liang, 2019, Future scenario of China's downstream oil supply chain: An energy, economy and environment analysis for impacts of pipeline network reform, *Journal of Cleaner Production*, 232, 1513-1528.
- M. Yuan, H. Zhang, B. Wang, L. Huang, K. Fang, Y. Liang, 2020, Downstream oil supply security in China: Policy implications from quantifying the impact of oil import disruption, *Energy Policy*, 136, 111077.
- W. Zhang, Z. Li, Q. Liao, H. Zhang, B. Wang, S. Huang, N. Xu, Y. Liang, 2019, A Stochastic Linear Programming Method for the Reliable Oil Products Supply Chain System With Hub Disruption, *IEEE Access*, 7, 124329-124340.
- T. Zheng, B. Wang, M.A. Rajaeifar, O. Heidrich, J. Zheng, Y. Liang, H. Zhang, 2020, How government policies can make waste cooking oil-to-biodiesel supply chains more efficient and sustainable, *Journal of Cleaner Production*, 263, 121494.

Markov Decision Process to Optimise Long-term Asset Maintenance and Technologies Investment in Chemical Industry

Hon Huin Chin*, Bohong Wang, Petar Sabev Varbanov, Jiří Jaromír Klemeš

*Sustainable Process Integration Laboratory – SPIL, NETME Centre, Faculty of
Mechanical Engineering, Brno University of Technology – VUT Brno, Technická
2896/2, 616 69 Brno, Czech Republic*

Abstract

The decisions on synthesising a process network are often to optimise the payback periods based on investment cost. In addition to the core investment and the cost of used resources, the long-term reliable operation of the process is also crucial. Given available states and technologies of the assets, this study aims to identify the long-term optimal asset planning policy. Markov Decision Process (MDP) is a promising tool in identifying the optimal policy under different states of the assets or equipment. The failure probability of the unit is modelled with the ‘bathtub’ model and each of the condition states are incorporated in the MDP. The decisions to implement the redundant units in the process with variety of technologies are allowed. This paper applied the MDP into an equivalent Mixed Integer Non-linear Programming (MINLP) to solve for the optimal long-term assets decision and the maintenance policy. The applicability of the method is tested on a real case study from Sinopec Petrochemical Plant. The capital and expected operational cost that accounts for equipment maintenance for an infinite time horizon are determined.

Keywords: Markov Decision Process, Asset Optimisation, Reliability, Maintenance Planning, Mixed Integer Non-linear Programming (MINLP)

1. Introduction

The availability, reliability and maintenance planning are significant factors in the project economics when designing the process plant in the early stage. Goel et al. (2003) mentioned that the operating and the capital cost should be affected by the inherent availability of the system. The recent review by Chin et al. (2020) analysed different strategies of maintenance considering individual process plants.

Vassiliadis and Pistikopoulos (2001) derived a MINLP model with a two-step strategy to identify optimal maintenance policies in continuous process operations. The first step involves generating structural design based on availability threshold values and maintenance time, while the second step involves operative-failed state-space model to determine the optimal policies. Several studies on improving the equipment availability can be found in Godoy et al. (2015) for hydropower or Süle et al. (2019) for petrochemical system. industry or Andiappan et al. (2019) in biomass trigeneration system. Ye et al. (2018) provided a general mixed-integer framework for selecting the

standby processing units to achieve optimal availability and cost. Ye et al. (2020) recently applied the Markov Decision Process (MDP) approach in synthesising the design optimisation, storage sizing and maintenance policy for the air separation process.

A variety of studies using MINLP frameworks to address asset availability issue have been proposed. However, there are still lacking system availability design considering different asset technologies, i.e. different reliability parameters or different sizes. The expected benefits of installing different asset technologies should be analysed by considering their long-term effects on the system efficiency and availability. Markov Decision Process (Bellman, 1957) is a framework that evaluates the optimal policies under different equipment states by optimising the long-term benefits (value functions) of each state. This method provides suggestions on actions for the equipment regardless of the equipment initial states. This means the optimal actions determined should be carried out for a specific equipment state to achieve expected long-term benefits.

Realising from the research gaps identified, this work aims to adapt the MDP formulation from Ye et al. (2020) to model the system availability design. The inspection intervals, the selection of the optimal technologies and the redundancy design are incorporated into the MDP with MINLP formulation. The stage interdependencies are considered within the state transitioning probability matrix.

2. Methodology

As most of the chemical processes are production systems with serial structures, a generic serial process structure is used to evaluate the reliability of the system. Each stage $i \in \mathbf{I}$ is dedicated to the subsystem/individual process unit with parallel units. The parallel units or redundant units are used as the standby so that when the operating unit is failed, the standby unit can be used immediately as a replacement to avoid downtime. One of the goal of this study is to determine the optimal number of redundant components in each stage. For each stage there are finite numbers of redundant design $j \in \mathbf{J}$, for which $j=2$ indicates there are 2 units installed in the stage. Each subsystem can also be implemented with assets technology $z \in \mathbf{Z}$ that has different reliability parameters, capacity or efficiency. Figure 1 shows the overall structure of the model.

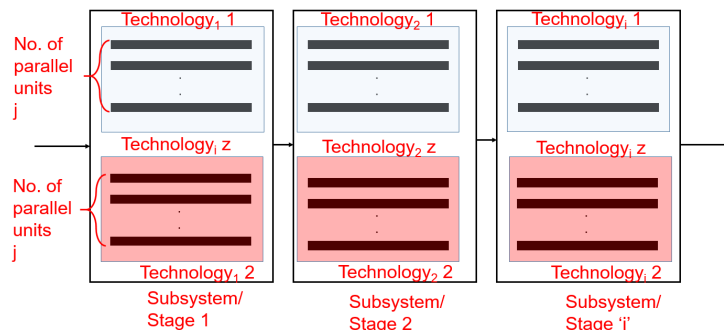


Figure 1: A generic structural representation of the problem

In this work, the condition of the equipment is modelled with the 'bathtub' model as the asset undergoes different period where their failure rates vary with time. As proposed by

Barlow and Proschan (1975), the deterioration of the asset is usually distinguished into three periods: ‘Infant’, ‘Stable’ and ‘Worn-out’.

Along with the three states as mentioned, the condition state of the equipment also includes states ‘Stopped’ and ‘Stand-by’. These states form the finite sets of condition states of the equipment $s \in \mathbf{S}$. Under different states, there are finite sets of possible actions $a \in \mathbf{A}$ that can be performed on the equipment: ‘None’, ‘Inspect’, ‘Maintain’ or ‘Stop’. The common approach to model asset maintenance plain using Markovian approach is to formulate the Continuous-time Markov Decision Process. For this approach, all of the equipment states in each stage are incorporated into the model. The approach is comprehensive; however, the applicability is hindered by the curse of dimensionality. Consider just two states exist for a piece of equipment: ‘Operating’ or ‘Failed’, if there are total of ‘k’ subsystems in the process, the number of available states is 2^k , and this is excluding the redundant units. If there are ‘n’ redundant units installed in each subsystem, the problem grows exponentially huge as the number of available states increases, as shown in Ye et al. (2020).

To overcome this issue, the MDP can be modelled in another way. The state-action representation can be modelled with each equipment in each stage and can be linked together with their transitioning probability, adopted from Ye et al. (2020). Figure 3 shows the state diagram for a single stage, with two redundant units installed in this work. It is assumed that the interdependency between units is affected by the failure of other subsystem or stage. For example, if one of the unit in any stage is failed, the other stages are not functional as well due to the serial structure. If any redundant is used, the stages go into ‘Standby’ state and resume to functional using the redundant equipment.

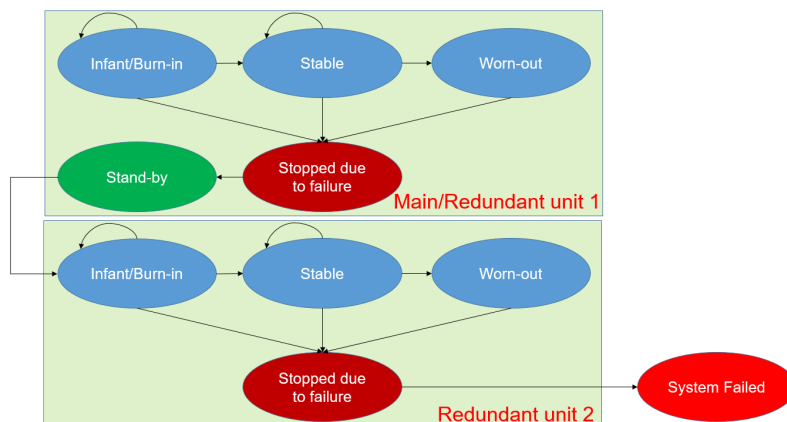


Figure 2: A state diagram MDP for a single stage, with two redundant units

3. Model formulation

The most common approach to solve a MDP is based on the Bellman Expected Optimality equation (Bellman, 1957) using value iteration or policy iteration. As the problem becomes complicated, these two approaches are not suitable and the problem can be converted to an optimisation formulation. The expected long-term cost to be minimised for the MDP is shown in Eq (1), with the constraint in Eq (2). These two equations are the basis for equivalent optimisation model for a MDP.

$$\sum_{i \in I, j \in J, z \in Z} y_{i,j,z} Cap_{i,j,z} + \sum_{i \in I, j \in J} \sum_{s \in S} \pi_{i,j}(s) v_{i,j}(s) \quad (1)$$

$$v_{i,j}(s) \geq \sum_{s' \in S} P_{i,j}(s, a, s') \sum_{z \in Z} y_{i,j,z} R_{i,j,z}(s, a, s') + \gamma \sum_{s' \in S} P_{i,j}(s, a, s') v_{i,j}(s') \quad (2)$$

$\forall i \in I, j \in J, s \in S, a \in A$

The variable $y_{i,j,z}$ is the binary variable indicating the selection of redundant numbers ‘j’ and technology ‘z’ for stage ‘i’; $Cap_{i,j,z}$ is the capital cost for the stage ‘i’ with ‘j’ numbers of redundant units using technology ‘z’. $\pi_{i,j}(s)$ is the probability distribution of state ‘s’ for the stage ‘i’ that is determined by the model. The variable $v_{i,j}(s)$ is the value function of each state ‘s’ in the MDP which it represents the ‘discounted’ rewards in state ‘s’. The rewards in this case is the maintenance cost for the system. $P_{i,j}(s, a, s')$ is the state transitioning probability from state ‘s’ with action ‘a’ to state ‘s’’, where this corresponds to the stage ‘i’ and redundant design ‘j’. $R_{i,j,z}(s, a, s')$ is the instant rewards when transitioning from state ‘s’ to state ‘s’ by action ‘a’, with selected asset technology ‘z’. As the problem involves minimisation of the total cost, the instant reward parameters are the instant maintenance cost associated with the actions. The γ is the discount factor which is fixed at 0.1 for this work. The constraint for $y_{i,j,z}$ is fixed with Eq (3), where only one redundant design and one technology can be selected. Eq (4) represents only one action is selected only for state ‘s’, for stage ‘i’ provided redundant number ‘j’ is selected. $x_{i,j}(s, a)$ is a binary variable indicating the selection of action ‘a’ in state ‘s’, for stage ‘i’ and redundant ‘j’.

$$\sum_{j \in J, z \in Z} y_{i,j,z} = 1 \quad \forall i \in I \quad (3)$$

$$\sum_{a \in A} x_{i,j}(s, a) = \sum_{z \in Z} y_{i,j,z} \quad \forall i \in I, j \in J, s \in S \quad (4)$$

Eqs (5) to (7) below indicate the constraints for the stationary probability distribution of the MDP, $\pi_{i,j}(s)$. Eq (6) is introduced as the sum of probability distribution for any state ‘s’ should exist only if the redundant unit ‘j’ and technology ‘k’ is selected. $\pi_{i,j}^{policy}(s, a)$ indicates actions ‘a’ is preferred in the state ‘s’ if its value > 0. Eq (7) indicates the total selected policy for a single state is equal to the probability distribution of state ‘s’.

$$\sum_{s \in S, a \in A} x_{i,j}(s, a) \pi_{i,j}^{policy}(s, a) P_{i,j}(s, a, s') = \pi_{i,j}(s') \quad \forall i \in I, j \in J, s' \in S \quad (5)$$

$$\sum_{s \in S} \pi_{i,j}(s) = \sum_{z \in Z} y_{i,j,z} \quad \forall i \in I, j \in J \quad (6)$$

$$\sum_{a \in A} x_{i,j}(s, a) \pi_{i,j}^{policy}(s, a) = \pi_{i,j}(s) \quad \forall i \in I, j \in J, s \in S \quad (7)$$

The states transition probability and the instant rewards adapted from Ye et al. (2020) are defined by Eqs (8) to (9). In this work, the maintenance cost involves a certain fixed cost plus the downtime cost due to maintenance duration. The inspection cost is based on fixed labour cost, and the shutdown cost is calculated with the production loss cost due to stopping the operation. The probability from ‘stopped’ to ‘standby’ is if more than one redundant unit is used.

$$P_{i,j}(s, a, s') = \{1 - e^{-\left(\frac{T_{infant}}{\theta}\right)^{b_1}}\} \quad \text{if } (s, a, s') = (Infant, None, Stopped), \quad e^{-\left(\frac{T_{in}}{\theta}\right)} \quad (8)$$

$$R_{i,j}(s, a, s') = \{C_{maintain} \quad \text{if } a = 'Maintain' \text{ and } s \neq 'Stopped' \quad C_{inspect} \quad (9)$$

4. Case study and results

The methodology is applied to a real case study from Sinopec Luoyang Petrochemical Plant from Hu et al. (2009). It is a naptha reforming reaction system that process 2,500 t/d of refined naptha, with subsystems of reactors, heat exchangers, compressors and pumps and absorbers. The full documentation of the maintenance records, as well as the inspection, shutdown and maintenance cost parameters for the equipment in the plant can be found in the reference source. The case study allows maximum of three redundant units to be installed in each equipment position. Different technologies are available by varying the shape parameter of the failure function by multiplying a random number, i.e. $\mu_1 \times b_2$, where μ_1 is a random number between 0 and 1. The capital cost for each technology of the unit is also assumed to be varied by the equation: $(1 - \mu_2) \cdot Base\ capital_i$, where μ_2 is a random number between -0.5 and 0.5. The MINLP model is solved using GAMS software with random number seed generator set as zero. The total expected optimal cost is \$ 2.4 M, with long-term risk cost as \$ 0.19 M/y (\$ 0.25 M/y from Süle et al. (2009)). The inspection intervals for each equipment range from 2 to 4 months.

5. Conclusion

A discrete-time Markov Decision Process has been proposed to evaluate the long-term economic benefits of different assets technologies and redundant unit allocation in a chemical process. The discrete decisions and the inspection periods are translated to MINLP model to optimise the long-term discounted cost. The method is tested on a real case study from Sinopec petrochemical plant. This approach is regardless to the initial state of the subsystem, which means that the optimal policies identified can be followed regardless of the states of the equipment. However, the approach suffers computational complexity due to the high number of binary variables and non-linear terms in the transitioning probability. This might be inconvenient if quick maintenance decision is to be made as data are updated. Future research can be focused on the linearisation or enumeration algorithm of the structural network to reduce the binary variables during operational optimisation, while generating multiple near-optimal solutions for industrial practitioners. The ‘bathtub’ model also could not represent all type of equipment accurately. Future research should consider real data-driven estimation of the failure rate.

Acknowledgement

The funding from the project 'Sustainable Process Integration Laboratory – SPIL funded by EU' CZ Operational Programme Research and Development, Education, Priority1: Strengthening capacity for quality research (Grant No. CZ.02.1.01/0.0/0.0/15_003/0000456) is gratefully acknowledged.

6. References

- R.E. Barlow, F. Proschan, 1975. *Statistical theory of reliability and life testing: probability models*. Hold, Rinehart and Winston (HRW), New York, USA.
- R. Bellman, 1957. A Markovian decision process. *Journal of Mathematics and Mechanics*, 6(5), 679-684.
- H.H. Chin, P.S. Varbanov, J.J. Klemeš, M.F.D. Benjamin, R.R. Tan, 2020. Asset Maintenance Optimisation Approaches in the Chemical and Process Industries-A Review. *Chemical Engineering Research and Design*, 164, 162-194.
- E. Godoy, S.J. Benz, N.J. Scenna, 2015. An optimization model for evaluating the economic impact of availability and maintenance notions during the synthesis and design of a power plant. *Computers & Chemical Engineering*, 75, 135-154.
- H. Goel, J. Grievink, M. Weijnen, 2003. Integrated optimal reliable design, production, and maintenance planning for multipurpose process plants. *Computers & Chemical Engineering*, 27, 11, 1543-1555.
- H. Hu, G. Cheng, Y. Li, Y. Tang, 2009. Risk-based maintenance strategy and its applications in a petrochemical reforming reaction system. *Journal of Loss Prevention in the Process Industries*, 22, 4, 392-397.
- Z. Süle, J. Baumgartner, G. Dörgö, J. Abonyi, 2019. P-graph-based multi-objective risk analysis and redundancy allocation in safety-critical energy systems. *Energy*, 179, 989-1003.
- C.G. Vassiliadis, E.N. Pistikopoulos, 2001. Maintenance scheduling and process optimization under uncertainty. *Computers & Chemical Engineering*, 25, 2-3, 217-236.
- Y. Ye, I.E. Grossmann, J.M. Pinto, S. Ramaswamy, 2018. Markov Chain MINLP Model for Reliability Optimization of System Design and Maintenance. 13th International Symposium on Process System Engineering- PSE 2018, Computer Aided Chemical Engineering, July 1-5 2018, San Diego, California, USA.
- Y. Ye, I.E. Grossmann, J.M. Pinto, S. Ramaswamy, 2020. Integrated optimization of design, storage sizing, and maintenance policy as a Markov decision process considering varying failure rates. *Computers & Chemical Engineering*, 142, 107052.

Cost, environmental and exergy optimization of hydrogen and methane supply chains: application to Occitania region, France

Eduardo Carrera, Catherine Azzaro-Pantel

Laboratoire de Génie Chimique, Université Toulouse, CNRS, INPT, UPS, Toulouse, France

Abstract

This paper presents a methodological framework for the design of Hydrogen and Methane Supply Chains (HMSC), mainly focusing on Power-to-Gas (PtG) with low-carbon sources to spur the energy transition. The scientific objective is to perform mono and multi-objective optimizations of the HMSC to provide effective support for the study of deployment scenarios, with the Occitania region (France) as a case study. The formulation developed is based on a Mixed Integer Linear Programming (MILP) approach with augmented ϵ -constraint, implemented in the GAMS environment following a multi-period approach (2035-2050). The three objectives to be minimized are the Total Annual Cost, the Total Global Greenhouse Gas emissions, and Total Exergy Lost/Destroyed, which are related to the whole HMSC over the entire period studied. The results show that hydrogen can be competitive with SMR with carbon rates below 270 €/tonCO₂, whereas other drivers (i.e., ambitious policies to reduce GHG emissions) are expected for synthetic methane. Furthermore, coupling with other energy systems and processes is essential to increase the exergetic performance of HMSC.

Keywords: Power-to-Gas, hydrogen, methanation, MILP, exergy analysis, supply chain optimization.

1. Introduction

Power-to-Gas (PtG) is a promising solution to drive the penetration of low-carbon energy sources into the energy mix. These systems allow the use of electrical power, and CO₂ to produce gas (i.e., hydrogen or methane). Faced with the intermittency of renewable energy sources, they provide a low-carbon fuel supply and promote CO₂ recovery (Götz et al., 2016). Despite their advantages, PtG systems face some challenges. Typical Levelized Cost of Energy (LCOE) values for hydrogen and methane from PtG are stated at 64-74 €/MWh (2.1-2.5 €/kgH₂), and 95 to 150 €/MWh in 2030 (E&E consultant et al., 2014), respectively. In this regard, other options such as Steam Methane Reforming and natural gas are economically more advantageous (IEA - International Energy Agency, 2019)). In addition, the factors that favor the use of low-carbon technologies are decisive, since GHG emissions for SMR and natural gas are approximately 327 kgCO₂-eq/MWh and 202 kgCO₂-eq/MWh, respectively (Reiter and Lindorfer, 2015). For hydrogen produced from PtG, typical GHG emissions values range around 18-90 kgCO₂-eq/MWh and 22-191 kgCO₂-eq/MWh for renewable hydrogen and methane, respectively (Reiter and Lindorfer, 2015).

Another issue to be addressed is the destruction and loss of exergy throughout the Hydrogen and Methane Supply Chains (HMSC). Exergy is destroyed or lost when a process is irreversible (second law of thermodynamics) or due to by-products (Dincer

et al., 2014). Thus, the overall exergy efficiency of the HMSC can vary considerably depending on the feedstock and technologies involved (Khosravi et al., 2018).

Given the set of technologies, energy sources, and services involved, energy supply chains have been extensively studied, particularly for hydrogen (Li et al., 2019). However, only a few studies (Mesfun et al., 2017), (Bramstoft et al., 2020) consider simultaneously several fuel demands through PtG concepts. This issue has also been tackled in (Carrera Guilarte and Azzaro-Pantel, 2020) by also considering the geographical and logistical aspects of the entire HMSC over a multi-period horizon. Many researchers have carried out exergy analyses especially for power generation and cogeneration (Dincer et al., 2014). To the best of our knowledge, very few studies have been developed on the exergy analysis of a HMSC based on PtG concepts. The scientific objective of this study is to propose a comprehensive optimization model that can determine the design and deployment of an HMSC with total cost, GHG emissions and exergy loss as objective functions to be minimized. The main contribution of this study is thus to address the typical features of the HMSC (i.e., primary resources, production units, hydrogen transportation, storage, and refueling stations), considering the evolution of its configuration over time according to these different criteria. The paper is divided into three sections following this introduction. The methods and tools are presented in Section 2 with a focus on the problem formulation, presentation of the case study, the key points of the mathematical model for the HMSC, and the solution strategy. Finally, the main results (i.e., mono and multi-objective optimizations, and decision variables) are analyzed in Section 3, followed by the conclusions and perspectives presented in Section 4.

2. Methods and tools

2.1. Problem formulation

The problem formulation is based on prior studies, and follows the guidelines proposed by (Carrera Guilarte and Azzaro-Pantel, 2020; De-León Almaraz et al., 2015). It can be summarized as follows:

Given:

- Spatio-temporal demands for hydrogen and methane, availability of feedstocks.
- Characteristics of each technology.

Assuming:

- A division of the territory into grids.
- SMR, PtG, and natural gas are used to satisfy the methane and hydrogen demands.
- Direct injection of the methane produced into the natural gas network.
- Limited availability of primary energy and CO₂ sources.
- Possibility of importing primary resources.
- Possibility of transporting hydrogen only between grids by tanker-trucks for liquid hydrogen transportation.
- A constant carbon price over the period studied.

Determine:

- The location, number, capacity, and rate of production, storage and transport units as well as of refueling stations for hydrogen and methane.
- Transport flows between the grids.

Subject to:

- The satisfaction of the hydrogen and methane demands.
- Conservation of mass and energy, and other physical laws.

Objectives to be minimized either separately such as:

- Total Annual Cost.
 - GHG emissions.
 - Total Exergy Loss and Destruction.
- or with a combination through ε -constraint method.

2.2. Case study

The case study is the Region of Occitania, located in the south of France. Its energy transition strategy is framed within the REPOS (Positive Energy Region, in French) scenario (Région Occitanie / Pyrénées-Méditerranée, 2017). Hydrogen demand has been determined based on the expected use of fuel cell electric vehicles (FCEVs), for buses, private and light-good-vehicles. It corresponds to 5 TWh in 2050. The demand for methane includes the residential, industry, transport, and service sectors (1 TWh in 2050) (ADEME, 2018). A five-year period is used for time discretization (2035 to 2050). The territory is divided into 13 grids that correspond to the departments of this region. An average inter-district distance has been considered for transport purpose. Three sources of renewable energy have been identified, namely solar, wind, and hydro-power. Hydrogen is produced in the form of gas. Cryogenic storage is assumed for hydrogen with liquid transportation by tanker-trucks. Synthetic methane is assumed to be produced from hydrogen in a catalytic methanation process. Different sizes of electrolyzers, SMR, and catalyst reactors are considered. The CO₂ sources stem from methanization and gasification processes. The cost of electricity varies depending on the source. The natural gas needed to produce hydrogen by SMR or to meet part of the demand for methane is imported from abroad. A discount rate of 5.25% (E&E consultant et al., 2014), and an average constant carbon price of 270 €/ton_{CO₂} (Li et al., 2019) have been adopted respectively.

2.3. Key items of model development

The methodological framework is based on Mixed Integer Linear Programming (MILP). The formulation aims to satisfy the demands of both energy vectors “i” (hydrogen and methane), through fossil or renewable energy sources “b”, considering the potential of each grid “g”. The availability of energy source “e” and CO₂ “c”, production of hydrogen “p”, storage option “s”, transport option “l”, methane production type “q”, and refueling station “f” for each period “t” are taken into account. Only the key points are discussed below.

2.3.1 Cost objective function

The Total Annual Cost (*TAC*, €) (Eq. (1)) takes into account the capital cost (*FCC_{it}*, €/year), operating cost (*FOC_{it}*, €/year), the Annual Carbon Price (*ACP_{it}*, €/year), the cost of imported natural gas (*INGC_{it}*, €/year) and the discount rate (*r*):

$$TAC = \sum_{it} \left(\frac{MFCC_{it} + MFOC_{it} + ACP_{it} + INGC_{it}}{(1+r)^t} \right) \quad (1)$$

2.3.2 Environmental objective function

The Total Greenhouse Gas emissions (TGHG, gCO₂-eq) (see Eq. (2)) generated by the different echelons of the supply chain have been considered as the environmental impact objective function, with the following contributions all expressed in gCO₂-eq/year: *GHG_{it}* related to the hydrogen (*HGHG_{it}*), and methane (*MGHG_{it}*) production, the hydrogen storage (*SGHG_{it}*), hydrogen transport (*TGHG_{it}*), and methane import (*IMGHG_{it}*).

$$TGHG = \sum_{it} (HGHG_{it} + SGHG_{it} + TGHG_{it} + MGHG_{it} + IMGHG_{it}) \quad (2)$$

2.3.3. Exergy objective function

The Total Exergy Loss and Destruction (TELD) corresponds to the exergy objective function (Eq. (3)). Each flow of mass and energy in the processes involved in HMSC is identified as a resource, product, or waste. Subsequently, the exergy destroyed and lost in each process was calculated considering the typical exergetic yield of each stage using the values presented in (Khosravi et al., 2018). The lost/destroyed exergy related to hydrogen ($HELD_{it}$), and methane ($MELD_{it}$) production, hydrogen storage ($SELD_{it}$), hydrogen transport ($TrELD_{it}$), and methane import ($IMELD_{it}$) can be expressed as follows (all expressed in MWh/year:

$$TELD = \sum_{it} (HELD_{it} + SELD_{it} + TrELD_{it} + MELD_{it} + IMELD_{it}) \quad (3)$$

2.4. Solution strategy

The model is solved by CPLEX 12 in the GAMS environment. An Intel Xeon E3-1505MV6, 3.00 GHz computer with 32 GB RAM was used. A mono-objective analysis of each criterion (TAC, TGHG, and TELD) is first performed for a better understanding of the problem. A multi-objective optimization strategy with the three objective functions is then conducted to reach the Pareto front using the augmented ϵ -constraint method (AUGMECON)(Mavrotas, 2009). Since AUGMECON uses the lexicographical optimization, a priority is established for the objective functions. In this case, TAC is first ranked, then TGHG, and finally TELD. The modified-TOPSIS (m-TOPSIS), a variant of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is applied to choose the best compromise configuration from the Pareto front, assigning equal weight to each optimization criterion.

3. Results and discussion

The optimal values of the decision variables of the mono-objective approach for each objective function (TAC, TGHG, and TELD), and the TOPSIS solution are presented in Table 1. The production, storage and transport units are installed progressively throughout the period. The minimization of exergy implies the absence of hydrogen transport, due to the large amount of exergy destroyed. The LCOE for hydrogen is between 130-285 €/MWh, which is the acceptable range of values to be economically viable (Hydrogen Europe, 2018). However, only the “min TGHG” and TOPSIS solutions propose methane production from methanation, where a significant reduction of CO₂-eq emissions is imposed. For “min TAC” and “min TELD”, the methane demand is met through natural gas, since its cost is half that of methane produced by methanation. The GHG for both vectors corresponds to the range of values reported from the literature (see Section 1), and varies according to the energy sources used in each period. The exergy lost and destroyed over the period varies between 55 and 92%, mainly due to the exergetic yield associated with energy production (i.e., 11 and 30% for PV and wind power plants, respectively). The HMSC configuration in 2050 can be visualized in Figure 2. The main source of electricity used is solar followed by wind. Due to the capacity factor associated with each energy source, the HMSC is decentralized, with on-grid production of hydrogen and methane. Depending on the scenario and decision criteria applied, the total hydrogen demand is fully satisfied by electrolysis while 68% of the methane demand is provided by methanation and the remaining part by natural gas import. The main

characteristics and computational effort of the solution found by the optimization model are summarized in Table 2, corresponding to each mono-objective optimization and the TOPSIS solution chosen.

Table 1. Optimization results of the HMSC

Solutions	min TAC				min TGHG				min TELD				TOPSIS			
	Year	2035	2040	2045	2050	2035	2040	2045	2050	2035	2040	2045	2050	2035	2040	2045
Number of H ₂ production facilities	119	181	246	246	196	283	402	402	96	98	133	133	153	238	351	351
Number of H ₂ storage facilities	28	42	57	57	24	32	46	46	58	58	64	64	24	32	46	46
Number of methane production facilities	0	0	0	0	13	15	16	16	0	0	0	0	2	7	16	16
Number of H ₂ transport units	7	8	10	14	4	4	5	6	0	0	0	0	8	12	11	12
CH ₄ LCOE (€/MWh)	94.4				190				94.4				162			
H ₂ LCOE (€/MWh)	142				144				208				144			
CH ₄ GHG _t (gCO ₂ /MWh)	201	201	201	201	94.7	95.5	113	104	201	201	201	201	187	198	198	198
H ₂ GHG _t (gCO ₂ /MWh)	84.9	89.2	108	150	95.9	100	99.7	103	382	384	386	387	93.3	104.7	130	154
φ _t	0.83	0.85	0.86	0.84	0.9	0.92	0.91	0.9	0.55	0.55	0.56	0.56	0.83	0.84	0.84	0.84
Objective functions																
TAC (M€)	3158				3930				4256				3303			
TGHG (x10 ¹² gCO ₂ -eq)	2.71				1.99				6.67				2.16			
TELD (TWh)	108				202				23				97			

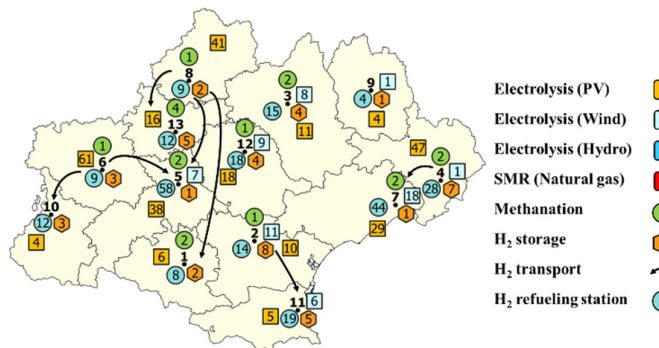


Figure 2. Network structure of the HMSC in 2050 for TOPSIS solution

Table 2. Computational results for the optimization

Single variables	Discrete variables	Single equations	min TDC		min TGHG		min TELD		TOPSIS	
			CPU (s)	GAP	CPU (s)	GAP	CPU (s)	GAP	CPU (s)	GAP
98,245	23,816	814,640	7200	0.007	8	0.000	25	0.000	7200	0.008

4. Conclusion and perspectives

In this paper, a general methodology for the design and deployment of an HMSC was developed and used for a baseline case study devoted to the Occitania region in France. Each mono-objective minimization proposes a drastically different configuration. The LCOE for hydrogen and methane was obtained from the optimization strategy. The results show that with electricity prices close to 31 €/MWh and a carbon price below 270 euros/ton CO₂, hydrogen can play an important role as an energy carrier. As for synthetic methane, its cost remains prohibitive to play a significant role in the energy mix. The total satisfaction of the expected methane demand through methanation could only be justified

in a scenario of massive reduction of greenhouse gas emissions. In addition, coupling this PtG ecosystem with other energy systems is necessary to reduce the exergy losses (i.e., with the cement industry or with methanization processes).

The methodology and the results found can be useful to enrich the decision-making process, especially to promote the deployment of specific technologies. The model is general enough to incorporate other criteria. The methodological framework paves the way for further extensions to other types of processes and energy vectors, creating a synergetic effect with other energy systems, thus exploring the scale-up potential for increased efficiency and cost reduction.

References

- ADEME, 2018. Un mix de gaz 100 % renouvelable en 2050 ? Angers.
- Bramstoft, R., Pizarro-Alonso, A., Jensen, I.G., Ravn, H., Münster, M., 2020. Modelling of renewable gas and renewable liquid fuels in future integrated energy systems. *Appl. Energy* 268, 114869. <https://doi.org/10.1016/j.apenergy.2020.114869>
- Carrera Guilarte, E., Azzaro-Pantel, C., 2020. A Methodological Design Framework for Hybrid “Power-to-Methane” and “Power-to-Hydrogen” Supply Chains: application to Occitania Region, France, in: *Proceedings of the 30 European Symposium on Computer Aided Process Engineering Th (ESCAPE30)*. pp. 679–684. <https://doi.org/10.1016/B978-0-12-823377-1.50114-2>
- De-León Almaraz, S., Azzaro-Pantel, C., Montastruc, L., Boix, M., 2015. Deployment of a hydrogen supply chain by multi-objective / multi-period optimisation at regional and national scales. *Chem. Eng. Res. Des.* 104, 11–31. <https://doi.org/10.1016/j.cherd.2015.07.005>
- Dincer, I., Midilli, A., Kucuk, H. (Eds.), 2014. *Progress in Exergy, Energy, and the Environment*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-04681-5>
- E&E consultant, HESPUL, Solagro, ADEME, 2014. Etude portant sur l’hydrogène et la méthanation comme procédé de valorisation de l’électricité excédentaire [WWW Document]. URL <http://www.grtgaz.com/fileadmin/engagements/documents/fr/Power-to-Gas-etude-ADEME-GRTgaz-GrDF-complete.pdf>
- Götz, M., Lefebvre, J., Mörs, F., McDaniel Koch, A., Graf, F., Bajohr, S., Reimert, R., Kolb, T., 2016. Renewable Power-to-Gas: A technological and economic review. *Renew. Energy* 85, 1371–1390. <https://doi.org/10.1016/j.renene.2015.07.066>
- Hydrogen Europe, 2018. *Hydrogen, enabling a zero emission Europe - Technology Roadmaps Full Pack*. Brussels.
- IEA - International Energy Agency, 2019. *The Future of Hydrogen* [WWW Document]. URL <https://www.iea.org/reports/the-future-of-hydrogen>
- Khosravi, A., Koury, R.N.N., Machado, L., Pabon, J.J.G., 2018. Energy, exergy and economic analysis of a hybrid renewable energy with hydrogen storage system. *Energy* 148, 1087–1102. <https://doi.org/10.1016/j.energy.2018.02.008>
- Li, L., Manier, H., Manier, M.A., 2019. Hydrogen supply chain network design: An optimization-oriented review. *Renew. Sustain. Energy Rev.* 103, 342–360. <https://doi.org/10.1016/j.rser.2018.12.060>
- Mavrotas, G., 2009. Effective implementation of the ϵ -constraint method in Multi-Objective Mathematical Programming problems. *Appl. Math. Comput.* 213, 455–465.
- Mesfun, S., Sanchez, D.L., Leduc, S., Wetterlund, E., Lundgren, J., Biberacher, M., Kraxner, F., 2017. Power-to-gas and power-to-liquid for managing renewable electricity intermittency in the Alpine Region. *Renew. Energy* 107, 361–372. <https://doi.org/10.1016/j.renene.2017.02.020>
- Région Occitanie / Pyrénées-Méditerranée, 2017. *Scénario région à énergie positive de la Région Occitanie / Pyrénées-Méditerranée*.
- Reiter, G., Lindorfer, J., 2015. Global warming potential of hydrogen and methane production from renewable electricity via power-to-gas technology. *Int. J. Life Cycle Assess.* 20, 477–489. <https://doi.org/10.1007/s11367-015-0848-0>

Planning of non-conventional gas field development with parametric uncertainties

Jui-Yuan Lee,^{a,*} Pei-Fan Bai,^a Raymond E. H. Ooi,^b Dominic C. Y. Foo,^b
Raymond R. Tan^c

^a*Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, 1, Sec. 3, Zhongxiao E. Rd., Taipei 10608, Taiwan, R.O.C.*

^b*Department of Chemical and Environmental Engineering, University of Nottingham Malaysia, Broga Road, 43500 Semenyih, Selangor, Malaysia*

^c*Chemical Engineering Department, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines*

Abstract

This paper presents a mathematical optimisation model for the planning of non-conventional gas field development, considering multiple impurities in natural gas streams and uncertainties in gas quality and technological parameters. The model consists mainly of mass balance equations and fuzzy constraints. A modified literature example is presented to illustrate the proposed approach.

Keywords: natural gas sweetening, mathematical programming, fuzzy optimisation.

1. Introduction

With growing demand for natural gas (NG) as a low-carbon energy source, developing non-conventional gas fields has become common in the oil and gas industry. NG streams produced from such fields are sour with high concentrations of CO₂ and H₂S, requiring treatment or “sweetening” before processing and downstream applications. Planning the development of non-conventional gas fields involves determining how much gas should be produced and treated at each source, and which gas sweetening technology should be used. Process integration techniques have been extended to address the gas field planning problem, which is analogous to the effluent treatment problem.

The gas field planning problem can be solved using pinch analysis and mathematical programming. Foo et al. (2016) developed pinch-based graphical and optimisation approaches to determine the minimum extent of CO₂ removal and the minimum sweet gas supply. Parand et al. (2018) later developed a pinch-based algebraic technique that overcomes the limitations of the graphical technique of Foo et al. (2016). Lee et al. (2018) developed a multi-period mathematical model that considers time-varying gas production and demand. However, none of these previous works has considered other impurities such as H₂S in NG streams as well as uncertainties in technical and economic parameters.

In this paper, a mathematical programming model is developed for optimal planning of gas field development. The formulation considers multiple impurities in NG streams, and incorporates fuzzy optimisation to allow for parametric uncertainties. An example is used to illustrate the proposed approach.

2. Problem Statement

The problem addressed in this paper is formally stated as follows. Given:

- A set of gas sources $i \in \mathbf{I}$ (e.g. gas fields). The NG streams produced from these sources contain a set of impurities $k \in \mathbf{K}$ (e.g. CO₂ and H₂S), and need to be treated, or sweetened, before proceeding to the gas processing plant (GPP).
- A set of gas sweetening technologies $j \in \mathbf{J}$ (e.g. amine absorption and membrane separation). Each technology is characterised by its product (NG) recovery factor, impurity removal ratio and cost.

The objective is to determine the optimal strategy that minimises the total operating cost associated with (a) gas sweetening to meet the product specifications and (b) gas transport for processing, while considering uncertainties in cost and performance parameters.

3. Model Formulation

Figure 1 shows a schematic diagram of gas sweetening at source i . The gas from source i can be partially treated with available gas sweetening technologies. The purified stream is sent to the GPP along with the bypass stream, while the reject stream may be used for enhanced oil recovery (EOR). Two scenarios are considered. Scenario 1 represents the deterministic case assuming that all cost and performance parameters are known precisely, while scenario 2 accounts for parametric uncertainties using fuzzy optimisation.

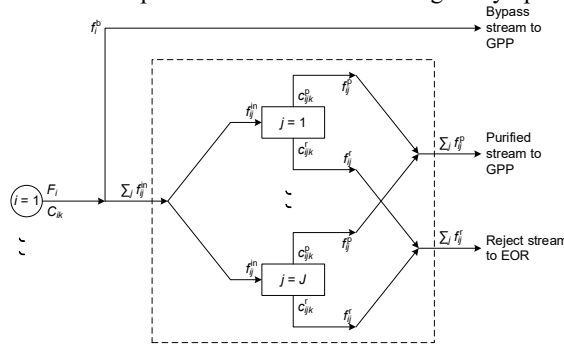


Figure 1. Schematic of Gas Sweetening

3.1. Scenario 1: Deterministic Case

Eq. (1) describes the flowrate balance for source i .

$$\sum_{j \in \mathbf{J}} f_{ij}^{\text{in}} + f_i^{\text{b}} \leq F_i \quad \forall i \in \mathbf{I} \tag{1}$$

where f_{ij}^{in} is the inlet flowrate to the gas sweetening unit using technology j at source i ; f_i^{b} is the bypass flowrate of gas from source i ; F_i is the available gas flowrate of source i . Eq. (2) states that at each source only one gas sweetening technology can be used.

$$\sum_{j \in \mathbf{J}} y_{ij} \leq 1 \quad \forall i \in \mathbf{I} \tag{2}$$

where y_{ij} is a binary variable indicating if technology j is used at source i . Eq. (3) sets an upper limit on f_{ij}^{in} , while ensuring that no gas is treated by unused technologies ($f_{ij}^{\text{in}} = 0$ if $y_{ij} = 0$).

$$f_{ij}^{\text{in}} \leq F_i y_{ij} \quad \forall i \in \mathbf{I}, j \in \mathbf{J} \tag{3}$$

Eqs. (4) and (5) describe the flowrate and impurity balances around the gas sweetening unit using technology j at source i , respectively.

$$f_{ij}^{\text{in}} = f_{ij}^{\text{p}} + f_{ij}^{\text{r}} \quad \forall i \in \mathbf{I}, j \in \mathbf{J} \quad (4)$$

$$f_{ij}^{\text{in}} C_{ik} = f_{ij}^{\text{p}} c_{ijk}^{\text{p}} + f_{ij}^{\text{r}} c_{ijk}^{\text{r}} \quad \forall i \in \mathbf{I}, j \in \mathbf{J}, k \in \mathbf{K} \quad (5)$$

where f_{ij}^{p} and f_{ij}^{r} are the flowrates of the purified and reject streams, respectively; C_{ik} is the concentration of impurity k in the gas from source i ; c_{ijk}^{p} and c_{ijk}^{r} are the impurity concentrations of the purified and reject streams, respectively.

The gas sweetening process is also characterised by the impurity removal ratio (RR_{jk}) in Eq. (6) and the product recovery factor (RF_j) in Eq. (7).

$$f_{ij}^{\text{p}} c_{ijk}^{\text{p}} = f_{ij}^{\text{in}} C_{ik} (1 - RR_{jk}) \quad \forall i \in \mathbf{I}, j \in \mathbf{J}, k \in \mathbf{K} \quad (6)$$

$$f_{ij}^{\text{p}} \left(1 - \sum_{k \in \mathbf{K}} c_{ijk}^{\text{p}} \right) = RF_j f_{ij}^{\text{in}} \left(1 - \sum_{k \in \mathbf{K}} C_{ik} \right) \quad \forall i \in \mathbf{I}, j \in \mathbf{J} \quad (7)$$

Eqs. (8) and (9) impose flowrate and concentration constraints for gas processing.

$$\sum_{i \in \mathbf{I}} f_i^{\text{b}} + \sum_{i \in \mathbf{I}} \sum_{j \in \mathbf{J}} f_{ij}^{\text{p}} \geq F^{\text{GP}} \quad (8)$$

$$\sum_{i \in \mathbf{I}} f_i^{\text{b}} C_{ik} + \sum_{i \in \mathbf{I}} \sum_{j \in \mathbf{J}} f_{ij}^{\text{p}} c_{ijk}^{\text{p}} \leq F^{\text{GP}} C_k^{\text{max}} \quad \forall k \in \mathbf{K} \quad (9)$$

where F^{GP} is the required gas processing flowrate; C_k^{max} is the maximum allowable concentrations of impurity k for gas processing.

The objective function in this scenario is to minimise the total operating cost associated with gas sweetening and gas transport (f_{TOC}), as given in Eq. (10).

$$\min f_{\text{TOC}} = \sum_{i \in \mathbf{I}} \sum_{j \in \mathbf{J}} OC_j f_{ij}^{\text{in}} \delta + TC \left(\sum_{i \in \mathbf{I}} D_i f_i^{\text{b}} \delta + \sum_{i \in \mathbf{I}} \sum_{j \in \mathbf{J}} D_i f_{ij}^{\text{p}} \delta \right) \quad (10)$$

where OC_j is the operating cost of technology j ; δ is the operating duration; TC is the gas transport cost coefficient; D_i is the distance from source i to the GPP.

The model for scenario 1 (Eq. (10) subject to Eqs. (1)-(9)) is a mixed integer nonlinear programme (MINLP) due to bilinear terms in Eqs. (5)-(7) and (9).

3.2. Scenario 2: Planning under Uncertainties

In this scenario, uncertainties in C_{ik} , OC_j and RR_{jk} are considered and handled using fuzzy optimisation. For conservative (low-risk) planning, higher values of the estimated impurity concentration and operating cost are more desirable, as given in Eqs. (11) and (12). Similarly, a lower value of the estimated removal ratio is more desirable, as given in Eq. (13).

$$C_{ik} \geq C_{ik}^{\text{L}} + \lambda (C_{ik}^{\text{U}} - C_{ik}^{\text{L}}) \quad \forall i \in \mathbf{I}, k \in \mathbf{K} \quad (11)$$

$$OC_j \geq OC_j^{\text{L}} + \lambda (OC_j^{\text{U}} - OC_j^{\text{L}}) \quad \forall j \in \mathbf{J} \quad (12)$$

$$RR_{jk} \leq RR_{jk}^{\text{U}} - \lambda (RR_{jk}^{\text{U}} - RR_{jk}^{\text{L}}) \quad \forall j \in \mathbf{J}, k \in \mathbf{K} \quad (13)$$

where fuzzy parameters C_{ik} , OC_j and RR_{jk} are redefined as variables; superscripts L and U denotes lower and upper limits, respectively; λ represents the overall degree of satisfaction of these fuzzy constraints. The value of λ ranges from 0 (unsatisfactory) to 1 (completely satisfactory), as stated in Eq. (14).

$$0 \leq \lambda \leq 1 \quad (14)$$

With parametric uncertainties, the objective of minimising the total operating cost in Eq. (10) is reformulated as an additional fuzzy constraint in Eq. (15), which indicates that a low value of f_{TOC} is desirable.

$$f_{\text{TOC}} \leq f_{\text{TOC}}^{\text{U}} - \lambda (f_{\text{TOC}}^{\text{U}} - f_{\text{TOC}}^{\text{L}}) \quad (15)$$

The objective function in this scenario is to maximise the overall degree of satisfaction of the fuzzy constraints, as given in Eq. (16).

$$\max \lambda \quad (16)$$

The model for scenario 2 consists of Eqs. (1)-(9) and (11)-(16) and is also an MINLP. In the following section, a modified literature example is solved to illustrate the proposed approach. The MINLP models for both scenarios are implemented and solved in GAMS using solver BARON. Results were obtained with negligible processing time.

4. Illustrative Example

Table 1 shows the data for the example, which is adapted from Foo et al. (2016). A total of four gas fields are to be developed. The maximum allowable concentrations of CO₂ and H₂S for gas processing and downstream applications (e.g. power generation) are 3% and 5 ppm, respectively. Amine absorption and membrane separation are available for gas sweetening. Table 2 shows the cost and performance parameters. It is assumed that the GPP is designed to process 1000 MMscfd of gas to meet the demand. The transport cost is taken to be 0.003 USD/MMscf/km.

Two scenarios are analysed in this example. Scenario 1 assumes that all the parameters are known precisely as given in Tables 1 and 2, while scenario 2 considers uncertainties in CO₂ concentrations, operating costs and CO₂ removal ratios.

Table 1. Data for the Gas Sweetening Example

Source	Flowrate (MMscfd)	Concentration (%)		Distance to GPP (km)
		CO ₂	H ₂ S	
Gas field 1	300	10	0.1	400
Gas field 2	450	20	0.2	300
Gas field 3	500	30	0.5	350
Gas field 4	600	35	1	150

Table 2. Cost and Performance Data for Gas Sweetening Technologies

Technology	Operating cost (USD/MMscf)	Product recovery factor (%)	Removal ratio (%)	
			CO ₂	H ₂ S
Amine absorption	0.4	92	90	99.9
Membrane separation	0.7	96	95	99.99

4.1. Results for Scenario 1

Scenario 1 aims to minimise the total operating cost associated with gas sweetening and transport. Solving the corresponding MINLP model gives the minimum cost of 559,044 USD/y. Table 3 shows the technologies used and the associated stream flowrates. Amine absorption is used at gas fields 1, 2 and 4 because of its lower operating cost. Membrane separation is needed at gas field 3 for its higher removal ratios. Gas fields 1 and 2 produce higher quality gases (with lower CO₂ and H₂S concentrations) and are fully developed; in

other words, the flowrates of gas produced are equal to the available amounts. Gas fields 3 and 4 produce lower quality gases and are only partially developed. Furthermore, the flowrate of gas produced from field 3 is less than that from field 4 because of the longer distance from gas field 3 to the GPP. All the gases produced are completely treated and there are no bypass streams. These results may be used to determine the sizes of gas sweetening units and pipelines. Figure 2(a) illustrates the optimal gas sweetening system for scenario 1.

Table 3. Technology Selection and Stream Flowrates for Scenario 1

Gas field	Technology	Flowrate (MMscfd)			
		Inlet stream	Purified stream	Reject stream	Bypass stream
1	Amine	300	251.124	48.876	0
2	Amine	450	339.373	110.627	0
3	Membrane	264.904	180.717	84.186	0
4	Amine	366.755	228.785	137.969	0

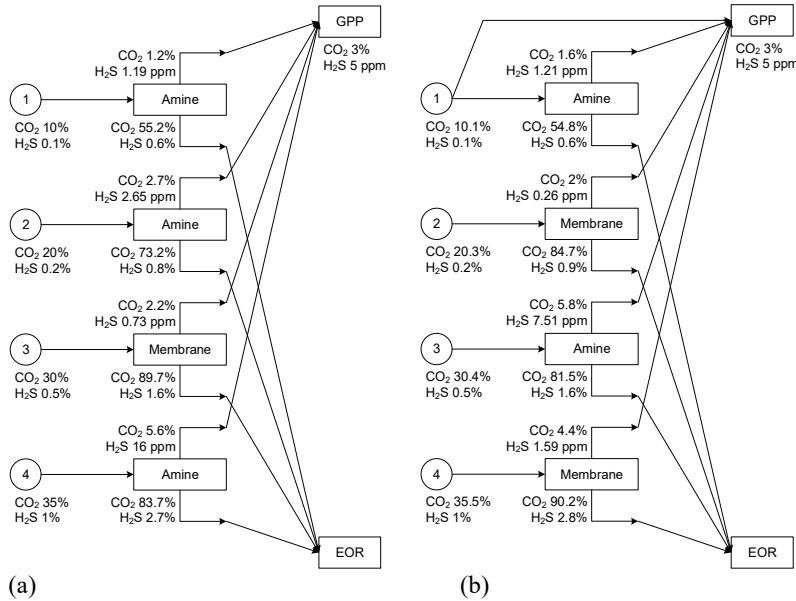


Figure 2. Gas Sweetening Systems for (a) Scenario 1 and (b) Scenario 2

4.2. Results for Scenario 2

In scenario 2, parametric uncertainties are considered, and the objective is to maximise the overall degree of satisfaction of the fuzzy constraints for a conservative, low-risk planning. The fuzzy ranges for the uncertain parameters in this example are given in Tables 4 and 5. In addition, the minimum cost of 559,044 USD/y in scenario 1 is taken as the lower limit for the total operating cost, while the upper limit of 752,091 USD/y is determined by solving the model for scenario 1 with operating cost coefficients at the upper limits and CO₂ removal ratios at the lower limits.

Solving the MINLP model for scenario 2 gives the maximum λ value of 0.568, meaning that each fuzzy constraint is at least 56.8% satisfied. The total operating cost is estimated

at 642,484 USD/y, which corresponds to a 14.9% increase from the lower limit. Table 6 shows the technologies used and the associated stream flowrates. Amine absorption is used at gas fields 1 and 3, and membrane separation at gas fields 2 and 4. Fields 1 and 2 with higher quality gases are 98.8% and 91.6% developed, respectively. With lower quality gases, field 3 is only 13% developed due to its long distance to the GPP, while field 4 is fully developed because it is the closest to the GPP. Figure 2(b) illustrates the optimal gas sweetening system for scenario 2.

Table 4. Fuzzy Limits for Impurity Concentrations

Source	CO ₂ concentration (%)	H ₂ S concentration (%)
Gas field 1	9–11	0.09–0.11
Gas field 2	18–22	0.18–0.22
Gas field 3	27–33	0.45–0.55
Gas field 4	31.5–38.5	0.9–1.1

Table 5. Fuzzy Limits for Operating Costs and Removal Ratios

Technology	Operating cost (USD/MMscf)	CO ₂ removal ratio (%)
Amine absorption	0.4–0.48	85–90
Membrane separation	0.7–0.84	90–95

Table 6. Technology Selection and Stream Flowrates for Scenario 2

Gas field	Technology	Flowrate (MMscfd)			
		Inlet stream	Purified stream	Reject stream	Bypass stream
1	Amine	296.371	248.606	47.765	3.629
2	Membrane	412.392	321.394	90.998	0
3	Amine	65.006	43.856	21.15	0
4	Membrane	600	382.515	217.485	0

5. Conclusion

A mathematical programming model for gas field development planning has been developed in this paper. The formulation considers multiple impurities in NG streams, with fuzzy constraints account for parametric uncertainties. A modified literature example was solved to illustrate the proposed approach. Future work will consider a more comprehensive energy system with hydrogen production and utilisation.

References

- D.C.Y. Foo, R.E.H. Ooi, R.R. Tan, J.-Y. Lee, 2016, Process Integration Approaches to Optimal Planning of Unconventional Gas Field Development, *Chemical Engineering Science*, 150, 85-93
- J.-Y. Lee, R.E.H. Ooi, D.C.Y. Foo, R.R. Tan, 2018, A Mathematical Technique for Multi-period Planning of Unconventional Gas Field Development, *Computer Aided Chemical Engineering*, 44, 1243-1248
- R. Parand, D.C.Y. Foo, R.E.H. Ooi, R.R. Tan, J.-Y. Lee, 2018, An Algebraic Targeting Approach for Optimal Planning of Gas Sweetening Problem in Non-conventional Gas Field Development, *Process Safety and Environmental Protection*, 120, 248-255

Effect of the lignin extraction process on the economics of a woody-based biorefinery

Juan Camilo Solarte-Toro,^a Javier A. Arrieta-Escobar,^b Brunelle Marche,^b
Carlos A. Cardona Alzate^{a*}

^aInstituto de Biotecnología y Agroindustria, Departamento de Ingeniería Química, Universidad Nacional de Colombia, Manizales, Caldas, Zip Code: 170003, Colombia

^bERPI (Équipe de Recherche des Processus Innovatifs), Université de Lorraine, 8 rue Bastien Lepage, 54000 Nancy Cedex, France

ccardonaal@unal.edu.co.

Abstract.

Woody biomass has been proposed as a versatile raw material to be upgraded in energy vectors and value-added products. Nevertheless, most studies have focused on the cellulose and hemicellulose conversion reducing the potential of lignin as a value product that can be commercialized or upgraded. Thus, this paper aims to evaluate the influence of the lignin extraction process on the economic performance of a woody-based biorefinery addressed to produce lignin and levulinic acid as products. Kraft, Soda, and Organosolv extraction methods were assessed and compared. After a simulation procedure and economic assessment using experimental information and kinetic models reported in the open literature, the results show the Kraft process as the most suitable option to produce both lignin and levulinic acid. The simulation process results show that the organosolv process requires higher processing scales than Kraft and Soda processes to reach economic feasibility. In conclusion, the further valorization of lignin in other high value-added products is necessary to increase the economic performance of extracting lignin. Moreover, the inclusion of high added-value products such as levulinic acid reduces processing scale and increases the possibility of implementing this type of process.

Keywords: Biorefineries, Lignin extraction, Process Design, Woody Biomass, Levulinic acid, Economic assessment.

1. Introduction.

Woody biomass has been defined as a potential raw material to be upgraded through either thermochemical or chemical pathways. Thermochemical processing of woody biomass has been preferred over other biomass sources. This renewable resource has some advantages related to easy handling, high energy content, and relative-high fixed carbon content. Indeed, woody biomass has supplied 9% of the energy consumption in 2019. Despite the suitability of woody biomass for energy production purposes, several studies have focused on converting this raw material into a series of value-added products through the biorefinery concept. Nevertheless, most of these studies have been addressed to convert the cellulose and hemicellulose, leaving aside the lignin content of woody biomass (Solarte-Toro et al., 2021).

Lignin is considered the most important renewable aromatic source in the world. Moreover, this biopolymer can be upgraded into a series of value-added products (e.g., vanillin, benzene, toluene, xylene, quinones, eugenol) able to be used in the chemical, pharmaceutical, cosmetic, and textile industries. Lignin isolation can be done in several ways, such as mechanical, chemical, physical, and enzymatic. However, the most common methods applied at the industrial level involves the use of chemical agents. In this way, extracted lignins can be classified as sulfur lignins and sulfur-free lignins, depending on the chemical agents used to isolate lignin. Sulfur free lignins are obtained through Organosolv and Soda, while sulfur lignins are obtained via Kraft and lignosulfonate pulping. The chemical composition of woody biomass before and after implementing the lignin extraction methods has been reported in the open literature (Tian et al., 2017). Nevertheless, few studies have been focused on the economic analysis of these processes. For instance, Carvajal et al. (2016) report the comparison of the stand-alone lignin extractions processes. These authors proposed the Kraft process as the most economical process for lignin extraction. Even so, these authors do not include the lignin isolation step. Moreover, few studies have analyzed the effect of the lignin extraction process on the economic performance of a biorefinery. In this way, this paper aims to evaluate the influence of the lignin extractions process on the economic performance of a woody-based biorefinery addressed to produce lignin and levulinic acid as products.

2. Methodology.

2.1. Conceptual design and process description.

The proposed biorefinery involves two processing stages. The first stage is related to the lignin extraction process. Kraft, Soda, and Organosolv processes were simulated. Kraft pulping conditions were active alkali 26 %, sulfidity 30 %, solid to liquid ratio 1:4, process temperature 172 °C, and residence time 2 hours (Przybysz et al., 2018). In contrast, Soda pulping was performed at 93 °C using an 8 %w/v NaOH solution. The soda process was simulated considering a residence time of 2 hours (Saber et al. 2020). Finally, the organosolv process considered a 60 %v/v ethanol solution and 1.25% w/w sulfuric acid solution as a catalyst. The process temperature and residence time were 180 °C and one h, respectively (Meng et al. 2020). For Kraft and Soda processes, the lignin recovery was made using CO₂ and H₂SO₄ as precipitating agents (Öhman et al. 2013). Finally, the organosolv lignin recovery was made following the downstream processing described by Solarte-Toro et al. (2018). The levulinic acid production stage involves (i) cellulose degradation and (ii) glucose conversion. The cellulose degradation process was achieved considering a process temperature of 220 °C for 30 min at an initial cellulose loading of 29 %w/w. The remaining solid was recycled to increase glucose concentration in the outlet liquid stream. Rich-glucose stream was upgraded to levulinic acid at 160°C, 5.5 MPa, and catalyst (Amberlyst 70) loading of 0.035 kg/L (Weingarten et al., 2012a). The residence time was less than 3 minutes. Humins and a liquid mixture composed of levulinic acid, formic acid, and water are the glucose decomposition products. Humins are separated by filtration, and levulinic acid is purified by distillation. The distillation process was designed to produce a levulinic acid solution at industrial grade.

2.2. Simulation procedure.

The simulation of the three scenarios using poplar (*Populus sp.*) as lignocellulosic feedstock was done using the simulation software Aspen Plus v9.0 (Aspen Technology

Inc. USA). The feedstock flow rate was set at 1 t/h on a wet basis. The vapor-liquid equilibrium in the simulation was estimated through the gamma-phi method by selecting the Non-Random Two Liquids (NRTL) activity coefficient model and the Peng-Robinson equation of state. The thermodynamic properties of cellulose, hemicellulose, and lignin were specified in the simulator. Temperature-dependent properties were introduced as reported by the National Renewable Energy Laboratory (NREL). Besides, constant thermodynamic properties such as enthalpy of formation and free energy of formation were estimated using the equations reported by Peduzzi et al. (2016). The chemical characterization of raw poplar (*Populus sp.*) on a dry basis is given in Table 1.

Table 1. Chemical characterization of poplar (*Populus sp.*)

Compound	Share %
Cellulose	51.10
Hemicellulose (Xylan)	21.40
Lignin	22.70
Extractives	4.40
Ash	0.40
Moisture	45.00

The chemical characterizations of poplar (*Populus sp.*) after Kraft, Soda, and Organosolv processes were used to estimate the mass balances of the lignin extraction process in each scenario. Thus, experimental results reported in the open literature were used as input data. Kraft process was simulated considering the data given by Buzala et al. (2017), Soda process was assessed according to the yields stated by Saber et al. (2020), and Organosolv lignin extraction was designed involving the information reported by Meng et al. (2020). Lignin recovery from black liquor was simulated following the process described by Öhman et al. (2013). The Kraft and Soda processes flow diagrams were based on the schemes reported by (Carvajal et al., 2016). The Organosolv process flowsheet was based on the reported figure by Solarte-Toro et al. (2018). The levulinic acid production process was simulated in three stages. The first stage comprised the cellulose hydrolysis to produce glucose. The mass balances of this stage were estimated considering the experimental cellulose conversion reported by Weingarten et al. (2012b). The second stage is related to glucose dehydration. The kinetic parameters (i.e., pre-exponential factor and activation energy) estimated by Weingarten et al. (2012a) were introduced to model a continuous stirred tank reactor (CSTR) in the simulation software. The third stage is addressed to the levulinic acid purification by distillation. Topologic thermodynamics was applied to identify the best way to separate levulinic acid mixture. Shortcut methods were used to estimate the number of theoretical stages, reflux ratio, and distillate to feed ratio. Finally, the information mentioned above was used to specify the rigorous distillation module in Aspen Plus to estimate material and energy balances.

2.3. Economic evaluation.

The economic assessment of the proposed scenarios was done using the Aspen Process Economic Analyzer v9.0. This software was used to estimate the total capital investment (TCI) of the proposed processes. The profitability analysis was accomplished following the methodology reported by Towler and Sinnott (2013). The cost of raw materials, utilities, and labor were used as input data (see Table 2). Besides, the following conditions were involved in the economic analysis: (i) the straight-line depreciation method with a salvage value of 10% was applied, (ii) the cash flow analysis was done considering a corporate tax rate value of 28% and an interest rate value of 1.64%, (iii) the proposed

biorefineries are conceived as continuous processes working 8360 hours in a year, (iv) The proposed scenarios consider the creation of eight operator jobs and two supervisor positions, and (v) the project lifetime was set in 10 years. Finally, all the proposed scenarios were compared using the net present value (NPV) and payback period (PBP) as economic metrics. This metrics estimation was done to assess the biorefinery profitability and the best extraction method to reach economic feasibility.

Table 2. Input data used to estimate the cash flow of the proposed scenarios.

Item	Price	Units	Item	Price	Units
Poplar	0.0358	USD/kg	MP steam	0.0081	USD/kg
NaOH	0.37	USD/kg	Electricity	0.076	USD/kwh
Na ₂ S	0.95	USD/kg	Operator labour	22.66	USD/h
Ethanol	0.92	USD/L	Supervisor labour	45.71	USD/h
H ₂ SO ₄	0.064	USD/kg	Organosolv lignin	0.60	USD/kg
Process water	2.95	USD/m ³	Kraft lignin	0.48	USD/kg
Cooling water	0.45	USD/m ³	Soda lignin	0.51	USD/kg
LP steam	0.0079	USD/kg	Levulinic acid	8.56	USD/kg

3. Results and Discussion.

The yields of lignin and levulinic acid production are summarized in Table 3.

Table 3. Production yields of lignin and levulinic acid production.

Scenario	Lignin extraction process	Lignin yield (kg/kg db.)	Levulinic acid yield (kg/kg db.)
1	Kraft	0.276	0.229
2	Soda	0.020	0.239
3	Organosolv	0.244	0.216

The lignin yield was estimated considering the resulting mass balances of each scenario. Kraft process offers a high lignin yield on a dry basis than the other options since more lignin can be isolated from the raw material. However, this yield is similar to the lignin extraction yield of the organosolv process. Indeed, both processes allow separating more than 80% of the lignin content in the raw material. In contrast, Soda extraction did not have a good mass performance since less than 10% of the lignin was isolated. This low extraction yield can be improved by increasing the temperature of the process. For instance, Lehto and Alén (2013) evaluated the lignin removal of wood chips using an 8 % w/w NaOH solution at different temperatures. The highest lignin removal (i.e., 72.5 %) was found at 150°C. Levulinic acid production yields are similar in all three scenarios since the cellulose content is not upgraded in monomeric sugars. The highest levulinic acid yield was obtained in scenario 2 (i.e., Soda process). Nevertheless, this value should be analyzed carefully since the biorefinery objective is not accomplished due to the low lignin production. In this way, the low yield of the lignin extraction in the Soda process makes the biorefinery a stand-alone process. The economic results suggest the Soda process as the most feasible option since the payback period is the lowest value reported in Table 4. Nevertheless, the low lignin production did not contribute to the process cash flow. Thus, the best option to upgrade woody biomass involving a lignin extraction stage is the Kraft process. Organosolv process did not reach economic feasibility at the proposed scale. Thus, a sensitivity analysis was carried out to find the Minimum

Processing Scale for Economic Feasibility (MPSEF) (Serna-Loaiza et al., 2018). After this analysis, the MPSEF of the organosolv process is about 2 t/h. Thus, higher scales are required to obtain positive cumulative cash over the lifetime of the project.

Table 4. Economic metrics estimated using a feedstock mass flow rate of 1 t/h.

Scenario	TCI (M.USD*)	PBP (years)	Cumulative cash over the lifetime of the project (M.USD*)	MPSEF (t/h)
1	4.51	6.61	1.58	0.95
2	3.55	4.04	3.67	0.89
3	5.75	0.00	-21.61	2.00

M.USD*: Millions of dollars.

In contrast, the proposed scale gave a positive cash flow at the end of the project in scenarios 1 and 2. Regarding the TCI, the Soda process has the lowest value since the capital investment required to isolate lignin contributes less than 5 % of the TCI. This low share in the lignin isolation process is attributed to the low input flow to this stage after the Soda process. Similarly, the lignin isolation process in the Kraft process accounts for 20 % of the TCI. This difference is attributed to the material flow of black liquor obtained after the Soda and Kraft processes. Finally, the Organosolv process is the most expensive since this lignin extraction method involves an additional step related to the ethanol recovery after lignin isolation. This stage (i.e., lignin isolation + ethanol recovery) accounts for more than 35% of the TCI of scenario 3. Therefore, the Organosolv process requires higher scales to reach economic feasibility. The economic evaluation results allowed us to determine the effect of the lignin extraction process on the economic performance of a woody-based biorefinery. Indeed, the organosolv process decreases the economic feasibility of the process, while Soda and Kraft processes have similar positive behaviors. Even so, the upgrading of lignin to other high value-added products can increase the economic performance of the process and reduce the required raw material mass flow.

4. Conclusions.

The comparison of three methods to extract lignin was done to evaluate the effect of these processes on the economic performance to produce lignin and levulinic acid as products. The product yield obtained in the simulation procedure allows proposing Kraft and Organosolv processes as the best ways to isolate lignin from poplar (*Populus sp.*). The soda process requires further research to find optimal conditions to increase lignin extraction. Regarding the economic perspective, the Kraft process offers the best economic performance since high lignin and levulinic acid production yields are allowed. Nevertheless, the low cost of lignin compared to other possible lignin-derived products limits the biorefinery potential. Indeed, the production of high-valued added compounds such as vanillin and eugenol plus the valorization of cellulose and hemicellulose to platform products can boost the real implementation of woody biomass applications at the industrial level using feasible and reliable raw materials input.

References

- Buzala, K.P., Kalinowska, H., Przybysz, P., Malachowska, E., 2017. Conversion of various types of lignocellulosic biomass to fermentable sugars using kraft pulping and enzymatic hydrolysis. *Wood Sci. Technol.* 51, 873–885.
<https://doi.org/10.1007/s00226-017-0916-7>

- Carvajal, J.C., Gómez, Á., Cardona, C.A., 2016. Comparison of lignin extraction processes: Economic and environmental assessment. *Bioresour. Technol.* 214, 468–476. <https://doi.org/10.1016/j.biortech.2016.04.103>
- Lehto, J., Alén, R., 2013. Alkaline pre-treatment of hardwood chips prior to delignification. *J. Wood Chem. Technol.* 33, 77–91. <https://doi.org/10.1080/02773813.2012.748077>
- Meng, X., Bhagia, S., Wang, Y., Zhou, Y., Pu, Y., Dunlap, J.R., Shuai, L., Ragauskas, A.J., Yoo, C.G., 2020. Effects of the advanced organosolv pretreatment strategies on structural properties of woody biomass. *Ind. Crops Prod.* 146, 112144. <https://doi.org/10.1016/j.indcrop.2020.112144>
- Öhman, F., Theliander, H., Tomani, P., Axegard, P., 2013. Method for separating lignin from black liquor. US 008486224 B2.
- Peduzzi, E., Boissonnet, G., Maréchal, F., 2016. Biomass modelling: Estimating thermodynamic properties from the elemental composition. *Fuel* 181, 207–217. <https://doi.org/10.1016/j.fuel.2016.04.111>
- Przybysz, K., Malachowska, E., Martyniak, D., Boruszewski, P., Howska, J., Kalinowska, H., Przybysz, P., 2018. Yield of pulp, dimensional properties of fibers, and properties of paper produced from fast growing trees and grasses. *BioResources* 13, 1372–1387. <https://doi.org/10.15376/biores.13.1.1372-1387>
- Saber, M., Karimi, K., Nasr, M., Kumar, R., 2020. Structural modification of pine and poplar wood by alkali pretreatment to improve ethanol production. *Ind. Crop. Prod.* 152, 112506. <https://doi.org/10.1016/j.indcrop.2020.112506>
- Serna-Loaiza, S., Carmona-Garcia, E., Cardona, C.A., 2018. Potential raw materials for biorefineries to ensure food security: The Cocoyam case. *Ind. Crops Prod.* 126, 92–102. <https://doi.org/10.1016/j.indcrop.2018.10.005>
- Solarte-Toro, J.C., González-Aguirre, J.A., Poveda Giraldo, J.A., Cardona Alzate, C.A., 2021. Thermochemical processing of woody biomass: A review focused on energy-driven applications and catalytic upgrading. *Renew. Sustain. Energy Rev.* 136. <https://doi.org/10.1016/j.rser.2020.110376>
- Solarte-Toro, J.C., Romero-García, J.M., López-Linares, J.C., Ramos, E.R., Castro, E., Alzate, C.A.C., 2018. Simulation approach through the biorefinery concept of the antioxidants, lignin, and ethanol production using olive leaves as raw material. *Chem. Eng. Trans.* 70. <https://doi.org/10.3303/CET1870155>
- Tian, D., Chandra, R.P., Lee, J.S., Lu, C., Saddler, J.N., 2017. A comparison of various lignin-extraction methods to enhance the accessibility and ease of enzymatic hydrolysis of the cellulosic component of steam-pretreated poplar. *Biotechnol. Biofuels* 10, 1–10. <https://doi.org/10.1186/s13068-017-0846-5>
- Towler, G.P., Sinnott, R.K., 2013. *Chemical engineering design. Principles, practice, and economics of plant and process design*, First. Ed, Chemical Engineering Design. Elsevier. <https://doi.org/10.1016/B978-0-08-096659-5.00022-5>
- Weingarten, R., Cho, J., Xing, R., Conner, W.C., Huber, G.W., 2012a. Kinetics and reaction engineering of levulinic acid production from aqueous glucose solutions. *ChemSusChem* 5, 1280–1290. <https://doi.org/10.1002/cssc.201100717>
- Weingarten, R., Conner, W.C., Huber, G.W., 2012b. Production of levulinic acid from cellulose by hydrothermal decomposition combined with aqueous phase dehydration with a solid acid catalyst. *Energy Environ. Sci.* 5, 7559–7574. <https://doi.org/10.1039/c2ee21593d>

Economic and environmental assessment of bio-hydrogenated diesel production process from waste cooking oil integrated with reforming process

Suksun Amornraksa^a, Lida Simasatitkul^{b,*}, Parinya Khongprom^c, Akekapob Sektaweelarp^b, Suttichai Assabumrungrat^d

^a*The Sirindhorn International Thai-German Graduate School of Engineering (TGGS), King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand*

^b*Department of Industrial Chemistry, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand*

^c*Department of Chemical Engineering, Faculty of Engineering, Prince of Songkla University, Songkla 90110, Thailand*

^d*Center of Excellence in Catalysis and Catalytic Reaction Engineering, Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand*

lida.s@sci.kmutnb.ac.th

Abstract

This work focuses on the performance analysis and life cycle assessment of two BHD production processes from waste cooking oil coupled with a reformer. The first process produces hydrogen from gas flare from the oil and gas industry, while the second process produces hydrogen from by-product propane. It was revealed that the second process requires less hydrogen by 15.27%, compared to the first process. However, its heating and cooling utilities were 22% and 10% higher than the first process. The economic assessment showed that the first process has a net present value of 2.39 times higher than the second one. Life cycle assessment is evaluated in terms of carbon footprint and other environmental impacts by using LCSof program. It was found that the carbon footprint released by the second process was higher than the first process by 0.45 kgCO₂ eq. The second process was also found to have higher environmental impacts in most categories. Thus, it can be concluded that the first process is preferable.

Keywords: Bio-hydrogenated diesel, Waste cooking oil, Process simulation, Economic assessment, Environmental assessment

1. Introduction

Bio-hydrogenated diesel (BHD) is considered a potential alternative energy source to replace petroleum diesel. The production process involves the hydrotreating of vegetable oil to enhance long-chained hydrocarbons. Four main reactions are involved in the process, which are hydrogenation, decarboxylation, decarbonylation, and hydrodeoxygenation. The reactions take place in a liquid phase under severe conditions (i.e., 350°C of temperature, 10 MPa of pressure in the presence of Ni catalyst) (Srifa et al., (2014)).

The gaseous by-products obtained are typically CO, CO₂, methane, and propane. One major problem of this process is that it requires a very high hydrogen to oil ratio (1000 Ncm³/cm³) to operate to complete the reaction (Hsu et al., 2018). It results in excessive unreacted hydrogen to handle and thus a high production cost.

Many researchers were concerned with the process design and evaluated the BHD process under economic criteria. Cheah et al. (2017) reported that raw material cost had the highest impact on profitability and selling price. Furthermore, the vegetable oil price, hydrogen gas price, and plant capacity were sensitive to the net present value (NPV), so waste cooking oil was preferable. Environmental impact assessment is another essential issue because energy utilization and waste disposal affect the environment and human health. The assessment of the BHD process is typically performed through a life cycle assessment (LCA). Boonrod et al. (2017) studied the processing and environmental impacts of raw materials, including crude palm oil, palm oil fatty acid distillate, and fatty acid methyl ester. The results showed that fatty acid methyl ester was the most suitable raw material due to low GHG emission.

There were only a few studies that concerned about hydrogen sources for the BHD production process. For example, Sungnoen et al. (2015) studied the combination of biodiesel, glycerol reforming and BHD process. The integrated process of hydroprocessing, gasification, and Fischer-Tropsch of by-product propane from hydroprocessing was performed for the bio-jet fuel (Alherbawi et al., 2020). However, none of them considered the technical, economic, and environmental aspects altogether. This work proposes two BHD production processes from waste cooking oil integrated with a reforming process. The first process produces hydrogen from a gas flare from the oil and gas industry, which is assumed as a waste but contains a considerable amount of hydrogen. The second process, however, produces hydrogen from the by-product propane. The best process would be considered in terms of economics and environmental impacts.

2. Methodology

The waste cooking palm oil used in this work consisted of 82.67% of triglycerides, 7.6% of free fatty acids, 2% of diglycerides, and 7.8% of dimers (J. Riera, and R. Codony (2000)). The composition of fatty acid content was following Srifa et al. (2014). The composition of gas flare was following Emam (2015). Simple triglyceride consisting of the same three fatty acids was assumed. A Non-random two-liquid model (NRTL) was used to predict thermodynamic properties. The sequence of unit operations in the model was arranged through thermodynamic properties analysis. The model simulation of both processes is performed by using the Aspen plus program. An economic analysis was then evaluated in terms of net present value (NPV), total production cost, payback period, and internal rate of return (IRR). Finally, the environmental impacts of both processes were analyzed by LCA.

2.1 BHD process with gas flare reforming

The flowsheet of the BHD process with gas flare reforming is shown in Figure 1(a). A gas flare from oil and gas industry is fed, together with steam, into a reformer, followed by a water-gas shift reactor with a temperature of 450°C. The effluent stream is sent to a pressure swing adsorption to obtain pure hydrogen (99.95%). It is then mixed with the excess hydrogen recirculated back from the process and waste cooking palm oil and enter

a hydrotreating reactor, operating with an H_2/oil ratio of $1000 \text{ Ncm}^3/\text{cm}^3$, a pressure of 50 bar, and a temperature of $300 \text{ }^\circ\text{C}$. The product is sent to a flash separator, operating at 20 bar and $130 \text{ }^\circ\text{C}$, to separate the vapor phase (i.e., methane, propane, H_2 , CO , and CO_2) and the liquid product. The separated gases are sent back to the reactor while the liquid product is sent to the first distillation column to remove the remaining gas and water. The bottom product containing the BHD is purified further in the second distillation column to produce the BHD with high purity (99wt% on dry basis).

2.2 BHD process with propane reforming

The flowsheet of the BHD process with propane reforming is shown in Figure 1(b). The configuration of this process is similar to the first process. However, the hydrogen feed used in this process is partly produced from the by-product propane from the hydrogen separator from an external source. Specifically, the gas stream from the flash separator is sent through an expander to reduce the pressure to 1 atm before sending it to a reformer. Then, it is followed by a water-gas shift reactor with a temperature of 450°C before recirculating back to the process.

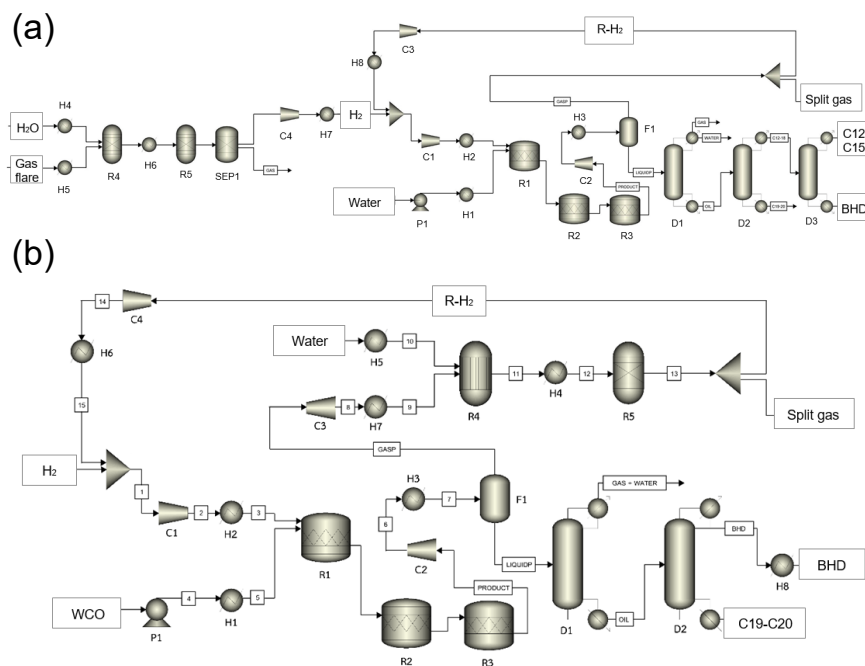


Figure 1 (a) stand-alone BHD process (b) gas-flare reforming (c) BHD process with by-product propane reforming

3. Results and Discussions

3.1 Process performance analysis

The process performance was determined in terms of BHD yield, H_2 consumption, and utility requirements. They are compared and illustrated in Table 1. The amount of waste

cooking oil feed at 56048 kg/hr and the BHD purity of 99 wt% were given for both processes. It can be seen that both processes provided the same yield of BHD at 75.16%. However, the second process could save hydrogen consumption by 15.27%, compared to the first process. For utility consumption, the second process used a higher heating utility by 22% due to the presence of the propane reforming process. Likewise, the cooling utility of the second process was also higher by 10% because of the larger volume of gases to be processed after the flash separator.

Table 1 Process performance of BHD processes

	BHD production with gas flare reforming	BHD production with by- product propane reforming
BHD yield (%)	75.16	75.16
H ₂ consumption (kg/hr)	1925	1631
Heating utility (x 10 ⁸ Btu/hr)	3.2	4.1
Cooling utility (x 10 ⁸ Btu/hr)	3.6	4.12

3.2 Economic assessment

Process data such as plant capacity, equipment sizing, amount of raw materials and chemicals, and utility consumptions were used to calculate the economics of the process. The total production cost for both BHD processes is summarized and shown in Figure 2. It was revealed that the raw material cost of both processes was a dominant factor of total production cost, taking about 82% - 90% of the total. The use of pure hydrogen as feed in the second process resulted in its raw material cost much higher than the first process. Although some hydrogen can be produced from the by-product propane, it is still not enough to compensate for the high cost of pure hydrogen feed (99.95%). The capital cost was represented in terms of the depreciation cost and, of which, about 40.56% came from a gas compressor. The presence of a gas compressor also required very high energy to operate, increasing the operating cost of the process. Another source of high energy-intensive units included the distillation columns and the heat exchanger used for preheating the feeds before entering the reformer and reactor. Comparing with a stand-alone BHD process (baseline scenario), the first process and the second process can reduce the cost by 18.96% and 2.75%, respectively.

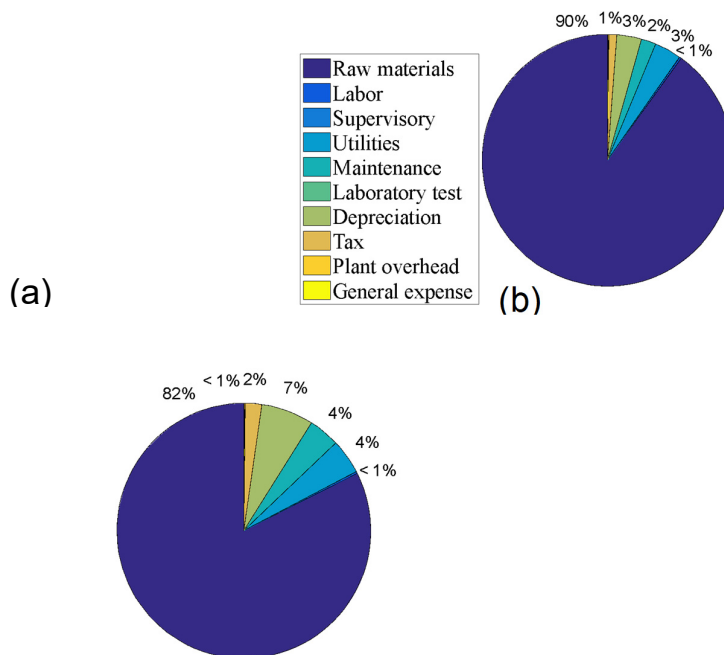


Figure 2 Total production cost contribution of (a) BHD process with gas flare reforming (b) BHD process with by-product propane reforming

The economic viability of the processes was evaluated through net present value (NPV), return on investment (ROI), and payback period. They were calculated and presented in Table 2. Noted that the plant lifetime of 10 years was assumed, and the income cash flow was derived from the selling price of the BHD at 1.1 \$/kg. It was found that both processes delivered the profit when the BHD selling price was higher than the cost of the raw materials by approximately 1.8 times. In conclusion, the first process offered better economics in all aspects. The main reason is that the second process used pure hydrogen as an additional feed, which is very costly. Therefore, it is clear to this point that to make the BHD process more competitive, a low-cost hydrogen supply source is necessary.

Table 2 Economic indicators of both processes

	BHD production with gas flare reforming	BHD production with by-product propane reforming
Return on investment	0.23	0.19
NPV (\$x10 ⁸ /year)	3.55	1.34
Payback period (year)	4.35	5.21

3.3 Environment impact assessment

In this section, the environmental impacts of both BHD processes were investigated by using LCSofT software. The gate-to-gate system boundary was assumed. In addition to carbon footprint, other environmental impact indicators were evaluated, namely, human toxicity noncarcinogenic impact (HTNC), Global warming potential (GWP), acidification potential (AP), human toxicity carcinogens (HTC), photochemical oxidation potential (PCOP), photochemical ozone formation, aquatic toxic potential (ATP), ecotoxicological potential (ET), particulate matter, and ozone depletion potential (ODP). The results were normalized in the same scale and illustrated in Figure 3. It was found that the carbon footprint of the first and second processes was 0.77 and 1.22 kg CO₂ eq., respectively which were higher than the baseline scenario (0.31 kg CO₂ eq.). The propane reformer was the most carbon footprint producer because it consumed the highest energy and released the most CO₂, accounting for 37.43% of the total carbon footprint. Heat exchangers and distillation columns were also the primary sources of CO₂. It was revealed that the reaction section was the main contributor to many environmental impact indicators in the first process. The gas flare reforming section notably contributed to ozone depletion potential (ODP) while, in the second process, the propane reforming contributed specifically to ET, HTNC, and ODP. HTNC was found to be the most significant environmental impact of the process. The hydrotreating section generates the highest amount of HTNC, followed by the separation section. The separation section, however, generates a relatively small amount of it. For example, in the first process, the HTNC generated from the hydrotreating, reforming, and separation sections were 1.5, 0.8, and 0.1 kg toluene eq., respectively.

4. Conclusions

The feasibility analysis of BHD processes from waste cooking oil integrated with a reformer was performed. Hydrogen sources from gas flare from the oil and gas industry and by-product propane were considered. Both processes were analyzed based on

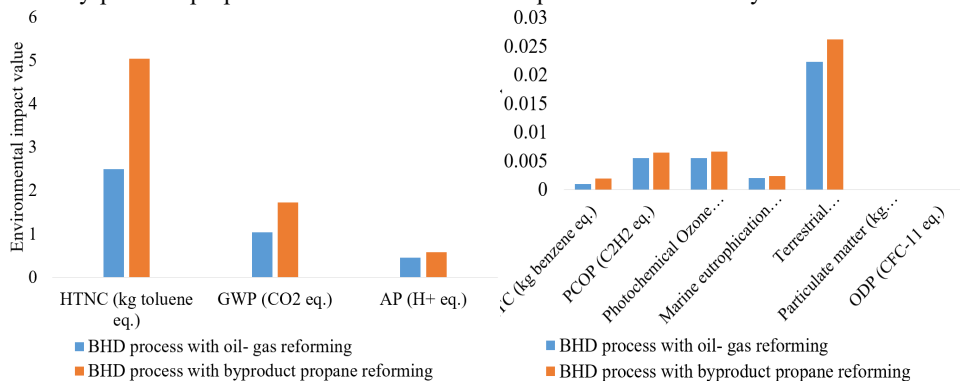


Figure 3 Environmental impacts BHD process with oil – gas reforming and BHD process with by-product propane reforming

economics and environmental impact indicators. The BHD process that used hydrogen from the gas flare was preferable. It consumed not only fewer utilities but also its economics and environmental impacts were more superior.

Acknowledgment

The authors would like to acknowledge the support of LCSoft program from PSE for SPEED. Support from Faculty of Applied Science (KMUTNB) fund (the grant No.642067)

References

- A. Srifa, K. Faungnawakij, V. Itthibenchapong, N. Viriya-empikul, T. Charinpanitkul, S. Assabumrungrat. (2014). *Journal Bioresource Technology*, 158, 81-90.
- B. Boonrod., C. Prapainainar., P. Narataruksa, A. Kantama, W. Saibautrong, K. Sudsakorn, T. Mungcharoen, P. Prapainainar. (2017). *Journal of Cleaner Production*, 142, 1210- 1221.
- B. Sungnoen, K. Ngaosuwana, S. Assabumrungrat. (2015). *International Journal of Mining, Metallurgy & Mechanical Engineering*, 3, ISSN 2320 – 4060.
- E. A. Emam. (2015). Gas flaring in industry. *Journal Petroleum and Coal*, 57, 535-555. G. Crahams and Head of STOA Team. (2000). *Recycled cooking oils : assessment of risks for public health*. European parliament, Division industry, Research and Energy
- K.-H. Hsu, W.-C. Wang, Y.-C. Liu (2018). *Energy*, 164, 99-111.
- K.W. Cheah, S. Yusup, H. Singh, Y. Uemura, H.L. Lam (2017). *Journal of Environmental Management*, 203, 950-961.
- M. Alherbawi, G. McKay, H.R. Mackey, T. Al-Ansari (2020). *Energy Conversion and Management*, 229, 113662.
- J. Riera, R. Codony (2000) *European parliament*, PE 289.889/Fin.St.

Modelling and simulation of a multiple feedstock integrated biorefinery for the production of aviation biofuel and other biofuels

Araceli Guadalupe Romero-Izquierdo,^a Fernando Israel Gómez-Castro,^a Claudia Gutiérrez-Antonio,^b Salvador Hernández,^a Massimiliano Errico^c

^a *Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Campus Guanajuato, Universidad de Guanajuato, Noria Alta s/n, Guanajuato, Gto. 36050, México.*

^b *Facultad de Ingeniería, Universidad Autónoma de Querétaro, Campus Amazcala, Carretera a Chichimequillas s/n km. 1, El Marqués, Qro. 76225, México.*

^c *Department of Green Technology, Faculty of Engineering, University of Southern Denmark, Campusvej 55, Odense, 5230, Denmark*
claudia.gutierrez@uaq.mx

Abstract

Renewable aviation fuel is the most promising alternative to reduce emissions in the aviation sector; its production is technically feasible, but not its cost respect to fossil jet fuel. Then, novel pathways need to be developed. One option is the processing through a biorefinery scheme, where biomass is completely converted to several products. Thus, in this work the computer-aided design and analysis of an integrated biorefinery for the conversion of *Jatropha curcas* fruit, castor bean plant, microalgae biomass and cooking oil is presented; the biorefinery is planned to produce aviation biofuel and other biofuels. The conversion processes are modelled and simulated using Aspen Plus. The assessment of the biorefinery is realized through the estimation of total annual cost, released CO₂ emissions, and the net gross profit. Results indicated that biojet fuel represents 33% of net gross profit, with 5.52 kW invested per kW of energy delivered by the products.

Keywords: biorefinery, biojet fuel, process simulation, Aspen Plus.

1. Introduction

Besides the recent challenges on human health, satisfying the energy demand is still a priority theme. The estimated energy requirements for 2040 will be 25% more than the amount required in 2019, and an increase up to 130% has been forecasted for 2050 in the main sectors involved in the energy consumption, such as the transport sector (BP, 2019). Among the transport ways, the aviation sector has presented the major growth rate, projecting an increase up to 40% for 2040 (Zhang et al, 2020). To achieve its sustainable development, the use of biojet fuel has been identified as the most appropriate strategy (Gutiérrez-Antonio et al, 2016). The biojet fuel can be obtained from any kind of biomass through different conversion routes (Gutiérrez-Antonio et al, 2020). However, its production is still not competitive regarding the fossil jet fuel. For this reason, research about novel alternatives to produce biojet fuel is highly needed. In this context, integrated biorefineries are considered a feasible option to produce a variety of products, including

chemicals, power, heat and steam from biomass (Energy, 2019). Few works about biorefineries to produce biojet fuel have been reported (Tongpun et al, 2019; Romero-Izquierdo et al, 2019; Neves et al, 2020); in most of them, only one kind of biomass is used. Nevertheless, the use of mixtures of raw materials could help to strengthen the industrial implementation of the production of this biofuel. Therefore, in this work an integrated biorefinery scheme to produce biojet fuel is presented, using *Jatropha curcas* (JC) fruit, castor bean (CB) plant, microalgae biomass (MA) and waste cooking oil (WCO). For this biorefinery, all the fractions of the raw materials (husk, oil, etcetera) are used, achieving the diversification of processes and biofuels. The computer-aided design of the biorefinery scheme is realized using Aspen Plus. The assessment of the scheme is performed through the estimation of total annual cost (TAC), the counting of CO₂ emissions, as well as the estimation of net gross profit of the products obtained. These estimations allow to analyse the biorefinery scheme by the use of energetic and environmental indicators, detecting the main opportunity areas for enhancing the scheme.

2. Process design and simulation

The biorefinery feedstock is defined as a mixture of JC seeds (19 wt%), CB seeds (58.5 wt%), MA (*Chlorella vulgaris*) after harvesting stage (22.4 wt%), and WCO without solid traces (0.1 wt%). The amount of each raw material was defined based on the potential cultivation (Zamarripa-Colmenero et al, 2009), the productivity of microalgal strain (Vera-Morales et al, 2017), and the recollection data in México (Romero-Izquierdo et al, 2019), respectively. The biorefinery was decomposed in four processing zones: oil extraction (Z1), press cake processing (Z2), oil conversion (Z3), and biojet fuel separation (Z4). The flow diagrams for Z1 and Z2 are presented in Figure 1, while in Figure 2 are showed the Z3 and Z4. The process starts in Z1, wherein the oil extraction is performed. The mechanical extraction of JC and CB seeds is carried out in two stages; heating up to 100°C using saturated steam, and mechanical pressing by *Crusher* module to obtain vegetable oil and press cake. In the case of MA, ultrasound technology is used. The oil is separated from press cake using a *Split* module as filter. Then, 20 wt% of JC press cake and 20 wt% of MA press cake are fed to anaerobic digestion process (AD) in Z2. The degradation of biomass in AD was described by the Buswell equation (Huun-Nguyen et al, 2014) and performed at 55°C and 1 atm, using a *RStoic* reactor module. The obtained biogas is used to produce electricity by a gas turbine (*Comp* module). The JC and MA press cake remaining enter two separated hydrolysis and fermentation processes (SHF), using two consecutive reactors modelled by *RStoic* modules. The operating conditions of acid hydrolysis are 121°C, 1 bar and H₂SO₄ solution 1 wt%; whilst the fermentation reactor operates at 30°C, 1 bar with an urea-to-biomass mass ratio of 0.1 (Ho et al, 2013; Dimian and Sorin, 2008). Roughly 7% ethanol yield is obtained, along with other co-products (furfural, lactic acid, succinic acid, glycerol). Ethanol is purified through 3 distillation columns (*RadFrac* modules) to reach 98 % mass purity. The first column (38 stages) separates the CO₂; while in the second column the ethanol is preconcentrated until 90% mass purity. Finally, in the last column (15 stages), glycerol is used as extractive entrainer to break the ethanol-water azeotrope. 70 wt% of ethanol is sent to ATJ process, and 30 wt% to transesterification processes. Z2 continues with the pyrolysis of CB press cake, modelled with the *RBatch* module at 623 K, using the kinetic model reported by Santos et al (2015). The biochar is removed by a *Filter* module; the condensable gases are the bio-oil product, whilst the non-condensable gases could be used as heating source.

In ATJ process, 3 reactors are modelled with *RStoic* modules: 1) ethanol dehydration at 450 °C and 11.4 bar (Ristic et al, 2017); 2) ethylene oligomerization at 120°C and 35 bar (Heveling et al, 1998); 3) hydrogenation at 100 °C and 15 bar (Gounder et al, 2011).

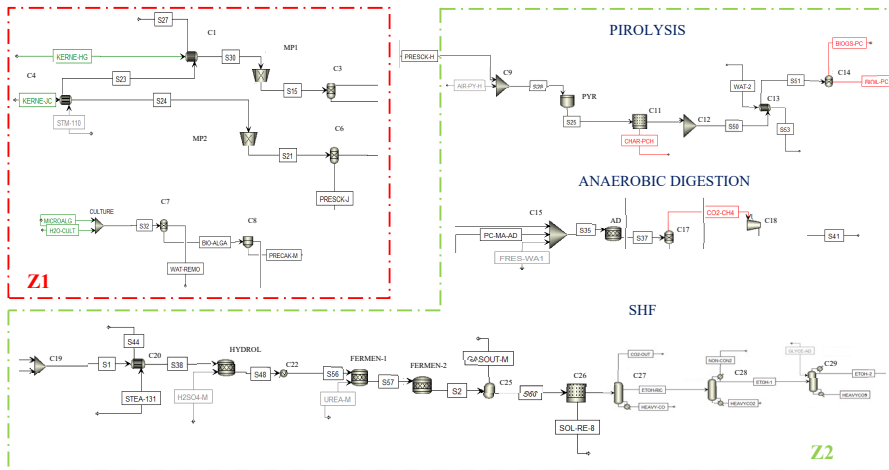


Figure 1. Flow diagram for Z1 and Z2 from Aspen Plus.

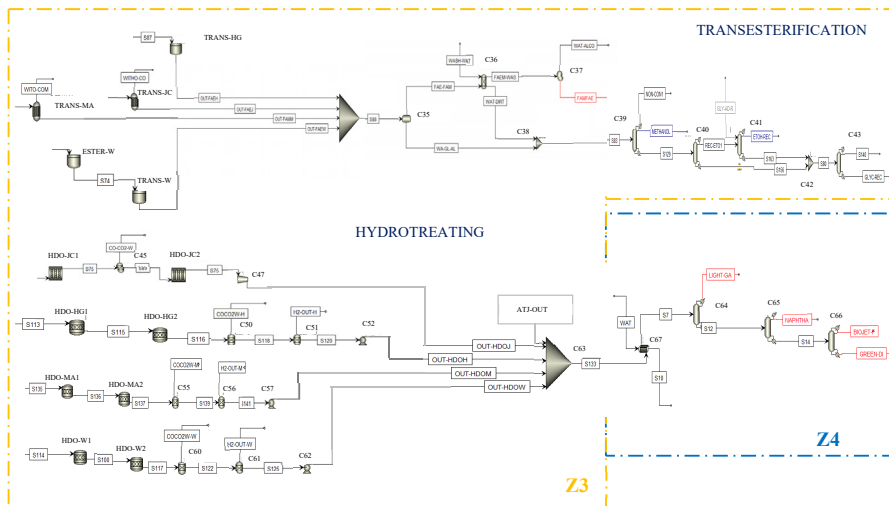


Figure 2. Flow diagram for Z3 and Z4 from Aspen Plus.

In Z3, 30 wt% of the oil flow is devoted to biodiesel production, and 70 wt% to the hydrotreating process. The production of biodiesel is carried out by transesterification reaction at 50 °C and 1 bar, using *REquil* modules for JC and MA, whilst *RBatch* modules are used for CB and WCO. For JC, CB and WCO the reaction uses ethanol (Abduh et al, 2013; Nivea et al, 2006; Romero-Izquierdo et al, 2019); for MA, the methanol is the reactant (Farooq et al, 2016). The obtained biodiesel is a mixture of ethyl and methyl esters. The separation of biodiesel starts with a *Decanter* module at 15 °C, followed by a wash-column and a *Flash* module to remove the excess of water. Then, the methanol,

ethanol and glycerol are separated using 4 distillation columns, modelled by *RadFrac* modules, reaching 0.958, 0.974 and 0.857 of mass purity, respectively. In the hydroprocessing (HDO), modules *RStoic* are used for CB oil, MA oil and WCO. The HDO for CB oil operates at 320 °C and 80 bar (Liu et al, 2015); while it operates at 350 °C and 50 bar for MA oil (Verma et al, 2011) and WCO (Romero-Izquierdo et al, 2019). Regarding to JC oil, two consecutive reactors (*RPlug* modules) at 80 bar are implemented, operating at 320 °C and 420 °C, respectively (Gutiérrez-Antonio et al, 2016). In order to reach 1 bar after the hydrotreatment for each oil, 4 turbines are used. After the turbines, the output streams from HDO are mixed with the hydrocarbons stream from ATJ process in Z4. This mixture is fed in a train of 3 distillation columns, modelled by the *RadFrac* module, reaching for each product 99% mass purity. In the first column (24 stages), light gases (C1-C4) are separated; while in the second one (38 stages), naphtha is purified (isoC5-C7). Finally, in third column (91 stages), biojet fuel (isoC8-C16) at top, and green diesel (C17-C21) at bottom are obtained. It is important to mention that sensibility analysis for decreasing the energy requirements were applied at separation zones (Z3 and Z4), varying reflux ratio and stages number. The biorefinery scheme is modelled with Peng-Robinson for the reactive zones, NRTL for SHF process, and BK10 for the Z4.

3. Economic and environmental assessment

The economic and environmental assessment for the biorefinery is realized through the total annual cost (TAC) and the counting of released CO₂ emissions. The TAC involves two items: the capital cost, calculated by Aspen Economics, adding 18% and 61% due to contingencies, installation fees and equipment maintenance, respectively (Turton et al, 2012); and the operating cost, which includes utilities cost (heating and cooling), raw material cost (JC, CB seeds, MA, hydrogen, ethanol, methanol) and additional reagents cost (glycerol, urea, H₂SO₄, etc), excluding catalyst cost and filters. The CO₂ emissions due to the production of steam and electricity are considered, and they are calculated through the methodology presented by Gutiérrez-Antonio et al (2016). On the other hand, the gross profit is calculated with the estimated product volume from simulation results, and the market price of each product: light gases, naphtha, biojet fuel, green diesel, bio-oil, bio-char (Romero-Izquierdo, 2020). These variables allow to analyse the biorefinery with two proposed indicators: IE, energetic indicator defined as invested energy used for heating, regarding to energy delivered by the products, and IA, environmental indicator defined as CO₂ emissions regarding to total product mass obtained from the biorefinery.

4. Results

According to results, the biorefinery generates 1% of its electricity requirements, thus, 81.6% of TAC is due to external power consumption. Moreover, the microalgae cost represents 85% of total raw materials cost. The major contributor to the TAC is the operating cost, whilst the capital cost represents less than 1%. In Table 1 is presented the TAC estimated for the biorefinery scheme. On the hand, the net gross profit estimated for obtained products is presented in Table 2. As it can be seen the greater contributors to the net gross profit are biodiesel with 37.7 % and biojet fuel with 33 %, which indicates its important position as useful biofuels in the market. Regarding CO₂ emissions, 6,579.44 MTon CO₂/year and 59.56 MTon CO₂/year are due to steam and electricity requirements. Thus, 99% of total CO₂ emissions are due to heating utilities. Finally, with regard IE, per each kW of total energy delivered by the products, 5.52 kW are invested in the processing

(by heating), which means that the energy investment for this biorefinery is 5.52 times higher than delivered energy by the products. Likewise, for IA, per each kg of total product mass, 1.51 ton of CO₂ are released; thus, the CO₂ emissions by the steam and electricity requirements are high, regarding to the products obtained; but the use of turbines to produce electricity in situ is a good strategy to diminish the total CO₂ emissions released by the processing.

Table 1. TAC for biorefinery scheme.

Operating Cost (million USD/year)					
Heating	Cooling	Additional reagents	Raw material	Electricity	Total Operating Cost
\$35.188	\$121.638	\$6,360.067	\$23,816.602	\$134,682.177	\$165,015.674
Annual Capital Cost (USD/year)					
Equipment Cost		A1	A2	Total Capital Cost	
\$67.205		\$12.096	\$40.995	\$24.059	
TAC (million USD/year): \$165,039.734					

Table 2. Net gross profit from biorefinery scheme.

Products	kg/h	USD/kg	USD/h	million USD/year
Light gases	20,646.56	\$0.46	\$9,560.67	\$81.265
Naphtha	60,770.14	\$0.47	\$28,258.12	\$240.193
Green diesel	65,657.22	\$0.96	\$62,719.16	\$533.112
Biodiesel	145,322.54	\$1.06	\$153,999.83	\$1,308.998
Biojet fuel	224,737.68	\$0.60	\$134,949.04	\$1,147.066
	GJ/h	USD/GJ		
Bio-oil	1,173.17	\$13.78	\$16,161.56	\$137.373
Bio-char	382.70	\$7.38	\$2,824.34	\$24.006
Total (million USD/year)				\$3,472.018

5. Conclusions

A biorefinery scheme with multiple feedstocks has been presented in this work for the production of biojet fuel and other biofuels. In this biorefinery, the microalgae cost represents 85% of the total raw materials cost, whilst the electricity cost is the major factor affecting the TAC (81.6%). In addition, 99% of total CO₂ emissions are due to steam requirements. Regarding to net gross profit, 33% corresponds to biojet fuel, thus its price in the market occupies the second place among the biofuels obtained in this proposal. Finally, 5.52 kW are invested (by heating) per kW of energy delivered by the products, whilst 1.51 ton of CO₂ per each kg of total product mass are released; thus, it is necessary to apply strategies to minimize the energy requirements and diminishing the CO₂ emissions. Nevertheless, this biorefinery scheme is a good alternative to produce biofuels, mainly biojet fuel, from the complete use of the selected biomasses.

6. Acknowledgments

Acknowledgment to CONACyT by the financial support, grants 239765 and 279753.

References

- A. C. Dimian, C. Sorin-Bildea, 2008, *Cehmical Process Design: Computer-Aided Case Studies*, Wiley.
- A. Farooq, et al, 2016, Transesterification of oil extracted from different species of algae for biodiesel production, *African Journal of Environmental Science en Technology*, 7, 358-364.

- A. G. Romero-Izquierdo, et al, 2019, Development of a biorefinery scheme to produce biofuels from waste cooking oil, *Computer Aided Chemical Engineering*, 46, 289-294.
- A. G. Romero-Izquierdo, 2020, Diseño, modelado y simulación de un esquema de biorefinería para el aprovechamiento integral de mezclas de materias primas renovables, PhD Thesis, Universidad de Guanajuato, México.
- A. Zamarripa-Colmenero, et al, 2009, Biocombustibles perspectivas de producción de biodiésel a partir de *Jatropha curcas* L. en el trópico de México, INIFAP.
- B. P. Energy Outlook, 2019, BP's Energy Outlook, Consulted (31-October-2019), Web site: <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2019.pdf>.
- C. Gutiérrez-Antonio, et al, 2016, Simultaneous energy integration and intensification of the hydrotreating process to produce biojet fuel from *jatropha curcas*, *Chemical Engineering and Processing: Process Intensification.*, 110, 134-145.
- C. Gutiérrez-Antonio, et al, 2021, Production processes of renewable aviation fuel, Elsevier, 1st Edition.
- D. Verma, et al, 2011, Aviation fuel production from lipids by a single-step route using hierarchical mesoporous zeolites, *Energy Environmental Science*, 4, 1667-1671.
- E. Energy Efficiency and Renewable Energy, 2019, Integrated biorefineries, Consulted (31-October-2019), Web site: <https://www.energy.gov/eere/bioenergy/integrated-biorefineries>.
- H. Zhang, et al, 2020, Prospects and perspectives foster enhanced research on bio-aviation fuels, *Journal of Environmental Management*, 274, 111214.
- H. Huun-Nguyen, et al, 2014, Energy potential from the anaerobic digestion of food waste in municipal solid waste stream of urban areas in Vietnam, *International Journal of Energy and Environmental Engineering*, 5, 365-374.
- J. Heveling, et al, 1998, Catalysts and conditions for the highly efficient, selective and stable Ni-AlSBA-15 mesoporous catalysts, *Applied Catalysts A General*, 173, 1-9.
- J. M. Vera-Morales, 2017, Obtención de aceite a partir de microalgas mediante métodos de bajo consumo energético, Thesis Dissertation, Universidad Autónoma de Querétaro.
- M. Y. Abduh, et al, 2013, Biodiesel synthesis from *Jatropha curcas* L. oil and ethanol in a continuous centrifugal contactor separator, *European Journal Lipidic Science Technology*, 115, 123-131.
- M. Ristic, M. Pacolli, 2017, Oligomerization of ethylene and ethanol into fuel through heterogeneous catalysis, Lund University.
- N. A.V. Santos, et al, 2015, Kinetic study of pyrolysis of castor bean presscake: an alternative use for solid waste arising from the biodiesel production, *Energy and Fuels*, 29, 2351-2357.
- N. L. D. S. Nivea, et al, 2006, Optimization of biodiesel production for castor oil, *Twenty-Seventh Symposium on Biotechnology for Fuels and Chemicals*, 405-414.
- P. Tongpun, et al, 2019, Techno-economic analysis of renewable aviation fuel production: from farming to refinery processes, *Journal of Clean Production*, 226, 6-17.
- R. C. Neves, et al, 2020, A vision on biomass to liquids thermochemical routes in integrated sugarcane biorefineries for biojet fuel production, *Renewable and Sustainable Energy Reviews*, 119, 109607.
- R. Gounder, E. Iglesia, 2011, Catalytic hydrogenation of alkenes on acidic zeolites: mechanistic connections to monomolecular alkane dehydrogenation reactions, *Journal Catalysts*, 277, 36-45.
- R. Turton, et al, 2012, Analysis, synthesis and design of chemical processes, 4th edition, Prentice Hall, New Jersey, 951-982.
- S. H. Ho, et al, 2013, Bioethanol production using carbohydrate-rich microalgae biomass as feedstock, *Bioresource Technology*, 135, 191-198.

Coordinated Management of Agricultural Waste and High-tech Value Bioproducts

Heriberto Alcocer-Garcia,^a Juan G. Segovia-Hernandez,^{a*} Eduardo Sanchez-Ramirez,^a Ana Gabriela Romero-García,^a Yicheng Hu,^b and Victor M. Zavala,^b

^a*University of Guanajuato, Noria Alta S/N, Guanajuato, GTO 36050, Mexico*

^b*University of Wisconsin-Madison, 1415 Engineering Dr, Madison, WI 53706, USA*

Abstract

The use of biomass has been associated mainly with the energy or biofuels. However, the production of high value-added products such as chemical building blocks, show a feasible alternative. However, the coordination of all the players within a market is complex because is necessary ensuring that the balance of profits generated will be greater than production and logistic cost, becoming an optimization problem. Therefore, we propose a coordination framework for managing lignocellulosic biomass from agricultural residues in a scalable manner by orchestrating biomass exchange, transportation, and transformation into value-added products, specifically: Levulinic acid and furfural. In the proposed framework, suppliers and consumers of biomass and derived products as well as transportation and technology providers bid into a coordination system that is operated by an independent system operator. All locations, prices for all biomass types and derived products are obtained by the operator by solving a dispatch problem that maximizes the social welfare and that balances supply and demand across a given geographical region. Through the coordinated framework, cleared transportation flows, locations and technologies involved are obtained to meet demand. The production of Levulinic acid was 330 Kt/year and 394 Kt/year of Furfural. Which represent 3% of global demand of raw materials for methyltetrahydrofuran. On the other hand, the benefit generated to biomass suppliers exceeds \$27 million / year, which would represent an increase in social welfare in the area, specifically to farmers. The framework allows to identify guidelines of the market behaviour considering all market players, with these guidelines is possible propose actions to be able to project a more competitive market for all products and that favors all market players.

Keywords: Coordination Framework, Agricultural Waste, Levulinic Acid, Furfural

1. Introduction

The concept of Biorefinery originated in the 1990's as a result of fossil fuel shortages and increasing trends in the use of biomass as a renewable raw material for the production of non-food products (Popa and Volf, 2018). The mismanagement and disposal of agricultural waste represents not only a danger to the environment, but also a lost economic opportunity. Specifically, there are various biomass processing technologies to generate value-added products from organic waste such as building block chemicals, fuels, and electricity. Unfortunately, the economic viability of waste processing routes depends to a large extent on the scale, transport costs and composition of agricultural waste (which vary depending on waste). Moreover, decision-makers are often unaware of the potential uses of waste streams and, therefore, of their inherent value. Another obstacle associated with waste management is the lack of cost-effective alternatives to

collect and transport large quantities over long distances. In Mexico, government regulations and incentives for the use of biomass are currently in place, however, incentives have not been able to exceed production and logistics costs. Therefore, there is no waste management network that guarantees a strong market, sustained growth for the urban and agricultural sectors. Another important aspect is that in Mexico studies on the use of biomass have focused on its use to generate energy or biofuels, for example: Quiroz-Ramirez et.al (2017) performs a optimal planning of biomass for biobutanol production and Santibañez-Aguilar et .al (2019) proposes the mathematical model for planning the use of biomass for energy production systems, both works leaving aside the potential biomass use as raw material for high added value products such as Levulinic acid and Furfural.

This work proposes a coordinated framework for the management of lignocellulosic biomass from agricultural waste (corn stover, wheat straw and sorghum stover) in a scalable way by organizing the exchange, distribution, transport and transformation of biomass into value-added bioproducts, specifically: Levulinic acid and Furfural. These high value-added products were selected because they are precursors of methyltetrahydrofuran, which is an additive that can be mixed with gasoline and that according to what was reported by Grand View Research (2015) has a demand of more than 20,000 kilo tons. The state of Guanajuato was taken as a case study, which is the second state in Mexico with the highest generation of agricultural waste that contribute 10.2 % of the national generation. Within the proposed framework, biomass suppliers, technologies involved in processing and transport suppliers participate in a coordinated system that is solved by an independent system operator (ISO). The operator solves a dispatch problem that maximizes social welfare and balances supply and demand in a specific geographical area, for this, the operator obtains all the locations, the prices of different biomass and derived products. Coordination allows the management of complex constraints and interdependencies arising from the transport and physicochemical transformations of biomass into value-added bioproducts.

2. Case of Study

Ministry of Agriculture, livestock, Rural Development, Fishing and Food (SAGARPA) shows that Guanajuato is the second state in Mexico with the highest generation of agricultural, which makes it an important sample to implement the methodology of the coordinated framework. Of the 3,830,305 tons by year of agricultural residues, corn stover represents 35 %, sorghum straw 41 %, wheat straw 17 % and barley straw 7 %. So, for this case study, only the three most abundant types of agricultural residues were considered. Production data were obtained from open data from the Agri-Food and *que* Fisheries Information Service (SIAP). To obtain the amount of biomass generated per amount of crop, we use the ratio reported by McIlveen-Wright et al. (2013) and biomass composition was taken within the ranges reported by Isikgor (2015). In addition, were used agricultural residues prices and transport cost reported by Santibañez-Aguilar et.al (2019). The optimization problem was formulated using 10-year data (2009-2018), each year was proposed as a scenario, all the scenarios were solved simultaneously. The main products are furfural, which is obtained from hemicellulose, obtaining 0.27 kg of furfural / kg of hemicellulose, and Levulinic acid, which is obtained from cellulose, with a yield of 0.49 kg of Levulinic acid / kg of cellulose. These conversions were proposed based on what was reported in previous works by Luo et. al (2019) and Reunanen et.al (2013). Biorefineries location was strategically selected, the criteria for selection were: its proximity to roads, provision of services for industries of this type and its proximity to

available biomass. In each location, the implementation of agricultural waste transformation technologies to Furfural and Levulinic acid was proposed. Levulinic acid and furfural production cost was considered based on the work of Gozan et. al (2018) and Cai et. al (2014), where the impact of the scale economy is clearly showed, considering plants between 295 tonne/day and 701 tonne/day, prices of 2159 USD and USD 1430 USD per tonne of biomass are obtained, respectively.

3. Coordination Framework

We consider a system that comprises a set of geographical locations (nodes) N , products P , suppliers S , consumers \mathcal{D} , transportation providers \mathcal{L} , and transformation (technology) providers \mathcal{T} . Products comprise different waste stream types and derived products, transportation providers offer alternatives (e.g., hauling, railway, pipelines) to move products between locations, and technology providers offer alternatives to process products to produce other higher value products.

The management system proposed is operated by an ISO that collects bidding information from all participants (costs, capacities, and transformation factors) to obtain optimal allocations of product supply, demand, transportation, and transformation services. The ISO determines these allocations by solving a dispatch problem that finds optimal transportation and transformation pathways for waste and derived products that maximize the social welfare and that balance supply and demand for all products across a geographical region. For reasons that will become apparent, this dispatch problem can be seen as a market clearing problem. The cleared transportation providers create a transportation network that connects nodes in the system that perform exchange of products. An efficient management system is expected to clear suppliers and providers that offer services at low costs and will give preference to consumers with higher bidding costs. The clearing problem also aims to find prices that are used to properly remunerate suppliers and providers to cover their service costs and to charge consumers for the service provided.

3.1. Dispatch Formulation

Given the bidding information associated with each supplier α^s , consumer α^d , transportation provider α^f , technology α^ξ and maximum capacity of biomass \bar{s} , demand \bar{d} , transportation provider \bar{f} , processing $\bar{\xi}$ the ISO solves the clearing problem Eq. (1) to find allocations. These allocations maximize the social welfare Eq. (1a) and satisfy the physical conservation laws Eq. (1b), and capacity constraints Eq. (1c)-(1f). Maximizing the social welfare function maximizes the demand served and minimizes the costs of supply, transportation, and transformation. The conservation laws are also known as the balancing constraints or market clearing constraints. The first term in parenthesis is the total input flow for product p into node n (given by supply flows and transportation flows entering the node). The second term in parenthesis is the total output flow of product p from node n (given by the demand flows and transportation flows leaving the node). The third term is the generation/consumption rate of product p in all technologies located at node n . In a work published by Apoorva et al. (2019) the complete dispatch formulation and fundamental properties of the coordination framework are described.

$$\max_{(s,d,f,\xi)} \sum_{j \in \mathcal{D}} \alpha_j^d d_j - \sum_{i \in \mathcal{S}} \alpha_i^s s_i - \sum_{\ell \in \mathcal{L}} \alpha_\ell^f f_\ell - \sum_{t \in \mathcal{T}} \alpha_t^\xi \xi_t \quad (1a)$$

$$\text{s.t.} \left(\sum_{i \in \mathcal{S}_{n,p}} s_i + \sum_{\ell \in \mathcal{L}_{n,p}^m} f_\ell \right) - \left(\sum_{j \in \mathcal{D}_{n,p}} d_j + \sum_{\ell \in \mathcal{L}_{n,p}^{out}} f_\ell \right) + \sum_{t \in \mathcal{T}_n} \gamma_{t,p} \xi_t = 0, (n,p) \in \mathcal{N} \times \mathcal{P}, (\pi_{n,p}) \quad (1b)$$

$$0 \leq s_i \leq \bar{s}_i, i \in \mathcal{S} \quad (1c)$$

$$0 \leq d_j \leq \bar{d}_j, j \in \mathcal{D} \quad (1d)$$

$$0 \leq f_\ell \leq \bar{f}_\ell, \ell \in \mathcal{L} \quad (1e)$$

$$0 \leq \xi_t \leq \bar{\xi}_t, t \in \mathcal{T} \quad (1f)$$

4. Results

Optimization problem includes: 51 nodes, 1277 biomass suppliers and 8 different products (raw materials and derived products). The 46 municipalities of the state of Guanajuato and their agricultural production for the period 2009-2018; four installation locations for conversion technologies and a single point of demand for the products resulting from the transformation was considered. To give an idea of the logistical complexity involved, the optimization problem is a linear programming problem that contains more than 24801 decision variables and 4860 restrictions.

For this case study, 60% of the biomass of the state of Guanajuato was used. This would produce around 434 Kt/year of Levulinic acid and 487 Kt/year of Furfural. However, through the implementation of the coordinated framework it shows, that having an availability of 60% biomass, the best solution found only uses 80% of this biomass, it occurs because the wheat straw processing cost is affected by the scale economy. If wheat straw were to be considered, a bigger technology for processing should be proposed. Therefore, in this work, a production of 330 Kt / year of Levulinic acid and 394 Kt / year of Furfural was obtained, which represents 3.0% of global demand of raw materials for the MTHF.

Table 1 shows the prices at which the market clears (clearing prices), these prices are calculated independently for each node present in the problem. As a result, we obtained a price range where our market still exists and all market players have benefit. In this case, the clearing prices for wheat straw are zero since they do not participate in the market. Figure 1 shows the cleared transportation flows in Guanajuato State for different biomass and derived products, in Figure 1c no flows appear because wheat straw does not participate in the market.

Table 1. Clearing prices

Product	Clearing prices		
	Theoretical Price	Minimal Price	Maximum Price
Corn stover	18.33	8.18	21.01
Sorghum straw	14.96	5.24	17.61
Wheat straw	16.66	0	0
Furfural	1700.00	1685.71	1700.00
Levulinic acid	11023	11008.71	11023

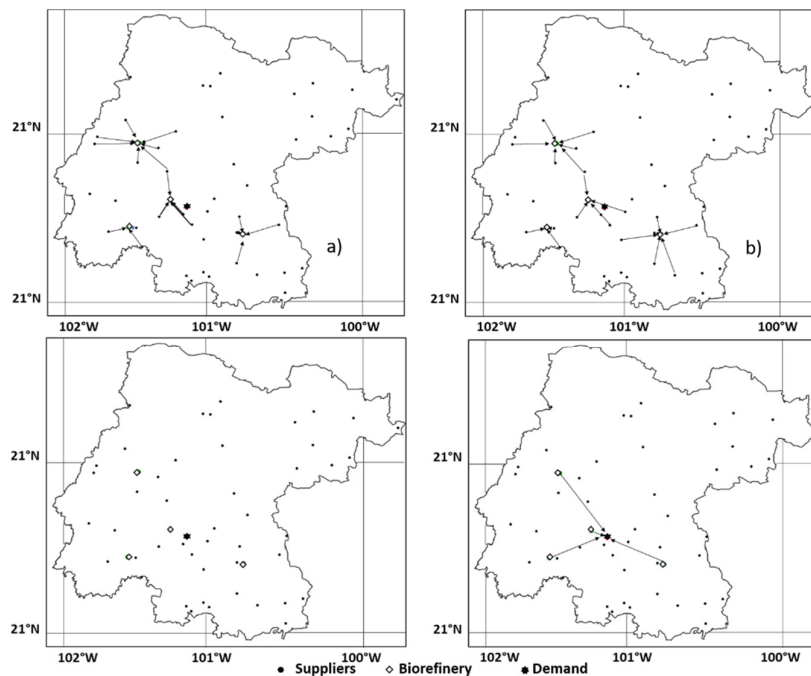


Figure 1. Cleared transportation flows in Guanajuato State for different biomass and derived products: A) Corn stover, B) Sorghum Straw, C) Wheat Straw and D) Derived products.

In Figure 1, not all suppliers are involved in the market, this due to process cost and mainly to transportation cost. The main suppliers are in municipalities near the biorefineries, in the center and south of the Guanajuato state. It may seem that there are very few suppliers involved, however it is necessary to mention that in each municipality there may be up to 10 corn stover suppliers and 10 sorghum stover suppliers. The municipalities involved in the market are: Abasolo, Acambaro, Apaseo el Alto, Celaya, Cortazar, Guanajuato, Irapuato, Jaral del Progreso, León, Penjamo, Pueblo Nuevo, Purísima del Rincon, Romita, Salamanca, Salvatierra, San Francisco del Rincon, Silao de la Victoria, Valle de Santiago and Villagran. The profit generated through the implementation of the coordinated system is $1.84E+03$ million USD/year, this profit is large due to the price of Levulinic acid. Total revenue is $4.49E+03$, this revenue is associated to sales. Although these results are encouraging, it is also important to mention that it has a direct impact on the region since the total transportation cost is 5.79 million USD/year, the total supply cost is $2.76E+01$ million USD/year.

5. Conclusions

We presented a coordinated system to facilitate the management of agricultural waste in a scalable way by coordinating the exchange, transport, and transformation of biomass into value-added products. The framework operates as a coordinated marketplace under which waste suppliers and consumers, as well as transport and technology, provide

bidding services to an independent system operator. The prices of waste and derived products in different geographical locations are obtained by solving a dispatch problem that maximizes social welfare and balances products throughout the region. Coordination enables you to handle complex interdependencies between products and locations. We demonstrate that the system offers prices and allocations that satisfy the fundamental economic and efficiency properties expected of a competitive market. We also show that the proposed system provides a systematic framework for monetizing environmental impacts, health impacts, and remediation benefits. In addition, prices reveal the true value of waste streams and capture spatiotemporal variations that help prioritize areas and reveal the need for investment in processing technologies, transportation, facility relocation, and seasonal storage. The proposed geographic framework is scalable in the sense that it can provide open access that encourages transactions between many small and large players by allowing coordination with other infrastructures. Additionally, this coordinated framework represents a valuable contribution to the visualization of the market in Mexico resulting from the use of biomass to obtain building block chemicals, such as Levulinic acid and Furfural, since to date, the efforts made have focused in the production of energy and biofuels. A coordinated framework will become increasingly necessary as the human population grows and mobilizes, and as the availability of resources becomes more uncertain.

References

- A. M. Sampat, Y. Hu, M. Sharara, H. Aguirre-Villegas, M. Ruiz-Mercado, R. A. Larson, and V. M. Zavala, (2019), Coordinated management of organic waste and derived products” *Comput Chem Eng*, 128:352–363.
- C. M. Cai, T. Zhang, R. Kumar, and C. E. Wyman, 2014, Integrated furfural production as a renewable fuel and chemical platform from lignocellulosic biomass. *J. Chem. Technol. Biotechnol.* 89(1), 2-10.
- D.R. Mellveen-Wright, Y. Huang, S. Rezvani, D. Redpath, M. Anderson, A. Dave, and N.J. Hewitt, 2013, A technical and economic analysis of three large scale biomass combustion plants in the uk. *APPL ENERG*, 112:396 – 404,.
- F. H. Isikgor and C. R. Becer, 2015, Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polym. Chem.*, 6:4497–4559.
- J. E. Santibañez-Aguilar, D.F. Lozano-García, F.J Lozano, and A. Flores-Tlacuahuac, 2019, Sequential use of geographic information system and mathematical programming for optimal planning for energy production systems from residual biomass. *Ind. Eng. Chem*, 58(35):15818–15837.
- J.J. Quiroz-Ramírez, E. Sánchez-Ramírez, J.G. Segovia-Hernández, S. Hernández, J.M. Ponce-Ortega, 2017, Optimal Selection of Feedstock for Biobutanol Production Considering Economic and Environmental Aspects, *ACS Sustain. Chem. Eng.*, 5, 4018 – 4030.
- J. Reunanen, P. Oinas, and T. Nissinen, 2013, Process for recovery of formic acid. US Patent 8,530,695.
- M. Gozan, B. Ryan, and Y. Krisnandi, 2018, Techno-economic assessment of levulinic acid plant from sorghum bicolor in indonesia. In *IOP Conf. Ser. Mater. Sci. Eng.*, volume 345.
- V. Popa and I. Volf, 2018, Biomass as renewable raw material to obtain bioproducts of high tech value, Elsevier, xv-xvii
- Y. Luo, Z. Li, X. Li, X. Liu, J. Fan, J.H. Clark, and C.Hu, 2019, The production of furfural directly from hemicellulose in lignocellulosic biomass: A review. *Catalysis Today*, 319:14 – 24, SI: Biomass Valorization.

A design and optimization framework for (bio-) chemical process based on exergo-economic and environmental aspects.

Jaime David Ponce-Rocha,^a Martín Picón-Núñez,^a Ana Carvalho,^b Andreia Santos,^b Fernando I. Gómez-Castro,^a Ricardo Morales-Rodríguez^{a*}

^a*Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Campus Guanajuato, Universidad de Guanajuato, Noria Alta S/N, Guanajuato, Gto., 36050, México.*

^b*Centro de Estudos de Gestao do IST (CEG-IST), Av. Rovisco Pais, Lisbon, 1049-001, Portugal.*

ricardo.morales@ugto.mx

Abstract

Currently, chemical processes must be designed and optimized considering multicriteria objectives focused on sustainable aspects. These targets can be effectively achieved via a systematic framework, capable to guide the users to take the best decision in terms of given objectives. Thus, the present work develops and implements a framework through four main steps that use multiple computer-aided tools and methods, linked to optimal indicators such as exergy loss, total annual cost and environmental factors and scores. The implementation of the methodology used a case study consisting of the ABE separation and purification scheme, using liquid-liquid extraction and conventional distillation techniques, as well as the evaluation of five extracting agents. The framework allowed to enhance the performance of all scenarios with respect the base case.

Keywords: Sustainable process design, Exergo-economic analysis, Aspen Plus-Matlab Interface, Multi-Objective Genetic Algorithms, Life Cycle Impact Assessment.

1. Introduction

Process system engineering has been successfully applied to address various product-process design problems present in the chemical and biochemical/biological industry, where the integration of computer-aided tools and systemic frameworks based on the optimization of processes in operational and design terms, it is usually a key factor for the selection of the best candidate products-processes (Gani, 2009). The current environmental challenges related to emissions, waste management and energy efficiency of the systems have motivated the search for economically feasible systems with a lower environmental impact, in this sense, the implementation of the exergo-economic analysis provides an adequate route to satisfy these targets (Akchiche et al., 2020). Additionally, exergo-environmental analysis has provided a useful approach for lower environmental impact processes design, moreover, the combination of the exergy, economic and environmental analysis approach in energy-intensive systems, has shown superior

economic and environmental benefits, compared with conventional analyses for process design (Li et al., 2019). Therefore, based on the current state of the art, the objective of this work focuses on the development and implementation of a systematic framework to determine a sustainable optimal process design based on exergetic, environmental and, economic indicators.

2. A design and optimization framework for (bio-)chemical process based on sustainability aspects

The framework for the design and optimization of (bio-)chemical processes based on sustainability aspects is presented in Figure 1. This framework includes four systematic main steps, which can be summarized as follows: 1) Problem definition, data collection, optimization techniques, available technology and, the review of methods and indicators focused on the sustainability of the product and process design. 2) Integration of thermodynamic aspects and rigorous systems designs through the use of commercial process simulator (e.g. Aspen Plus), in addition to the implementation of the model and sustainable multi-objective functions for optimization into a programming environment (e.g. Matlab), generating a simulation-optimization platform that provides a set of results that satisfy all problem constraints. 3) Selection of methods for the evaluation of the environmental impacts in process and products design, with the aim of implementing a better metric for the selection of sustainable processes. 4) Integration of graphical tools with the aim of facilitating the analysis of the results, the selection and implementation of the best scenario in the selected case study.

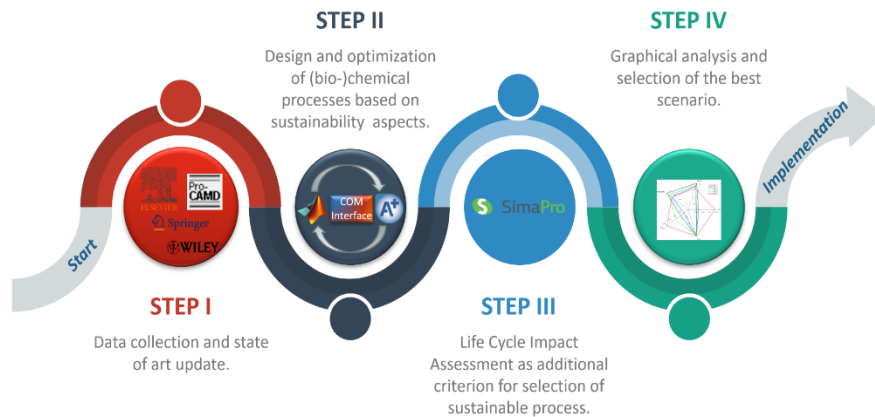


Figure 1. A design and optimization framework for (bio-)chemical process based on sustainability aspects.

3. Case study

The framework was evaluated using the separation and purification of Acetone-Butanol-Ethanol (ABE) from a fermentation broth mixture. The feed stream used in this study was 20,393.64 kg/h with a mass composition of: Acetone (0.1116), Butanol (0.2922), Ethanol (0.0784), Water (0.5138), and Carbon Dioxide (0.0040). A hybrid separation process (liquid-liquid extraction and conventional distillation) was selected, where five extracting agents (EA) were evaluated.

4. Framework description

4.1. Data collection and state of art update.

The first step was to define the case study. In a previous work some authors identified the advantages and disadvantages in the ABE broth mixture separation and purification, including general characteristics of biochemical processes as: multi-production process, diluted products, azeotropes presence, etc. These aspects were successfully addressed by implementing optimal schemes based on total annual cost (TAC) and environmental targets, including environmental, friendlier and economic extracting agents, on the separation and purification schemes (SPS) (Ponce-Rocha et al., 2018).

4.2. Design and optimization of (bio-)chemical processes based on sustainability aspects.

Nowadays, a commercial process simulator (e.g., Aspen plus) allows process design using a rigorous thermodynamic description, through the determination of suitable model and binary interaction parameters. Once the system components were established, the selection of the thermodynamic models was carried out, the NRTL and Hayden O'Connell modification were selected to describe the liquid-vapor equilibrium, respectively (Morales-Espinosa et al., 2017). The SPS were carried out based on heuristic rules, the equipment design was done using shortcut methods, rigorous methods and, sensitivity analysis to improve the design of columns. The selected extracting agents to be evaluated were hexyl acetate (HEX), heptyl acetate (HEP), a HEX-HEP mixture, 2-ethyl-1-hexanol (2-ETH-1-OL), and dichloromethane (DCM). Due to the volatilities of extracting agents, two different SPS were generated as shown in Figure 2.

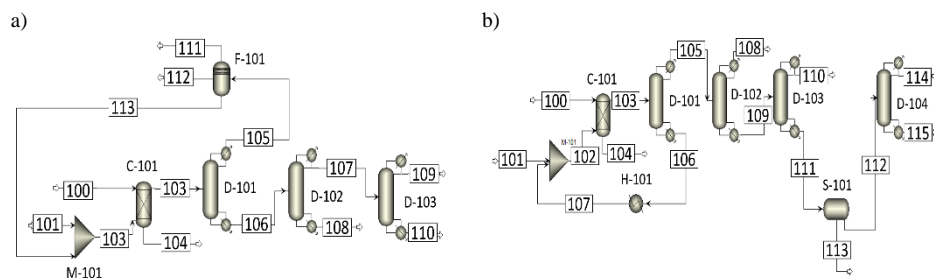


Figure 2. a) DCM process configuration (ELL+3DC columns); b) HEX, HEP, HEX-HEP, and 2-ETH-1-OL process configuration (ELL+4DC columns).

An inherent characteristic in the optimization of chemical processes involves the presence of multiple objectives, unfortunately, commercial simulators lack of multi-objective optimization (MOO) modules. Thus, an alternative to tackle those issues have been done by integrating a simulation-optimization platform, using a computational interface (e.g., COM) into a programming environment. Regarding to the optimization part, the use of black-box optimization methods do not require an algebraic model for their description, only iterative evaluations of the process variables and conditions, therefore, multi-objective genetic algorithms (MOGA) are an excellent tool for making this task (Muñoz et al., 2017).

4.2.1. Optimization problem formulation

In this work the multi-objective optimization is represented as following in Equation 1:

$$\underset{x}{\text{Min}} \quad \underline{Z} = [Z_k(\underline{x})] \quad k = 1, 2, \dots, k. \quad \text{s.t.} \quad \mathbf{h}(\underline{x}) = 0; \mathbf{g}(\underline{x}) \leq 0 \quad (1)$$

where \underline{Z} represents a vector with k -objectives functions, \underline{x} is the vector of state variables (stages number, feed stages, reflux ratios, distillate or bottoms rates, and the inlet flow of the extracting agent). $\mathbf{h}(\underline{x})$ and $\mathbf{g}(\underline{x})$ are the vectors of model equality and inequality constraints associated with mass purities (P) and mass recoveries (R). (Equation 2-3 for HEX, HEP, HEP-HEX and, 2-ETH-1-OL. Equations 4-5 for DCM).

$$[P_{EA}, P_A, P_B, P_E] \geq [0.999, 0.950, 0.995, 0.900] \quad (2)$$

$$[R_{EA}, R_A, R_B, R_E] \geq [0.999, 0.995, 0.000, 0.992] \quad (3)$$

$$[P_{EA}, P_A, P_B, P_E] \geq [0.995, 0.990, 0.995, 0.990] \quad (4)$$

$$[R_{EA}, R_A, R_B, R_E] \geq [0.980, 0.990, 0.999, 0.000] \quad (5)$$

In this study, four objective functions were selected to evaluate and optimize the sustainable process configurations: the energy requirements, exergy losses, TAC and, the E factor.

The energy requirements (ER) associate the ratio between total heat requirements and the amount of product purified in each case study, as described in Equation 6.

$$ER = \frac{\text{Energy requeriments [MJ]}}{\text{Products [kg]}} \quad (6)$$

Exergy losses (σT_0) were calculated considering the total exergy losses given by the heating and cooling processes (Q), separation (Sep) and, concentration changes ($Conc$) in all the equipment (j) and streams processes (i) (see Equation 7).

$$\sigma T_0 = \Delta \dot{E}x_{In} - \Delta \dot{E}x_{Out}$$

$$\sigma T_0 = \sum_{j=1}^n \sum_{i=1}^m [\Delta \dot{E}x_{Q,j} + \Delta \dot{E}x_{Sep,i} + \Delta \dot{E}x_{Conc,i}] \quad (7)$$

Total annual cost (TAC) illustrated in Equation (8), establishes the equipment investment cost and payback period (adjusted to five years), and the sum of annualized cost of utilities, referred to extraction agent lost, cooling and heating services.

$$TAC = \frac{\text{Investment cost}}{\text{Payback period}} + \text{Utilities} \quad (8)$$

The E factor (E) determinates the waste and products ratio (see Equation 9). Further, the streams with water composition greater than 99% (mass) were not considered as waste streams.

$$E_{factor} = \frac{\text{Waste [kg]}}{\text{Products [kg]}} \quad (9)$$

3.2.1 Metaheuristic optimization strategy

The generated alternatives were evaluated and optimized using a MOGA: Gamultiobj (a variant of NSGA-II) integrated in Matlab. Each optimization considered 50 generations, 200 individuals, and a crossover fraction of 0.8.

4.3. Life Cycle Impact Assessment as an additional criterion for selection of sustainable process.

The life cycle analysis (LCA) evaluation has been considered as one of the most appropriate methodologies to estimate the environmental impacts of products and processes. This methodology is comprised of four main stages: definition of objective and scope, inventory analysis (LCI), impact evaluation or Life cycle impact assessment (LCIA), and interpretation of results. The LCIA methods (e.g. ReCiPe 2008) allow to estimate a single scores (SS) capable to describe the total environmental impact, providing an objective criterion for the comparison of systems with similar characteristics (Santos et al., 2018).

4.4. Graphical analysis and selection of the best scenario.

Finally, the previous results were compiled and selected according to the best SPS for each extracting agent using a graphical analysis to represent the four sustainable objective functions.

5. Results

The optimal results that accomplished the established constrains were selected for each extracting agent (EA), and subsequently, the total environmental impact was measured by life cycle impact assessment using ReCiPe 2008 method. Table 1 shows the most relevant indicators for base and optimal scenarios obtained for each solvent in MOO and LCIA. The best scenarios for TAC, σT_0 , ER, E and SS_{LCIA} were when using 2-ETH-1-OL, DCM, HEP, HEP and HEP, respectively.

Table 1. Summary of results of the evaluated extracting agents.

EA	Scenario	TAC [M\$/year]	σT_0 [MW]	ER [MJ/kg- ABE]	E-factor	SS_{LCIA}
HEX	Base	13.63	2.99	12.93	0.51	0.28
	Optimal	10.27	2.18	9.13	0.47	0.26
HEP	Base	11.72	3.01	10.35	0.28	0.23
	Optimal	8.09	1.95	6.99	0.25	0.21
HEP- HEX	Base (0.5-0.5)	12.49	2.93	11.14	0.35	0.24
	Optimal (0.41-0.59)	9.10	2.02	7.83	0.37	0.24
2-ETH-1- OL	Base	14.88	4.04	15.52	0.79	0.31
	Optimal	7.67	1.98	8.12	0.77	0.29
DCM	Base	26.63	1.79	7.28	1.28	0.22
	Optimal	26.14	1.77	7.00	1.28	0.22

Figure 3 shows the optimal sustainability indicators for each extracting agent given in Table 1. This graph allows a visual and clearer comparison between the optimal results

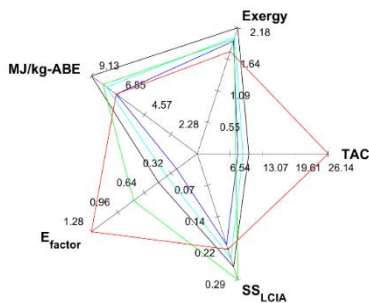


Figure 3. Optimal results using the different EA.

for each extraction agent. In this sense, the use of HEP as extracting agent provides the best scenario for separation and purification of ABE mixture under the selected sustainability targets.

Furthermore, analyzing all indicators in Table 1, it is possible to observe a major benefit in the sustainable indicators for each extraction agent optimal case (TAC, σT_0 , ER, and E-factor). In addition, these improvements minimize or alternatively keep an identical environmental impact.

6. Conclusion

The systematic steps and selected tools and methods integrated in this framework, allowed to improve most of indicators for each SPS. The results illustrated that the best SPS was found when using HEP as EA, which permitted to minimize substantially the TAC (31%), exergy losses (35%), energy requirements (32%), E-factor (11 %) and LCIA (9%) with respect to the base scenario; in addition to major sustainability benefits compared with the other indicator's values reported by each EA in Table 1. Therefore, the process design and optimization based on sustainable indicators can be applied successfully in (bio-) chemical and chemical processes with environmental and energetic problems.

7. Acknowledgments

Jaime D. Ponce-Rocha acknowledges the National Council of Science and Technology (CONACyT) for the financial support on the development of this project.

8. References

- M. Akchiche, J.-L. Beauquin, S. Sochard, S. Serra, J.-M. Reneaume & P. Stouffs, 2020, Sustainable Exergoeconomic Optimization of Petroleum Production Systems. *Computer Aided Chemical Engineering*, 48, 1447–1452.
- R. Gani, 2009, Modelling for PSE and product-process design, *Computer Aided Chemical Engineering*, 27, 7–12.
- X. Li, Y. Zhang, L. Fang, Z. Jin, Y. Zhang, X. Yu, X. Ma, N. Deng, & Z. Wu, 2019, Energy, exergy, economic, and environmental analysis of an integrated system of high-temperature heat pump and gas separation unit, *Energy Conversion and Management*, 198, 111911.
- N. Morales-Espinosa, E. Sánchez-Ramírez, J. J. Quiroz-Ramírez, J. G. Segovia-Hernández, F. I. Gómez-Castro, & R. Morales-Rodríguez, 2017, A Framework for an Optimized Sustainable Product and Process Design: Acetone-Butanol-Ethanol Separation and Purification, *Computer Aided Chemical Engineering*, 40, 697–702.
- C. A. Muñoz, D. Telen, P. Nimmegeers, L. Cabianca, F. Logist, & J. Van Impe, J., 2017, Investigating practical aspects of the exergy based multi-objective optimization of chemical processes, *Computer Aided Chemical Engineering*, 40, 2173–2178.
- J. D. Ponce-Rocha, E. Sánchez-Ramírez, J. G. Segovia-Hernández, F. I. Gómez-Castro, & R. Morales-Rodríguez, 2018, Optimized sustainable molecular and purification process design framework: acetone-butanol-ethanol case study. *Computer Aided Chemical Engineering* 44, 385-390.
- A. Santos, A. Barbosa-Póvoa, & A. Carvalho, 2018. Life cycle assessment of pulp and paper production-A Portuguese case study, *Computer Aided Chemical Engineering*, 43, 809-814.

Integrating Process Simulation and Life Cycle Assessment to Evaluate the Economic and Environmental Performance of Algae Biodiesel

S. Zapata-Boada^a, M. Gonzalez-Miquel^{a,b}, M. Jobson^a, R. Cuellar-Franca^{a,*}

^a*Department of Chemical Engineering and Analytical Science, University of Manchester, Manchester, M13 9PL, United Kingdom*

^b*Departamento de Ingeniería Química Industrial y del Medio Ambiente, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Madrid, 28006, España*

Abstract

This study aims to develop a systematic methodology through the combination of process simulation and life cycle assessment (LCA) to comprehensively evaluate biodiesel production from algae biomass in terms of the minimum biodiesel selling price (MBSP) and the environmental impacts of algae biodiesel. Downstream processing stages (lipid extraction with hexane, lipid refining through esterification with glycerol, and conversion to biodiesel with an alkali-based transesterification) were designed and simulated in Aspen Plus V8.8. Then, a discounted cash flow rate of return analysis was used to obtain the MBSP. Finally, an LCA was performed to obtain the environmental impacts of algae biodiesel and identify the hot-spots in the processing stages. The MBSP was found to be 8.95 USD per gal, around 3.5 times higher than the average price of fossil diesel in 2020 for the US. The primary energy demand (PED) and climate change (CC) impacts of algae biodiesel were estimated at 1.54 MJ/MJ biodiesel and 96 g CO₂eq/MJ biodiesel, resulting 27% and 2% higher, respectively, than the reference values for fossil diesel. Lipid extraction was identified as the hotspot of downstream processing, accounting for 47% of PED and 60% of CC of the process. Hexane production and evaporation losses were also identified as major contributors to other environmental impact categories, including human toxicity, eutrophication, and photochemical ozone formation. This methodology can be used to evaluate alternative processing options for algae biodiesel, such as using bio-based solvents (terpenes) for algae lipid extraction, to contribute to more sustainable and environmentally friendly processes.

Keywords: lipid extraction, solvent selection, techno-economic analysis.

1. Introduction

Biofuels production from algae is a significant area of interest within the field of biorefining because algae offer several advantages over conventional feedstocks for biofuel production, including high productivity, high content of lipids, cultivation in saline and wastewater sources, and production of high-value coproducts. Despite these advantages, biofuel production from algae is not cost-competitive due to several challenges in the processing steps (Quinn and Davis, 2015). Among these, algae lipid extraction is a major bottleneck in biodiesel's commercial production, as it has a significant impact on the economic viability, energy consumption, and environmental performance of the process (Harris et al., 2018). Conventional processes for lipid

extraction rely on fossil-based volatile organic solvents, such as hexane, which introduce negative environmental impacts due to their hazardous, volatile, and flammable nature. Alternatives to hexane have been proposed, such as bio-based solvents (terpenes) for lipid extraction (Dejoye Tanzi et al., 2013). However, information is lacking about the process performance and sustainability of non-conventional solvents for lipid extraction.

Techno-economic analysis combined with environmental sustainability tools, such as life cycle assessment (LCA), have proven to be useful for evaluating the performance and identifying improvement opportunities of novel processes, including biodiesel production from algae (DeRose et al., 2019). The minimum selling price of algae biofuels has been estimated and reported widely in the literature (Davis et al., 2018). Most previous LCA of algae biofuels focused only on the primary energy demand and climate change impacts without considering potential impacts related to the use of hexane in the lipid extraction stage, such as human toxicity and photochemical oxidation. Besides, only a few LCA studies (Passell et al., 2013) identified hotspots in the downstream processing of algae biofuels, with the objective of improving its economic and environmental performance. In an effort to better understand the impact of the lipid extraction stage, and particularly the benchmark solvent hexane, on the economic viability and environmental sustainability of algae biodiesel production, this study presents a systematic methodology that combines process simulation and life cycle assessment to provide a robust evaluation of the economic and environmental performance of algae biodiesel on a life cycle basis.

2. Methodology

The proposed methodology, illustrated in Figure 1, consists of four iterative steps: 1) problem definition, 2) process design and simulation, 3) economic analysis, and 4) life cycle assessment. Data required and outcomes of each step are included in Figure 1. The application of this methodology allows the identification of the hotspots in algae biodiesel production, which can be modified to improve the economic and environmental performance of the product.

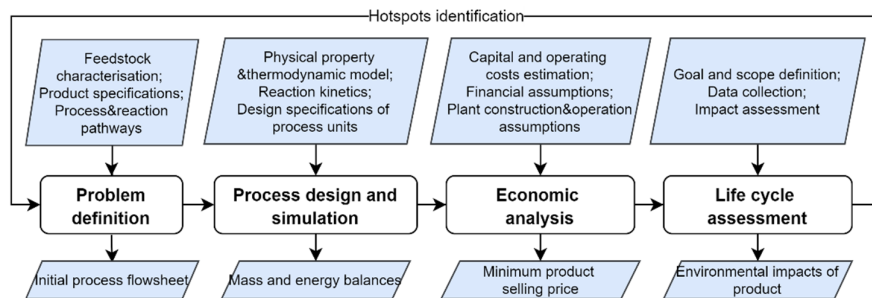


Figure 1. Proposed methodology to evaluate economics and environmental impacts of algae biodiesel

Figure 2 shows an overview of algae biodiesel's life cycle, including the processing stages and its combustion. The upstream processing stages comprise algae cultivation and harvesting/dewatering processes. *Nannochloropsis salina* was selected as the algae strain due to its demonstrated potential for biodiesel production, known composition data (Yao et al., 2015), and refining and conversion kinetics data (Silva et al., 2013). It is assumed that the algae are cultivated in an open pond system and fed with a pure stream of CO₂, which was assumed to be captured from a coal-fired power plant using monoethanolamine (MEA). Fertilisers (diammonium phosphate and ammonia) are added

to the cultivation ponds to support algae biomass growth. The biomass is then routed to the harvesting/dewatering stage, where it is concentrated from 0.5 g/L to 200 g/L with settling tanks and a centrifuge unit (Davis et al., 2016). The downstream processing stages comprise lipid extraction, refining, and conversion to biodiesel. Lipid extraction includes cell disruption (with a high-pressure homogeniser), extraction with the solvent hexane, and solvent recovery. In the lipid refining stage, the free fatty acid content of the extracted lipids decreases to < 1 wt% through an esterification reaction with glycerol (Silva et al., 2013). The refined lipids are converted to biodiesel with an alkali-based transesterification reaction with methanol in the lipid conversion stage. An anaerobic digestion/combined heat and power unit (AD/CHP) is included in the flowsheet to generate heat and power from the waste biomass (lipid-extracted algae) and waste glycerol and to recover nutrients for algae cultivation in the AD effluent stream.

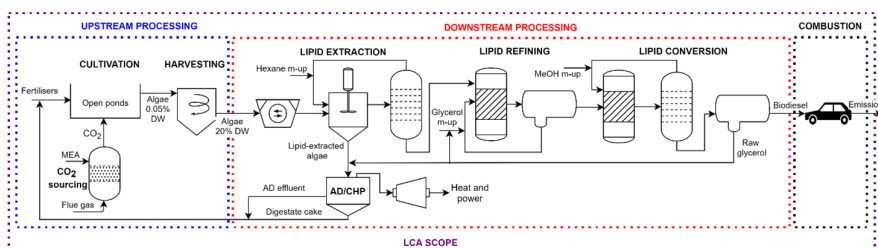


Figure 2. Overview of the life cycle of algae biodiesel

The plant's objective is the continuous production of 10,000 t y⁻¹ of biodiesel that meets the quality specifications given by the ASTM D675 standard. The water content in the algae biomass is 80 wt%. The lipid content of the dry biomass is 30%, consisting of a mixture of triglycerides (85%) and free fatty acids (15%). Due to the presence of polar components and the nonideal thermodynamics of the liquid mixtures that need to be separated, the NRTL activity coefficient model is used to model phase equilibrium, and the Dortmund UNIFAC method is used to estimate missing binary interaction parameters. The lipid extraction process is modelled assuming a counter-current extraction column with five equilibrium stages, at 25°C and 1 atm. The solvent-to-dry feed mass ratio is assumed to be 5 to 1 (Davis et al., 2014). The EXTRACT block in Aspen Plus V8.8 (an equilibrium-stage model) is used to model the extraction process. In this model, the distribution coefficients are calculated with an activity coefficient model capable of representing two liquid phases (NRTL); adiabatic operation is assumed. The process design for biodiesel production from the extracted lipids proposed by Dimian and Kiss (2019) was adapted and used to guarantee that the biodiesel product meets the quality specifications. The minimum biodiesel selling price (MBSP, in USD per gal) is calculated using a discounted cash flow rate of return analysis, assuming a 10% internal rate of return (IRR) after taxes (Davis et al., 2014). Capital investment costs are estimated using the factorial method of cost estimation and Hand installation factors (Towler and Sinnott, 2020). Operating costs are calculated based on the material and energy balance of the process obtained via the simulation. The cost of the algae biomass (*Nannochloropsis salina*) that enters the lipid extraction process is taken from Davis et al. (2016).

The LCA is conducted according to ISO 14040/44 guidelines. The goal is to estimate the life cycle environmental impacts of production and combustion of biodiesel derived from algae, considering a "cradle-to-grave" scope, as illustrated in Figure 2. The functional unit is defined as "the production and combustion of one MJ of biodiesel." The inventory data

for the upstream processing stages are obtained from Davis et al. (2018), while those for the downstream processing stages are obtained from mass and energy balances resulting from process simulation. The emissions resulting from biodiesel combustion are obtained from GREET (Argonne National Laboratory, 2019). All the background data, e.g. grid electricity, provision of heat from natural gas, production of chemicals for upstream and downstream processing (MEA, fertilisers, hexane, methanol, NaOH, glycerol), and wastewater treatment, are sourced from the Ecoinvent database V3.5 (Ecoinvent, 2019). The LCA modelling is carried out in the software GaBi v9.5. The environmental impact categories are selected based on the recommendations of the International Reference Life Cycle Data System Handbook (EC-JRC, 2010) and are estimated using the ReCiPe2016 impact assessment method (Huijbregts et al., 2017).

3. Results and Discussion

The minimum biodiesel selling price was estimated at 8.95 USD per gal. This value is around 3.5 times the average price of fossil diesel in 2020 for the US (2.55 USD per gal). The cost of algae biomass and hexane accounted for 76% and 2%, respectively, of the biodiesel price. An increase in the recovery of lipids from algae biomass, e.g. using other solvents with higher extraction capacity, could reduce algae biomass's required feed rate to achieve the desired biodiesel production, thus decreasing the biodiesel selling price. Besides, many by-product opportunities for lipid-extracted algae, such as the production of bioplastics and high value-added molecules, should be explored to improve the economics of the process. Primary energy demand (PED), climate change (CC), human toxicity (HTNC), eutrophication (FE), and photochemical ozone formation (POFE) impacts of algae biodiesel are 1.54 MJ/MJ, 96 g CO₂-eq/MJ, 66 g 1,4-DCB-eq, 37.8 mg P-eq., and 0.36 g NOx-eq., respectively, with a large contribution (86% to PED, 92% to CC, 92% to HTNC, 96% to FE, and 39% to POFE) of the upstream processing stages. These stages are highly energy-intensive and involve the use of chemicals (MEA for CO₂ capture) and fertilisers (for cultivation), which explains their high contribution to the environmental impacts of algae biodiesel. In terms of downstream processing of algae biodiesel, lipid extraction is an environmental hotspot, accounting for 49% of PED, 61% of CC, 52% of HTNC, 56% of FE, and 97% of POFE. Figure 3 shows the individual contributions to the environmental impacts of the lipid extraction process. The energy required for the process (electricity and heat) is the main driver for PED, CC, and FE. Meanwhile, hexane production and evaporation losses are the main drivers for human toxicity and photochemical oxidation.

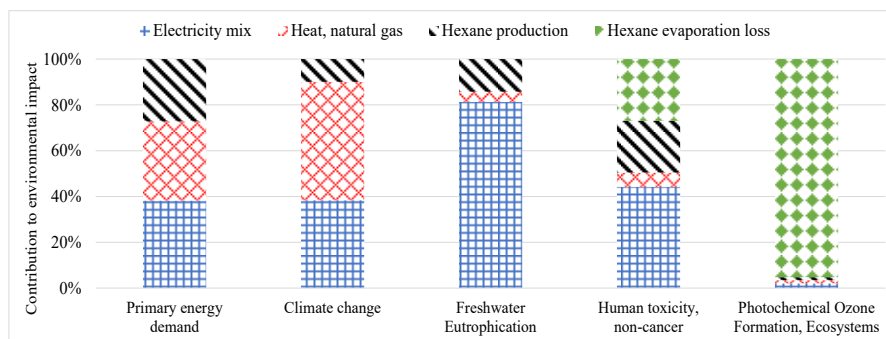


Figure 3. Individual contributions to the environmental impacts of the lipid extraction process

The replacement of hexane with more sustainable solvents could improve the environmental performance of algae biodiesel. The lipid extraction process was simulated using the same five-stage cascade referred to earlier, for four solvents: hexane and the terpenes limonene, cymene and pinene. Table 2 summarises key findings based on the simulation results.

Table 1. Performance of lipid extraction process using alternative solvents (hexane and terpenes)

Solvent	Lipid recovery (%)	Energy requirement (MJ/kg lipid)	Solvent intensity (kg solv/kg lip)	Operating costs (million USD/year)	GHG emissions (kg CO ₂ -eq)	Organic emissions (kg/h)
Hexane	99.8	7.61	0.10	22.41	635.80	19.98
Limonene	100	10.51	0.08	28.57	880.39	0.48
Cymene	99.7	10.26	0.08	30.00	854.77	0.48
Pinene	96.4	11.43	0.09	26.10	922.54	0.96

Figure 4 shows the results of the performance indicators of terpenes, normalised with respect to hexane. The health, safety and environmental (HSE) ranking of the solvents (hazardous: 1; problematic: 0.5, recommended: 0), obtained from the CHEM21 solvent selection guide (Prat et al., 2016), is also included in Figure 4. As seen, terpenes offer advantages over hexane in terms of HSE performance, organic emissions, and solvent intensity. However, terpenes have higher energy requirements, operating costs, and GHG emissions than hexane. Therefore, using terpenes may result in higher PED and CC impacts of algae biodiesel, but lower human toxicity and photochemical oxidation. In order to evaluate and understand these trade-offs, a full LCA of algae biodiesel using terpenes for lipid extraction is required.

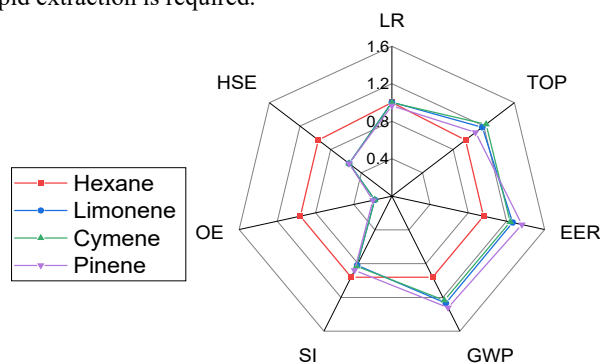


Figure 4. Comparison of performance of lipid extraction solvents [LR: Lipid recovery; TOP: Total operating costs; EER: Equivalent energy requirement; GWP: Global warming potential; SI: Solvent intensity; OE: Organic emissions; HSE: CHEM21 ranking]

4. Conclusions

This study presents a systematic methodology combining process simulation and life cycle assessment to comprehensively evaluate biodiesel production from algae biomass in terms of economic and environmental sustainability criteria. Lipid extraction is identified as the environmental hotspot of downstream processing stages due to its high energy consumption and the use of the solvent hexane, which is an important driver for eutrophication, human toxicity, and photochemical ozone formation. The evaluation of the use of terpenes in the lipid extraction process, applying several performance indicators, showed that terpenes could offer advantages in terms of HSE performance,

organic emissions, and solvent intensity, but higher energy requirements, operating costs, and GHG emissions. Future work will carry out a complete life cycle analysis of biodiesel production from algae biomass using bio-based solvents for lipid extraction. Also, a systematic methodology to screen solvents for biorefinery processes considering techno-economic and environmental sustainability criteria on a life cycle basis will be developed.

References

- Argonne National Laboratory 2019. GREET v1.3.0.13520.
- Davis, R., Kinchin, C., Markham, J., Tan, E., Laurens, L., Sexton, D., Knorr, D., Schoen, P. & Lukas, J. 2014. Process Design and Economics for the Conversion of Algal Biomass to Biofuels: Algal Biomass Fractionation to Lipid- and Carbohydrate-Derived Fuel Products. ; National Renewable Energy Lab. (NREL), Golden, CO (United States).
- Davis, R., Markham, J., Kinchin, C., Grundl, N., Tan, E. C. D. & Humbird, D. 2016. Process Design and Economics for the Production of Algal Biomass: Algal Biomass Production in Open Pond Systems and Processing Through Dewatering for Downstream Conversion. ; National Renewable Energy Lab. (NREL), Golden, CO (United States).
- Davis, R. E., Markham, J. N., Kinchin, C. M., Canter, C., Han, J., Li, Q., Coleman, A., Jones, S., Wigmosta, M. & Zhu, Y. 2018. 2017 Algae Harmonization Study: Evaluating the Potential for Future Algal Biofuel Costs, Sustainability, and Resource Assessment from Harmonized Modeling. United States.
- Dejoye Tanzi, C., Vian, M. & Chemat, F. 2013. New procedure for extraction of algal lipids from wet biomass: a green clean and scalable process. *Bioresour Technol*, 134, 271-5.
- DeRose, K., DeMill, C., Davis, R. W. & Quinn, J. C. 2019. Integrated techno economic and life cycle assessment of the conversion of high productivity, low lipid algae to renewable fuels. *Algal Research*, 38.
- Dimian, A. C. & Kiss, A. A. 2019. Eco-efficient processes for biodiesel production from waste lipids. *Journal of Cleaner Production*, 239, 118073.
- Ecoinvent. 2019. *Database 3.5* [Online]. Available: <https://www.ecoinvent.org/database/database.html> [Accessed 2020].
- EC-JRC. 2010. International Reference Life Cycle Data System (ILCD) Handbook - General guide for Life Cycle Assessment - Detailed guidance.
- Harris, J., Viner, K., Champagne, P. & Jessop, P. G. 2018. Advances in microalgal lipid extraction for biofuel production: a review. *Biofuels, Bioproducts and Biorefining*, 12, 1118-1135.
- Huijbregts, M. A. J., Steinmann, Z. J. N., Elshout, P. M. F., Stam, G., Verones, F., Vieira, M., Zijp, M., Hollander, A. & van Zelm, R. 2017. ReCiPe2016: a harmonised life cycle impact assessment method at midpoint and endpoint level. *The International Journal of Life Cycle Assessment*, 22, 138-147.
- Passell, H., Dhaliwal, H., Reno, M., Wu, B., Ben Amotz, A., Ivry, E., Gay, M., Czartoski, T., Laurin, L. & Ayer, N. 2013. Algae biodiesel life cycle assessment using current commercial data. *Journal of Environmental Management*, 129, 103-111.
- Prat, D., Wells, A., Hayler, J., Sneddon, H., McElroy, C. R., Abou-Shehada, S. & Dunn, P. J. 2016. CHEM21 selection guide of classical- and less classical-solvents. *Green Chemistry*, 18, 288-296.
- Quinn, J. C. & Davis, R. 2015. The potentials and challenges of algae based biofuels: a review of the techno-economic, life cycle, and resource assessment modeling. *Bioresour Technol*, 184, 444-452.
- Silva, C., Soliman, E., Cameron, G., Fabiano, L. A., Seider, W. D., Dunlop, E. H. & Coaldrake, A. K. 2013. Commercial-Scale Biodiesel Production from Algae. *Industrial & Engineering Chemistry Research*, 53, 5311-5324.
- Towler, G. P. & Sinnott, R. K. 2020. *Chemical engineering design*, Kidlington, Oxford, United Kingdom, Butterworth-Heinemann, an imprint of Elsevier.
- Yao, L., Gerde, J. A., Lee, S.-L., Wang, T. & Harrata, K. A. 2015. Microalgae Lipid Characterization. *Journal of Agricultural and Food Chemistry*, 63, 1773-1787.

Optimization of biogas production from cattle manure by anaerobic digestion using a gradient-based algorithm

Tina Kegl*, Anita Kovač Kralj

*University of Maribor; Faculty of Chemistry and Chemical Engineering, Smetanova 17,
2000 Maribor, Slovenia*

tina.kegl@um.si

Abstract

One of the most promising technologies for production of biogas from organic waste is anaerobic digestion (AD). As the experiments related to AD are time-consuming, labour intensive, and expensive, numerical simulations are important, because they provide a useful tool towards developing optimized and stable AD processes. This paper focuses on the optimization of the AD process in order to maximize biogas production from cattle manure by determining optimal values of process parameters. For this purpose, a modified mathematical model has been developed to simulate AD at various temperatures and pH values by considering appropriate inhibitions, detailed pH and temperature effects on microbial growth, and temperature dependencies of liquid-gas mass transfer coefficients, dissociation constants, and Henry's law coefficients. This model was applied for the simulation of AD in a batch reactor and verified with experimental data. By using this model in an optimization procedure, the optimal values of pH and temperature, both acting as design variables, were determined. By engaging a gradient-based optimizer, based on an adaptive approximation method, it was shown that a reliable AD model can be used very efficiently in gradient-based optimization. The optimization process proved to be efficient, stable, and starting-point-independent; the engagement of numerical differentiation did not cause any problems. The results show that at optimal process conditions the production of CH₄ is significantly improved in comparison to conditions often given in the literature. The results obtained confirm the usefulness of the proposed approach, which can easily be adapted or upgraded for complex substrates and other reactor types.

Keywords: anaerobic digestion, cattle manure, biogas, numerical simulation, optimization.

1. Introduction

Ever increasing energy demands, depletion of fossil fuels, and waste accumulation are the main forces for development of environmentally friendly waste-to-energy technologies. One of the most promising technologies is the production of biogas from animal manure (organic waste) by the anaerobic digestion (AD) process (Khalil et al. 2019; Rasapoor et al. 2020). As experiments related to AD are time-consuming, labor intensive, and expensive, numerical simulations are important, because they provide a

useful tool for AD process understanding and optimization (Ma et al. 2016). Till today various more or less sophisticated mathematical AD models have been developed in order to simulate biochemical, chemical, and physical processes by considering the near-actual composition of both, biogas and complex substrate (Maharaj et al. 2018; Tsapekos et al. 2018; Sun et al. 2019; Kovalovszki et al. 2020; Kegl and Kovač Kralj, 2020). These AD models are based either on ADM1, which uses the indirect Chemical Oxygen Demand (COD) for component concentrations computation, developed by the International Water Association (IWA) Task Group for Mathematical Modelling of Anaerobic Digestion Processes (Batstone et al. 2002) or on the BioModel with mass-based unit system, developed by Angelidaki et al. (1993). The objective in this study is to: (i) develop and test a modified mathematical model, based on the BioModel, to simulate AD at various temperature and pH values by considering appropriate inhibitions, detailed pH and temperature effects on microbial growth, and temperature dependencies of liquid gas mass transfer coefficients, dissociation constants, and Henry's law coefficients; (ii) attach a highly-efficient gradient-based optimizer and implement a procedure to optimize the AD process conditions by utilizing numerical differentiation; (iii) perform AD process conditions optimization from various starting points to estimate optimization process efficiency, stability and sensitivity to possible problems resulting from using derivatives obtained by numerical differentiation.

2. Mathematical model of anaerobic digestion

The proposed AD batch reactor model, which is presented in detail by Kegl and Kovač Kralj (2020), is based on biochemical reactions. It consists of a system of 21 differential equations of a microbial kinetic model and mass balance equations in the liquid, liquid-gas and gas phases. The most important equations are presented in the following, Eqs. (1)-(9). It must be pointed out that in the mass balance equations for insoluble c_{is} (gL^{-1}) and soluble c_s (gL^{-1}) organic compounds of cattle manure, volatile fatty acids (VFA) inhibition of the hydrolysis process is also considered, Eqs. (1)-(2). Furthermore, for the determination of CO_2 and CH_4 concentration in the liquid phase, $c_{\text{CO}_2,1}$ (gL^{-1}) and $c_{\text{CH}_4,1}$ (gL^{-1}), the liquid-gas transfer rate must be taken into consideration, Eqs. (7)-(8).

$$\frac{dc_{is}}{dt} = -K_0 \frac{K_{I,VFA}}{K_{I,VFA} + c_{VFA}} c_{is} \quad (1)$$

$$\frac{dc_s}{dt} = 0.55K_0 \frac{K_{I,VFA}}{K_{I,VFA} + c_{VFA}} c_{is} - Y_{s/A} \mu_A c_A \quad (2)$$

$$\frac{dc_{ac}}{dt} = Y_{ac/A} \mu_A c_A + Y_{ac/AP} \mu_{AP} c_{AP} + Y_{ac/AB} \mu_{AB} c_{AB} - Y_{ac/M} \mu_M c_M \quad (3)$$

$$\frac{dc_{pr}}{dt} = Y_{pr/A} \mu_A c_A - Y_{pr/AP} \mu_{AP} c_{AP} \quad (4)$$

$$\frac{dc_{but}}{dt} = Y_{but/A} \mu_A c_A + Y_{but/AB} \mu_{AB} c_{AB} \quad (5)$$

$$\begin{aligned} \frac{dc_{am}}{dt} = & 0.301 K_0 \frac{K_{I,VFA}}{K_{I,VFA} + c_{VFA}} c_{is} - Y_{am/A} \mu_A c_A - Y_{am/AP} \mu_{AP} c_{AP} - \\ & Y_{am/AB} \mu_{AB} c_{AB} - Y_{am/M} \mu_M c_M \end{aligned} \quad (6)$$

$$\begin{aligned} \frac{dc_{CO_2,l}}{dt} = & Y_{CO_2/A} \mu_A c_A + Y_{CO_2/AP} \mu_{AP} c_{AP} - Y_{CO_2/AB} \mu_{AB} c_{AB} + \\ & Y_{CO_2/M} \mu_M c_M - r_{l-g,CO_2} \end{aligned} \quad (7)$$

$$\frac{dc_{CH_4,l}}{dt} = Y_{CH_4/AP} \mu_{AP} c_{AP} + Y_{CH_4/AB} \mu_{AB} c_{AB} + Y_{CH_4/M} \mu_M c_M - r_{l-g,CH_4} \quad (8)$$

$$\frac{dV_{biogas}}{dt} = \frac{RTV_{liq}}{(p_{total} - p_w)} \left(\frac{r_{l-g,CO_2}}{M_{CO_2}} + \frac{r_{l-g,CH_4}}{M_{CH_4}} \right) \text{ and } \frac{dV_{CH_4}}{dt} = \frac{RTV_{liq} r_{l-g,CH_4}}{p_{total} M_{CH_4}}$$

(9)

where μ_{bac} and c_{bac} , $bac \in \{A, AP, AB, M\}$ are specific growth rates (day^{-1}) and mass concentrations (gL^{-1}) of acidogens, propionate degrading acidogenic bacteria, butyrate degrading acidogenic bacteria, and methanogenic bacteria; $Y_{i/bac}$ are yields coefficients (g of i^{th} component per g of bacteria cell); K_0 is the non-inhibited hydrolyzed rate constant (day^{-1}); $K_{I,VFA}$ is the inhibition constant of volatile fatty acids (VFA) (gL^{-1}); p_{total} and p_w are the total and saturated vapour pressures (atm); c_{ac} , c_{pr} , c_{but} , and c_{am} are concentrations of acetate, propionate, butyrate, and ammonia, respectively (gL^{-1}); V_{biogas} is the volume of produced biogas (L); V_{liq} is the liquid volume of the reactor (L); T is the temperature (K); R is the gas constant ($\text{atmLmol}^{-1}\text{K}^{-1}$); M_{CO_2} and M_{CH_4} are the molecular masses of CO_2 and CH_4 (gmol^{-1}), and r_{l-g,CO_2} and r_{l-g,CH_4} are dynamic liquid-gas transfer rates of CO_2 and CH_4 , where the liquid-gas equilibrium is reached for both gaseous components in accordance with Henry's law.

3. Optimization

In this optimal design problem of the AD system, the objective function g_0 is defined in such a way to promote maximal production of biogas, while the constraint functions g_i

and g_2 are related to the CH_4 to biogas volume ratio, Eq. (10). The defined objective function maximizes the biogas volume, V_{biogas} , without taking any consideration of the time needed to obtain it.

$$g_0 = -V_{\text{biogas}}, \quad g_1 = \frac{V_{\text{CH}_4}}{V_{\text{biogas}}} - 0.8, \quad g_2 = 0.65 - \frac{V_{\text{CH}_4}}{V_{\text{biogas}}} \quad (10)$$

The design variables are only pH value and temperature, meanwhile there are 21 response variables. The values of pH and temperature design variables were normalized by Eq. (11).

$$x_1 = \frac{(\text{pH} - \text{pH}_{\min})}{(\text{pH}_{\max} - \text{pH}_{\min})} \quad \text{and} \quad x_2 = \frac{(T - T_{\min})}{(T_{\max} - T_{\min})} \quad (11)$$

where $\text{pH}_{\min} = 5.5$ and $\text{pH}_{\max} = 8.5$ are prescribed minimal and maximal pH values of the AD process, while $T_{\min} = 25^\circ\text{C}$ and $T_{\max} = 65^\circ\text{C}$ are prescribed minimal and maximal temperatures of the AD process.

Both the described mathematical model of the AD process and the optimization algorithm were coded in the C# language. The system of differential equations was solved by the 3th order Runge Kutta method. A gradient based approximation method optimizer, which is well described in Kegl & Oblak, (1997), was engaged to solve the NLP optimization problem. This method essentially generates an approximate NLP problem in each optimization cycle. The algorithm uses the history of provided design derivatives of the objective and constraint functions in order to enhance the quality of the approximation. To compute the design derivatives involved in the NLP functions, numerical derivation was used. The derivatives computation was parallelized in order to boost computational efficiency.

4. Numerical results and discussion

First of all, the described mathematical model was already verified (Kegl & Kovač Kralj, 2020). The needed input data for the optimization are cattle manure composition, given in Table 1, for a bioreactor with a liquid volume of 1 L.

Table 1. Feedstock characteristics

Variable	Value	Variable	Value	Variable	Value	Variable	Value
$c_{\text{is}} (\text{gL}^{-1})$	30.4	$c_{\text{pr}} (\text{gL}^{-1})$	2.3	$c_{\text{CO}_2} (\text{gL}^{-1})$	0.0	$c_{\text{AP}} (\text{gL}^{-1})$	0.01
$c_{\text{s}} (\text{gL}^{-1})$	5.4	$c_{\text{but}} (\text{gL}^{-1})$	0.2	$c_{\text{CH}_4} (\text{gL}^{-1})$	0.0	$c_{\text{AB}} (\text{gL}^{-1})$	0.01
$c_{\text{ac}} (\text{gL}^{-1})$	4.5	$c_{\text{am}} (\text{gL}^{-1})$	2.0	$c_{\text{A}} (\text{gL}^{-1})$	0.13	$c_{\text{M}} (\text{gL}^{-1})$	0.8

The design variable and objective function variation histories obtained during the optimization process are presented in Figure 1 for various initial values.

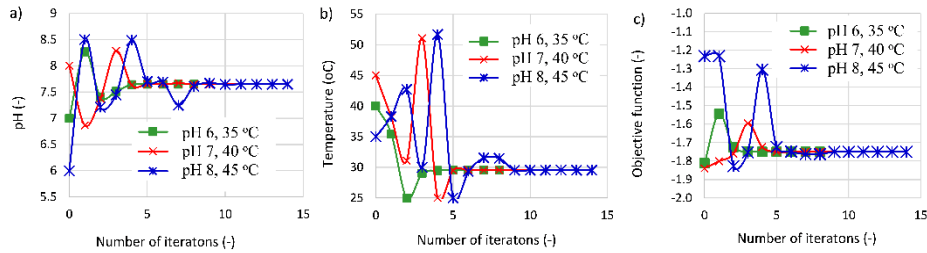


Figure 1. Optimization histories with various initial values of design variables: (a) pH, (b) temperature, and (c) objective function

Within a maximum of 14 iterations, the minimal value of the objective function was reached at the optimal values of pH 7.7 and temperature of 30 °C; all constraints were fulfilled. As can be seen from Figure 1, the optimal values of these design variables are practically the same for any initial values of pH and temperature. Moreover, the value of the objective function after optimization is the same regardless of the chosen initial values of design variables, Figure 1.

Figure 2 shows the: (a) formation and consumption of acetate, propionate, and butyrate, and (b) volumes of produced gases before and after optimization. Regarding the VFA concentrations, propionate and butyrate convert to acetate; therefore, acetate concentration is higher than the concentration of other VFA. As we can see in Figure 2, the concentration of acetate does not approach zero at initial state, while at the optimal state the whole acetate is converted to CH₄ and other components. With respect to the obtained production rate of CH₄, CO₂, and total biogas, the total volume of biogas produced at optimal conditions is higher by about 10% compared to initial conditions, Figure 2. At optimal conditions, the volume of CO₂ produced is decreased by about 23% with respect to initial conditions; meanwhile, the volume of the obtained CH₄ is more than 1.4 times higher than at initial conditions.

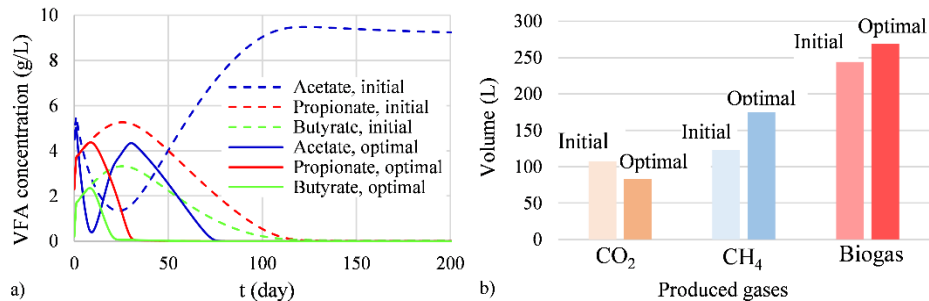


Figure 2. Acetate, propionate, and butyrate concentration (a) and gas volumes (b) before and after optimization.

5. Conclusions

Optimization of the anaerobic digestion process of cattle manure within a batch reactor was investigated. For this purpose, a modified mathematical model of the AD process was developed and the obtained results were compared with the results delivered by

existing mathematical models and with the experimental data. Both the described mathematical model of the AD process and the optimization algorithm were coded in the C# language to enable efficient numerical simulation. According to the numerical results, the following conclusions can be made:

- The model can be successfully engaged as a state equation in the optimization problem formulation and solution, where the AD conditions such as pH and temperature can be optimized in order to maximize the biogas production; the model can also be easily extended to describe certain processes more accurately, without affecting the optimization procedure layout.
- The used gradient-based approximation method optimizer is an attractive tool to solve AD process optimization problems; under the considered circumstances, it proved to deliver a stable and starting-point-independent optimization.
- The obtained optimization results show that the optimal values of design variables, 7.7 pH and 30 °C, were reached within 14 iterations for various initial values of design variables; this confirms that this optimization procedure is very efficient; since the numerical derivatives computation can be easily parallelized, the CPU times are also relatively low.

Acknowledgements

The authors are grateful for the financial support of the Slovenian Research Agency (PhD research fellowship contract No. 1000-18-0552 and core research funding No. P2-0032).

References

- I. Angelidaki, L. Ellegaard, BK. Ahring, 1993, A mathematical model for dynamic simulation of anaerobic digestion of complex substrates: Focusing on ammonia inhibition, *Biotechnology and Bioengineering*, 42, 159-166. <https://doi.org/10.1002/bit.260420203>.
- DJ. Batstone, J. Keller, I. Angelidaki, SV. Kalyuzhnyi, SG. Pavlostathis, A. Rozzi, WTM. Sanders, H. Siegrist, VA. Vavilin, 2002, *Anaerobic Digestion Model No. 1*, London: IWA Task Group.
- M. Kegl, MM. Oblak, 1997, Optimization of mechanical systems: on non-linear first-order approximation with an additive convex term, *Communications in Numerical Methods in Engineering*, 13, 13-20. [https://doi.org/10.1002/\(SICI\)1099-0887\(199701\)13:1<13::AID-CNM33>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1099-0887(199701)13:1<13::AID-CNM33>3.0.CO;2-E).
- T. Kegl, A. Kovač Kralj, 2020, Multi-objective optimization of anaerobic digestion process using a gradient-based algorithm, *Energy Conversion and Management*, 226, 113560. <https://doi.org/10.1016/j.enconman.2020.113560>.
- M. Khalil, MA. Berawl, R. Heryanto, A. Rizalie, 2019, Waste to energy technology: The potential of sustainable biogas production from animal waste in Indonesia, *Renewable and Sustainable Energy Reviews*, 105, 323-331. <https://doi.org/10.1016/j.rser.2019.02.011>.
- A. Kovalovszki, L. Treu, L. Ellegaard, G. Luo, I. Angelidaki, 2020, Modeling temperature response in bioenergy production: Novel solution to a common challenge of anaerobic digestion, *Applied Energy*, 263, 114646. <https://doi.org/10.1016/j.apenergy.2020.114646>.

- AYW. Ma, BKW. Cheung, KCM. Kwok, M. Cai, PKH. Lee, 2016, Recent advances of anaerobic digestion for energy recovery. In: OP. Karthikeyan, K. Heimann, SS. Muthu, editors. Recycling of solid waste for biofuels and bio-chemicals, Singapore: Springer Nature, 87-126.
- BC. Maharaj, MR. Mattei, L. Frunzo L, MR. van Hullebusch, G. Esposito, 2018, ADM1 based mathematical model of trace element precipitation/dissolution in anaerobic digestion process, Bioresource Technology, 267, 666-676. <https://doi.org/10.1016/j.biortech.2018.06.099>.
- M. Rasapoor , B. Young , R. Brar , A. Sarmah , WQ. Zhuang , S. Baroutian, 2020, Recognizing the challenges of anaerobic digestion: Critical steps toward improving biogas generation, Fuel, 261, 116497. <https://doi.org/10.1016/j.fuel.2019.116497>.
- H. Sun, A. Kovalovszki, P. Tsapekos, M. Alvarado-Morales, A. Rudatis, S. Wu, R. Dong, PG. Kougias, I, Angelidaki, 2019, Co-digestion of Laminaria digitata with cattle manure: A unimodal simulation study of both batch and continuous experiments, Bioresource Technology, 276, 361-368. <https://doi.org/10.1016/j.biortech.2018.12.110>.
- P. Tsapekos, PG. Kougias , S. Kuthiala , I. Angelidaki, 2018, Co-digestion and model simulations of source municipal organic waste with cattle manure under batch and continuously stirred tank reactor, Energy Conversion and Management, 159, 1-6. <https://doi.org/10.1016/j.enconman.2018.01.002>.

Photosynthetic Bioplastics Production with Cyanobacteria by Coupled Growth-Production Mutants

Romina Lasry Testa, Vanina Estrada, Claudio Delpino and M. Soledad Diaz*

Planta Piloto de Ingeniería Química (PLAPIQUI), Universidad Nacional del Sur (UNS)-CONICET, Camino La Carrindanga Km 7, Bahía Blanca 8000, Argentina

*Corresponding Author's e-mail: sdiaz@plapiqui.edu.ar

Abstract

In this work, we study PHB direct production (without a heterotrophic step) using *Synechocystis* sp PCC 6803 within a Systems Metabolic Engineering approach. We consider a GENome-scale Model (GEM) of the cyanobacteria, developed in previous work, and use it to couple growth with PHB production by finding genetic interventions (*knock-outs*) that make PHB production essential for growth. This is achieved by formulating a bilevel-programming problem, in which the *knock-outs* are represented by binary variables related to the genes of the GEM. After reformulation based on duality theory, a Mixed Integer Linear Programming (MILP) problem is obtained and solved in GAMS. A mutant that completely couples biopolymer production with growth in *Synechocystis* is obtained, and its flux distributions under exponential growth conditions are analysed by Flux Balance Analysis (FBA). Numerical results provide useful insights on sustainable photosynthetic bioplastics production with cyanobacteria.

Keywords: MILP, cyanobacteria, PHB, metabolic modelling

1. Introduction

The growing concern on climate change and environmental plastic pollution in recent years has led to the study of biodegradable and sustainable plastics based on biomass to replace conventional fossil-based plastics (García Prieto et al., 2017). Besides being produced from fossil fuels, when these polymers are incinerated, they generate carbon dioxide (CO₂) emissions that contribute to greenhouse effect, and also produce accumulation of non-biodegradable microplastics when they are built up in continental and marine environments and landfills (Kamravamanesh et al., 2018).

Polyhydroxyalcanoates (PHAs) are considered as an alternative to fossil-based plastics and are accumulated by a variety of microorganisms as a carbon and energy source (Ramos et al., 2019). Within this family, Poly(3-hydroxybutyrate) (PHB) is the most investigated bioplastic because it presents similar characteristics to polypropylene, and it is actually being commercially produced by heterotrophic bacteria, like *Cupriavidus*

necator y *Escherichia coli*. However, its production is limited due to the high production and carbon sources costs.

In this context, cyanobacteria become potential candidates for PHB production as they can grow on atmospheric CO₂, light and minimal nutrients, reducing substrate costs and contributing to lower atmospheric CO₂ concentrations. *Synechocystis* sp. PCC68003 (*Synechocystis*) is a model organism among the wide group of cyanobacteria and it has the capacity to accumulate PHB naturally under nutrient limitation conditions (Wu et al., 2001). One Systems Biology approach to study PHB production with *Synechocystis* is coupling biopolymer production with biomass production, so that it becomes essential for growth. This is achieved by identifying genetic interventions (*knock-outs*) *in silico*. We consider a bilevel programming formulation (Lasry Testa et al., 2019a), that represents these gene *knock-outs* through binary variables. In this work, we use this formulation to study the possibility of coupling PHB production with growth.

2. Methods

2.1. Genome Scale Metabolic Network

We consider a GENome-scale Model (GEM) which contains 784 reactions, divided into reversible, irreversible and exchange reactions, and 535 metabolites (Lasry Testa et al., 2019a). The model contains a detailed description of the photosynthesis, and it includes both of the reactions catalyzed by the RuBisCO enzyme: Carboxylation and photorespiration. It also contemplates the possibility of taking up carbon as CO₂ and bicarbonate (HCO₃⁻), and it allows for the interconversion of CO₂ from intracellular decarboxylation into HCO₃⁻ (Nogales et al., 2012). PHB is naturally produced by *Synechocystis* as a carbon and energy storage compound, along with glycogen, and it is accumulated as an intracellular metabolite, but to analyze its production we included a fictitious PHB excretion reaction for modelling purposes.

We consider photoautotrophic carbon limited growth conditions, to which end all organic carbon sources, like glucose, are fixed to zero. The uptake of the other inorganic nutrients (Nitrogen, Phosphorus) required is unrestricted. Inorganic carbon is limited with a maximum of 3.7 mmol gDW⁻¹ h⁻¹, which is the maximum experimental uptake identified for *Synechocystis*. The final model was turned into the GAMS Data eXchange (GDX) format to allow the use in the GAMS (GAMS Development Corporation, Washington, DC) environment.

2.2. Coupled growth-PHB production

To search for a mutant that couples PHB production to growth we formulated a bilevel programming problem which searches for the minimum set of genetic interventions needed to achieve the objective. The mathematical structure of this optimization problem is that of a MILP with the requirement that a subset of its variables is the solution to an "inner" LP. This inner LP, represented by constraints of Eq. (2) to Eq. (7), can be solved by a flux state (v) which is mass balanced (constraint XX) and fulfills a minimum growth requirement (Eq. (3)). Some reactions rates are required to be fixed at a certain value (Eq. (4)), and constant upper and lower bounds are established for each reaction (Eq. (6)). For the set K of candidate reactions, which are the reactions not

essential for growth under the photoautotrophic conditions considered, these upper and lower bounds can be set both to a value of 0, if the corresponding element in vector y is set to 1 (Eq. (4) and Eq. (5)). This vector is a variable of the outer problem, which requires these variables to be binary, therefore transforming the outer optimization problem in a MILP. The inner LP objective function minimizes the flux through the target reaction to be coupled, and a constraint on the outer problem on this value (Eq. (1)) requires it to be greater than a minimum. The amount of interventions, represented by the sum over this binary vector is minimized, as the outer MILP objective function. Therefore, the inner optimization problem guarantees that a series of interventions, represented by blocking flux through certain reactions by the outer problem, raise the minimum value of the flux through the target reaction from 0, pointing out that in the intervened flux space, the target reaction is essential to growth. For more detail in this formulation refer to (Lasry Testa et al., 2019a).

$$\min_{v,y} \sum_k y_k \quad (1)$$

$$v_{PHB} \geq v_{PHB}^{\min}$$

s.t.

$$\left. \begin{array}{l} \min v_{PHB} \\ v \\ \text{s.t.} \\ v_{biomass} \geq v_{biomass}^{\min} \end{array} \right\} \quad (2)$$

$$\left. \sum_j S_{i,j} v_j = 0 \quad \forall i \in M \right\} \quad (3)$$

$$\left. v_j = v_j^{fixed} \quad \forall j \in R^{fixed} \right\} \quad (4)$$

$$v_j \geq LB_j (1 - y_k) \quad \forall (j,k) \in K \quad (5)$$

$$v_j \leq UB_j (1 - y_k) \quad \forall (j,k) \in K \quad (6)$$

$$LB_j \leq v_j \leq UB_j \quad \forall j \in R \quad (7)$$

$$\sum_k y_k \geq 1 \quad (8)$$

$$v_j \in \mathbf{R}, y_k \in \{0,1\}$$

This bilevel problem, reformulated into a single level MILP, was solved in GAMS (GAMS Development Corporation, Washington, DC) in a PC with an Intel Core i7-4790 de 3,60 GHz processor and 8 GB RAM. To evaluate the production rates of the obtained mutants we used FBA (Savinell and Palsson, 1992).

3. Results and Discussion

We used the bilevel programming approach described to search for mutants that couple PHB production with growth in *Synechocystis*. We fixed minimal PHB production (Eq. (1)) to a value of 0.01 mmol PHB gDW⁻¹ h⁻¹ and minimal biomass production required (Eq. (2)) to 0.01 h⁻¹ to avoid a solution with zero growth rate.

With these conditions we obtained a mutant that effectively couples PHB production with growth, for which 14 genetic interventions are required. Due to the Gen-Protein-Reaction associations this *knock-outs* correspond to the deletion of 18

reactions from the GEM. The genetic interventions to obtain the mutant are presented in Table 1.

The mutant presents a growth rate of 0.0625 h^{-1} and a PHB production rate of $0.1059 \text{ mmol PHB gDW}^{-1} \text{ h}^{-1}$. The maximum growth rate of wild-type *Synechocystis* can be obtained by FBA of the GEM with biomass objective function, this value is of 0.089 h^{-1} for the photoautotrophic conditions considered (Lasry Testa et al., 2019b), and comes with no PHB production associated. We also obtained the maximum PHB production possible by running a FBA with PHB production as the objective function, resulting in a flux of $0.925 \text{ mmol PHB gDW}^{-1} \text{ h}^{-1}$ for no biomass production. Therefore, the maximum growth rate of the mutant is a 29.7% lower than the maximum growth rate of the wild type, and the PHB production rate represents and 11.5% of the maximum for no biomass production.

The PHB production pathway includes a NADPH consuming reaction, so the intervention strategy obtained is mostly related with the NADP/NADPH balance of the cell. We see in Table 1 that the *knock-outs* of genes *slr1755* and *sll1561* avoid the production of NADH to favour the use of other NADPH producing pathways. The gene *sll1342* codifies for the NADPH dependent glyceraldehyde-3-phosphate dehydrogenase, that competes with PHB production. Finally, the knock-out of gene *slr1239* avoids the possibility of the interconversion of NADPH to NADH. The knock-out of *slr0301* and *slr0783* favour flux to be directed down to the lower glycolysis, leading to an increase in Acetyl-CoA with is the precursor of PHB production. Moreover, we find the *knock-out* of the acetate membrane transporter (*gs20*) that also contributes to a rise in Acetyl-CoA production, avoiding the loss of carbon through the excretion of Acetate, whose production has to be directed completely to Acetyl-CoA production. We also find that in the mutant the TCA cycle is interrupted by the *knock-out* of gene *slr0018* which is related to the interconversion of (S)-Malate into Fumarate. As this may lead to a Fumarate overflow, the strategy also selects the Fumarate excretion reaction as another *knock-out*. The gene related to this excretion has not been identified yet, but there is experimental evidence on the capacity of *Synechocystis* to excrete this metabolite to the culture media (Du et al., 2019).

Table 1. *Knock-outs* required to obtain coupled growth-PHB mutant.

Genes	E.C. Number	Reaction
<i>slr1840</i>	2.7.1.165	$\text{ATP}(\text{cyt}) + \text{D-Glycerate}(\text{cyt}) \rightarrow \text{ADP}(\text{cyt}) + 2\text{-Phospho-D-glycerate}(\text{cyt})$
<i>slr1755</i>	1.1.1.94	$\text{sn-Glycerol 3-phosphate}(\text{cyt}) + \text{NAD}^+(\text{cyt}) \leftrightarrow \text{Glycerone phosphate}(\text{cyt}) + \text{NADH}(\text{cyt}) + \text{H}^+(\text{cyt})$
<i>slr1239</i>	1.6.1.2	$\text{NADPH}(\text{cyt}) + \text{H}^+(\text{cyt}) + \text{NAD}^+(\text{cyt}) \leftrightarrow \text{NADP}^+(\text{cyt}) + \text{H}^+(\text{cyt}) + \text{NADH}(\text{cyt})$
<i>slr0783</i>	5.3.1.1	$\text{D-Glyceraldehyde 3-phosphate}(\text{cyt}) \leftrightarrow \text{Glycerone phosphate}(\text{cyt})$
<i>slr0301</i>	2.7.9.2	$\text{ATP}(\text{cyt}) + \text{Pyruvate}(\text{cyt}) + \text{H}_2\text{O}(\text{cyt}) \rightarrow \text{AMP}(\text{cyt}) + \text{Phosphoenolpyruvate}(\text{cyt}) + \text{P}_i(\text{cyt})$
<i>slr0214</i>	2.1.1.37	$\text{S-Adenosyl-L-methionine}(\text{cyt}) + \text{DNA cytosine}(\text{cyt}) \rightarrow \text{S-Adenosyl-L-homocysteine}(\text{cyt}) + \text{DNA 5-methylcytosine}(\text{cyt})$
<i>slr0018</i>	4.2.1.2	$(\text{S})\text{-Malate}(\text{cyt}) \leftrightarrow \text{Fumarate}(\text{cyt}) + \text{H}_2\text{O}(\text{cyt})$

		$L\text{-Glutamate 5-semialdehyde(cyt)} + NAD^+(cyt) + H_2O(cyt) \rightarrow$ $L\text{-Glutamate(cyt)} + NADH(cyt) + H^+(cyt)$
<i>sll1561</i>	1.2.1.88	$(S)\text{-1-Pyrroline-5-carboxylate(cyt)} + NAD^+(cyt) + 2 H_2O(cyt) \rightarrow$ $L\text{-Glutamate(cyt)} + NADH(cyt) + H^+(cyt)$
		$L\text{-1-Pyrroline-3-hydroxy-5-carboxylate(cyt)} + NAD^+(cyt) + 2 H_2O(cyt) \rightarrow$ $L\text{-erythro-4-Hydroxyglutamate(cyt)} + NADH(cyt) + H^+(cyt)$
	1.5.5.2	$L\text{-Proline(cyt)} + PQ(tlm) \rightarrow (S)\text{-1-Pyrroline-5-carboxylate(cyt)} +$ $PQH_2(tlm)$
		$L\text{-Proline(cyt)} + PQ(cym) \rightarrow (S)\text{-1-Pyrroline-5-carboxylate(cyt)} +$ $PQH_2(cym)$
<i>sll1359</i>	1.2.2.1	$Formate(cyt) + 2 Ferricytochrome\ b1(cyt) \rightarrow CO_2(cyt) + 2$ $Ferrocycytochrome\ b1(cyt) + 2 H^+(cyt)$
<i>sll1342</i>	1.2.1.59	$D\text{-Glyceraldehyde 3-phosphate(cyt)} + Orthophosphate(cyt) +$ $NADP^+(cyt) \leftarrow 3\text{-Phospho-D-glyceroyl phosphate(cyt)} + NADPH(cyt) +$ $H^+(cyt)$
<i>sll0006</i>	2.6.1.4	$Glycine(cyt) + 2\text{-Oxoglutarate(cyt)} \leftrightarrow Glyoxylate(cyt) +$ $L\text{-Glutamate(cyt)}$
<i>slr0381</i>	4.4.1.5	$(R)\text{-S-Lactoylgutathione(cyt)} \leftrightarrow Glutathione(cyt) + Methylglyoxal(cyt)$
<i>g823</i>	-	$Fumarate(cyt) \rightarrow Fumarate(ext)$
<i>g820</i>	-	$Acetate(cyt) \rightarrow Acetate(ext)$

Only for comparison purposes, we simulated the described mutant in a batch bioreactor model that was implemented in gPROMS (PSEnterprise Ltd., 2020). The model included equations for biomass (Eq. (9)) and PHB production (Eq. (10)) and also an equation to account for light limitation (Eqs. (11) and (12)), a typical effect produced in photobioreactors due to the biomass shade effect (Delpino et al., 2014).

$$\frac{dX}{dt} = \mu_{max} * fI * X \quad (9)$$

$$\frac{dPHB}{dt} = v_{PHB} X \quad (10)$$

$$f(I) = \frac{I_0}{I_{opt}} \exp\left(1 - \frac{I_0}{I_{opt}}\right) \quad (11)$$

$$I_0 = 100 * \frac{1 - \exp(-K_{ext} * p * X)}{K_{ext} * p * X} \quad (12)$$

We fixed the values of μ_{max} and v_{PHB} to the values previously reported for the coupled mutant. K_{ext} , p and I_{opt} were fixed to the values $1.29 \text{ L gDW}^{-1} \text{ m}^{-1}$, 0.3 m and $125 \mu\text{E m}^{-2} \text{ h}^{-1}$, respectively (Laiglecia et al., 2013). We considered a biomass initial concentration of 0.078 gDW/L and set a simulation time of 48 hours to assure that the culture was in exponential growth phase. The model solution provides a final PHB concentration of $2.33 \text{ mmol}^{-1} \text{ L}^{-1}$ and a biomass concentration of 1.37 gDW/L . Considering the molecular weight of the PHB monomer of 84 mg/mmol , the production is equivalent to 200.38 mg PHB/L and the productivity results in a value of $100.19 \text{ mg PHB L}^{-1} \text{ day}^{-1}$.

4. Conclusions

In this work we have obtained a mutant that successfully couples PHB production to growth in *Synechocystis* by formulating a bilevel programming problem that identifies gene knockouts, which are represented by binary variables. Although the number of genetic interventions is high, CRISPR technologies (clustered regularly interspaced short palindromic repeats) are currently being developed and applied for simultaneous manipulation of multiple genes in cyanobacteria (Behler et al., 2018). Numerical results show productivity indexes of around 40% lower than the best one obtained experimentally (Wang et al., 2018) under nutrient limitation conditions. Further studies are needed to assess if coupled growth-production strategies are potentially suitable for direct large-scale PHB production through photosynthesis.

References

- Behler, J., Vijay, D., Hess, W.R., Akhtar, M.K., 2018. CRISPR-Based Technologies for Metabolic Engineering in Cyanobacteria. *Trends Biotechnol.* 36, 996–1010.
- Delpino, C., Estrada, V., Laglecia, J., Vidal, R., Florencio, F., Guerrero, M.G., Diaz, M.S., 2014. Dynamic Flux Balance Analysis in Cyanobacteria for ethanol production with Simultaneous Optimization Approaches. *Dynamic Flux Balance Analysis in Cyanobacteria for Ethanol Production with Simultaneous Optimization Approaches*, Proceedings of the 24th European Symposium on Computer Aided Process Engineering - ESCAPE 24. Elsevier.
- Du, W., Jongbloets, J.A., Guillaume, M., Van De Putte, B., Battaglini, B., Hellingwerf, K.J., Branco Dos Santos, F., 2019. Exploiting Day- And Night-Time Metabolism of *Synechocystis* sp. PCC 6803 for Fitness-Coupled Fumarate Production around the Clock. *ACS Synth. Biol.* 8, 2263–2269.
- Kamravamanesh, D., Lackner, M., Herwig, C., 2018. Bioprocess engineering aspects of sustainable polyhydroxyalkanoate production in cyanobacteria. *Bioengineering* 5, 1–18.
- Laglecia, J., Estrada, V., Vidal, R., Florencio, F.J., 2013. Dynamic Flux Balance Analysis of a Genetic Engineered Cyanobacterium for Ethanol Production. *Chem. Eng. Trans.* 32, 955–960.
- Lasry Testa, R., Delpino, C., Estrada, V., Diaz, M.S., 2019a. In silico strategies to couple production of bioethanol with growth in cyanobacteria. *Biotechnol. Bioeng.* 1–13.
- Lasry Testa, R., Delpino, C., Estrada, V., Diaz, M.S., 2019b. Bioethanol Production with Cyanobacteria by a Two-Stage Fermentation Strategy. *Comput. Aided Chem. Eng.* 46.
- Meixner, K., Kovalcik, A., Sykacek, E., Gruber-brunhumer, M., Zeilinger, W., Markl, K., 2018. Cyanobacteria Biorefinery — Production of poly (3-hydroxybutyrate) with *Synechocystis* salina and utilisation of residual biomass. *J. Biotechnol.* 265, 46–53.
- Nogales, J., Gudmundsson, S., Knight, E.M., Palsson, B.O., Thiele, I., 2012. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc. Natl. Acad. Sci.* 109, 2678–2683.
- PSEnterprise Ltd., 2020. gPROMS Advanced User Guide-release 2.3. Process Systems Enterprise Ltd., London
- Savinell, J.M., Palsson, B.O., 1992. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J. Theor. Biol.* 154, 421–454.
- Wang, B., Xiong, W., Yu, J., Maness, P.C., Meldrum, D.R., 2018. Unlocking the photobiological conversion of CO₂ to (R)-3-hydroxybutyrate in cyanobacteria. *Green Chem.* 20, 3772–3782.

Parametric analysis of mammalian cell (GS-NS0) culture performance for advanced mAb biopharmaceutical manufacturing

Wil Jones, Dimitrios I. Gerogiorgis*

Institute of Materials and Processes (IMP), School of Engineering, University of Edinburgh, The Kings Buildings, Edinburgh, EH9 3FB, United Kingdom

**D.Gerogiorgis@ed.ac.uk*

Abstract

Monoclonal antibodies (mAbs) are an example of therapeutic biomolecules which have been used for treatments of cancer, autoimmune diseases and other ailments. Numerous mAbs are currently being given market approval and the total mAb market value is predicted to increase for the foreseeable future. Traditionally, mAbs are manufactured via secretion from mammalian cell cultures such as Chinese Hamster Ovaries (CHO) or myeloma cell lines e.g. GS-NS0. Technoeconomically successful bioreactor operation is inherently dynamic and must invariably rely on limited state measurement accessibility. In all cases, culture productivity and economic incentive must be ensured simultaneously, thus implying a pressing need for model-based elucidation of optimal operation policies. Dynamic simulation is an established tool in bioprocess systems engineering which can accelerate process intensification and scale-up, for capital and operational cost reduction. The present study extends uncertainty characterization beyond recently used response envelopes by firstly presenting dynamic simulation results for a GS-NS0 culture, and then performing a parametric sensitivity analysis so as to systematically deduce the impact of model parameters on the said key output variables which govern technoeconomic appeal.

Keywords: Dynamic simulation; biotherapeutics; monoclonal antibodies/mAb; GS-NS0.

1. Introduction

Monoclonal antibodies (mAbs) encompass a wide spectrum of biotherapeutics which have gained great attention for their potential in cancer and autoimmune disease therapies. The first therapeutic mAb which received approval by the United States Food and Drug Administration (FDA) was *muromonab-CD3* in 1986; since that historic milestone, many (78) more mAbs have been approved for production and commercial use, boosting the phenomenal mAb market growth to a valuation of \$115.2 billion in 2018 (Lu et al., 2020). Traditionally, industrial manufacturing of mAbs at production scale requires the use of mammalian cell cultures in a batch or fed-batch fermentation, whose raw effluent is then fed to a sequence of downstream separation units (e.g. centrifugation, chromatography, and others) in order to isolate the active biopharmaceutical agent (Rodrigues et al., 2010). Acquiring a mechanistic understanding of how feed and culture media affect upstream metabolism is critical in order to first capture biological complexity in a tractable fashion and then identify viable, industrially valuable optimisation prospects (Chen et al., 2016). Credible modelling enables dynamic simulation and rapid technoeconomic evaluation of *in vivo* bioprocess operation policies without costly, arduous experimental campaigns, if validation is performed and relevance ensured (Shirahata et al., 2019; Diab et al., 2020).

A study published by Kiparissides et al. (2015) has employed several *in silico* techniques to quantitatively describe mAbs manufacturing from a GS-NS0 mammalian cell culture via batch and fed batch reactors. A novel, concise dynamic model and its parameterisation based on Monod kinetics has been presented but also duly validated against three batch mAb production experiments, each of which had considered a different feed composition. The purpose of the present paper is to use the same dynamic model in order to probe the industrial potential of GS-NS0 cells by means of an extensive local sensitivity analysis. Key batch reactor outputs (substrate consumption, viable cell concentration, mAb yield) and state sensitivities vs. parameter levels are obtained, indicating remarkable differences among parameters (and sometimes policies) with respect to techno-economic projections.

2. Dynamic model for mAb production from a GS-NS0 cell culture

The dynamic model for GS-NS0-based mAb production of Kiparissides et al. (2015) is summarised in Table 1, with Monod kinetics describing all metabolic reactions occurring. Eqs. (1)-(3) define volume V , viable cell concentration X_V and dead cell concentration X_D , respectively. Eqs. (4)-(9) portray metabolite and product concentrations evolution for glucose [GLC], glutamate [GLU], ammonia [AMM], lactate [LAC], glutamine [GLN], and mAb [mAb], respectively. Both glucose and glutamate are assumed to be growth-limiting substrates, hence their combined effect modelled as additive. Of the 20 model parameters, 18 are published (Kiparissides et al., 2015); the normalised total cell number and lag potential ($X_{T,NORM}$ and δ_{NORM}) have been computed on the basis of published data therein. The lag potential (δ) is a parameter describing the time required by cells for adaptation to the environment, before entering an exponential growth phase (Kiparissides et al., 2015).

Table 1: Dynamic model for mAb production (Kiparissides et al., 2015).

$$\frac{dV}{dt} = F_{in} - F_{out} \quad (1)$$

$$\frac{dX_V}{dt} = \frac{(-X_V \frac{dV}{dt} + (\mu - \mu_d)X_V V - F_{out}X_V)}{V} \quad (2)$$

$$\frac{dX_D}{dt} = \frac{(-X_D \frac{dV}{dt} + \mu_d X_V V - k_{lys} X_D V - F_{out}X_D)}{V} \quad (3)$$

$$\frac{d[GLC]}{dt} = \frac{(-[GLC] \frac{dV}{dt} - Q_{GLC} X_V V + F_{in}[GLC]_{in} - F_{out}[GLC])}{V} \quad (4)$$

$$\frac{d[GLU]}{dt} = \frac{(-[GLU] \frac{dV}{dt} - Q_{GLU} X_V V + F_{in}[GLU]_{in} - F_{out}[GLU])}{V} \quad (5)$$

$$\frac{d[AMM]}{dt} = \frac{(-[AMM] \frac{dV}{dt} + (Q_{GLN} Y_{AMM,GLN}^{ACTUAL} X_V V - F_{out}[AMM]))}{V} \quad (6)$$

$$\frac{d[LAC]}{dt} = \frac{(-[LAC] \frac{dV}{dt} + Q_{LAC} X_V V - F_{out}[LAC])}{V} \quad (7)$$

$$\frac{d[GLN]}{dt} = Q_{GLC} X_V - Q_{GLN} X_V \quad (8)$$

$$\frac{d[mAb]}{dt} = \frac{(-[mAb] \frac{dV}{dt} + (\frac{\mu_{GLU}}{Y_{X,mAb}} + m_{mAb} \frac{[GLC]}{\mu}) - F_{out}[mAb])}{V} \quad (9)$$

3. Dynamic simulation of mAb production

Dynamic simulation of mAb production has been performed on the basis of the foregoing model to validate our code vs. three published experiments (Kiparissides et al., 2016). Fig. 1 illustrates concentration and cell density trajectories for each of the three cases. The mAb maintenance term (m_{mAb}) varies among the published three experiments, being $6.244 \cdot 10^{-10}$, $9.944 \cdot 10^{-10}$ and $2.044 \cdot 10^{-10}$ mg cell⁻¹ h⁻¹ for these Expts. 1, 2, 3 respectively. Furthermore, for our dynamic model validation we assume that there is no correlation between seeding density and ammonia yield on glutamine, hence $Y_{AMM, GLN}^{ACTUAL} = Y_{AMM, GLN}$. All published dynamic simulation results and trends have been successfully reproduced. Experiments with high lag times produce mAb at the lowest concentrations (Fig. 1a, 1d), while $[mAb]$ production rate is a constant (albeit variant) derivative in all Expts. (Fig. 1d). High $[mAb]$ productivity is correlated here with high initial viable cell concentration $[X_v]$. Interestingly, the highest $[mAb]$ production occurs for a culture inoculated with the least total amount of $[GLU]+[GLC]$ substrates (Expt. 2), implying that starving the GS-NS0 culture of nutrients may prove beneficial towards boosting cellular mAb productivity. Glucose oversupply may have induced excessive $[LAC]$ (and inhibited $[mAb]$) production as by-products harm cell growth and productivity (Gagnon et al., 2011; Fan et al., 2015). Viable cell concentration $[X_v]$ clearly peaks earlier for lower lag potential (Fig. 1a), with peak magnitude increasing as a function of initial glutamate concentration. Finally, high lactate concentrations occur for high initial glucose substrate concentrations (Fig. 1b-1c).

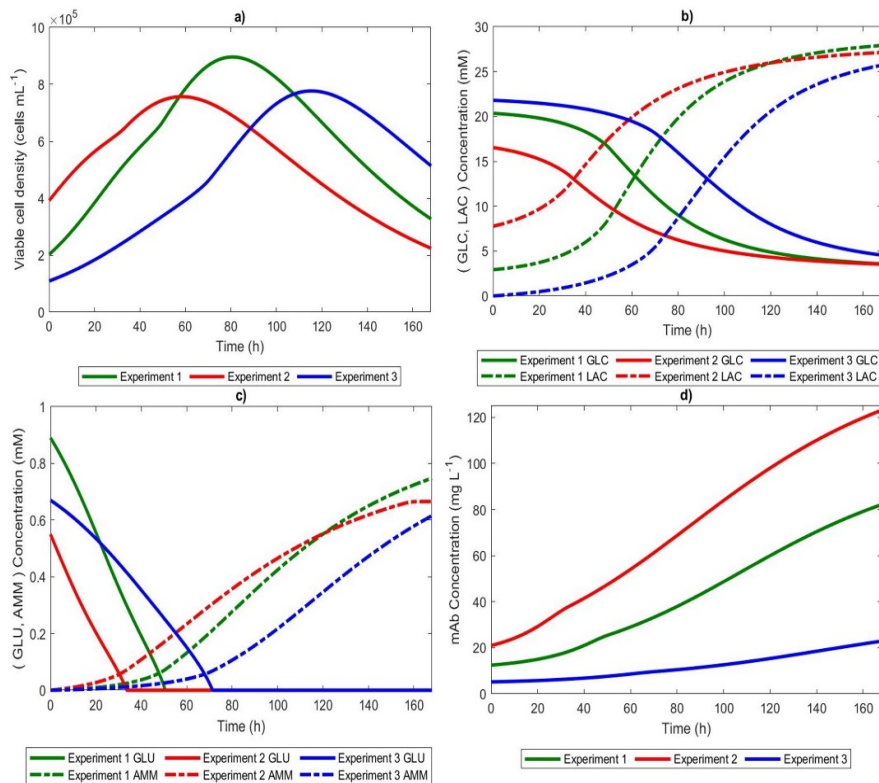


Figure 1: Dynamic simulation of key state variables for three published experiments.

4. Parametric analysis methodology

Dynamic simulation results clearly indicate that numerous factors significantly influence GS-NS0 batch culture stability and performance in regard to mAb production efficiency. Because of this intrinsic and multi-faceted variability, the purpose of this paper is to conduct a (hitherto not attempted) parametric sensitivity analysis in order to establish the influence of model parameters on process state variables at the end of each experiment. The first state variable, V (Eq. 1) remains constant in all experiments and is discounted. Our purpose is to visualise high sensitivities (e.g. > 1) which imply that a relatively small parameter change (or estimation inaccuracy) induces considerable output state variable value variation (especially $[mAb]$) at batch completion ($t = t_f$). The latter in turn signifies that small parameter estimation errors may snowball into gross model output errors, thus rendering model fidelity questionable and technoeconomic projections highly unreliable. The parametric sensitivity analysis relies on computing a sensitivity index (S) via Eq. (10) for each state (x) and parameter (P), with respective changes ∂x and ∂P (Kreyszig, 1999).

$$S = \left[\frac{\left(\frac{\partial x}{x} \right)}{\left(\frac{\partial P}{P} \right)} \right] \quad (10)$$

Four different sensitivity analysis cases have been considered and computed for as many parametric differential variation levels, $(\partial P/P) = \{-0.50, -0.05, 0.05, 0.50\}$ or $\pm 5\%, \pm 50\%$. The sensitivity outputs (S) have been normalised in every case for visual clarity, but also to comparatively appreciate the relative effect of each parameter on each state variables. The normalised sensitivity index $\bar{S}_{x,P}$ is given by the sensitivity for state variable x and parameter P , $S_{x,P}$, divided by the respective maximum observed, $S_{x,MAX}$ (Kreyszig, 1999).

$$\bar{S}_{x,P} = \frac{S_{x,P}}{S_{x,MAX}} \quad (11)$$

The initial experimental conditions remained identical to those used in section 3. Only one parameter was varied at a time, meaning the sensitivity analysis was local in nature.

5. Local parametric sensitivity analysis

Sensitivities for six states ($[mAb]$, $[X_V]$, $[X_D]$, $[GLC]$, $[AMM]$, $[LAC]$) are given in Fig. 2. A clear and remarkably consistent reversal trend emerges: for a contraction of sensitivity magnitude in case of negative parameter changes ($-5\%, -50\%$), we observe an expansion (of different magnitude) for the respective positive changes ($+5\%, +50\%$), and vice versa. All state variables are highly sensitivity to lag potential parameters (δ_{NORM} , k_{LAG} , $X_{T,NORM}$), emphasising the importance of a nutritious environment towards GS-NS0 proliferation. The state of highest technoeconomic interest $[mAb]$ is sensitive to only certain parameters ($K_{G,GLC}$, K_{LLAC} , $Y_{LAC,GLC}$, $\mu_{MAX,GLC}$, $X_{T,NORM}$) for all experiments (Fig. 2a), i.e. precisely to those affecting cell ability to consume glucose and/or produce lactate (Chen et al, 2012). Viable cell concentration $[X_V]$ is most sensitive to $Y_{LAC,GLC}$ and K_{LLAC} , in agreement with the fact that high lactate concentration adversely affects cell growth (Gagnon et al., 2011). Dead cell concentration $[X_D]$ displays much lower sensitivities to most parameters, and is relatively insensitive to cell lysis (k_{LYS}) compared to the lag potential parameters (Fig. 2c). Glucose concentration $[GLC]$ is affected by 5 parameters (>0.2) (Expts. 2-3 differ a lot). Ammonia concentration $[AMM]$ has a different sensitivity profile vs. all previous states: several (at least 8 of 20) parameters affect its final value, but also Expt. 2 seems to induce key differences vs. Expts 1+3, possibly as $[AMM]$ rises much faster in that case (Fig. 1c).

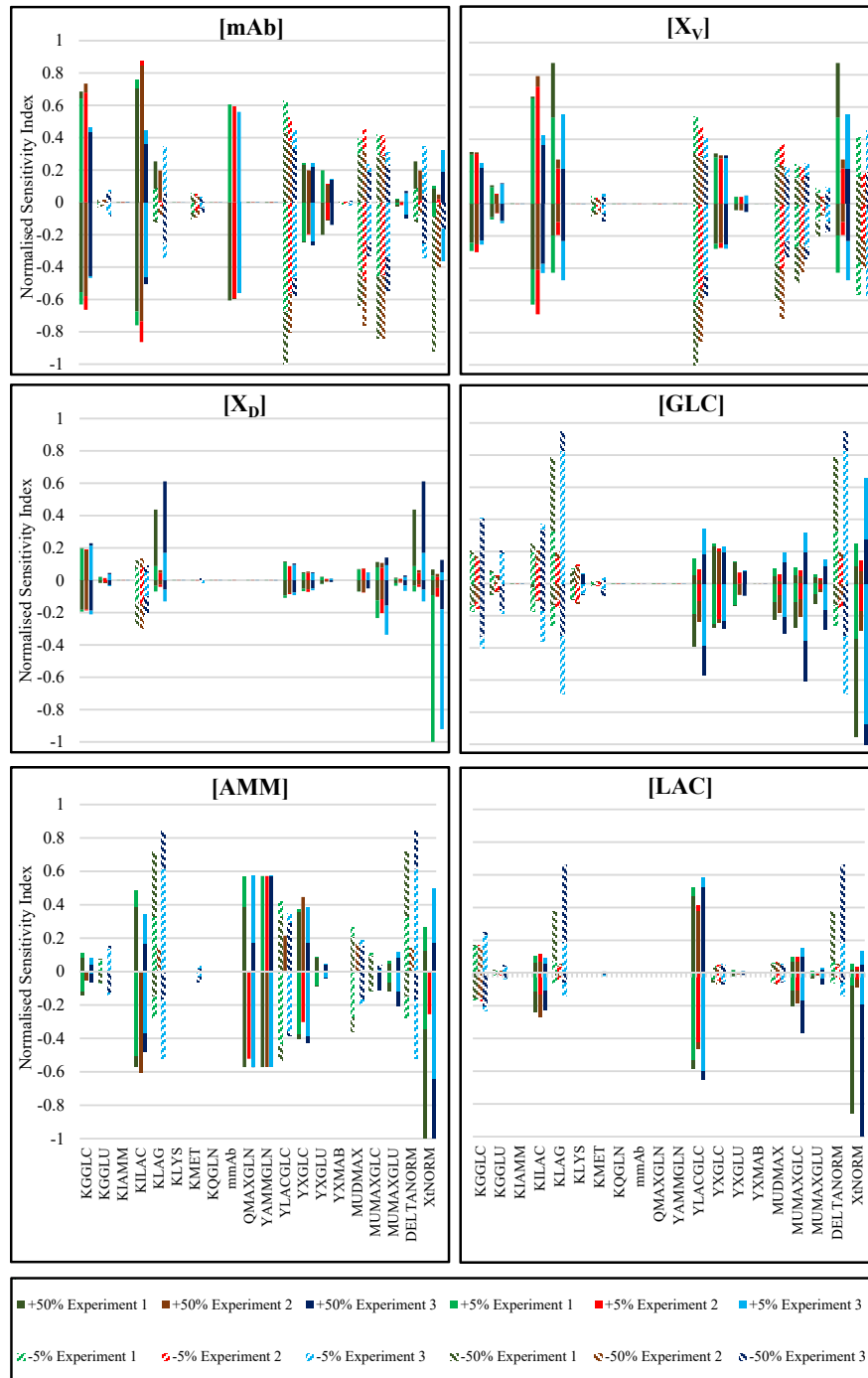


Figure 2: Local sensitivity analysis for different final state variable values.

The high $[AMM]$ sensitivity both to max. glutamine uptake rate ($Q_{MAX,GLN}$) and yield from glutamine consumption ($Y_{AMM,GLN}$) agrees with the observation that ammonia levels are correlated with glutamine and asparagine in mammalian cell cultures (Zhang et al., 2016). Lactate concentration $[LAC]$ sensitivity to its yield from glucose consumption ($Y_{LAC,GLC}$) is high, due to its key role (and inefficiency) in glucose metabolism (Gagnon et al., 2011). The said parameter ($Y_{LAC,GLC}$) influences lactate $[LAC]$ a lot more than glucose $[GLC]$ for all three Expts., illustrating that glycolysis can be more efficient towards securing energy.

6. Conclusions

Dynamic simulation and parametric sensitivity analysis of GS-NS0 culture usage for biotherapeutic mAbs manufacturing has been pursued on the basis of a published model. High target product concentration $[mAb]$ has been correlated with high viable biomass availability $[X_V]$ and low initial cumulative substrate ($[GLU]$ and $[GLC]$) concentrations. The parametric sensitivity analysis is illustrated for several key model output variables: from an operational viewpoint aimed at biopharmaceutical manufacturing improvements, GS-NS0 culture modifications decreasing $Y_{LAC,GLC}$, μ_{MAX} , $\mu_{MAX,GLC}$ and $X_{T,NORM}$ (as well as increasing $K_{G,GLC}$ and $K_{L,LAC}$) are expected to boost the final target $[mAb]$ concentration. Judicious substrate $[GLC]$ provision and suppression of lactate $[LAC]$ production over time emerge as key recommendations, beyond obvious (e.g. $X_{T,NORM}$) conditioning needs. Further to bioreactor modelling, process optimisation must also consider separation units, so that high-fidelity submodels in the latter credibly describe their output and plant yield.

Acknowledgements

The authors gratefully acknowledge the Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Partnership (DTP) PhD Scholarship awarded to W.J., as well as a Royal Society Short Industrial Fellowship (2020-21) awarded to D.I.G.

References

- C. Chen, H. Le, C. Goudar, 2016. Integration of systems biology in cell line and process development for biopharmaceutical manufacturing, *Biochem. Eng. J.*, **107**: 11-17.
- F. Chen, Z. Ye, L. Zhao et al., 2012. Correlation of antibody production rate with glucose and lactate metabolism in CHO cells, *Biotechnol. Lett.* **34**(3): 425-432.
- S. Diab, S. Badr, H. Sugiyama, D.I. Gerogiorgis, 2020. Dynamic simulation and visualisation of pH modulated fed-batch fermentation for mAb production from CHO cell cultures, *Comput. Aided Chem. Eng.*, **47**: 1657-1662.
- Y. Fan, I. Jimenez Del Val, C. Müller et al. 2015. A multi-pronged investigation into the effect of glucose starvation and culture duration on fed-batch CHO cell culture. *Biotechnol. Bioeng.* **112**(10): 2172-2184.
- M. Gagnon et al., 2011. High-end pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures. *Biotechnol. Bioeng.* **108**(6): 1328-1337.
- A. Kiparissides, E. N. Pistikopoulos, A. Mantalaris, 2015. On the model-based optimization of secreting mammalian cell (GS-NS0) cultures, *Biotechnol. Bioeng.* **112**(3), 536-548.
- E. Kreyszig, 1999, *Advanced Engineering Mathematics*, Wiley.
- R. Lu, Y. Hwang, I. Liu et al., 2020. Development of therapeutic antibodies for the treatment of diseases, *J. Biomed. Sci.*, **27**(1), 1-30.
- M. Rodrigues, A. Costa, M. Henriques et al., 2010. Technological progresses in monoclonal antibody production systems, *Biotechnol. Progress*, **26**, 332-351.
- H. Shirahata, S. Diab, H. Sugiyama, D.I. Gerogiorgis, 2019. Dynamic modelling, simulation and economic evaluation of two CHO cell-based production modes towards developing biopharmaceutical manufacturing processes, *Chem. Eng. Res. Des.*, **150**, 218-233.
- L. Zhang, W. Zhang, C. Wang et al., 2016. Responses of CHO-DHFR cells to ratio of asparagine to glutamine in feed media, *Bioresour. Bioprocess.* **3**(1): 1-12.

A technoeconomic assessment of an on-site biocrude production from sewage sludge in Qatar's wastewater treatment plants

Mohammad Alherbawi,^a Tareq Al-Ansari,^{a,b,*} Hamish R. Mackey,^a Gordon McKay^a

^aDivision of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.

^bDivision of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar.

* talansari@hbku.edu.qa

Abstract

The generation of sewage sludge in the State of Qatar is increasing with the rapid growth of its population. Around 300,000 m³ of wet sludge is generated every year. Although the sludge is being treated to meet the international standards, its final disposal into landfills raises concerns associated with the limited land areas in the country, as well as the possible contamination of soil and groundwater. On the other hand, the hydrothermal liquefaction (HTL) technology is attracting more and more attention at the expense of other thermochemical waste valorisation pathways including incineration, gasification and pyrolysis. Since it can tolerate the high-water content of wet feedstocks, while the HTL biocrude has a comparable heating value to that of crude petroleum and has the potential to be co-processed in existing oil refineries. Therefore, this study investigates the technoeconomic aspects of an on-site biocrude production from sewage sludge at wastewater treatment plants in Qatar via hydrothermal liquefaction. The system is modelled using Aspen Plus, while an economic assessment is conducted with the aid of Aspen Process Economic Analyser (APEA). The results indicate a significant biocrude yield of up to 38.2 wt.% (on dry and ash-free basis). In addition, based on the economic feasibility evaluation, the produced biocrude achieved a competitive minimum selling price of 52.6 US\$/bbl, which is below the past 5-year global average price of crude petroleum. As such, the incorporation of biocrude into the oil refining process can be a promising alternative for oil-refining countries, especially at times of petroleum prices hikes.

Keywords: Qatar, Biocrude, Sewage sludge, Hydrothermal liquefaction, Treated sewage effluent.

1. Introduction

Qatar is an oil producing country and the world's second largest exporter of natural gas (Britishpetroleum, 2019). Along its prominent gas facilities, two state-owned oil refineries operate in Qatar, which are: Qatar Petroleum Refinery that mainly processes crude petroleum at a capacity of 137,000 barrels per stream day (bpsd), and Laffan Refinery that utilises field condensate obtained from natural gas facilities with a capacity of 292,000 bpsd (Qatarpetroleum, 2020).

However, the country is striving to incorporate renewables into its energy mix to mitigate carbon emissions and prolong its petroleum reserves' lifetime. For this purpose, several local feedstocks, as well as processing technologies have been investigated (Al-Ansari et al., 2020; Alnousse et al., 2018; Elkhailifa et al., 2019), which emphasised that municipal, agricultural and industrial wastes are considered as the most abundant feedstock for the production of alternative liquid and gaseous fuels. Meanwhile, the presence of refining facilities in the country and the recent ASTM certification of bio-oils' co-processing with fossil crude, these factors grant an advantage to the oil-producing biomass processing pathways such as pyrolysis and hydrothermal liquefaction (HTL) over other technologies.

Pyrolysis is a thermal decomposition of biomass in inert media to produce bio-oil and biochar at (300-700 °C). However, biomass must undergo a drying step prior to pyrolysis (Alherbawi et al., 2021a). In contrast, hydrothermal liquefaction is often utilised to decompose the high-moisture biomass using water at subcritical conditions, which mimics the crude petroleum formation underground (Alherbawi et al., 2020). The HTL process mainly produces biocrude oil which is of a better quality and with a higher net heating value as compared to pyrolysis oil (Peterson et al., 2008).

Amongst the different available wet biomass in Qatar, sewage sludge can be considered as one of the most fitting feedstocks for HTL process due to its high moisture and carbon contents. As such, no additional water is required to create a slurry feed.

With the rapidly growing population in Qatar, the quantity of generated sludge has significantly increased by 3.5 folds through 2010-2015 (MDPS, 2018), which is deemed to be a growing concern. Whereby, sludge treatment and disposal account for up to 50% of total operating costs and nearly 40% of the net GHG emissions associated to the operations of wastewater treatment plants (WWTPs) (Callegari and Capodaglio, 2018).

As such, considering alternatives to sludge disposal is becoming more urging to reduce costs and carbon footprint, as well as to valorise the sludge into value-added products.

Currently, there are 23 WWTPs in Qatar as presented in Figure 1. However, three of them account for nearly 75% of total wastewater treatment capacity, named: Doha North (30%), Doha South (22%) and Doha West (22%). Therefore, this study proposes an on-site utilisation of sewage sludge to produce biocrude via HTL in Qatar. The sludge from the three main WWTPs is to be utilised, whereby, “Doha South WWTP” is selected as the hub to establish the HTL plant for being the closest to Qatar Petroleum (QP) refinery (~45 km). While the sludge from the other two main WWTPs is transported by tanker trucks to the HTL plant. Whereas produced biocrude is assumed to be subsequently upgraded at QP refinery by co-processing with crude petroleum (not evaluated in this study). The HTL process is modelled using Aspen Plus® software, while an economic analysis is conducted to evaluate the feasibility of biocrude production with the aid of Aspen Process Economic Analyser (APEA).



Figure 1: WWTPs and oil refineries in Qatar (Data adapted from (MDPS, 2018)).

2. Methodology

2.1 Model Development

Advanced system process engineering (ASPEN PLUS V.9) is used to simulate the hydrothermal liquefaction process. All subsystems are simulated assuming to have steady-state and isothermal processes. Besides, char is assumed to comprise carbon only, while the nitrogen and sulfur contents of feedstock are released in the form of NH_3 and H_2S respectively or as nitrogen and sulfur-containing hydrocarbons (Alherbawi et al., 2021b).

Initially, the wet sludge is defined as a non-conventional component considering the proximate and elemental analysis adapted from a previous study by Shen et al. (2001). While the typical Qatari sludge's moisture content of 82% is considered (MDPS, 2018).

No water is further added to the slurry as the solid content of sludge does not exceed 18%. Sodium hydroxide is added to the slurry with a 1 wt.% load to enhance the potential of biocrude formation (Cao et al., 2017).

The stream is then pumped to 150 bars using two consecutive pumps, then passed into a counter-current heat exchanger. The initial decomposition (hydrolysis) stage is simulated using an “RYield” aspen unit, by which the biomass is broken-down into conventional components using a Fortran code (Alnousse et al., 2018).

The conventional stream is then processed into the main HTL reactor, which is simulated at 300 °C and 150 bars using an “RGibbs” unit, which functions following the principle of Gibbs free energy minimisation. The possible biocrude, gas and aqueous phase compositions are adapted from previous studies (Tekin et al., 2013).

Finally, the outflow stream is fed back into the heat exchanger to recover the heat before separating the solid phase using a solid separator. While the remaining volatile stream is fed into a three-stage flash drum, yielding three phases including gas, biocrude and an aqueous phase. The process flow diagram is presented in Figure 2.

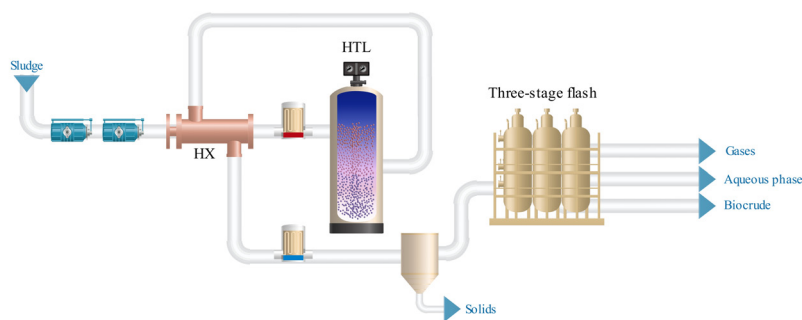


Figure 2: Process flow diagram of HTL process.

2.2 Economic Assessment

Aspen Process Economic Analyser (APEA) is utilised to evaluate the economics of a WWTP’s on-site HTL facility in Qatar. A plant lifespan of 30 years and a start-up period of 6 months are considered. The plant’s feed capacity is fixed at 222,000 m³ of wet sludge, which roughly represents the quantity of sludge generated from the three main WWTPs of Qatar.

As for capital expenses, the main HTL reactor cost is adapted from published reports (Zhu et al., 2014), while the costs of other smaller units (i.e., heaters, pumps, etc.) are evaluated by APEA based on the defined size and operating parameters. All costs are re-evaluated for the base year of analysis (2019) using the chemical engineering plant cost index (CEPCI).

The utilities costs, sludge handling charges and wastewater discharge costs are quoted locally (Gulfimes, 2020; Kahrama, 2020). While the operating and maintenance costs are evaluated by APEA. The minimum selling price of biocrude is evaluated in (\$/bbl) based on the traditional levelised cost of energy (LCOE) formula, with a discount rate of 20%:

$$LCOE \left(\frac{U.S.\$}{bbl} \right) = \frac{\sum_1^t ((K+O\&M+R)(1+d)^{-t})}{\sum_1^t ((P(1+d)^{-t})} \quad (1)$$

Where, (K) is the annualised capital expenses, (O&M) is the operating and maintenance costs, (R) is the raw materials prices, (d) is the discount rate, (P) is the biocrude yield and (t) is the plant lifespan.

3. Results and Discussion

3.1 Process Yield

The model results indicate a considerable biocrude yield of 38.22 daf.% (on feed dry and ash-free basis), and around 14 daf.% of hydrochar as illustrated in Figure 3. In addition, a mixture of gases is obtained including CO₂, CO, H₂, CH₄ and NH₃, with carbon dioxide being the dominant gas. While the aqueous phase contains dissolved organics including alcohol, phenols, as well as carbon dioxide.

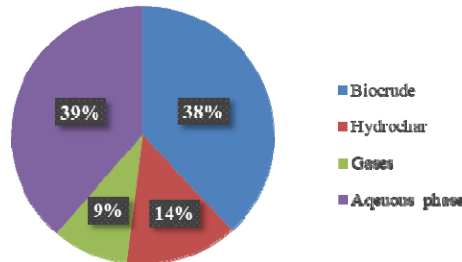


Figure 3: HTL products composition relative to dry and ash-free feed (%).

3.2 Economic Assessment

A summary of the plant's economic assessment is illustrated in Figure 4. The estimated project's capital expenses (CAPEX) is ~13.5 M\$ (million U.S. Dollars), while the annual operating expenses (OPEX) is 2.55 M\$. The breakdown of CAPEX indicates that purchased equipment contribute to the biggest portion (~41%) of the cost, followed by equipment settings (11%). As for OPEX, the catalysts and labour costs account for the highest shares of 30% and 26% respectively, followed by plant overhead (14%). Since the process is mostly operated on-site, the cost of feedstock is lowered down to only ~11% of total costs.

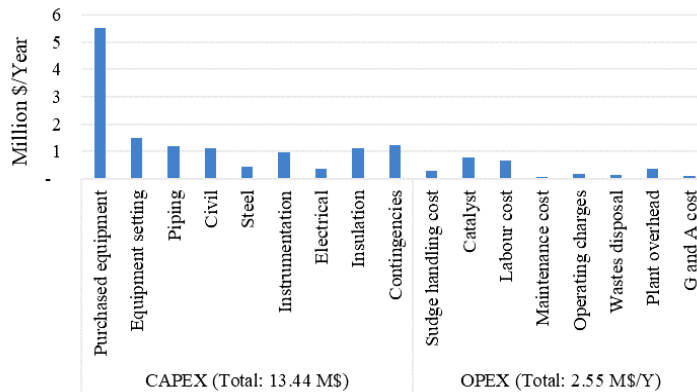


Figure 4: Breakdown of project's capital and operating expenses.

In addition, the achieved minimum selling price (MSP) of biocrude is at 52.6 \$/bbl, which is (3-19%) below the global crude petroleum price for the base year of analysis (2019) as presented in Figure 5. The achieved MSP of biocrude is also slightly below the past 5-year average crude price of 52.9 \$/bbl (Macrotrends, 2020). The main reason behind achieving a competitive MSP is the ‘on-site’ utilisation of sludge, which significantly reduces the collection and handling costs of feedstock.

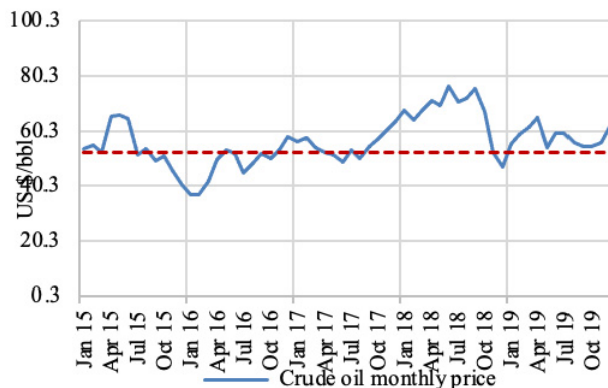


Figure 5: Biocrude MSP (this study) as compared to global crude prices.

4. Conclusion

The on-site valorisation of sewage sludge at WWTPs proved to be a promising pathway to produce biocrude, which can be co-processed with crude petroleum at the existing conventional refineries. In this study, the HTL yielded a significant biocrude composition while the achieved biocrude’s MSP is competitive and can feasibly replace portion of crude oil, which in return helps to prolong Qatar’s oil reserves lifetime and mitigate its carbon footprint.

References

- A. A. Peterson, F. Vogel, R. P. Lachance, M. Fröling, J. Antal Michael, J. W. Tester, 2008, Thermochemical biofuel production in hydrothermal media: A review of sub- and supercritical water technologies. *Energy Environ. Sci.* 1, 32–65.
- A. Alnoss, G. McKay, T. Al-ansari, 2018, Optimum Utilization of Biomass for the Production of Power and Fuels using Gasification, in: 28th European Symposium on Computer Aided Process Engineering. Elsevier Masson SAS, pp. 1481–1486.
- A. Callegari, A. Capodaglio, 2018, Properties and Beneficial Uses of (Bio)Chars, with Special Attention to Products from Sewage Sludge Pyrolysis. *Resources* 7.
- Britishpetroleum, 2019, BP Statistical Review of World Energy.
- GulfTimes, 2020, Ashghal sets fee for services. URL <https://www.gulf-times.com/story/653326/Ashghal-sets-fee-for-services> (accessed 8.13.20).
- Kahrama, 2020, Utilities tariff. URL <https://www.km.com.qa/CustomerService/Pages/Tariff.aspx> (accessed 1.1.20).
- K. Tekin, S. Karagöz, S. Bektaş, 2013, Effect of sodium perborate monohydrate concentrations on product distributions from the hydrothermal liquefaction of Scotch pine wood. *Fuel Process. Technol.* 110, 17–23.

- L. Cao, C. Zhang, H. Chen, D. C. W. Tsang, G. Luo, S. Zhang, J. Chen, 2017, Hydrothermal liquefaction of agricultural and forestry wastes: state-of-the-art review and future prospects. *Bioresour. Technol.* 245, 1184–1193.
- L. Shen, V.H.B.V. HB, H. M. Yan, D. K. Zhang, 2001, Pyrolysis of putrescible garbage and sewage sludge mixtures in the OFS process, in: *Proceedings of the 6th World Congress of Chemical Engineering*.
- Macrotrends, 2020, Crude Oil Prices - 70 Year Historical Chart. URL <https://www.macrotrends.net/1369/crude-oil-price-history-chart> (accessed 11.2.20).
- M. Alherbawi, A. AlNouss, G. McKay, T. Al-Ansari, 2020, Optimum Utilization of Jatropha Seedcake Considering the Energy, Water and Food Nexus, in: *30 European Symposium on Computer Aided Process Engineering*. Elsevier, pp. 229–234.
- M. Alherbawi, G. McKay, H. R. Mackey, T. Al-Ansari, 2021a, Jatropha curcas for jet biofuel production: Current status and future prospects. *Renew. Sustain. Energy Rev.* 135, 110396.
- M. Alherbawi, A. AlNouss, G. McKay, T. Al-Ansari, 2021b, Optimum sustainable utilisation of the whole fruit of Jatropha curcas: An energy, water and food nexus approach. *Renew. Sustain. Energy Rev.* 137, 110605.
- MDPS, 2018, Water Statistics in the State of Qatar 2017.
- Qatarpetroleum, 2020. QP Refinery Briefing. URL <https://qp.com.qa/en/QPActivities/QPOperations/Pages/QPRefinery.aspx> (accessed 11.2.20).
- S. ElKhalifa, A. AlNouss, T. Al-Ansari, H. R. Mackey, P. Parthasarathy, G. McKay, 2019. Simulation of Food Waste Pyrolysis for the Production of Biochar: A Qatar Case Study, in: *29 European Symposium on Computer Aided Process Engineering*. Elsevier, pp. 901–906.
- T. Al-Ansari, A. AlNouss, N. Al-Thani, P. Parthasarathy, S. ElKhalifa, G. McKay, M. Alherbawi, 2020, Optimising Multi Biomass Feedstock Utilisation Considering a Multi Technology Approach, in: *30 European Symposium on Computer Aided Process Engineering*. Elsevier, pp. 1633–1638.
- Y. Zhu, M. J. Bidy, S. B. Jones, D. C. Elliott, A. J. Schmidt, 2014, Techno-economic analysis of liquid fuel production from woody biomass via hydrothermal liquefaction (HTL) and upgrading. *Appl. Energy* 129, 384–394.

Production costs of advanced biofuels using a multi-component learning curve model

Paraskevi Karka^a, Filip Johnsson^a, Stavros Papadokonstantakis^{a*}

^a*Chalmers University of Technology, Department of Space Earth and Environment,
Division of Energy Technology, Hörsalsvägen 7B, 41296 Gothenburg, Sweden*

**stavros.papadokonstantakis@chalmers.se*

Abstract

The production costs of advanced biofuel options are currently higher than those of their fossil fuel equivalents. Capital Expenditures (CAPEX) for the production of liquid biofuels for road, aviation and marine transport sectors have a significant contribution to the overall production cost, together with the feedstock cost. It is, therefore, important to estimate the potential for cost reduction through R&D and experience in assembling a growing number of respective plants (i.e., from first-of-a kind (FOAK) to the Nth plant (NOAK)), which comprise a mix of established and innovative technological components. This could provide valuable information to stakeholders for the expected investment costs to meet European Commission goals in 2050.

This study adopts a methodological framework based on the “learning curve theory” to estimate cost reduction as a result from the experience of technology implementation, in terms of numbers or capacity of units implemented. This work applies the learning theory as a multicomponent analysis, which requires a systematic decomposition of the entire production process to identify established and innovative technological components that can be analysed in detail using the corresponding technoeconomic data.

The analysis showed that CAPEX reduction in the range of 10-25% could be expected to reach capacities corresponding to NOAK plants in 2050. To reach further CAPEX reduction of 40%, for example, would require higher cumulative annual growth rate to achieve two orders of magnitude increase of cumulative installed capacity. This corresponds to hundreds of GWs or equivalently some hundreds or thousands of large-scale plants to meet the goal of 20-25% transportation fuels consumption to be covered by advanced biofuels in 2050.

Keywords: biofuels deployment, CAPEX reduction, learning curve, TRL increase.

1. Introduction

The implementation of processes for advanced biofuels is a multidimensional issue affected by a wide range of technical, economic aspects and policy aspects. According to the study of IEA (2020) CAPEX values of bioprocesses based on lignocellulosic biomass have a significant contribution to the overall production cost, similar to the feedstock cost (around 40% each), while other operating and maintenance costs typically cover 15%-20% of the overall production cost.

Considering that many of these processes have reached TRL demonstration scale (e.g. gasification and 2nd generation ethanol) and/or satisfactory efficiency, the scope of CAPEX reduction due to experience gain when shifting from the first-of-a kind (FOAK) to the Nth plant (NOAK) could be a driver for long term deployment making them competitive to fossil based equivalents, and meeting the European Commission goals of

13%-24% contribution of liquid biofuels in the energy consumption for transportation in 2050 (EUROPEAN COMMISSION, 2018).

The learning curve (LC) theory is an important tool for estimating technical change related to the evolution of technologies and provides a systematic link between decreasing production costs and cumulative volume. It has found empirical support in a wide range of industrial sectors and technologies. In the energy field, the experience curve concept is so far used mostly to describe learning for modular products, such as wind turbines (Junginger M, et al., 2005), and solar photovoltaic (PV) modules (Mauleón I., 2016). Learning effects have been also investigated for bioenergy systems (Junginger et al, 2006), large-scale production of ethanol from sugarcane in Brazil (Goldemberg, 2003), the development of natural-gas-fired combined cycle power plant (Claeson Colpier and Cornland, 2002) and the case of renewable fuels (Detz et al, 2018). Most recently, an estimation of cost reduction potential of biofuels was presented by IEA (2020) using the LC theory from a top-down perspective based on data gathered from pilot and demonstration plants and interviews with experts.

The aim of this study is to develop a transparent modular framework for the application of the LC methodology for advanced biofuels from a bottom-up perspective. This is achieved by using a systematic decomposition of the advanced fuels production paths into a mix of innovative and established technologies that allows for the investigation of potential CAPEX reduction under various scenarios of installed capacity evolution. These scenarios may be influenced by policy targets and market estimates and refer to both the process as a whole and the independent evolution of some of its technological components.

2. Methodology

2.1. The learning curve framework

Technology learning is commonly modelled as a single factor approach to provide the way that production costs are reduced by a constant fraction for doubling of cumulative production). The single factor relationship is commonly expressed through a power-law function between production costs and cumulative production (Nemet 2006; Trappey et al., 2016)

$$C(Q_t) = C(Q_0) \cdot \left[\frac{Q_t}{Q_0}\right]^{-b} \quad (1)$$

where Q_t is the cumulative production, b is the positive learning parameter, related to the LR learning rate ($LR = 1 - 2^{-b}$), $C(Q_t)$ is the unit cost of production at Q_t , (i.e., operating cost and annualised investment cost, in the more general form), $C(Q_0)$ and Q_0 are respectively the cost and cumulative production at an arbitrary starting point.

The component learning approach (Ferioli et al., 2009, Upstill and Hall, 2018) extends the single factor model by not applying the methodology at the overall process level but as a sum of the costs of its individual technological components. Such an approach would be of importance when applying learning to biomass processes which consist of groups of unit operations (i.e., process components) with different maturity. In this case, the technology learning relationship may be expressed for each component individually, as follows:

$$C(Q_t) = \sum C(Q_{0i}) \cdot \left[\frac{Q_t}{Q_0}\right]^{-b(i)} = C_{01} \left[\frac{Q_{t1}}{Q_{01}}\right]^{-b(1)} + C_{02} \left[\frac{Q_{t2}}{Q_{02}}\right]^{-b(2)} + \dots + C_{0n} \left[\frac{Q_{tn}}{Q_{0n}}\right]^{-b(n)} \quad (2)$$

where $b(i)$ is positive learning parameter for technological component i , $C(Q_t)$ is the unit cost of production at cumulative production Q_t , Q_0 is the cumulative production at an arbitrary starting point, C_{0i} is the cost and Q_{0i} is the cumulative production of component

i at an arbitrary starting point, and $Q_{t,i} = (Q_{t-1,i} + CAGR_i \cdot Q_{t-1,i}) \cdot DF$. $CAGR_i$ is the cumulative annual growth rate of the corresponding technological component and DF is a decline factor as proposed by Detz, et al. (2018), equal to 0.98 for the particular calculation.

2.2. Cost reduction methodology in the case of advanced biofuels

The method is applied for the case of seven liquid biofuels (methanol, liquefied methane, jet fuels, diesel, kerosene, ethanol, butanol) produced from thermochemical and biochemical pathways. The chosen fuels and production technologies are selected as their technology readiness level (TRL) ranges from 6 to 9, (Figure 1).

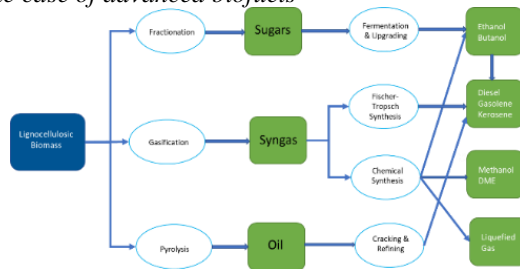


Figure 1. Production pathways for the liquid biofuels of the current study

To apply the multicomponent learning approach for the case of advanced biofuels, their production is modelled by a pathway consisting of a sequence of unit processes which produce the desired processes and replace equivalent fossil-based products. Each pathway of biofuels may consist both of purely biobased processes which convert biomass into intermediate products and conventional parts for upgrading purposes.

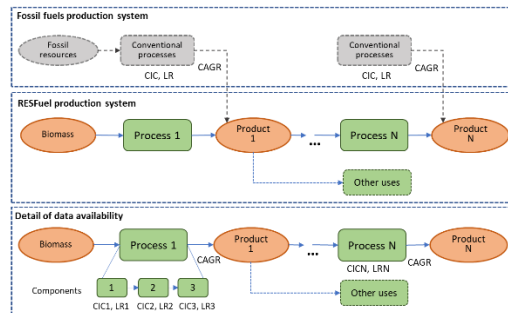


Figure 2. Decomposition of biobased pathway into unit processes for subsequent division into technological components (see Step 2).

Thus, the application of the learning curve theory to assess cost reduction potential through learning by doing is composed of three steps.

Step 1: The technology pathway is divided into process steps (first decomposition level represented by the intermediate box of Figure 2) resulting in a specific intermediate product (syngas, pyrolysis oil, etc.). Then, each process step is divided into components representing elementary technological steps (second decomposition level represented by the lower box of Figure 2). These components are those presented in Eq. 2. For example, in the case of methanol two process steps are identified in the first decomposition level (i.e., biomass gasification and methanol synthesis), while the decomposition of each process steps into technological components is presented in Table 1. This two-level process decomposition facilitates an efficient use of available data with respect to LR and CAGR values, while considering the technological maturity of the various components and maintaining the bottom-up nature of the proposed application of the technological learning framework. This is further explained in Step 2.

Step 2: For each time point t , defined as one year starting from 2020, the specific investment of each technological component i , $C(Q_t,i)$, is calculated from the analysis of cost reduction based on the learning curve theory (Eq. 2). Regarding the selection of LR values, the ideal case would be that a process step contains a combination of identifiable

mature and less mature technological components and the LR of all component are known but this is not the case.

For this reason, the next two rules are followed. In case of missing LR values for these technological components, mature technologies (e.g., conventional steps) are assigned by default a low LR value of 0.05 and less mature steps a higher value of 0.15 (e.g. in the cases of LR₁, LR₂, LR₃ of Process 1 in lower box of Figure 2).

Table 1. Input values to LC method for methanol for baseline scenario (Zhu, et al., 2011)

Process steps	Technological components	CAPEX (Euro 2018)	Learning rate (LR)	Cumulative installed capacity (CIC), (MW)	Cumulative annual growth rate (CAGR)
Gasification step	Air separation unit	0.0	0.05	200	11%
	Feed prep and drying	33.0	0.05		
	Gasification with tar reforming, heat recovery, scrubbing	44.9	0.15		
	Syngas cleanup & compression	84.7	0.15		
Methanol synthesis	Methanol synthesis & purification	37.2	0.05	57040	7%
	Steam system and power generation	27.7	0.05		
	Remainder off-site battery limits	6.5	0.05		
	Total CAPEX	234.0			

If, however, LR values for the entire process are available (e.g., Detz et al., (2019)) and LR values for mature and less mature technologies in this process step are not available, then the LR value for the entire process step is used and not the default LR values mentioned above. For example, for the case of biomass to methanol, the first rule of addressing LR values to mature and less mature components was applied as shown in Table 1; the gasifier operation in the gasification process step was considered as less mature and all the other components as mature technologies in conventional systems. Cumulative installed capacities (CIC) are defined by considering information for demo or pilot plants in operation in Europe or globally or values mentioned in simulation reports as representative for FOAK plants. For example, CIC of 200 MW is assumed for the gasification step, while for the methanol synthesis step the global production capacity for methanol in the year 2018 is used regardless of its production as fossil or renewable (M. Alvarado, IHS Chem. Week, 2016, 10–11). CAGR values are generally lacking due to very limited and relatively very recent commercial plants in operation. CAGR values are defined based on scenarios depending on policy targets and estimations of the market evolution of the corresponding conventional fuels. The growth rate of advanced biofuels is subject to many uncertainties depending on current conditions for growth rates of fossil-based demand and their future role in the transportation mix. In this work, three scenarios are formulated:

The *Baseline Scenario* which assumes CAGR values equal to the growth rate of the corresponding market of the fuel, which is a conservative scenario not leading to “greening” of the transportation mix. *Scenario A* assumes marginally higher CAGR values than the growth rate of the corresponding market of the fuel. Thus, it does not lead to a significant share of the market in short- to mid-term and the corresponding “greening” achieved is not enough to satisfy environmental targets for the time horizon (i.e., 2030-2050). *Scenario B* assumes an annual growth that is considerably bigger than the growth rate of the corresponding market of the fuel to an extent that it can satisfy targeted shares of the market in the considered time horizon. *Scenario B* estimates the CAGR of the biobased fuel to achieve 20% of the production of the respective fossil-based fuel. This is in agreement with the scenarios of European Commission 2018 that refer to 13%-24% contribution of liquid biofuels in the energy consumption for transportation in 2050 (EUROPEAN COMMISSION, 2018).

Step 3: By adding the values of all process components, the total specific investment cost of the process step is calculated and consecutively of the whole technological pathway ($CAPEX(t)$) at the time point t and therefore also at $CIC_i(t)$ (or Qt_i according to Eq.2).

3. Results

CAPEX reduction for the biomass to methanol pathway through indirect gasification, according to the *Baseline Scenario* is presented in Figure 3, showing the extent to which the components with the greater LR values present higher cost reduction rate. With respect to all advanced fuels analysed in this study based on Scenario A, Figure 4 shows that CAPEX reductions from 10-25% can be expected assuming only a handful of plants installed.

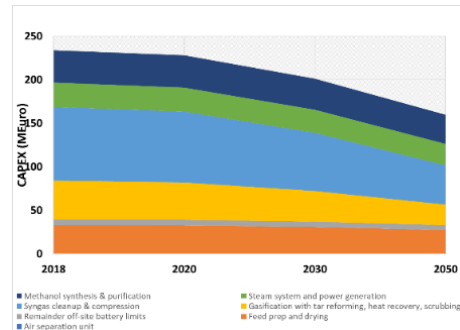


Figure 3. CAPEX reduction for methanol synthesis

If higher cumulative installed capacities are reached in 2050 (Scenario B, Figure 5), meeting the goal of 20-25% of transportation fuels consumption (to be covered by advanced biofuels), CAPEX reduction up to 40-50% can be expected for the new plants that will be built then. In this regard, it is important to remember that not all pathways reach similar CIC values in 2050, and thus the results should not be interpreted as a difference arising from the “potential to learn” or the status of the current maturity of the pathway. As an example, methanol and methane production via biomass gasification are both of similar technical maturity, the gasifier technology being the process component with higher technical learning potential. The reason that, in Figure 5, CAPEX of liquefied methane production appears to decrease faster than the methanol one is the very different CAGR values leading to CICs of 390 GW for liquefied gas to 53 GW of methanol. On the other hand, in Scenario A (Figure 4), where the CAGR values of methanol are higher than those of biogas, the opposite trend appears with respect to CAPEX decrease potential.

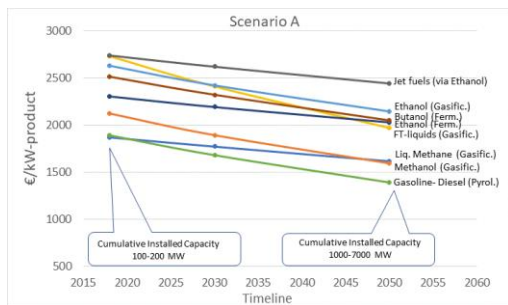


Figure 4 Scope of CAPEX reduction (2020-2050) of advanced biofuels from lignocellulosic biomass for a scenario of capacity annual growth rate marginally higher than the current fuel demands.

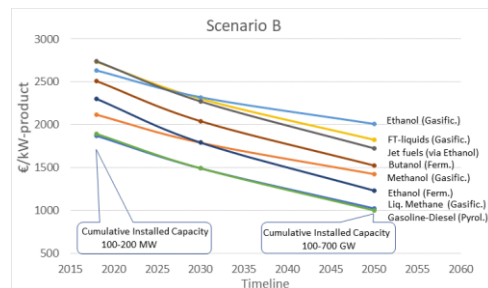


Figure 5 Scope of CAPEX reduction (2020-2050) of advanced biofuels from lignocellulosic biomass for a scenario of capacity annual growth rate meeting the goal of 20-25% transportation fuels consumption to be covered by advanced biofuels in 2050.

4. Conclusions

Notwithstanding the large uncertainties in the cost data, the learning curve theory can be useful to structure cost data, as a basis for estimating reductions as a function of installed

capacity. In the case of advanced biofuels, most of the technical learning is allocated in assembling the plants which consist, to a large extent, from mature technological components. It is, however, clear that substantial investment in large-scale plants is required in order to achieve CAPEX reduction of 40-50%, which together with favourable policies (e.g., access to capital, public/private partnerships, etc.) and feedstock cost factors, may increase the competitiveness of advanced biofuels. Two important methodological aspects refer to the degree of available information allowing or not a detailed decomposition of the pathway to process components, as well as the technical learning potential of these components and especially after which time point (or installed capacity) this can be assumed to attain near zero values. Disruptions (e.g. unforeseeable external events that may affect market conditions) have not been included in the modelling procedure. The method can become even more sophisticated by considering gradual or step changes in the learning parameters when a status of NOAK plant is reached to impose asymptotic behaviour faster and by incorporating sensitivity analysis for parameters of high uncertainty (e.g. LR and CAGR). It should be noted that technologies of advanced biofuels are only available in limited numbers. As a result, learning curve theory estimations cannot be based on a sufficient sample of similar systems. Secondly, there are very few full-scale plants available for advanced biofuel production on a commercial basis. This introduces significant uncertainties both in short- and long-term cost estimations of such plants. Additionally, if large facilities for fuel production are implemented, they will most likely affect the price of biomass leading to higher operating costs (or other market factors related with the role of biomass in the future energy system).

References

- EUROPEAN COMMISSION, 2018, IN-DEPTH ANALYSIS IN SUPPORT OF THE COMMISSION COMMUNICATION COM(2018) 773, A Clean Planet for all, A European long-term strategic vision for a prosperous, modern, competitive and climate neutral economy, Brussels
- Detz, R.J., J.N.H. Reek, and B.C.C. van der Zwaan. 2018. "The future of solar fuels: when could they become competitive?" *Energy Environ. Sci.*
- Pacific Northwest National Laboratory, (2011) Techno-economic Analysis for the Thermochemical Conversion of Biomass to Liquid Fuels (Indirect gasification)
- Feroli, F., Schoots, K and van der Zwaan; B.C.C. (2009): Use and Limitations of Learning Curves for Energy Technology Policy: a Component-Learning Hypothesis. *Energy Policy*,37, 2009, 2525-2535.
- Trappey AJC, Trappey CV, Tan H, Liu PHY, Li S-J, Lin L-C. The determinants of photovoltaic system costs: an evaluation using a hierarchical learning curve model. *J Clean Prod* 2016;112(Part2):1709–16. <http://dx.doi.org/10.1016/j.jclepro.2015.08.095>.
- Nemet Gregory F., Beyond the learning curve: factors influencing cost reductions in photovoltaics, *Energy Policy* 34 (2006) 3218–3232
- Upstill G, Hall P., Estimating the learning rate of a technology with multiple variants: The case of carbon storage, *Energy Policy* 121 (2018) 498–505
- Alvarado, M. . 2016. "IHS Chem. Week." 10–11.
- Junginger, Martin, Erika de Visser, Kurt Hjort-Gregersen, Joris Koornneef, Rob Raven, André Faaij, and Wim Turkenburg. 2006. "Technological learning in bioenergy systems." *Energy Policy*
- Mauleón I. Photovoltaic learning rate estimation: issues and implications. *Renew Sustain Energy Rev* 2016;65:507–24. <http://dx.doi.org/10.1016/j.rser.2016.06.070>.
- Junginger M, Faaij A, Turkenburg WC. Global experience curves for wind farms. *Energy Policy* 2005;33:133–50. [http://dx.doi.org/10.1016/S0301-4215\(03](http://dx.doi.org/10.1016/S0301-4215(03)

Simulation of the Biofuel Production Process from Organic Fraction of Municipal Solid Waste (OFMSW)

José Rocha-Rios,^a Arturo Pérez-Roman,^{a,b} Teresa Lopez-Arenas,^c Mauricio Sales-Cruz^c

^a*Centro de Alta Dirección en Ingeniería y Tecnología, Universidad Anáhuac México, Av. Universidad Anáhuac 46, 52786 Estado de Mexico, Mexico*

^b*Departamento de Ingeniería Agroindustrial, Universidad Autónoma Chapingo, km. 38.5 Carretera Mexico – Texcoco Chapingo, 56230 Estado de Mexico, Mexico*

^c*Departamento de Procesos y Tecnología, Universidad Autónoma Metropolitana – Cuajimalpa, Av. Vasco de Quiroga 4871, 05348 Mexico City, Mexico*

Abstract

The primary management of the organic fraction of municipal solid waste (OFMSW) was commonly composting, landfill or incineration, while now the anaerobic digestion of OFMSW offers the possibility of a clean and renewable energy source. This work presents the modelling and simulation of an OFMSW biorefinery to produce biogas, fertiliser and bioethanol. The conceptual design of the biorefinery begins with an anaerobic digestion to degrade OFMSW and produce biogas. The resulting solid fibres and the effluent are sent to the subsequent stages to produce fertiliser from the solid waste and anhydrous bioethanol from the fermentation broth. The process simulations are based on an anaerobic biorefinery capable to treat 1,000 tonnes per day of OFMSW generated in CDMX. The main results show the yields are 73 kg ethanol, 54 kg biogas and 59 kg fertiliser per tonne of dry waste.

Keywords: biorefinery, biofuel production, process simulation, municipal solid waste.

1. Introduction

The production and management of municipal solid waste (MSW) is a critical global problem. In general, the organic fraction of municipal solid waste (OFMSW) is a common name for heterogeneous waste mixtures from residential, commercial, partly industrial and urban areas. It is made up of different organic and inorganic fractions such as food, vegetables, paper and wood. Despite the variability in its composition, the OFMSW constitutes the highest percentage of the solid waste which can be broken down into simpler compounds by anaerobic microorganism (Bilgili et al., 2009). In the past decades, OFMSW's primary management was generally composting, landfill or incineration. Recently, OFMSW has become a promising candidate for energy recovery through the production of biofuels, such as biogas, bioethanol, biohydrogen, biodiesel, etc. (Campuzano and Gonzalez-Martinez, 2016). Despite the fact that several reviews have been published both to determine the composition (Campuzano and Gonzalez-Martinez, 2016) as well as the challenges and opportunities of biorefineries (i.e., Sawatdeenarunat et al., 2016; Dahiya et al., 2018), very few studies report aspects of modelling, simulation and design of these biorefineries.

Treatment processes for an integral OFMSW harnessing can be classified into three groups (Nizami et al, 2017): (1) thermochemical (e.g. pyrolysis or gasification), (2) physicochemical (e.g. hydrolysis or transesterification), and (3) biochemical (e.g. anaerobic digestion or fermentation). Each treatment can generate different products of added value and has advantages or disadvantages depending on the raw material and conditions of the place or country (Tsakalova et al., 2014). In particular, anaerobic digestion is considered the most technologically feasible method due to its many advantages, as it not only serves to efficiently manage and treat a huge amount of OFMSW, but also acts as a convenient source of unconventional energy (Wang et al., 2017).

Currently Mexico City has several transfer stations, classification plants, compactor plants, composting plants and landfills for the management of OFMSW. Thus, in this work an alternative to the management of OFMSW is proposed, through the design of a biorefinery to process 1,000 tons of OFMSW per day generated in Mexico City. The products generated from the OFMSW-biorefinery are biogas, fertiliser and bioethanol. Modelling and simulation tools are used for the design of the plant. The biorefinery begins with an anaerobic digestion where biogas is produced, and then several processing stages are followed to produce anhydrous ethanol as the main product and fertiliser as a fermentation by-product.

2. Methodology

2.1. Biorefinery conceptual design

The conceptual design of the biorefinery consists on seven stages, as shown in Figure 1. The biorefinery begins with an anaerobic digestion to degrade OFMSW and produce biogas. The resulting solid fibres and the effluent are sent to the subsequent stages of pretreatment by acid hydrolysis, saccharification by enzymatic hydrolysis, co-fermentation of sugars (pentoses and hexoses) to ethanol, evaporation and distillation to concentrate the ethanol solution, and finally ethanol dehydration by adsorption with molecular sieves. Fertiliser is obtained as a by-product of the sugar fermentation.

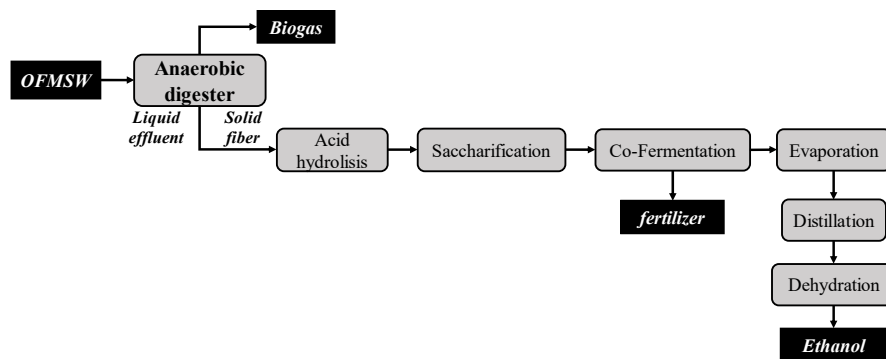


Figure 1. Conceptual design of the *OFMSW*-biorefinery

2.2. Implementation of the simulation models

The process flow diagram of the biorefinery was implemented in Aspen Plus® V.10 (see Figure 2), considering rigorous kinetics for the anaerobic reactor (Rajendran et al., 2014) and stoichiometric reactions for the hydrolysis, saccharification and fermentation reactors

(Aden et al., 2002). The thermodynamic properties of pure components and mixtures were estimated using the database and included methods, and the NRTL (Non-Random Two Liquid) model was employed as thermodynamic method.

Anaerobic digestion process of OFMSW was simulated using a model previously developed by Rajendran et al. (2014), which is based on the Anaerobic Digestion Model 1 (ADM1) and was validated against a variety of lab and industrial data. A total of 46 reactions were used in this model including inhibitions, kinetic rates, ammonia production, reactor volume, loading rate, and retention time. The biodigester was modelled in Aspen Plus as an RSTOIC block model (D-0), where hydrolysis occurs at 55 °C and 1 atm; together with a CSTR block model (D-1), where acidogenic, acetogenic and methanogenic reactions are carried out at 55 °C and 1 atm.

The pretreatment stage was simulated using two stoichiometric reactors (HIDROLI1, HIDROLI2), both operated at the same pressure and temperature conditions (190°C and no vapour fraction). The first reactor converts most of the hemicellulose into shorter chain polysaccharides through hydrolysis reactions (with 20% conversion), while the second reactor converts these polysaccharides into soluble sugars, mainly xylose, mannose, glucose, galactose and arabinose (with 92.5% conversion). Hydrolysis reactions also produce soluble lignin, acetic acid, and degradation products (mainly furfural and hydroxymethylfurfural).

The saccharification process was simulated in a stoichiometric reactor (SACHAR), where enzymatic hydrolysis reactions are carried out at 1 atm and 65 °C. The main reaction corresponds to the conversion of cellulose to glucose (with 90% conversion). And the co-fermentation process was simulated in a stoichiometric reactor adding *Zymomonas mobilis* as strain for the cultivation, where the conversion of both pentoses and hexoses are converted into ethanol (with 97% conversion), at 1 atm and 41 °C.

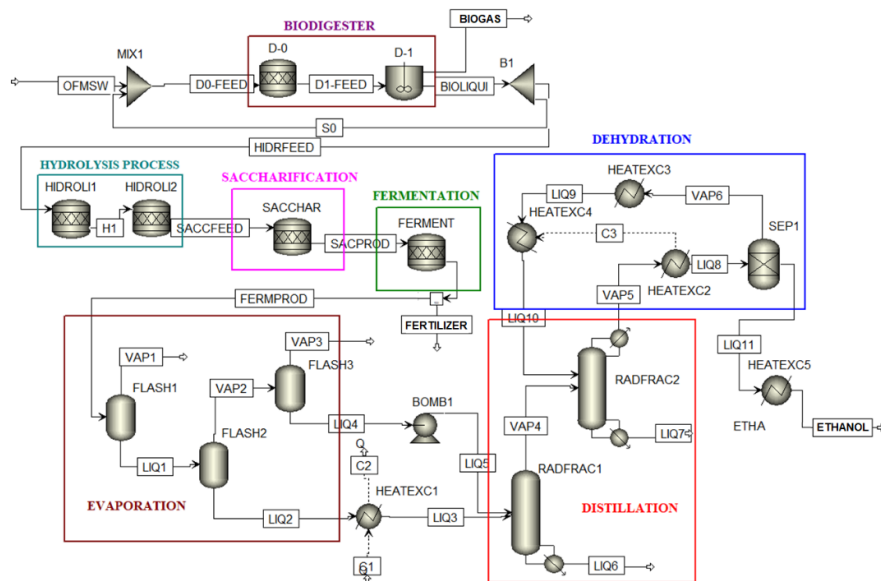


Figure 2. Process flow diagram for OFMSW-biorefinery

The outlet stream from the fermenter was treated by an evaporation section to remove considerable amounts of components such as CO₂, NH₃, H₂S, O₂ and H₂O, among others. Subsequently, two distillation columns (RADFRAC1, RADFRAC2) were simulated to concentrate the alcoholic broth near the azeotropic point. The theoretical number of equilibrium stages for the first column was 9, without condenser and with a molar reflux ratio in the reboiler of 0.23, the pressures of stages 1 and 9 were 1.53 and 1.67 atm, respectively. The theoretical equilibrium stages for the second rectifying column were 18, with molar reflux ratios of 2.3 and 0.23 for condenser and reboiler, respectively; the pressures of stages 1 and 18 were 1.36 and 1.63 atm, respectively. And finally, the distillate stream was sent to the dehydration process by adsorption with molecular sieves. This dehydration technology uses a zeolite bed in which the water molecules are strongly retained within the pores, while the ethanol molecules pass through the bed. The process consists of combined cycles of dehydration of ethanol and regeneration of the molecular sieve. The main stream reaches an ethanol purity of 99.5%, which is cooled and stored

3. Simulation results

For simulation purposes, the biorefinery was designed to process 1,000 t/d of wet waste, assuming an average OFMSW composition of Mexico City presented in Table 1, which was adapted from Campuzano and González-Martínez (2016). The composition and properties of OFMSW were implemented in Aspen Plus® using its extensive database of components and property estimation.

Table 1. Components and mass fraction of OFMSW.

Component	Molecule	Chemical formula	Mass fraction
Carbohydrates	Cellulose	C ₆ H ₁₀ O ₅	0.0471
	Hemicellulose	C ₅ H ₈ O ₄	0.0114
	Dextrose	C ₆ H ₁₂ O ₆	0.0234
	Starch	C ₆ H ₁₂ O ₆	0.0381
Protein	Protein	C _{4.13} H ₂₅ O ₇ N ₃ S	0.0212
	Keratin	C _{4.39} H ₈ O _{2.1} N	0.0127
Lipids	Triolein	C ₅₇ H ₁₀₄ O ₆	0.0098
	Tripalmitate	C ₅₁ H ₉₈ O ₆	0.0098
	SN-1-Palmito-2-Olein	C ₃₇ H ₇₀ O ₅	0.0098
	SN-1-Palmito-2-Linolein	C ₃₇ H ₆₈ O ₅	0.0098
Lignin	Lignin	Inert	0.0300
Ashes	Ashes	Inert	0.0739
Humidity	Water	H ₂ O	0.7030

Table 2 shows the results for the main streams of each stage of the process, where it can be seen how OFMSW is degrading as it passes through each stage of the process. The biogas yield was 0.538 kg/kg of dry waste. Being the most common use of biogas for heat

and electricity generation. While the fertiliser yield was 0.059 kg/kg of dry waste. The use of biofertilisers improves the fertilization of crops compared to chemical fertilisers and does not compromise the yield or quality of the crops (Dahiya et al., 2018). And finally, the ethanol produced was 0.073 kg/kg of dry waste. This is low compared to the reported values of 1.15-1.5 kg ethanol/kg dry waste (Sawatdeenarunat et al., 2016). The main reason is due to the low carbohydrate composition of the OFMSW obtained for Mexico City (Table 1). It is worth mentioning that the anhydrous ethanol is a product with high added value, which can be used mixed with gasoline as a fuel for transportation. Although the product yields obtained in this work are still low compared to those obtained with other lignocellulosic substrates, the complex nature of this substrate must be taken into account. Up to now, the simulation was focused on the conceptual design, but the next step will be to improve the operating conditions of each stage.

Table 2. Main streams for the OFMSW-biorefinery.

Stream	Stream Name	Flow (t/d)	Composition
Biodigester feed	DO-FEED	1,000	Given in Table 1
Hydrolysis feed	HIDRFEED	836.81	42.4% CH ₄ , 43.6% CO ₂ , 12.4% H ₂ O, 0.8 % Ethanol, 0.7 % H ₂ S, 0.1% others
Saccharification feed	SACCFEED	836.82	78.1% H ₂ O, 12.4% inert, 2% C ₅ H ₇ NO ₂ , 1.4% Ethanol, 1.1% Cellulose, 0.9% Xylose, 0.9 % Starch, 0.5% CH ₃ COOH, 2.6% others
Fermentation feed	SACPROD	830.35	78.5% H ₂ O, 12.5% inert, 2% C ₅ H ₇ NO ₂ , 2.0% Ethanol, 1.3% Glucose, 1.0% Xylose, 0.5 % CH ₃ COOH, 2.8% others
Evaporation feed	FERMPROD	830.78	78.5% H ₂ O, 12.5% inert, 2.7% Ethanol, 2% C ₅ H ₇ NO ₂ , 1.2% CO ₂ , 0.6% CH ₃ COOH, 2.6% others
Distillation feed	LIQ3	816.87	79.3% H ₂ O, 12.7% inert, 2.7% Ethanol, 2% C ₅ H ₇ NO ₂ , 0.6% CH ₃ COOH, 2.7% others
Distillation feed	LIQ5	0.30	90.9% H ₂ O, 8.5% Ethanol, 0.3% CH ₃ COOH, 0.3% others
Dehydration feed	LIQ8	102.98	69.2% H ₂ O, 29.3% Ethanol, 0.4% CH ₃ COOH, 1.1% others
Biogas	BIOGAS	159.77	42.4% CH ₄ , 43.6% CO ₂ , 12.4% H ₂ O, 0.8 % Ethanol, 0.7 % H ₂ S, 0.1% others
Fertiliser	FERTILIZER	17.59	93.5% <i>Z. mobilis</i> , 5.0% H ₂ O, 1.5% others
Ethanol	ETHANOL	21.55	99.95% H ₂ O, 0.05% Ethanol

4. Conclusions

The conceptual design and simulation of a biorefinery to convert OFMSW into value-added products have been presented as decision support for the management of OFMSW in Mexico City. The results show that its reuse is feasible for the production of biogas, bioethanol and fertiliser. An important problem for obtaining OFMSW bioproducts is their heterogeneity and variability in chemical composition, such that they could modify the results obtained. For example, when food waste is the major compound, biogas production would increase; while when lignocellulosic compounds predominate, the production of bioethanol and biofertilisers would increase.

We are currently working on the economic evaluation, as well as on the extension of this conceptual design proposal to produce other value-added products, such as the production of biodiesel using the liquid effluent from anaerobic digestion (with algae that can also reduce CO₂), lignin, recovery of solid waste generated in fermentation, and possible production of electricity by combustion of solid waste.

5. Acknowledgments

This work was supported by the Chapingo Autonomous University [institutional sponsorship agreement for teaching staff carrying out postgraduate studies], and the Anahuac University Mexico [CADIT industrial engineering grant] for Arturo Pérez-Roman, as well as the Project Fund for José Rocha-Ríos

References

- A. Aden, M. Ruth, K. Ibsen, J. Jechura, K. Neeves, J. Sheehan, B. Wallace, L. Montague, A. Slayton, J. Lukas, 2002, Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis for Corn Stover, NREL, Golden, CO, NREL/TP-510-32438. 154.
- M.S. Bilgili, A. Demir, G. Varank, 2009, Evaluation and modeling of biochemical methane potential (BMP) of landfilled solid waste: a pilot scale study. *Bioresource Technology* 100, 4976-4980.
- R. Campuzano, S. González-Martínez, 2016, Characteristics of the organic fraction of municipal solid waste and methane production: A review, *Waste Management*, 54, 3-12.
- S. Dahiya, A.N. Kumar, J.S. Sravan, S. Chatterjee, O. Sarkar, S.V. Mohan, 2018, Food waste biorefinery: Sustainable strategy for circular bioeconomy, *Bioresource Technology*, 248, 2-12.
- A.S. Nizami, M. Rehan, M. Waqas, M. Naqvi, O.K.M Ouda, K. Shahzad, R. Miandad, M.Z. Khan, M. Syamsiro, I.M.I. Ismail, Deepak Pant, 2017, Waste biorefineries: Enabling circular economies in developing countries, *Bioresource Technology*, 241, 1101-1117,
- K. Rajendran, H. R. Kankanala, M. Lundin, M. J. Taherzadeh, 2014, A novel process simulation model (PSM) for anaerobic digestion using Aspen Plus, *Bioresource Technology*, 168, 7-13.
- C. Sawatdeenarunat, D. Nguyen, K.C. Surendra, S. Shrestha, K. Rajendran, H. Oechsner, L. Xie, S.K. Khanal, 2016, Anaerobic biorefinery: Status, challenges, and opportunities, *Bioresource Technology*, 215, 304-313.
- M. Tsakalova, A. Yang, A.C. Kokossis, 2014, A Systems Approach for the Holistic Screening of Second Generation Biorefinery Paths for Energy and Bio-based Products, *Computer Aided Chemical Engineering*, 33, 205-210,
- X. Wang, M. Guo, K.H. van Dam, R.H.E.M. Koppelaar, C. Triantafyllidis, N. Shah, 2017, Waste-Energy-Water systems in sustainable city development using the resilience.io platform, *Computer Aided Chemical Engineering*, 40, 2377-2382.

Comparison of different processing routes for the valorisation of olive tree pruning wastes

Luis David Servián Rivas^a, Ismael Diaz^a, Manuel Rodriguez^a, María González-Miquel^a, Emilio J. González^a

^aDepartamento de Ingeniería Química Industrial y del Medioambiente, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, C/ José Gutiérrez Abascal 2, 28006 Madrid, Spain

luisdavid.servian.rivas@upm.es

Abstract

In this work, two different bioprocess plants are economically compared using Aspen Plus process simulations based on information from literature. Olive tree pruning (OTP) wastes are used as the raw material in the processes which are valorised through two different processing plants. One biorefinery scheme consists of an initial steam explosion pre-treatment to further obtain three final products: ethanol, xylitol and antioxidants (steam explosion plant = SEP). The other one consists of an organosolv (ethanol and water) pre-treatment operation to finally ferment glucose to ethanol, co-producing catechol from lignin depolymerization (organosolv plant = OP). Both plants show similar equipment costs (43-47 M€). However, the first plant (SEP) manufacture three marketable products with lower ethanol production (6364.35 m³/yr) than OP (8118 m³/yr). In comparison, the production of xylitol (999.36 tonne/yr) and antioxidants (1214.46 tonne/yr) as co-products in the SEP are much higher than the catechol production (124.05 tonne/year) in the OP. As a result, much of the organosolv resulting lignin is burned to produce energy. Considering all costs and incomes in both plants, the SEP would attain a profitability of 20%, while the OP would be -7%.

Keywords: Lignocellulosic biorefinery, Organosolv, Steam explosion, Technoeconomic analysis.

1. Introduction

Currently most of the fuels obtained from biomass sources in a commercial scale come from corn, sugarcane, wheat and other cereals, which means that most of the biofuels are obtained in first generation plants. On the contrary, second generation bio-products obtained from lignocellulosic biomass do not compete in the human food market, being one of the reasons why they have attracted many attention in the last decade (Haghighi et al., 2013).

In Spain, many works at different scales are being developed using olive tree pruning (OTP) wastes to obtain marketable products. That biomass is highly available (around 4.5 Mtonnes of OTP are generated each year), being burned in most of the cases to obtain thermal energy (Manzanares et al., 2017). However, different high-value products can be also synthesized from OTP wastes. For instance, antioxidants from olives are highly

demanding bio-products by society in food and cosmetic applications. They have been successfully extracted (Conde et al., 2009) from OTP wastes, as well as xylose which can be used as raw material for the production of both xylitol (Rivas et al., 2006) and ethanol. Obtaining these products usually requires an initial steam explosion or organosolv pre-treatments to separate lignin from fermentable compounds.

On the other hand, lignin is usually burned to obtain energy. Some authors (Toledano et al., 2012) showed that depolymerization of OTP lignin is feasible and some valuable compounds can be produced (p.e. catechol). Organosolv pre-treatment has shown the best performance to obtain exploitable lignin.

In this work, two biorefineries schemes are compared, both using OTP waste as raw material. One of the considered processes uses steam explosion as the pre-treatment operation, producing antioxidants, ethanol and xylitol, (lignin along with the rest of the organic waste will be burned to obtain thermal energy). The other plant uses organosolv in the pre-treatment step, producing ethanol and lignin which is further depolymerized to produce catechol. Aspen Plus software is used to model and simulate both plants and a techno-economic evaluation is carried out as shown in section 3.

2. Materials and Methods.

2.1. Feedstock

The input in both plants are 96.96 tonnes of OTP wastes per day. The composition (% dw) is the same described by Ballesteros et al. (2011): 28.0 cellulose (as cellulose); 20.6 hemicellulose (as xylan); 25.2 lignin; 2.1 acetic groups; 5.9 Ash (as CaO); and extractives with is a mixture of: 7.9 glucose, 0.1 arabinose, 0.1 mannose, 0.9 galactose, 4 mannitol and 3.8 antioxidant (hydroxytyrosol as the reference compound) (Conde et al. 2009).

2.2. Process description of plant A

In plant A, the pre-treatment has two steps; firstly, the OTP is sent to a solid-liquid extraction using water at 393.15 K and 5 bar. The water stream extracts 90% of extractives (Ballesteros et al., 2011). The remaining solid fraction is sent to a steam explosion process at 468.29 K and 14 bar with phosphoric acid (1%) as catalyst (Negro et al., 2014). Then a water insoluble solid fraction (WIS) and a liquid fraction with xylose and the degradation compounds as principal components (LiX), are obtained.

The liquid fraction with the extractives is cooled down to 308.15 K and streamed to a liquid-liquid extraction column with ethyl acetate (EtA) at 303 K, where 90% of the antioxidants are extracted by EtA. Antioxidants are then separated from EtA by a non-ionic adsorption process (Romero-García et al (2016)). All EtA is separated from water and other organic components by distillation and recirculated to the process, recovering 98% of EtA. The LiX is cooled down at 323 K and mixed with $\text{Ca}(\text{OH})_2$ to remove furans and phenolic compounds. Then, it is neutralized with H_2SO_4 . The detoxified LiX is then fermented at 303 K (75% of xylose is fermented to xylitol). The product is filtered to remove the yeast and other solids and evaporated to concentrate the xylitol at 313 K (this temperature avoids xylitol degradation). Xylitol is mixed with ethanol to reduce xylitol solubility and then purified by crystallization at 268 K (yield of 47 % with a purity of 99% (Rivas et al., 2006). The WIS stream is sent to a pre-saccharification and simultaneous saccharification and fermentation process (PSSF) for the production of bio-ethanol. Pre-saccharification is carried out at 323 K. Then the resulting stream is further cooled to 308 K and fermented to ethanol (70% of theoretical yield is obtained) (Ballesteros et al., 2011). The resulting ethanol stream is firstly purified in a beer column which

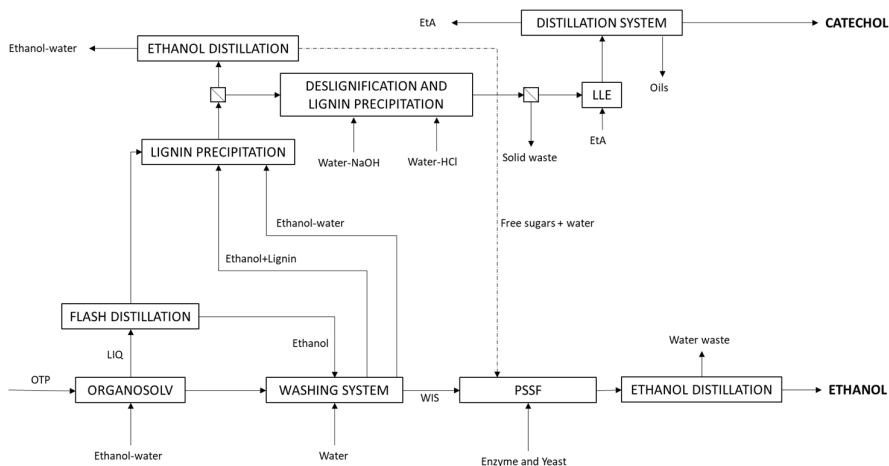


Figure 2. Block diagram of the modeled plant B

2.4. Process modelling and economic evaluation.

Both processes are simulated in Aspen Plus V11 software (Aspen Technology Inc, USA). Physical property data of cellulose, hemicelluloses, lignin and enzymes was taken from literature (Wooley, 1996). The non-random two liquid (NRTL) thermodynamic model was used in the process. Antioxidant are modelled as hydroxytyrosol, whose properties are calculated using the built-in property estimation models (group contribution approach from molecular structure). Missing binary interaction parameter are estimated using UNIFAC. To economically compare both plants, the Discounted Cash Flow Rate of Return (DCFRROR) and Net Present Value (NPV) are calculated for a 15-y period with a cost of capital of 15%. These parameters provide a useful way of comparing the performance of capital for different projects. Aspen Economic Analyzer V11 (Aspen Technologies Inc, USA) is used to calculate equipment cost of both plants. Raw materials and products prices are obtained from literature (Susmozas et al. 2019) and utilities consumptions are calculated following the methodology available in (Sinnot & Towler 2008).

3. Results and Discussion.

Table 1 shows the distribution of equipment cost, revenues and the major contributions of the operating costs. Plant B revenues do not compensate the estimated investment so the NPV is negative. In this plant, the depolymerization process accounts for the 18% of the equipment cost, not compensating by the revenues obtained from the production of catechol. The yield of catechol is quite low, just representing 0.31% of all products obtained on the delignification process, so most of the products obtained from delignification are burned to obtain thermal energy. Even more, organosolv is more expensive than steam explosion, because it requires more equipment to recover as much ethanol as possible.

On the other hand, Plant A is estimated to be profitable, with a calculated DCFRROR of 20% and NPV at year 15 of 11.54 M€. In both plants cooling water represents the highest utility consumption. That makes sense considering that both plants pre-treats OTP wastes at high temperature and the subsequent enzymatic hydrolysis takes place below 323 K.

Furthermore, the biological reactors operate at constant temperature so cooling water is also needed.

Table 1. Equipment cost, revenues, variable cost distribution and DCFROR

	Plant A	Plant B	
ISBL	51.95%	51.95%	ISBL
OSBL	20.78%	20.78%	OSBL
Engineering Cost	7.27%	7.27%	Engineering Cost
Contingency	10.91%	10.91%	Contingency
Total Fixed capital cost (€)	43.8 M€	47.9 M€	Total Fixed capital cost (€)
Equipment Cost			
Steam Explosion	9.83%	25.59%	Organosolv
Ethanol process	18.07%	16.99%	Ethanol process
Antioxidant process	12.66%	18.69%	Catechol process
Xylitol process	13.16%		
Water treatment	17.93%	18.46%	Water treatment
Power and heat generation	28.35%	20.27%	Power and heat generation
Total (€)	24.0 M€	24.9 M€	
Revenues			
Ethanol	16.27%	53.31%	Ethanol
Xylitol	13.96%	1.54%	Catechol
Antioxidant	58.86%		
Electricity	10.92%	45.14%	Electricity
Total (€/yr)	22.7 M€/yr	8.8 M€/yr	Total (€/yr)
Variable cost			
OTP	42.15%	30%	OTP
Process Water	23.29%	6.72%	Process Water
Cooling Water	11.70%	30.72%	Cooling Water
Ethyl Acetate	5.53%	23.55%	NaOH
Total (€/yr)	6.9 M€/yr	5.3 M€/yr	Total (€/yr)
DCFROR			
	20%	-4%	
NPV			
	11.5 M€	-25.4 M€	

4. Conclusion

In this work, we have compared two different OTP waste biorefinery schemes in terms of capital and operating costs, and potential revenues. Our results show that delignification is not yet a competitive route in comparison with the obtention of antioxidants and xylitol from OTP wastes. Current yields to catechol limit the potential of the delignification route (efficiency of the depolymerisation around 5%). Other high-value products are being developed from the organosolv resulting lignin, which would expand the range of marketable compounds, thus increasing the competitiveness of this route. On the other hand, high revenues have been estimated for the case of the steam explosion-based plant. In that biorefinery scheme, xylitol, ethanol and antioxidants are

simultaneously produced. Our results show that the most economically promising route is the production of antioxidants from OTP through a liquid-liquid extraction process with ethyl acetate.

References

- Ballesteros, I., Ballesteros, M., Cara, C., Sáez, F., Castro, E., Manzanares, P., Negro, M. J., & Oliva, J. M. (2011). Effect of water extraction on sugars recovery from steam exploded olive tree pruning. *Bioresource Technology*, *102*(11), 6611–6616. <https://doi.org/10.1016/j.biortech.2011.03.077>
- Conde, E., Cara, C., Moure, A., Ruiz, E., Castro, E., & Domínguez, H. (2009). Antioxidant activity of the phenolic compounds released by hydrothermal treatments of olive tree pruning. *Food Chemistry*, *114*(3), 806–812. <https://doi.org/10.1016/j.foodchem.2008.10.017>
- Haghighi, S., Hossein, A., & Tabatabaei, M. (2013). Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. *Renewable and Sustainable Energy Reviews*, *27*, 77–93. <https://doi.org/10.1016/j.rser.2013.06.033>
- Mabrouk, A., Erdocia, X., Alriols, M. G., & Labidi, J. (2018). Economic analysis of a biorefinery process for catechol production from lignin. *Journal of Cleaner Production*, *198*, 133–142. <https://doi.org/10.1016/j.jclepro.2018.06.294>
- Manzanares, P., Ruiz, E., Ballesteros, M., Negro, M. J., Gallego, F. J., López-Linares, J. C., & Castro, E. (2017). Residual biomass potential in olive tree cultivation and olive oil industry in Spain: Valorization proposal in a biorefinery context. *Spanish Journal of Agricultural Research*, *15*(3), 1–12. <https://doi.org/10.5424/sjar/2017153-10868>
- Negro, M. J., Alvarez, C., Ballesteros, I., Romero, I., Ballesteros, M., Castro, E., Manzanares, P., Moya, M., & Oliva, J. M. (2014). Ethanol production from glucose and xylose obtained from steam exploded water-extracted olive tree pruning using phosphoric acid as catalyst. *Bioresource Technology*, *153*, 101–107. <https://doi.org/10.1016/j.biortech.2013.11.079>
- Rivas, B., Torre, P., Domínguez, J. M., Converti, A., & Parajó, J. C. (2006). Purification of xylitol obtained by fermentation of corncob hydrolysates. *Journal of Agricultural and Food Chemistry*, *54*(12), 4430–4435. <https://doi.org/10.1021/jf053156x>
- Romero-García, J. M., Sanchez, A., Rendón-Acosta, G., Martínez-Patiño, J. C., Ruiz, E., Magaña, G., & Castro, E. (2016). An Olive Tree Pruning Biorefinery for Co-Producing High Value-Added Bioproducts and Biofuels: Economic and Energy Efficiency Analysis. *Bioenergy Research*, *9*(4), 1070–1086. <https://doi.org/10.1007/s12155-016-9786-3>
- Sinnot, R., & Towler, G. (2008). *Chemical Engineering Design: principles, practice and economics of plant and process design*.
- Susmozas, A., Moreno, A. D., Romero-García, J. M., Manzanares, P., & Ballesteros, M. (2019). Designing an olive tree pruning biorefinery for the production of bioethanol, xylitol and antioxidants: A techno-economic assessment. *Holzforchung*, *73*(1), 15–23. <https://doi.org/10.1515/hf-2018-0099>
- Toledano, A., Serrano, L., & Labidi, J. (2012). Process for olive tree pruning lignin revalorisation. *Chemical Engineering Journal*, *193–194*, 396–403. <https://doi.org/10.1016/j.cej.2012.04.068>
- Wooley, R. J. (1996). Physical Property Database for. *National Renewable Energy Laboratory*, 36.

The Use of GVL for Holistic Utilization of Biomass

Antreas Pateromichelakis, Melina Psycha, K. Pyrgakis, A. Kokossis*

National Technical University of Athens, Iroon Polytechniou 9, GR-15780, Greece

Abstract

This work presents solutions for the holistic utilization of lignocellulosic biomass based on the “lignin-first” concept. Biomass is fractionated by means of γ -Valerolactone (GVL) and Formaldehyde solvents to effectively extract and protect lignin and xylose ingredients. The process flowsheet and simulation of a large scale biomass fractionation technology – recently validated in laboratory scale by Shuai et al. (2016) – are developed to test and build performance in use of energy, water and materials. The biorefinery value chain additionally integrates chemistries for the production of platform chemicals and biofuels (furfural, levulinic acid and lignin-aromatics), while the C6 sugars fraction is partially converted into GVL to offset solvent losses. Finally, energy integration and techno-economic analysis resulted in up to 6% steam savings, 1.8 MW power cogeneration and 37 M€/yr net revenues for the overall biorefinery.

Keywords: γ -Valerolactone (GVL), Lignin-first, Biorefineries.

1. Introduction

Biomass fractionation is challenged by the effective recovery of cellulose, hemi-cellulose and lignin components. While organosolv technologies have been credited to extract xylan and glucan, processing and depolymerization of lignin structures are mostly inhibited by degradation of ether bonds and spontaneous repolymerization of C-C bonds resulting in random structural recombination of the polymer matrix. The limited recovery of functional lignin intermediates usually leads to misconceptions that the lignin-rich pulp is a process-waste or low value fuel. The technology proposed by Shuai et al. (2016) takes advantage of the “lignin-first” concept to stabilize and protect lignin from repolymerization during processing and extraction. The aim of this work is to face underutilization of the lignin content and test performance of large scale lignin-first biorefineries for the production of platform chemicals and high value biofuels.

2. Bottlenecks in lignin processing and extraction

Lignin-rich feedstocks and lignin aromatic content are underutilized due to processing bottlenecks during biomass fractionation stages, whose goal is to separate and transform lignin into functional units by cleavage of biopolymer bonds in between lignin polymer chains and intramolecular bonds within polymer chains. Biomass pre-treatment commonly takes place at relatively high temperature (>120 °C) and in strong acid or alkaline environments (>1% wt). Under these conditions, there is a featured effect related with cleavage of ether bonds (β -O-4, 4-O-5 etc.) in between monolignols and repolymerization by the formation of C-C bonds due to interactions between positively charged and electron-rich sites. This effect results in irreversible condensation of lignin and significantly reduced yields of aromatic monomers during subsequent processing.

The yield of desired lignin-aromatics is correlated with the ether bond content and is enabled by the use of catalysts. The use of unprotected lignin is responsible for the reduction of aromatic yields and the inhibition of catalytic activity due to the existence of C-C bonds in lignin structure.

The above processing bottlenecks have an apparent negative impact on the utilization of biomass lignin content and, thus, on process economics. This work exploits the “lignin-first” concept – recently validated in laboratory scale by Shuai et al. (2016) – as a practice to protect lignin and build high performance industrial scale biorefineries.

3. The “Lignin-first” concept

The implementation of lignin-first biomass fractionation aims to successfully extract lignin by inhibiting the formation of C-C bonds during pre-treatment and to protect lignin through the reaction system of Figure 1.a. The fractionation process utilizes a γ -Valerolactone (GVL)-water mixture as solvent, which (a) favors selectivities towards sugars, lignin and platform chemicals; (b) limits sugars and lignin degradation; (c) features limited solvent degradation in the selected process conditions; and (d) facilitates downstream separations due to properties (n.b.p. 219 °C) (Alonso et al., 2017; Questell-Santiago et al., 2018). In addition, this work examines the use of formaldehyde in biomass fractionation to further inhibit the formation of C-C bonds between monolignols by reacting with newly cleaved sites.

An additional effect of formaldehyde is the protection of xylose. Another major reaction that occurs during fractionation is the depolymerization of hemicellulose into xylose and its degradation into furfural and humins. The formation of furfural at this stage complicates downstream separations due to the formation of furfural-water azeotrope. Formaldehyde protects the xylose content by the formation of diformyl-xylose (reaction system of Figure 1.b), which is next converted into xylose by introducing an additional intermediate reaction stage before finally upgrading xylose into furfural end-product.

The protected (by GVL and formaldehyde) lignin is next hydrogenated in presence of Pd-Ru supported catalysts resulting in near theoretical lignin-aromatic yields of 50%, whereas corresponding yields based on unprotected lignin are typically expected in the range of 5-20%. Moreover, the use of GVL favors the formation of hydroxymethyls, which in turn yield in blends suitable for biojet and biodiesel use.

GVL has been selected as the primary solvent for the depolymerization of sugars due to its ability to deter degradation of biomass components; to effectively depolymerize hemicellulose into xylan; to inhibit humins formation; and to dissolve solid lignin. The latter, can be tuned by simple addition of water. The chemistry takes advantage of a GVL-water (80-20 wt%) mixture to dissolve biomass components, while lignin precipitation starts at higher water/GVL mass ratios (>1.5). The zeotropic GVL-water mixture also facilitates downstream processing and separations. Considerable improvements on the yields and quality of extracted lignin are attributed to the use of formaldehyde (3% w/w) as co-solvent to prevent lignin condensation, to hinder the formation of C-C linkages and to produce soluble lignin (Shuai et al., 2016). Based on the laboratory scale efficiencies of Shuai et al. (2016), process systems engineering methods are used to test and build performance of large scale lignin-first biorefineries.

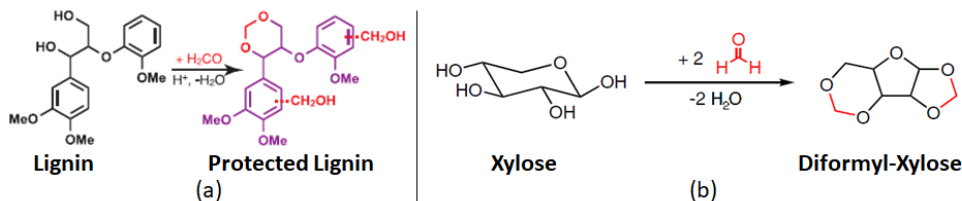


Figure 1. Lignin protection from GVL (a), xylose protection from formaldehyde (b)

4. Methodological steps

In this work, two alternative biorefinery scenarios have been developed for the production of xylose-based furfural, glucose-based levulinic acid and lignin-aromatics. The next objective is the evaluation of the effects of lignin exploitation on process economics with a purpose to identify most profitable solutions and set the basis of an expanded value chain. Both biorefinery cases (A & B) use GVL in biomass fractionation, while the second case (B) additionally uses formaldehyde as co-solvent.

The biorefinery plant (Figure 2) includes 5 Units: (U200) Biomass fractionation unit; (U300) Glucose-to-Levulinic acid unit; (U400) Xylose-to-Furfural unit; (U500) Lignin-to-aromatics unit; and (U600) Waste Water Treatment (WWT) unit. Energy integration and techno-economic analysis are next implemented both in contexts of each Unit and the overall biorefinery to estimate steam savings and extract conclusions regarding capital and operating expenses and profitability margins of each biorefinery case.

Mass and energy balances are calculated in the basis of 10 tn/h of white birch as feedstock, which features high lignin (18-25%) and ether (46-75%) bond contents. In Case A, each part of biomass requires 3.2-0.8-0.04 parts of GVL-water- H_2SO_4 in fractionation stage, while in Case B, 0.18 parts of formaldehyde are additionally used per part of biomass in feed. The process flowsheets (Figure 3) and mass and energy balances for each case (A/B) have been developed and calculated in Aspen plus (v8.6).

5. Results

5.1. Process flowsheeting and simulation

Figure 3 presents the process flowsheets of all biorefinery units (U200-U600 for Case A & B), which are interconnected according to the process diagram of Figure 2 and simulated as a single biorefinery plant in Aspen plus (v8.6). The simulation results are presented in Table 1 including key input/output mass balances for each unit and case.

The process flowsheets of biomass fractionation (U200) and Xylose upgrade (U400) are different for each Case A & B due to additional requirements in Case B, for the recovery of formaldehyde in U200B and the conversion of diformyl-xylose into xylose in U400B. The Units U300, U500 and U600 are the same for both Cases A and B.

Biomass fractionation (U200) is operated in a monophasic (liquid) reactor at 120 °C and 2 bar. GVL hydrolyses 81% of hemicellulose into xylose and dissolves 95% of lignin content; 93% of cellulose is recovered as solid pulp through a two-stage centrifugation system. The extracted lignin is next precipitated with the addition of water using a 6:1 water-GVL mixture. In Case B of U200, formaldehyde additionally reacts with the hydroxyl content of lignin driving the stabilization of dissolved lignin, while, next, formaldehyde is recovered through the distillation of U200B. The downstream units are developed for the conversion of intermediate sugars and lignin into platform chemicals.

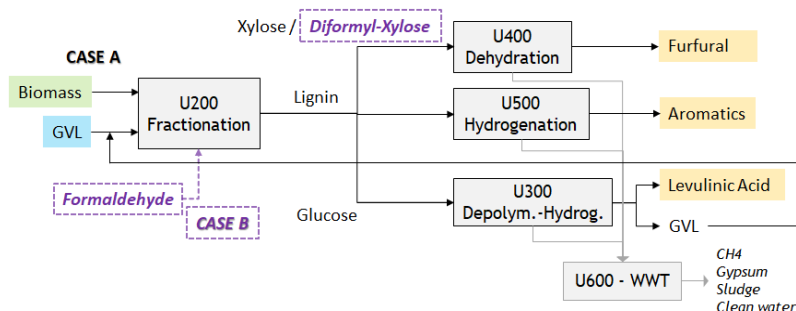


Figure 2. Lignin-first biorefinery diagram for Cases A and B.

In U400A, the extracted xylose-GVL mixture is concentrated through evaporation, while GVL is next recovered by means of distillation; in Case B (U400B), an intermediate reactor is introduced to transform diformyl-xylose into xylose, which is finally dehydrated into furfural by 90%. In U300, the cellulose-rich pulp is mixed with pure GVL, hydrochloric acid (HCl 37 wt%) and water forming an 80:20 wt% GVL-water mixture. Depolymerization of cellulose occurs at 150 °C and 3 bar, where cellulose depolymerizes into glucose, glucose isomerizes into fructose and fructose is hydrolyzed into 5-Hydroxymethylfurfural; the latter decomposes into levulinic acid and formic acid. The overall glucose to levulinic acid yield is 55%. The levulinic-formic acid mixture exiting the reactor is partially driven to a hydrogenation reactor for the conversion of levulinic acid into GVL, while the rest is separated by distillation resulting pure levulinic acid. U500 is responsible for the hydrogenation of lignin into aromatics in presence of THF solvent. In U500B, the treatment of protected lignin (by formaldehyde) yields in 10 parts of aromatics for each 100 parts of biomass – while in absence of formaldehyde (Case A), the yield in U500A is only 3 parts of aromatics – resulting in a significant impact on biorefinery economics. WWT is implemented for the recovery of contaminated process water and anaerobic digestion (AD) of organic residual content. U600 respectively achieves 97% and 90% water recovery in Case A & B, while methane production by AD is respectively estimated at 193 and 247 kg/hr.

5.2. Energy and economic analysis results

Total Site integration (Pyrgakis and Kokossis, 2019) is implemented to estimate utilities, steam savings and cogeneration targets for the overall lignin-first biorefinery. The Grand Composite Curves of all units are developed and combined resulting in the Site Sources and Sinks Profiles of Figure 4 for each Case A & B. The energy results of Total Site integration analysis are summarized in Table 2.

Figure 5 presents the breakdown of materials expenses for each case highlighting the need for less expensive biomass types. Concerning Case A, cheaper solvent alternatives

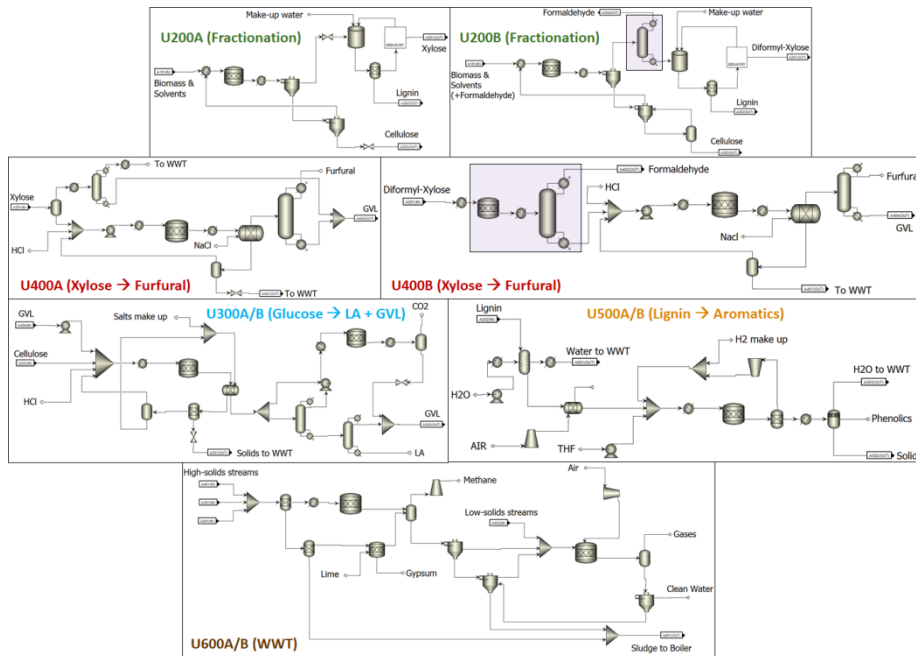


Figure 3. Process flowsheets of biorefinery units (Cases A and B) in Aspen plus v8.6.

Table 1: Inlet and outlet mass balances [tn/hr] for each Unit and each Case A & B.

	IN	OUT	IN	OUT	IN	OUT	IN	OUT	IN	OUT
Units in Case A	U200A		U300A		U400A		U500A		U600A	
Birch	10.0									
H ₂ O	10.8		0.6		3.5		3.4			17.9
GVL	32.0			32.0						
Xylose		2.3								
Cellulose		4.2								
Lignin		1.7								
Furfural						1.2				
Levulinic acid				1.1						
Aromatics							0.4			
Sulphuric acid	0.4									
NaCL			0.1		1.1					
THF					0.3					
Others			0.2		0.2				0.6	1.0
Units in Case B	U200B		U300B		U400B		U500B		U600B	
Birch	10.0									
H ₂ O	9.2		0.7		0.3		2.2			11.0
GVL	36.4			36.4						
Formaldehyde	1.8	0.5		0.9						
Cellulose		4.1								
Lignin		1.1								
Diformyl Xylose		2.7								
Furfural						1.2				
Levulinic acid				1.2						
Aromatics							0.9			
Sulphuric acid	2.4									
NaCL			0.1		3.5					
THF					0.2					
Others			0.2		0.3				0.6	5.4

should be considered instead of THF for lignin upgrade, highlighting the option for partial conversion of furfural output into THF solvent. In Case B, upgrading of lignin attributes to 34% of material costs (including formaldehyde), while the high cost for mineral salts results in the need for cost effective salt recovery systems. Table 2 summarizes the net revenues per product in each case. In the case of aromatics, catalytic hydrogenation of unprotected lignin in Case A should be reconsidered due to its negative economic impact. To the contrary, the exploitation of protected lignin in Case B favors profits from lignin-aromatics. Levulinic acid production appears similar economics in both cases. Finally, furfural production in Case A appears superior economics compared to Case B; however, the milder pre-treatment with formaldehyde in Case B preserves a higher cellulose fraction than that of Case A, which in turn yields in larger amounts of levulinic acid. Overall, the use of formaldehyde in Case B results in better process economics and higher net revenues than those of Case A by 34%.

6. Conclusions

The exploitation of unprotected lignin in Case A for the production of aromatics features a large relative difference in net revenue compared with Case B. In both cases, biomass cost is nearly half of total material expenses, thus cheaper biomass feedstocks should be considered, while the high cost of THF solvent justifies the need for partial conversion of furfural end-product into THF. In conclusion, the implementation of “Lignin-first” concept in presence of formaldehyde (Case B) for the fractionation of biomass is considered as the most profitable option.

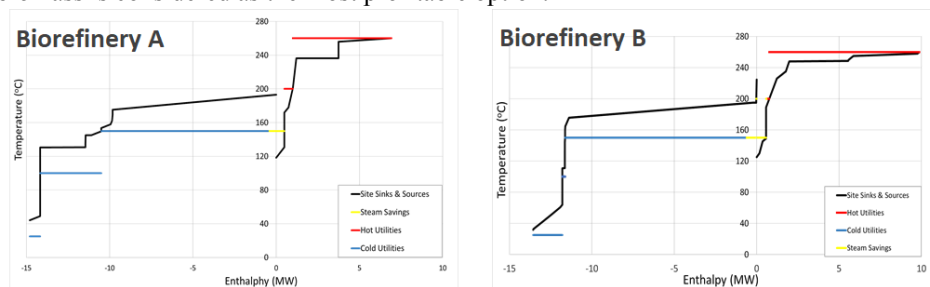


Figure 4. Site Sources and Sinks Profiles for each biorefinery scenario (Case A & B).

Table 2: Total Site integration energy results [MW].

	Case A	Case B		Case A	Case B
HP-260 °C	6.5	9.3	Steam Savings-200 °C	-	0.1
MP-150 °C	10	11	Steam Savings-150 °C	1	1.2
LP-100 °C	3.7	0.2	Power Cogeneration potential [MW]		
Cooling Water	0.6	1.8	Power Cogeneration	1.4	1.8

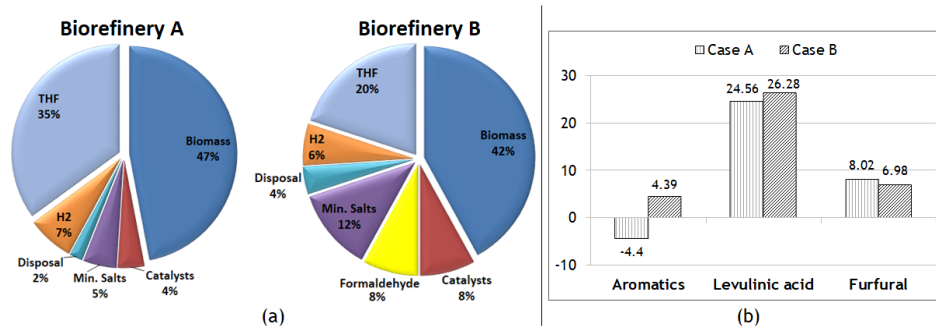


Figure 5. Breakdown of materials expenses (a), net revenue [M€/yr] per product (b)

Acknowledgments

The authors acknowledge financial support by the H2020-EU.1.3.3. Grant RENESING II (H2020-778332).

References

- L. Shuai, M.T. Amiri, Y.M. Questell-Santiago, F. Héroguel, Y. Li, H. Kim, R. Meilan, C. Chapple, J. Ralph, J.S. Luterbacher, 2016. Formaldehyde stabilization facilitates lignin monomer production during biomass depolymerization. *Science*, 354, 6310, 329-333.
- K.A. Pyrgakis, A.C. Kokossis, 2019. A Total Site Synthesis approach for the selection, integration and planning of multiple-feedstock biorefineries. *C&CE*, 122, 326-355.
- Y.M. Questell-Santiago, R. Zambrano-Varela, M.T. Amiri, J.S. Luterbacher, 2018. Carbohydrate stabilization extends the kinetic limits of chemical polysaccharide depolymerization. *Science Chem.*, 10, 1222-1228.
- D.M. Alonso, S.H. Hakim, S. Zhou, W. Won, O. Hosseinaei, J. Tao, V. Garcia-Negron, A.H. Motagamwala, M.A. Mellmer, K. Huang, C.J. Houtman, N. Labbé, D.P. Harper, C.T. Maravelias, T. Runge, J.A. Dumesic, 2017, Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Science Advances*, vol. 3, no.5, e1603301

Valorization of sugarcane bagasse to lactic acid: Life cycle assessment and Techno-economic evaluation in Indian scenario

Munagala Meghana,^{a,*} Yogendra Shastri,^a Kakasaheb Konde,^b Sanjay Patil^b

^a *Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, 400076, India*

^b *Department of Alcohol Technology and Biofuels, Vasantdada Sugar Institute, Manjari, Pune, 412307, India*

meghana.munagala3@gmail.com

Abstract

This work has performed Life Cycle Assessment (LCA) and Techno-Economic Analysis (TEA) to assess the sustainability of a novel process to convert sugarcane bagasse to lactic acid. Experimental results are combined with downstream process simulation using Aspen Plus[®] to generate process data. The LCA model is developed in OpenLCA, with a cradle-to-gate scope and 1 kg of LA as a functional unit. Ecoinvent[®] database is used for inventory, and ReCiPe Midpoint (H) methodology is chosen for impact assessment. Process economic approach is used to perform TEA, and the payback period and product costs are calculated. LCA results showed that climate change impact was 4.62 kg CO₂ eq. per kg LA, with the pretreatment stage identified as a hotspot. TEA results showed that LA product cost was \$3.27/kg, which reduced to \$2.9/kg if bagasse was free. Recycling of NaOH used in pretreatment can result in environmental and economic benefits of the bagasse based LA process.

Keywords: Life cycle assessment (LCA), sugarcane bagasse, lactic acid, techno-economic assessment, waste valorization

1. Introduction

Sustainable valorization of industrial and agricultural wastes is of utmost importance for a thriving circular bioeconomy (Stegmann et al., 2020). In this context, sugarcane bagasse, a lignocellulosic fibrous waste generated by the sugar industry, has gained significance. Increasing research trends in bagasse valorization can be attributed to its wide availability, affordability, and vast potential to produce a multitude of products such as biofuels, biochar, and biochemicals (Meghana and Shastri, 2020). In India, a leading country in sugarcane cultivation, bagasse is burned in the boiler to meet a sugar mill's energy demands. Large mills use to produce surplus electricity and sold to the grid. However, the valorization potential of lignocellulose resource like bagasse is not completely exploited in these current practices. Additionally, current handling practices of bagasse are not environmentally benign due to added emissions resulting from bagasse burning (Meghana and Shastri, 2020). Therefore, developing sustainable production routes that enable and utilize the lignocellulose nature of bagasse to produce value-added chemicals and biofuels is important.

Utilization of bagasse to produce lactic acid (LA), one of the platform chemicals with diverse applications in the food, pharmaceutical, cosmetic, and polymer industry (Alves de Oliveira et al., 2018), has garnered significant interest. Integrating environmental assessment and techno-economic assessment is a way forward for the successful development of bagasse-based routes to produce value-added chemicals like lactic acid in a sustainable manner.

In this study, detailed environmental and techno-economic assessment is performed for an upstream technology developed at Vasantdada Sugar Institute (VSI) to produce LA from bagasse in the Indian context. Reactive distillation is employed for separation and purification of LA and is simulated using Aspen Plus[®]. The life cycle assessment (LCA) framework is employed to identify the environmental hotspots, and techno-economic evaluation has been performed to assess the economic viability.

2. Process Development

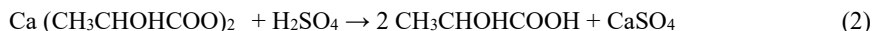
The methodology followed for developing the LA production route from bagasse is described in the following section. Upstream stages developed by VSI and downstream simulated using Aspen Plus[®] software are briefly discussed.

2.1. Process description

Detailed process technology developed by VSI is reported elsewhere (Nalawade et al., 2020). Upstream stages involve size reduction followed by alkali pretreatment with NaOH. In hydrolysis, cellulose and xylan present in the pretreated solid fraction are hydrolysed to fermentable sugars by Cellic CTec2 enzyme. The saccharified broth is filtered to remove the solid fraction, and the hydrolysate proceeds to fermentation. In fermentation, excess calcium carbonate (CaCO₃) is added to the hydrolysate to maintain the desired pH, along with yeast extract and inoculum. Glucose present in the broth is metabolized by *Bacillus coagulans* NCIM 5648 to produce L (+) lactic acid.



As shown in Eq. (1), LA present in the broth reacts with CaCO₃ to form calcium lactate. Acidification with sulfuric acid (H₂SO₄) liberates calcium lactate to form lactic acid, as shown in Eq. (2). The broth is then filtered to obtain a stream containing water and dilute concentrations of LA.

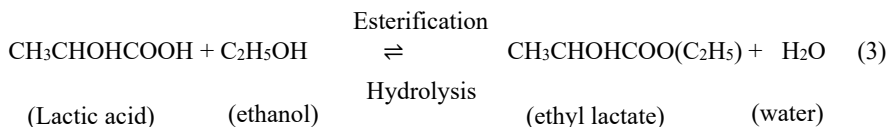


2.2. Downstream separation

In the downstream stage, the stream obtained after filtration is subjected to multi-effect evaporation to concentrate the LA to 40 %, followed by reactive distillation with ethanol in esterification column. In this column, LA reacts with ethanol to produce ethyl lactate and water, as shown in Eq. 3. Ethyl lactate, water, and unreacted ethanol obtained in the top stream are then fed to the hydrolysis column to undergo hydrolysis, as shown in Eq.

(3), to give 99.9 % pure lactic acid. The ethanol removed from the top is fed to the next column, where ethanol is separated by distillation and recycled.

The downstream separation of LA is simulated using process simulation software Aspen Plus[®]. Reaction kinetics and catalyst assumption data are taken from the literature (Su et al., 2013). UNIQUAC method is chosen as the property method, and evaporators are simulated using the flash model. RADFRAC is used to model the reactive distillation columns and DSTWU block for modeling the ethanol-water separation column.



3. Methodology

The LCA and TEA assessment has been performed for a LA facility with 104 tonnes per day production capacity, annexed with an existing sugar mill. The bagasse is assumed to be transported to the LA facility by trucks at a round trip distance of 100 km. Size reduction data is obtained from the literature (Shastri et al., 2012)(Moiceanu et al., 2019). Upstream stages are adjusted accordingly in a sequential batch process, so that downstream takes place in a continuous mode.

3.1. LCA Methodology

The methodology followed for LCA analysis is per ISO 14040. The goal of the LCA is to quantify the environmental impacts of bagasse based LA production facility annexed with an Indian sugar mill. The LCA model has cradle to gate scope, including stages of sugarcane farming, transportation, pretreatment, hydrolysis, fermentation, and downstream. The system boundary of the LCA analysis can be seen from Figure 1. 1 kg of LA is chosen as a functional unit.

Inventory data for farming (Mandade et al., 2016) and transportation (Murali and Shastri, 2019) specific to India are taken from literature. Economic allocation method is used to allocate these impacts for bagasse. Carbon sequestration of biomass and avoided impacts due to bagasse valorization are not included in the study. Experimental data provided inputs to the upstream stages, and process simulations are used for downstream data. Emissions data for producing the enzyme, an input to the hydrolysis stage, are adapted from literature (Sreekumar et al., 2020). The LCA model is developed in OpenLCA, and Ecoinvent[®] database is used for inventory. Recipe (H) mid-point indicators methodology is used for impact assessment.

3.2. TEA Methodology

An economic evaluation of the LA facility annexed to an existing sugar mill is performed to determine the product cost as reported in Table. 1. Equipment cost for pretreatment is calculated by designing the reactor using engineering design principles. Whereas the equipment cost for other upstream stages is adapted from the literature by scaling the cost

using the six-tenth rule and downstream from Aspen Plus® process economic analyzer. The facility is assumed to operate for 330 days. Bagasse is priced at \$54/tonne and costs of cooling water, natural gas, and boiler water are taken as \$0.54/m³, \$1.22/m³ and \$1.86/MMBTU, respectively. Electricity required for the LA facility is assumed to be supplied by the cogeneration plant in the sugar mill. The cooling tower assumed to be part of the sugar mill's infrastructure is not included in the costing.

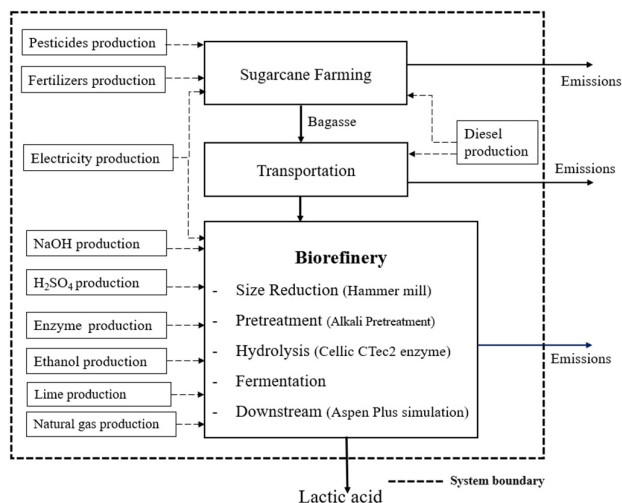


Figure 1: System boundary considered for LCA analysis of bagasse based LA production

Table.1: Economic assessment methodology followed for the LA facility

Economic Estimation Summary

Procurement Cost = 1.15 * Equipment Cost

Fixed Capital Cost (FCC) = Procurement Cost / 0.4

Working Capital = 0.15 * FCC

Total Capital Investment = Fixed Capital + Working Capital

Fixed Operating Cost (FOC): Maintenance = 0.05 * FCC; Operating labour (OL) = (0.5-0.15) * TOC; Laboratory Cost = 0.2 * OL; Supervision = 0.2 * OL; Plant Overheads = 0.5 * OL; Insurance = 0.01 * FCC; Capital Charges = (FCC * economic life of plant) /100; Local taxes = 0.02 * FCC; Royalties = 0.01 * FCC

Variable Operating Cost (VOC) = Raw material Cost + Utilities

Total Operating Cost (TOC) = Fixed Operating + Variable Operating Cost

General Overhead (GO) = 0.25 * TOC

Annual Production Cost = TOC + GO

4. Results and discussion

Environmental assessment results revealed that the life cycle climate change impact for the bagasse-based LA process was 4.62 kg CO₂ eq. per kg LA. Figure 2 shows the

pretreatment stage as a major environmental hotspot for GHG emissions. Adverse impacts associated with NaOH manufacturing, an input to the pretreatment stage, alone resulted in 2.66 kg CO₂ eq. emissions. Whereas, enzyme production and sugarcane farming resulted in 0.78 and 0.42 kg CO₂ eq. emissions respectively. Recently published LCA studies on corn-based LA (Ögmundarson et al., 2020) and vine based LA (Pachón et al., 2020) reported life cycle GHG emissions as 7.9 kg CO₂ eq. and 3.26 kg CO₂ eq., respectively. They also reported that the usage of chemicals such as lime, triethanolamine majorly contributed to GHG emissions. However, Nature works reported relatively lower emissions for corn-based LA production at 1.53 kg CO₂ eq. per kg of polylactic acid produced (Vink et al., 2010). This lower value can be attributed to the improved fermentation technology that required less CaCO₃ and absence of pretreatment stage.

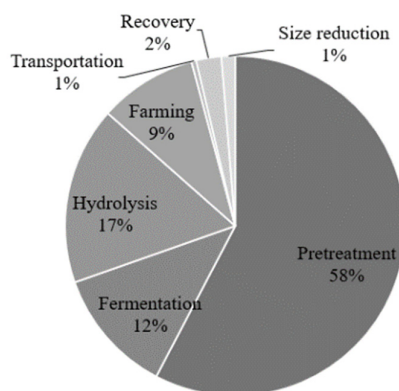


Figure 2: Stage-wise distribution of GHG emissions for bagasse based LA production

LCA analysis revealed water depletion and agricultural land occupation of the LA process as 18.89 m³ and 1.02 m²a, respectively.

From economic evaluation, the total capital investment and total annual operating cost of the LA facility were estimated at 69.86 Million \$ and 86.50 Million \$, respectively. The product cost for LA was \$3.27/kg (Rs.242/kg) when bagasse was not free, with hydrolysis and pretreatment stage contributing to 38 % and 36 % of this product cost. Higher costs associated with hydrolysis and pretreatment stage was due to enzyme and NaOH requirement. For a 10.32% internal rate of return and 6 years payback period, the LA price was found to be \$3.58/kg (Rs. 265/kg). The total capital investment value presented in this work agrees with a study based on an economic evaluation of LA biorefinery (Manandhar and Shah, 2020) upon scaling to the same production capacity.

If bagasse was available for free, LA production cost and LA price to achieve 6 years payback time with 10.32% internal rate of return were \$2.9/kg (Rs.215/kg) and \$3.21/kg (Rs.237/kg), respectively. Sensitivity analysis was also performed for key parameters to identify opportunities to reduce the life cycle impacts and production cost.

5. Conclusions

The production of LA from bagasse was found to be economically viable based on the proposed process, with the market price of commercially produced LA at \$ 3/kg. Solid loading in pretreatment, the amount of NaOH used for pretreatment, and enzyme quantity used in the hydrolysis stage were three significant factors affecting the production cost and the life cycle GHG emissions of the LA process. Recycling of NaOH used in the pretreatment process can result in environmental and economic benefits of the bagasse based LA process.

References

- Alves de Oliveira, R., Komesu, A., Vaz Rossell, C.E., Maciel Filho, R., 2018. Challenges and opportunities in lactic acid bioprocess design—From economic to production aspects. *Biochem. Eng. J.* 133, 219–239. <https://doi.org/10.1016/j.bej.2018.03.003>
- Manandhar, A., Shah, A., 2020. Techno-Economic Analysis of Bio-Based Lactic Acid. *Processes* 199, 8.
- Mandade, P., Bakshi, B.R., Yadav, G.D., 2016. Ethanol from Indian agro-industrial lignocellulosic biomass: an emergy evaluation. *Clean Technol. Environ. Policy* 18, 2625–2634. <https://doi.org/10.1007/s10098-016-1179-y>
- Meghana, M., Shastri, Y., 2020. Sustainable valorization of sugar industry waste: Status, opportunities, and challenges. *Bioresour. Technol.* 303, 122929.
- Moiceanu, G., Paraschiv, G., Voicu, G., Dinca, M., Negoita, O., Chitoiu, M., Tudor, P., 2019. Energy consumption at size reduction of lignocellulose biomass for bioenergy. *Sustain.* 11, 1–12. <https://doi.org/10.3390/su11092477>
- Murali, G., Shastri, Y.S., 2019. Biofuels Life cycle assessment based comparison of different lignocellulosic ethanol production routes production routes. *Biofuels*.
- Nalawade, K., Saikia, P., Behera, S., Konde, K., Patil, S., 2020. Assessment of multiple pretreatment strategies for 2 G L-lactic acid production from sugarcane bagasse. *Biomass conversion and biorefinery*.
- Ögmundarson, Ó., Sukumara, S., Laurent, A., Fantke, P., 2020. Environmental hotspots of lactic acid production systems. *GCB Bioenergy* 12, 19–38. <https://doi.org/10.1111/gcbb.12652>
- Pachón, E.R., Mandade, P., Gnansounou, E., 2020. Conversion of vine shoots into bioethanol and chemicals: Prospective LCA of biorefinery concept. *Bioresour. Technol.* 303.
- Shastri, Y.N., Rodriguez.L.F, Hansen.A.C, Ting.K.C, 2012. Impact of distributed storage and pre-processing on Miscanthus production and provision systems. *Biofuels, Bioprod. Biorefining* 6, 21–31. <https://doi.org/10.1002/bbb>
- Sreekumar, A., Shastri, Y., Wadekar, P., Patil, M., Lali, A., 2020. Life cycle assessment of ethanol production in a rice-straw-based biorefinery in India. *Clean Technol. Environ. Policy* 22, 409–422. <https://doi.org/10.1007/s10098-019-01791-0>
- Stegmann, P., Londo, M., Junginger, M., 2020. The circular bioeconomy: Its elements and role in European bioeconomy clusters. *Resour. Conserv. Recycl.* X 6, 100029.
- Su, C.Y., Yu, C.C., Chien, I.L., Ward, J.D., 2013. Plant-wide economic comparison of lactic acid recovery processes by reactive distillation with different alcohols. *Ind. Eng. Chem. Res.* 52, 11070–11083. <https://doi.org/10.1021/ie303192x>
- Vink, E.T.H., Davies, S., Kolstad, J.J., 2010. The eco-profile for current Ingeo® polylactide production. *Ind. Biotechnol.* 6, 212–224. <https://doi.org/10.1089/ind.2010.6.212>

Hybrid Cybernetic Modeling of the Microbial Production of Polyhydroxyalkanoates Using Two Carbon Sources

Stefanie Duvigneau^{a,*}, Robert Dürr^b, Lena Kranert^a, Annette Wilisch-Neumann^a, Lisa Carius^a, Rolf Findeisen^a, Achim Kienle^{a,b}

^a Otto-von-Guericke-University, Universitätsplatz 2, 39106 Magdeburg, Germany

^b Max-Planck-Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany

* stefanie.duvigneau@ovgu.de

Abstract

Compared to conventional petroleum-based polymers, biopolymers like polyhydroxyalkanoates are a promising alternative as raw material for manufacturing of plastics. One way to reduce the microbial PHA production costs is to use organic wastes from agriculture and food industries, which contain a wide range of carbon sources. Two organic waste carbon sources are fructose, present in fruit juice waste and acetate, present in waste from wine or biogas production. Simulations using kinetic models can help to reduce the range of optimal substrate combinations and thus prevent time-consuming and costly experimental investigations. This contribution outlines an extended hybrid cybernetic model for the polyhydroxybutyrate production in *Cupriavidus necator* growing on fructose, acetate, and nitrogen. Data from bioreactor experiments are used for the adjustment of the kinetic parameters and for the validation of the hybrid cybernetic model. The new model is able to precisely predict the dynamic behavior of hydroxybutyrate while co-feeding with fructose and acetate. Finally, our model is used in a computational study to analyze different initial carbon/ammonium ratios and oxygen partial pressures with respect to the maximum hydroxybutyrate concentration.

Keywords: Polyhydroxyalkanoates, Hybrid Cybernetic Modeling, *Cupriavidus necator*, carbon/ammonium ratio, oxygen partial pressure

1. Introduction

In recent decades, plastic pollution has come more and more into focus. Besides large plastic carpets, there is also a huge amount of microplastic beneath the ocean's surface. Due to the ingestion of these particles by aquatic animal species, our garbage ultimately ends up on our plate. In addition to personal rethinking towards reusable packaging materials and optimizing the recycling stream, investigations of less durable alternatives to conventional plastic raw materials are necessary. Polyhydroxybutyrate (PHB) from the group of polyhydroxyalkanoates (PHA) is one example of a suitable raw material that can be used to replace conventional plastic (Sabapathy *et al.*, 2020). PHA is biodegradable in the environment, non-toxic and bio-based since PHA can be produced using a wide variety of microorganisms. The best examined representative is *Cupriavidus necator* (*C. necator*), which was also used in this work. Compared to the production of conventional plastic raw material, PHA production is very expensive. The costs of the

process can be reduced by using inexpensive substrates (Riedel *et al.*, 2015), intelligent control strategies (Morabito *et al.*, 2019) or adapting production process conditions (Lopar *et al.*, 2013). Mathematical models can help to avoid costly experimental investigations by simulating a large number of possible process conditions and investigating the product yield.

This contribution presents a hybrid cybernetic model (HCM) with two carbon sources as inputs, which typically occur in waste from juice, wine or biogas production. The model is based on our previous approach in (Duvigneau *et al.*, 2020). In contrast to this, additional experimental data for a mixture of fructose and acetate as carbon sources and additional elementary modes (EMs) were considered. It is shown that the extended model is able to reproduce the experimental data significantly better. Further, the growth rate expression is extended to account explicitly for different oxygen partial pressures in the culture broth which has a major influence on the maximum concentration of hydroxybutyrate (HB). Finally, the HCM is used to correlate different carbon/ammonium (C/N) ratios and oxygen partial pressures (pO_2) and to evaluate them with regard to the maximum concentration of HB. The preliminary analysis with our mathematical model makes it possible to identify suitable initial substrate concentrations for precise experimental screening studies.

2. Experimental Methods

The strain *Cupriavidus necator* (H16, DSM 428) obtained from DSMZ GmbH Braunschweig was used in the presented experiments. The experimental data set with acetate as single carbon source was obtained as described in (Duvigneau *et al.*, 2020). For the other two data sets, the strain was cultivated in a DASGIP parallel bioreactor system (Eppendorf, Jülich) with 1 L working volume. The initial and cultivation conditions are summarized in Table 1. Ammonium concentrations were determined from supernatants of the samples using an enzymatic test kit (R-Biopharm AG, Darmstadt, Germany) and following the manufacturer's instructions.

Acetate, HB and cell dry weight (total biomass) concentrations were determined as described in (Duvigneau *et al.*, 2020). Fructose concentrations were determined with high performance liquid chromatography (HPLC, Agilent Infinity II 1290, Agilent, Waldbronn, Germany). For this, 10 μ L of the filtered supernatant was loaded on an RHM-Monosaccharide ion exchange column (Phenomenex, Torrance, USA) and eluted isocratically with 0.6 $mL \cdot min^{-1}$ DI-water at 80 $^{\circ}C$. The fructose peaks were detected with a refractive index detector (G7162A, Agilent, Waldbronn, Germany) at 40 $^{\circ}C$.

Table 1: Cultivation conditions

data set	[FRU(0), ACE(0), $NH_4Cl(0)$] in g/L	media	T in $^{\circ}C$	pO_2 in %
I	[22, 0, 1.7]	M81 ¹	30	70
II	[26, 4.4, 1.4]	M81 ¹	30	5

¹ Recipes for the Medium 81 can be found in (Franz *et al.*, 2011) or on the web page of the DSMZ.

3. Mathematical Model

Differential equations for the substrates (fructose, acetate, ammonium), HB content and the total cellular biomass are defined as follows:

$$\frac{d}{dt} \begin{bmatrix} x_{fru} \\ x_{ace} \\ x_N \end{bmatrix} = \mathbf{S}_S \mathbf{Z} \mathbf{r}_M \mathbf{c}, \quad \frac{dm_{HB}}{dt} = \mathbf{S}_{HB} \mathbf{Z} \mathbf{r}_M, \quad \frac{dc}{dt} = \mu c \quad (1)$$

The EM matrix in yield space for the substrates $\mathbf{S}_S\mathbf{Z}$ consists of the stoichiometric information of the metabolic network. The metabolic model was derived in (Duvigneau *et al.*, 2020) from models published in (Franz *et al.*, 2011) and (Yu and Si, 2004). The Monod-type kinetic in $r_{M,i}$ (see Equation (5)) expresses influence of substrate concentrations for each selected EM. Internal metabolites, except for internal storage material such as PHB, are assumed to have fast dynamics compared to the external substrates and do not need a model equation. The EM matrix in yield space for HB $\mathbf{S}_{HB}\mathbf{Z}$ describes normalized consumption and production of HB in yield space.

The growth rate μ consists of a rate vector \mathbf{r}_M , the EM matrix for the normalized biomass production $\mathbf{S}_c\mathbf{Z}$ and a proportional factor K_μ :

$$\mu = \mathbf{S}_c\mathbf{Z}\mathbf{r}_M K_\mu \quad (2)$$

The proportional factor K_μ can be calculated for each experimental setup with respect to the pO_2 in the culture broth by using the following correlation:

$$K_\mu = \frac{\mu(pO_2)}{\mu(70\%)} \quad \text{with} \quad \mu(pO_2) = -0.151pO_2 + 0.256 \quad (3)$$

Further, the use of enzymes for each chosen EM is characteristic for the hybrid cybernetic approach:

$$\frac{de}{dt} = \alpha + \mathbf{r}_{EM}b - \beta\mathbf{e}, \quad \text{where } e_i^{\text{rel}} = \frac{e_i}{e_i^{\text{max}}} \quad \text{with } e_i^{\text{max}} = \frac{\alpha_i + k_{e,i}}{\beta_i + k_{r,i}(\mathbf{S}_c\mathbf{Z})_i} \quad (4)$$

The increase of the enzyme level occurs via the constitutive enzyme synthesis rate α and the catalytically active part of the total biomass b . Monod-type kinetics \mathbf{r}_{EM} (Equation (5)) control the catalytically active biomass fraction. The term $-\beta\mathbf{e}$ represents the enzyme degradation by a protein turnover (β). Compared to the previous model description, the dilution terms $-\mu m_{HB}$ and $-\mu \mathbf{e}$ were excluded in Equation (1) and (4), respectively.

A general description of rates r_M and r_{EM} of the i^{th} EM is

$$r_{M,i} = v_i k_{r,i} e_i^{\text{rel}} r_{\text{core},i} \quad r_{EM,i} = u_i k_{e,i} r_{\text{core},i} \quad (5)$$

The rate r_{core} of the i^{th} EM is a multiplied Monod-type kinetic:

$$r_{\text{core},i} = \frac{X_1}{K_{X_1} + X_1} \cdots \frac{X_n}{K_{X_n} + X_n} \quad (6)$$

The number of negative yield coefficients (consumption terms) in each EM defines the number of factors given in Equation (6). For the regulation of enzyme synthesis and activity, cybernetic control variables u and v are introduced and calculated by cybernetic control laws as described in the publication of (Young and Ramkrishna, 2007).

The matrix of all normalized EMs $\mathbf{SZ} = [\mathbf{S}_S\mathbf{Z}, \mathbf{S}_{HB}\mathbf{Z}, \mathbf{S}_c\mathbf{Z}]$ was obtained in (Duvigneau *et al.*, 2020). The selection of EMs describing the co-feeding with fructose and acetate was done during parameter estimation of the kinetic parameters \mathbf{k}_r for those modes. For this, the set of 38 generating modes from yield analysis of the underlying metabolic model was used to find a sufficient EM candidate (Song and Ramkrishna, 2009; Duvigneau *et al.*, 2020).

The model was implemented and solved numerically in *MATLAB2019b*. Parameter adjustment of rates \mathbf{k}_r was performed using the *MATLAB* routine *fmincon* with a lower boundary zero and a multi-start approach with $n=1000$ by minimizing the following objective function

$$\text{ESS} = \sum_{i=1}^n \left(\frac{\mathbf{x}_{\text{exp}}(t_i) - \mathbf{x}_{\text{sim}}(t_i)}{\max(\mathbf{x}_{\text{exp}})} \right)^2. \quad (7)$$

Kinetic parameters \mathbf{k}_r are shown in Table 2. All other 49 parameters of the model were kept as described in (Duvigneau *et al.*, 2020).

Table 2: Kinetic parameters \mathbf{k}_r from three model-data combinations

Original [†]	$[0.26, 3.71, 0.07, 0, 0, 0, 0, 0, 0, 3.44, 1.51, 0.44, 0.44, 0.44, 0.25]^T$
Set A	$[0.26, 3.71, 0.07, 0.04, 28.48, 0.67, 0, 0, 0, 3.44, 1.51, 0.44, 0.44, 0.44, 0.25]^T$
Set B	$[0.08, 0.89, 0.04, 0.09, 0.15, 0.61, 0, 0, 0, 1.45, 0.43, 0.43, 3.15, 0.20]^T$

[†](Duvigneau *et al.*, 2020)

4. Results

Figure 1 (dashed lines) shows the dynamics for the substrates and products after simulation of the model in (Duvigneau *et al.*, 2020) applying the initial condition of data set II (acetate and fructose as carbon sources).

The model already delivers a qualitatively good result without the inclusion of EMs with both substrates as input.

After inclusion of these EMs and the kinetic parameters \mathbf{k}_r of these EMs (set A) the model simulation shows the consumption of both carbon sources at the same time (diauxic growth), but underestimates the final biomass by approx. 20 %. The simulation result with set A can be seen as solid lines in Figure 1.

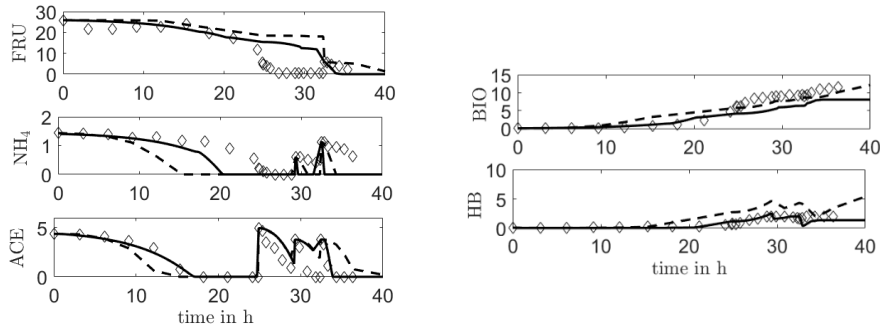


Figure 1: Model simulation and experimental data set II (fructose and acetate co-feeding, diamonds) using the model of (Duvigneau *et al.*, 2020) (dashed lines) and the adapted version with respect to the kinetic parameters \mathbf{k}_r of the EMs with acetate and fructose as co-substrates (solid lines, set A). Legend: FRU, fructose; NH_4 , ammonium; ACE, acetate; BIO, total biomass; HB, hydroxybutyrate. All concentrations are given in g/L.

In general, precise data sets with controlled and constant process conditions are required for adapting kinetic models (Carius and Findeisen, 2016). The data sets from (Duvigneau *et al.*, 2020) and data set II were obtained from different cultivation systems and hence, differ in their dynamic behavior. For this reason, data sets I and II as well as the data set with acetate as the single carbon source from our previous publication were selected for a new adjustment of the kinetic parameters \mathbf{k}_r . The data sets only differ in the available carbon source and $p\text{O}_2$ (see Table 1). A proportionality factor K_{μ} can be determined with Equation (3). This leads to an increased growth at lower $p\text{O}_2$. Furthermore, dilution terms in the differential equations for the HB content and the hybrid cybernetic enzymes have been eliminated (Equations (1) and (4)). The result of adapting the new model to the above-mentioned data sets can be seen in Figure 2. Now, the concentrations of HB and

biomass can be simulated more precisely. All model simulations show an overestimation for fructose between 25 and 30 h. This discrepancy between simulation and experimental data could be compensated by considering maintenance, energy generation or CO₂ due to the activity of the central metabolism.

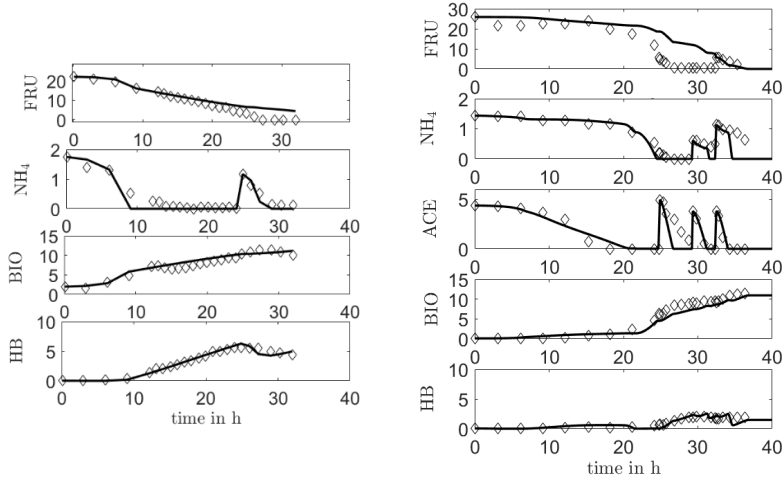


Figure 2: Model simulation and experimental data set I (left) and data set II (right) using the restructured and adapted model. Legend: FRU, fructose; NH₄, ammonium; ACE, acetate; BIO, total biomass; HB, hydroxybutyrate. All concentrations are given in g/L.

The new model can be used to analyze different initial conditions with respect to the maximum product concentration. Besides a well-chosen ration of fructose and acetate (FRU/ACE ratio), ammonium in the culture broth promotes the growth of non-PHB biomass. Thus, the C/N ratio can also be decisive. Figure 3 shows a simulation study using different C/N ratios and pO₂ while keeping the initial concentrations for the carbon sources as in data set II. The model study delivers one very interesting result: with decreasing pO₂, a higher C/N ratio can be selected, so that a higher maximum HB concentration is reached.

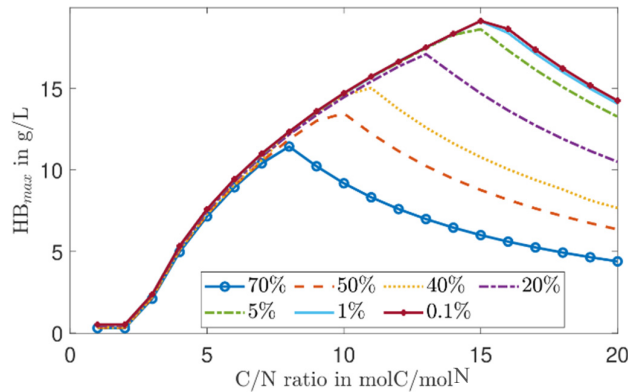


Figure 3: Model study with different carbon to ammonium ratios (C/N) and oxygen partial pressures (pO₂). Selected pO₂ levels are shown in the legend.

5. Conclusion

In the presented work an HCM was adapted to experimental data with fructose and acetate as carbon sources. Diauxic growth is one of the effects that can be described very well with a hybrid cybernetic approach (Song, Morgan and Ramkrishna, 2009). By changing the kinetic parameters, it is also possible to describe diauxic growth in the presence of fructose and acetate with the model presented in this manuscript. By including the pO_2 in the model structure further improvement of the model can be achieved.

Finally, we investigate different C/N ratios and pO_2 levels with respect to the maximum product yield. Therein, lower pO_2 values lead to higher HB concentrations. In addition, a different C/N ratio should be selected for each pO_2 level in order to achieve an optimal production yield. Future work will focus on an extension of the model study using different initial carbon concentrations, experimental validation of the results shown here and a comparison of the pO_2 effect with other factors e.g. temperature variations.

Acknowledgements We would like to acknowledge the EU-program ERDF (European Regional Development Fund) for funding the projects DIGIPOL and MAPO. Furthermore, we would like to thank Johannes Pohlodek for the support in the laboratory.

References

- Carius, L. and Findeisen, R. (2016) 'The impact of experimental data quality on computational systems biology and engineering', *IFAC-PapersOnLine*. Elsevier B.V., 49(26), pp. 140–146. doi: 10.1016/j.ifacol.2016.12.116.
- Duvigneau, S. et al. (2020) 'Hybrid Cybernetic Modeling of Polyhydroxyalkanoate Production in *Cupriavidus necator* using Fructose and Acetate as Substrates.', *Proceedings to IFAC World Congress 2020*.
- Franz, A. et al. (2011) 'Experimental and theoretical analysis of poly(β -hydroxybutyrate) formation and consumption in *Ralstonia eutropha*', *Biochemical Engineering Journal*, 55(1), pp. 49–58. doi: 10.1016/j.bej.2011.03.006.
- Lopar, M. et al. (2013) 'Five-step continuous production of PHB analyzed by elementary flux, modes, yield space analysis and high structured metabolic model', *Biochemical Engineering Journal*, 79, pp. 57–70.
- Morabito, B. et al. (2019) 'Multi-mode Model Predictive Control and Estimation for Uncertain Biotechnological Processes', *IFAC-PapersOnLine*. Elsevier B.V., 52(1), pp. 709–714. doi: 10.1016/j.ifacol.2019.06.146.
- Riedel, S. L. et al. (2015) 'Polyhydroxyalkanoates production with *Ralstonia eutropha* from low quality waste animal fats', *Journal of Biotechnology*. Elsevier B.V., 214, pp. 119–127. doi: 10.1016/j.jbiotec.2015.09.002.
- Sabapathy, P. C. et al. (2020) 'Recent developments in Polyhydroxyalkanoates (PHAs) production – A review', *Bioresour. Technology*. Elsevier, 306, p. 123132. doi: 10.1016/j.biortech.2020.123132.
- Song, H.-S., Morgan, J. A. and Ramkrishna, D. (2009) 'Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose', *Biotechnology and Bioengineering*, 103(5), pp. 984–1002.
- Song, H.-S. and Ramkrishna, D. (2009) 'Reduction of a set of elementary modes using yield analysis', *Biotechnology and Bioengineering*, 102(2), pp. 554–568.
- Young, J. D. and Ramkrishna, D. (2007) 'On the Matching and Proportional Laws of Cybernetic Models', *Biotechnology Progress*, 23(1), pp. 83–99.
- Yu, J. and Si, Y. T. (2004) 'Metabolic carbon fluxes and biosynthesis of polyhydroxyalkanoates in *Ralstonia eutropha* on short chain fatty acids', *Biotechnology Progress*, 20(4), pp. 1015–1024.

Microbial Production of Polyhydroxyalkanoates – Modeling of Chain Length Distribution

Robert Dürr^{a,*}, Stefanie Duvigneau^b, Achim Kienle^{a,b}

^a *Max-Planck-Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany*

^b *Otto-von-Guericke-University, Universitätsplatz 2, 39106 Magdeburg, Germany*

* duerr@mpi-magdeburg.mpg.de

Abstract

Compared to crude oil-based polymers, biopolymers like polyhydroxyalkanoates represent a promising raw material to produce plastics as they feature eco-friendly desirable properties such as biodegradability and non-toxicity. Moreover, polyhydroxyalkanoates can be produced by a wide range of microorganisms from organic material under phosphorus, nitrogen, or oxygen limitation. This motivates application of microbial production processes for valorization of organic wastes from agriculture and food industries, containing a wide range of carbon sources. Mathematical modeling of cellular metabolism as well as polymer formation is necessary for sophisticated process control and intensification. In this contribution, a multi-scale model for PHB production in *Cupriavidus necator* under limiting conditions is presented, which accounts for the dynamics of the chain length distribution and also the intracellular regulation of the cellular metabolism. While the dynamics of chain length distribution is described by a population balance model, cellular metabolism is characterized by a hybrid cybernetic model.

Keywords: Polyhydroxyalkanoates, Polymerization, Population Balance Modeling, Multiscale Modeling

1. Introduction

Due to the existing global waste problem, the demand for alternatives to conventional plastic is continuously increasing. Polyhydroxybutyrate (PHB) from the group of polyhydroxyalkanoates (PHAs) is bio-based and, above all, biodegradable and therefore represents a suitable alternative. In addition, PHAs can be produced by a large variety of microorganisms (Jendrossek and Pfeiffer, 2014; Singh et al., 2018). One common PHA producer is *Cupriavidus necator* (*C. necator*), which was also used in this work. In addition to precise measurements and predictions of overall PHA concentrations, the specific structure of the polymer chains in terms of chain length and composition, is an important property which is related to physical properties, like melting point and brittleness (Laycock et al., 2014). Thus, there is an increasing interest for production of PHAs of specific structure to meet certain product requirements.

In this context, multi-scale mathematical modeling of the cellular metabolism and polymer chain formation supports experimental investigations. Early modeling approaches used only little structural information from metabolic networks to explain the general dynamics of the polymer production (Heinzle and Lafferty, 1980). Starting in the

1990s, first structural information on reaction kinetics and metabolic regulation was added (Leaf and Srien, 1998). Further models used small metabolic reaction networks to describe PHB production more precisely for mixed microbial cultivations (Dias et al., 2005). Coupling of the thereby described dynamics on the reactor scale to chain formation models have been presented employing a population balance modeling framework to describe the dynamics of chain length distribution (CLD) (Penloglou et al., 2010, 2012a,b, 2017; Mantzaris et al., 2002). Those are derived from similar models characterizing chemical polymer formation in bulk solution or emulsion (Butté et al., 2002; Kiparissides, 2006; Vale and McKenna, 2005) and can be solved with established numerical techniques (Kumar and Ramkrishna, 1996; Krallis et al., 2008). Further extensions include modeling of block-copolymer chains (Krallis et al., 2008; McChalicher and Srien, 2007). Those models have been used for the development of control concepts (Penloglou et al., 2017; Mantzaris et al., 2001; Iadevaia and Mantzaris, 2006) for the design of polymers with desired properties.

A shortcoming of the previous approaches is that they lack detailed information on the cellular metabolism, in particular intracellular regulation. As an alternative to dynamic flux balance analysis, the hybrid cybernetic modeling (HCM) approach developed by Ramkrishna and coworkers (see (Ramkrishna and Song, 2018) and the references therein) can be used to take into account optimal and detailed metabolic regulation (resource allocation) with respect to existing substrates. This is achieved by introducing hybrid cybernetic control variables.

In this contribution, we will present a multi-scale model for PHB formation in *C. necator* which combines an HCM and a dynamic model for CLD dynamics. Therein, two different carbon sources, namely fructose and acetate, are considered (Duvigneau et al., 2020).

2. Model Formulation

The proposed multi-scale model consists of a state-of-the-art HCM (Duvigneau et al., 2020) characterizing the cellular metabolism and intracellular regulation which is complemented by a population balance model describing dynamics of the PHB chain length distribution Penloglou et al. (2017).

The HCM in (Duvigneau et al., 2020) represents a set of ordinary differential equations that describes the temporal evolution of concentrations of substrates fructose, acetate and ammonium but also total biomass concentration and intracellular PHB content

$$\begin{aligned} \frac{d}{dt}(x_{fru} \ x_{ace} \ x_N) &= S_S \cdot Z \cdot r_M \cdot c, \\ \frac{d}{dt}(c \ m_{PHB}) &= (S_c \cdot Z \cdot r_M \cdot c \ S_{HB} \cdot Z \cdot r_M - \mu \ m_{PHB}) \end{aligned} \quad (1)$$

with the Monod-type kinetics r_M and selected active modes (AMs) for substrates, biomass and biopolymer $S_S \cdot Z$, $S_c \cdot Z$, $S_{HB} \cdot Z$, which are determined by metabolic yield analysis (Song and Ramkrishna, 2009). The kinetic rates r_M are controlled by so-called cybernetic control variables which represent intra-cellular regulation of enzyme activity. Further, the HCM includes by the enzyme dynamics for each AM

$$\frac{de}{dt} = \alpha + r_{EM} \cdot b - \beta \cdot e - \mu \cdot e. \quad (2)$$

To account for the dynamics of the chain length distribution, the HCM is complemented

by a population balance model (Vale and McKenna, 2005). Here, polymers from HB-monomers are distinguished into active (living) and inactive (dead) polymer species, $[LP]_i$ and $[DP]_i$. The dynamics of their CLDs are described by

$$\begin{aligned}\frac{d[LP]_i}{dt} &= \delta(i-1)k_{ini}[MonSyn] + k_{m2}[Mon]([LP]_{i-1} - [LP]_i) - k_{term}[LP]_i \\ \frac{d[DP]_i}{dt} &= k_{term}[LP]_i - k_{dep}([DP]_i - [DP]_{i-1})\end{aligned}\quad (3)$$

where $[e]_i$ denotes the number density distribution of chains with distinct chain length i .

In contrast to Penloglou et al. (2017), the species of intermediate polymers is neglected in our formulation. The dynamics of the monomer concentration and the monomer-synthase complex are given by

$$\begin{aligned}\frac{d[Mon]}{dt} &= -k_1[Mon][Syn] - k_{m2}[Mon]\sum_{i=1}^{\infty}[LP]_i + [Mon]^+ \\ \frac{d[MonSyn]}{dt} &= k_1[Mon][Syn] - k_{ini}[MonSyn]\end{aligned}\quad (4)$$

Under the assumption, that both are not accumulated, i.e., monomers are directly polymerized and constant synthase concentration $k_{m1} = k_1[Syn] = const.$, steady states for both are derived as

$$[Mon]^* = \frac{[Mon]^+}{k_{m1} + k_{m2}\sum_{i=1}^{\infty}[LP]_i}, \quad [MonSyn]^* = \frac{k_{m1}[Mon]^+}{k_{ini}}\quad (5)$$

The total amount of PHB is now given by

$$\begin{aligned}\frac{1}{MW_{HB}}\frac{dc_{m_{HB}}}{dt} &= \frac{d}{dt}\left\{\sum_{i=1}^{\infty}i([LP]_i + [DP]_i)\right\} \\ &= k_{ini}[MonSyn]^* + k_{m2}[Mon]^*\sum_{i=1}^{\infty}[LP]_i - k_{dep}\sum_{i=1}^{\infty}[DP]_i \\ &\quad r_{PHB}^+ MW_{HB}^{-1} \quad r_{PHB}^- MW_{HB}^{-1}\end{aligned}\quad (6)$$

with the molecular mass of a HB-monomer MW_{HB} . The PHB-production and depletion rates r_{PHB}^+ and r_{PHB}^- can be derived from the HCM. Taking all above considerations into account the production rate of monomers $[Mon]^*$ and the depolymerization rate k_{dep} are given as

$$[Mon]^+ = \frac{r_{PHB}^+}{MW_{PHB}}, \quad k_{dep} = \frac{r_{PHB}^-}{MW_{PHB}\sum_{i=1}^{\infty}[DP]_i}\quad (7)$$

3. Numerical Solution

The overall model (1)-(6) consists of a large-scale system of ordinary differential equations. As the average chain length is in the order of 10^6 , simulation of each single chain length class is unfeasible. To scale the computational effort to an approachable scale the Fixed-Pivot-Method (Kumar and Ramkrishna, 1996; Saliakas et al., 2007) was

applied to divide the chain length scale into 100 different pivots. The resulting system of 220 ODEs was solved numerically in MATLAB using ode15s and the parameters reported in Duvigneau et al. (2020) for the HCM-model. Due to lack of experimental data of PHB CLDs, parameters of the polymerization model were chosen such that the obtained CLDs are in ranges reported for microbial PHB production (Kawaguchi and Doi, 1992).

4. Results

The simulation scenarios followed in this contribution have been presented first in Duvigneau et al. (2020) and are based on two different experimental runs with either acetate or fructose as carbon supplying substrate. In Fig. 1 the simulation results for substrates and overall biomass and PHB are shown along the experimental records. It is seen that the amount of PHB of the macroscopic HCM and the microscale polymerization model are the same. It can thereby be inferred that the applied numerical scheme provides an accurate solution for the latter. In Fig. 2, characteristics of the overall CLD (active +

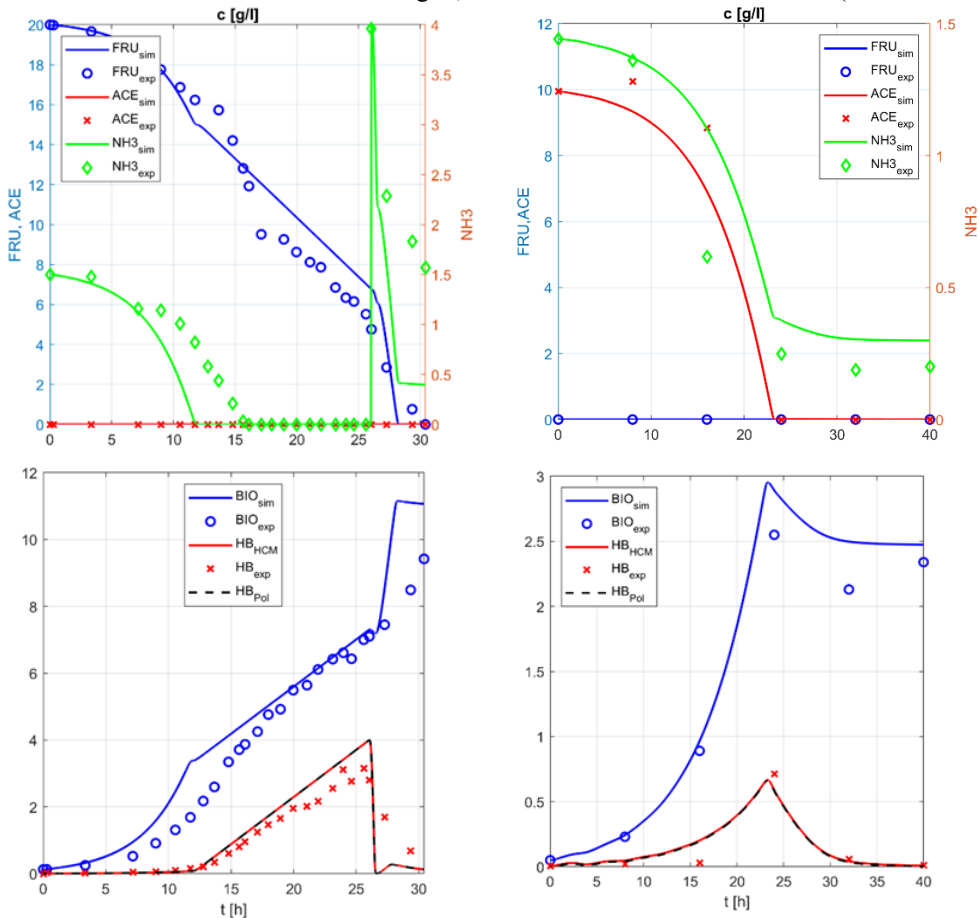


Figure 1: Macroscopic dynamics for scenario I (left, fructose) and scenario II (right, acetate)

inactive polymer chains) are depicted. It is seen that the addition of ammonium in the first scenario (fructose as the single carbon source) results in general PHB consumption, represented by a flattening CLD. However, for $t > 27h$ the kinetic model predicts a small increase of PHB for a short time range. This is also seen from the increased CLD in Figure 2. The model also indicates that for $t > 28.5h$ the CLD flattens out as PHB is metabolized and the chain depolymerize. For the second scenario (acetate as single carbon source), a similar effect is visible after acetate is depleted for $t > 24h$. The intracellular PHB is metabolized as ammonium is still present in the medium resulting in a flattening CLD. For both cases, the number average molecular mass is in the order of 10^6 while the polydispersity index is around 2 and thereby within ranges reported by Kawaguchi and Doi (1992).

5. Conclusion

In this contribution a multi-scale model for PHB formation was presented, which links a state-of-the-art HCM with a detailed description of the chain length distribution dynamics. To the best of our knowledge, this is the first multi-scale model which is able to describe PHB formation on two different carbon sources combining a medium scale metabolic model, optimal intracellular regulation and polymerization kinetics at the same time. The simulation delivers reasonable average chain lengths in the right order of magnitude and will be a valuable tool for future investigations.

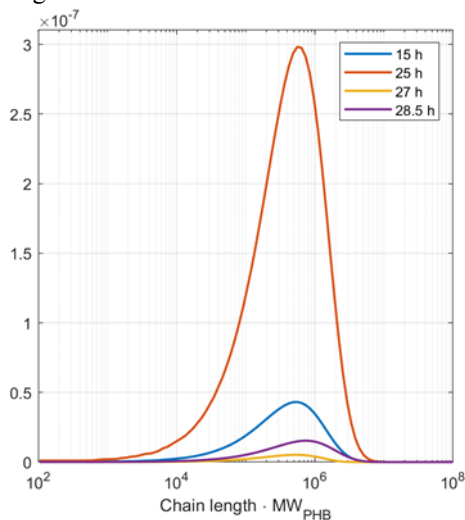


Figure 2: Overall chain length dynamics for scenario I (left, fructose) and scenario II (right, acetate)

Future work will be concerned with model calibration using own data from experiments with *C. necator*. Afterwards, the multi-scale model shall be used for optimization of substrate feeding strategies to obtain high PHB molecular weights. Furthermore, the model could be extended to include additional carbon sources and account for co-polymer formation.

Acknowledgements This work is partly funded by the European Regional Development Fund (ERDF) project "DIGIPOL". The financial support is hereby gratefully acknowledged.

References

- A. Butté, G. Storti, M. Morbidelli, 2002. Evaluation of the chain length distribution in free-radical polymerization, I. Bulk polymerization. *Macromolecular Theory and Simulations* 11 (1), 22–36.
- J. M. Dias, L. S. Serafim, P. C. Lemos, M. A. Reis, R. Oliveira, 2005. Mathematical modelling of a mixed culture cultivation process for the production of polyhydroxybutyrate. *Biotechnology and Bioengineering* 92 (2), 209–222.
- S. Duvigneau, R. Dürr, L. Carius, A. Kienle, 2020. Hybrid cybernetic modeling of polyhydroxyalkanoate production in *Cupriavidus necator* using fructose and acetate as substrates. Proceedings to the 21st IFAC World Congress (IFAC WC 2020), July 12 - 17, Berlin (Germany). 140–146. doi: 10.1016/j.ifacol.2016.12.116.
- E. Heinzle, R. M. Lafferty, 1980. A kinetic model for growth and synthesis of poly- β -hydroxybutyric acid (PHB) in *Alcaligenes eutrophus* H 16. *European Journal of Applied Microbiology and Biotechnology* 11 (1), 8–16.
- S. Iadevaia, N. V. Mantzaris, 2006. Genetic network driven control of PHBV copolymer composition. *Journal of Biotechnology* 122 (1), 99–121.
- D. Jendrossek, D. Pfeiffer, 2014. New insights in the formation of polyhydroxyalkanoate granules (carbonosomes) and novel functions of poly(3-hydroxybutyrate). *Environmental Microbiology* 16 (8), 2357–2373.
- Y. Kawaguchi, Y. Doi, 1992. Kinetics and Mechanism of Synthesis and Degradation of Poly(3-hydroxybutyrate) in *Alcaligenes eutrophus*. *Macromolecules* 25 (9), 2324–2329.
- C. Kiparissides, 2006. Challenges in particulate polymerization reactor modeling and optimization: A population balance perspective. *Journal of Process Control* 16 (3), 205–224, selected Papers from Dycops 7 (2004), Cambridge, Massachusetts.
- A. Krallis, D. Meimaroglou, C. Kiparissides, 2008. Dynamic prediction of the bivariate molecular weight–copolymer composition distribution using sectional-grid and stochastic numerical methods. *Chemical Engineering Science* 63 (17), 4342 – 4360.
- S. Kumar, D. Ramkrishna, 1996. On the solution of population balance equations by discretization - I. A fixed pivot technique. *Chemical Engineering Science* 51 (8), 1311–1332.
- B. Laycock, P. Halley, S. Pratt, A. Werker, P. Lant, 2014. The chemomechanical properties of microbial polyhydroxyalkanoates. *Progress in Polymer Science* 39 (2), 397 – 442.
- T. A. Leaf, F. Sreenc, 1998. Metabolic modeling of polyhydroxybutyrate biosynthesis. *Biotechnology and Bioengineering* 57 (5), 557–570.
- N. V. Mantzaris, A. S. Kelley, P. Daoutidis, F. Sreenc, 2002. A population balance model describing the dynamics of molecular weight distributions and the structure of PHA copolymer chains. *Chemical Engineering Science* 57 (21), 4643 – 4663.
- N. V. Mantzaris, A. S. Kelley, F. Sreenc, P. Daoutidis, 2001. Optimal carbon source switching strategy for the production of PHA copolymers. *AIChE Journal* 47 (3), 727–743.
- C. W. McChalicher, F. Sreenc, 2007. Investigating the structure–property relationship of bacterial PHA block copolymers. *Journal of Biotechnology* 132 (3), 296 – 302, international Symposium on Biological Polyesters 2006.
- G. Penloglou, C. Chatzidoukas, C. Kiparissides, 2012a. Microbial production of polyhydroxybutyrate with tailor-made properties: An integrated modelling approach and experimental validation. *Biotechnology Advances* 30 (1), 329 – 337, *Systems Biology for Biomedical Innovation*.
- G. Penloglou, E. Kretza, C. Chatzidoukas, S. Parouti, C. Kiparissides, 2012b. On the control of molecular weight distribution of polyhydroxybutyrate in *azohydromonas lata* cultures. *Biochemical Engineering Journal* 62, 39 – 47.
- G. Penloglou, A. Roussos, C. Chatzidoukas, C. Kiparissides, 2010. A combined metabolic/polymerization kinetic model on the microbial production of poly(3-hydroxybutyrate). *New Biotechnology* 27 (4), 358–367.

- G. Penloglou, A. Vasileiadou, C. Chatzidoukas, C. Kiparissides, 2017. Model-based intensification of a fedbatch microbial process for the maximization of polyhydroxybutyrate (PHB) production rate. *Bioprocess and Biosystems Engineering* 40 (8), 1247–1260.
- D. Ramkrishna, H.-S. Song, 2018. *Cybernetic Modeling for Bioreaction Engineering*. Vol. I. Cambridge University Press.
- V. Saliakas, C. Chatzidoukas, A. Krallis, D. Meimaroglou, C. Kiparissides, 2007. Dynamic optimization of molecular weight distribution using orthogonal collocation on finite elements and fixed pivot methods: An experimental and theoretical investigation. *Macromolecular Reaction Engineering* 1 (1), 119–136.
- A. K. Singh, L. Sharma, J. K. Srivastava, N. Mallick, M. I. Ansari, 2018. *Microbially Originated Polyhydroxyalkanoate (PHA) Biopolymers: An Insight into the Molecular Mechanism and Biogenesis of PHA Granules*. Springer International Publishing, Cham, pp. 355–398.
- H.-S. Song, D. Ramkrishna, 2009. Reduction of a set of elementary modes using yield analysis. *Biotechnology and Bioengineering* 102 (2), 554–568.
- H. Vale, T. McKenna, 2005. Modeling particle size distribution in emulsion polymerization reactors. *Progress in Polymer Science* 30 (10), 1019–1048.

Identifiability of metabolic flux ratios on carbon labeling experiments

Rafael D. de Oliveira^a, Caroline S. M. Nakama^{a,b}, Vânia Novello^c, José G. C. Gomez^c and Galo A.C. Le Roux^{a*}

^a Department of Chemical Engineering, Polytechnic School - University of São Paulo. Prof. Luciano Gualberto Ave., 380, 05508-900, São Paulo, Brazil

^b Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

^c Department of Microbiology, 1374, Institute of Biomedical Sciences, University of São Paulo, Prof. Lineu Prestes Ave., 05508-000, São Paulo, Brazil
galoroux@usp.br

Abstract

Systems biology is a field where Process System Engineering has an immense potential for contribution. The first step in applying Systems Biology to a given problem is to characterize potential microorganisms to be used for a transformation of interest. One basic characterization is to estimate what are the extents of the reactions that take place inside the cell, which is called flux analysis. In this contribution, we present an application of identifiability analysis to ¹³C-Metabolic Flux Ratio Analysis (¹³C-MFA), which is essentially a parameter estimation problem. A simple and precise way to assess if an experiment would give a unique solution for all flux ratio estimates is important for the successful use of ¹³C-MFA. In this work, the principal components (PC) of the Hessian matrix were evaluated to assess the identifiability of metabolic fluxes on labeling experiments. A recently proposed sparse PC calculation was used for that end and compared with classic PC. This approach was applied to the estimation of metabolic flux ratios of the central metabolism of *Pseudomonas aeruginosa* producing polyhydroxyalkanoates (PHA). The results indicated that sparse PC make it straightforward to interpret dependencies among fluxes, and that the substrate [6-¹³C]glucose is the best option for labeling experiments with *Pseudomonas aeruginosa*. The identifiability analysis also showed that the exchange rates of the PP pathway cannot be estimated uniquely, just some combinations of a reduced set of them are identifiable. In contrast, the flux through the Entner-Doudoroff Pathway can be estimated with high accuracy. The results presented here indicate that applying identifiability analysis to ¹³C-MFA can reduce experimental costs and increase estimation precision.

Keywords: Identifiability, ¹³C-MFA, Metabolism, Systems Biology, Metabolic Engineering

1. Introduction

Metabolic Engineering is a field that aims to assist in the development of new strains for sustainable production of valuable bioproducts. In essence, it consists of applying mathematical models to guide genetic modification in cells in order to obtain higher yields and productivity (Sauer, 2006). Among the tools that have been developed in the field, ^{13}C -Metabolic Flux Ratio Analysis (^{13}C -MFA) emerged as a powerful tool to describe the metabolism of a cell (Sauer, 2006) by estimating the flux distribution of metabolites. Figure 1 illustrates how this method works; first a carbon 13 labeled substrate is used in the culture medium (e.g. $[6-^{13}\text{C}]$ glucose), then the labeling goes through the internal metabolites until it reaches a bioproduct. The labeling pattern in the bioproduct is a function of the pathway that was used by the cell, and each pathway shuffles the labeling differently. Finally, the labeling in the bioproduct is measured by techniques such as Gas Chromatography-Mass Spectrometry (CG-MS) and Nuclear magnetic resonance (NMR). Using a metabolic network model, an estimation problem is formulated and solved and the flux distribution inside the cell estimated. Determining precisely what fluxes can be estimated from a given experiment architecture (i.e. labeling substrate and bioproducts measurements) is crucial due to the high experimental cost associated with carbon labeling experiments.

The Process System Engineering community has given important contributions to the identifiability of nonlinear model parameters, and identifiability studies on ^{13}C -MFA have also been conducted (van Winden et al., 2001; Kappelmann et al., 2016). However, most of them focus on structural identifiability and not on practical identifiability. Here, the well-known Principal Component Analysis (PCA) is applied to identify nonlinear relations between metabolic fluxes in ^{13}C -MFA. Also, a recently developed methodology to obtain sparse PC is applied (Nakama et al., 2020). As a case study, the identifiability of metabolic flux ratios of the central metabolism of *Pseudomonas aeruginosa* producing polyhydroxyalkanoates (PHA) is evaluated. PHA is a group of natural biodegradable polyesters and, to make its production economically feasible, their yield must be increased.

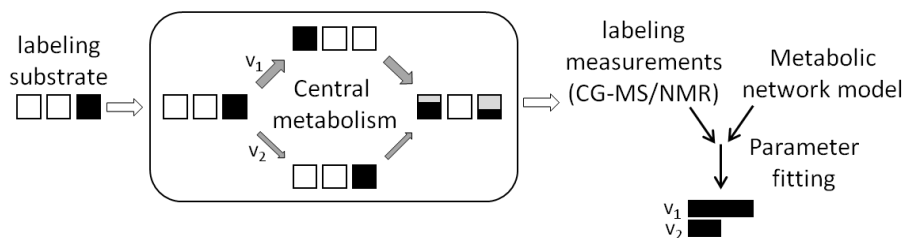


Figure 1: ^{13}C -MFA method flow diagram. The labeling substrate is metabolized by the cell that will shuffle the labeling pattern depending on the flux distribution. The labeling pattern of specific metabolites is measured by analytical techniques and using a metabolic network model the flux distribution can be estimated.

2. Methodology

The ^{13}C -MFA metabolic model consists of metabolite and carbon labeling (x) balances. The metabolic fluxes (v) were parameterized using the metabolite balances and

calculated in terms of the so-called free fluxes (v_{free}). The approach for parametrization of fluxes variables presented by Wiechert et al. (1999), where a bidirectional reaction has a net flux (v_{net}) and an exchange flux (v_{xch}), was also applied. The fluxes are then represented by

$$v = g(v_{free}), \quad v_{net} = v^{\rightarrow} - v^{\leftarrow}, \quad v_{xch} = \min(v^{\rightarrow}, v^{\leftarrow}) \quad (1)$$

The labeling balance (2) is a system of bilinear equations. By performing a linear transformation of the variables (Wiechert et al., 1999), it is possible to compute the sensitivities of the labeling of measured components with respect to the free fluxes in an explicit form (3).

$$f(x, x_{inp}, v_{free}) = 0, \quad y = h(x) \quad (2)$$

$$\frac{\partial y(x, v_{free})}{\partial v_{free}} = \frac{\partial y(x)}{\partial x} \cdot \frac{\partial x(v_{free})}{\partial v_{free}} \quad (3)$$

where x_{inp} is the labeling of the substrate and y the labeling of the measurements. PCA was applied to assess the identifiability of the parameters of the nonlinear model and select a subset of parameters that can be estimated with high accuracy (Vajda et al., 1989). The method consists in computing the Hessian matrix (H) obtained by the Gauss-Newton approximation, using normalized sensitivities. Then, its PCs are obtained with the Singular Value Decomposition (SVD) as follows:

where V is the matrix of PCs (eigenvectors of H) and Σ the matrix with the eigenvalues of H. The PCs of H represent linear combinations of the original parameters, which may indicate dependencies between parameters. However, because these matrices are normally dense, finding combinations with a reduced number of relevant parameters can be challenging. Recently, Nakama et al. (2020) developed a method to compute sparse PC that are orthogonal to the components associated with small eigenvalues, which can be particularly blamed for the large variance in the parameters. In this work, both PC and sparse PC were computed to evaluate the identifiability of metabolic fluxes. The ^{13}C -MFA metabolic model was implemented in MATLAB_{v2015a} and the sparse eigenvector decomposition was calculated in Julia.

3. Case study

To evaluate the efficiency of this new methodology, a small but challenging estimation problem of ^{13}C -MFA was selected. This case study is based on the work of Riascos et al. (2013), which consists of the estimation of metabolic flux ratios of the central metabolism of *P. aeruginosa* producing PHA with only few measurements available. Usually, the ^{13}C -MFA method is applied under growth conditions and the labeling measurements of amino acids are available (Sauer, 2006). However, PHA production by *P. aeruginosa* mainly occurs under non-growth conditions and, in this case, only the measurements of PHA mass isotopomer are available (Riascos et al., 2013). The metabolic network of *P. aeruginosa* used in this work is presented in Figure 2. It uses glucose as the sole carbon source and PHA is the only bioproduct. There are two

options to metabolize glucose into PHA, either by the Pentose Phosphate (PP) pathway, or by the Entner-Doudoroff (ED) pathway, which can also operate in a cyclic mode. Riascos et al.(2013) estimated the metabolic fluxes ratios on *P. aeruginosa* using data from an experiment with 80% of [U-¹³C]glucose and made the hypotheses that the ED pathway operates only in linear mode and the PP pathway without reversible reactions. However, those hypotheses are not necessarily valid and, in this case, these pathways would shuffle the labeling of PHA mass isotopomers (Kohlstedt and Wittmann, 2019).

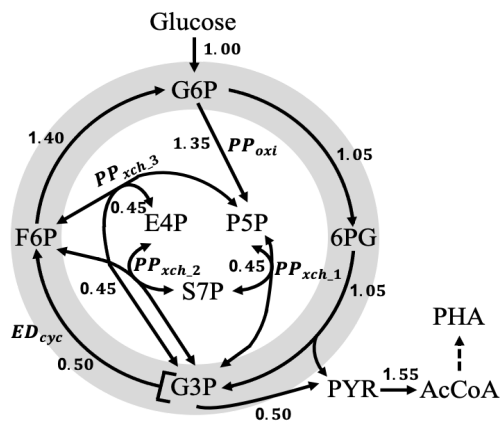


Figure 2: *Pseudomonas aeruginosa* metabolic network.

In this work, the effect of considering the aforementioned pathways in the fluxes ratio estimation problem is evaluated by an identifiability analysis. The network has 15 reactions (including three reversible reactions from the PP pathway), 9 internal metabolites, and glucose uptake is fixed at the unity. Consequently, the system has five degrees of freedom, that is, five free fluxes. The chosen free fluxes are: the net flux of the oxidative PP pathway (PP_{oxi}); the net flux of the cyclic mode of the ED pathway (ED_{cyc}); and the exchange fluxes of the PP pathway ($PP_{xch.1}$, $PP_{xch.2}$, and $PP_{xch.3}$).

4. Results

The results presented in this section are based on a flux distribution similar to the one used by Riascos et al. (2013), in which part of the flux through the ED pathway was directed to its cyclic mode (Figure 2). PCs associated with small eigenvalues implies in a large variance of the estimates. Based on the standard deviations of measurements presented by Riascos et al. (2013), eigenvalues lower than 10^{-4} are considered small (Vajda et al., 1989). The eigenvectors (v) and eigenvalues (σ) of the Hessian matrix for the [U-¹³C]glucose experiment are presented in Table 1. Only one of the eigenvalues is above the threshold, indicating that the problem has identifiability issues with dependencies between parameters. The only component that can be estimated with a certain confidence is the flux through the cyclic mode of the ED. After analyzing a few commercially available labeled glucose, substrate [6-¹³C]glucose presented the best results regarding parameter identification. As it can be seen in Table 1, only two components are below the threshold. Although the problem still presents identifiability

issues, more parameters can be estimated with higher confidence with this experiment. Obtaining a more precise and physically meaningful combination of the parameters from data is not always a simple task (Vajda et al., 1989). For this reason, the sparse eigenvectors of the hessian matrix were calculated.

Table 1: Eigenvectors (PC) and corresponding eigenvalues of the Hessian matrix.

[U- ¹³ C]Glucose					
	V ₁	V ₂	V ₃	V ₄	V ₅
PP _{oxi}	-0.19	-0.06	0.08	-0.97	-0.14
ED _{cyc}	-0.92	0.36	0.04	0.16	0.01
PP _{xch1}	-0.08	-0.33	0.83	0.04	0.44
PP _{xch2}	-0.08	-0.29	0.33	0.19	-0.87
PP _{xch3}	-0.33	-0.82	-0.45	0.06	0.15
σ^2	0.17	8e-18	9e-19	5e-19	2e-20
[6- ¹³ C]Glucose					
	V ₁	V ₂	V ₃	V ₄	V ₅
PP _{oxi}	-0.07	0.35	0.29	-0.69	0.56
ED _{cyc}	0.37	-0.86	0.16	-0.24	0.20
PP _{xch1}	0.59	0.21	-0.32	0.39	0.59
PP _{xch2}	0.59	0.21	-0.32	-0.50	-0.50
PP _{xch3}	0.41	0.22	0.82	0.25	-0.20
σ^2	60.7	0.69	0.004	9e-16	1e-16

Sparse PCs were obtained using the methodology proposed by Nakama et al. (2020). Components v_4 and v_5 were removed and the other components were sparsified, keeping the orthogonality to v_4 and v_5 . Sparse PCs for the [6-¹³C]glucose experiment using the complete network are presented in Table 2. Now some combinations of parameters can be easily identified; by analyzing component one, it is clear that the objective function only depends on the PP_{xch1}/PP_{xch2} combination, not on these fluxes separately. Component v_2 indicates that flux ED_{cyc} can be estimated individually. A similar analysis can be applied to component v_3 , and PP_{xch3}, and flux PP_{oxi} cannot be individually estimated from this experiment.

The metabolic network for *P. aeruginosa* can differ among strains, hence two other typical networks were also considered, the Δ ED_{cyc} and the Δ PP_{oxi} strains. For the former, in which the ED pathway cannot operate in cyclic mode (Table 2), PP_{oxi} could be

estimated separately, which was the original goal of Riascos et al. (2013). The relation among the exchange fluxes of the PP pathways practically remains unchanged; however, component v_1 now has no influence from PP_{xch3} . The other common variation of the *P. aeruginosa* network considered is the absence of the oxidative branch of the PP pathway (Table 2). In this case, the situation is similar to the complete network without the PP_{oxi} influence on the components.

Table 2: Sparse PC of the Hessian matrix for the [6- ^{13}C]glucose experiment.

	complete network			ΔED_{cyc}			ΔPP_{oxi}		
	v_1	v_2	v_3	v_1	v_2	v_3	v_1	v_2	v_3
PP_{oxi}	0.00	-0.33	0.34	0.00	1.00	0.00	x	x	x
ED_{cyc}	0.00	0.94	0.00	x	x	x	0.00	1.00	0.00
PP_{xch1}	-0.67	0.00	0.00	-0.71	0.00	0.00	-0.71	0.00	0.00
PP_{xch2}	-0.67	0.00	0.00	-0.71	0.00	0.00	-0.71	0.00	0.00
PP_{xch3}	-0.29	0.00	0.94	0.00	0.00	1.00	0.00	0.00	1.00

5. Conclusion

Identifiability analysis is an essential tool for research in which ^{13}C -MFA is applied. This analysis can reduce experimental costs and increase the precision of the estimated metabolic fluxes. Here, it was demonstrated that the application of a sparse PC decomposition can improve the identification of dependencies among parameters and the determination of estimable fluxes from available data. In the case study of *P. aeruginosa* producing PHA, the identifiability analysis indicated that the substrate [6- ^{13}C]Glucose allows for the estimation of the fluxes through the ED pathway in cyclic mode and the oxidative branch of the PP pathway for some networks. The method presented in this work enables the identification of parameter combinations that can be estimated from labeling experiments. This represents an advantage when compared to the methods presented in literature that only verifies whether a subset of parameters is identifiable, leading to a combinatorial problem (van Winden et al., 2001; Kappelmann et al., 2016).

Acknowledgments

The authors would like to express their acknowledgement to CAPES (001) and CNPq (308714/2019-9 and 312049/2018-8 fellowships) for their financial support.

References

- J. Kappelmann, W. Wiechert, S. Noack, Mar. 2016. Cutting the Gordian Knot: Identifiability of anaplerotic reactions in *Corynebacterium glutamicum* by means of ^{13}C -metabolic flux analysis. *Biotechnol. Bioeng.* 113 (3), 661–674.

- M. Kohlstedt, C. Wittmann, Mar. 2019. GC-MS-based ^{13}C metabolic flux analysis resolves the cyclic glucose metabolism of *Pseudomonas putida* KT2440 and *Pseudomonas aeruginosa* PAO1. *Metabolic Engineering*.
- C. S. M. Nakama, G. A. C. Le Roux, V. M. Zavala, Aug. 2020. Optimal constraint-based regularization for parameter estimation problems. *Computers & Chemical Engineering* 139, 106873.
- C. A. M. Riascos, A. K. Gombert, L. F. Silva, M. K. Taciro, J. G. C. Gomez, G. A. C. Le Roux, Jan. 2013. Metabolic pathways analysis in PHAs production by *Pseudomonas* with ^{13}C -labeling experiments. *Computer Aided Chemical Engineering* 32, 121–126.
- U. Sauer, Nov. 2006. Metabolic networks in motion: ^{13}C -based flux analysis. *Molecular Systems Biology* 2.
- S. Vajda, H. Rabitz, E. Walter, Y. Lecourtier, Sep. 1989. Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. *Chemical Engineering Communications* 83 (1), 191–219.
- W. A. van Winden, J. J. Heijnen, P. J. T. Verheijen, J. Grievink, Sep. 2001. A priori analysis of metabolic flux identifiability from ^{13}C -labeling data. *Biotechnol. Bioeng.* 74 (6), 505–516.
- W. Wiechert, M. Möllney, N. Isermann, M. Wurzel, A. A. d. Graaf, 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnology and Bioengineering* 66 (2), 69–85.

Metabolic Network Reconstruction from Time-Series of Concentration Data: Evaluation of the Numerical Matrices Methods

Severo Balasbas III,* Ivan Ivanov, Kai Sundmacher

^aProcess Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, D-39106 Magdeburg, Germany

**Corresponding author: sbalasbasiii@mpi-magdeburg.mpg.de*

Abstract

The Numerical Matrices Methods (NMM) are assessed regarding their potential for reconstructing the topology and kinetics of biochemical reaction networks featuring increasing complexity, from a single-enzyme system to a three-enzyme cyclic network. The eventual goal is to use this analysis framework for synthetic *in vitro* enzymatic networks for which kinetic data are available. The NMM use time-dependent concentration data of the chemical species to reconstruct the network and determine its kinetic parameters. Preliminary results indicate that the methods are able to identify the correct network structure and kinetics, but have problems dealing with rate constants deviating from each other by several orders of magnitude. This finding is particularly relevant for biochemical systems with highly variable kinetics. Future addressing of the problem will enable the NMM to accommodate fast reaction kinetics and aid the establishment of a more robust model-based design and optimization framework.

Keywords: metabolic engineering, enzymes, Numerical Matrices Methods, network reconstruction, biochemical kinetics

1. Introduction

There is an increasing interest in *in vitro* metabolic engineering due to the advantages of using cell-free ensembles of purified enzymes or cell lysates for the production of compounds of biotechnological interest (Dudley et al. 2015). Synthetic *in vitro* metabolic networks (*SivMeNs*) operating in cell-free systems are easier to manipulate and optimize due to the reduced complexity compared to living cells – the interconnectivity and cell-wide regulations of the latter are yet to be properly understood, even for model organisms such as *Escherichia coli*. The complete decoupling of the enzymatic pathway of interest from other cellular processes allows cell-free systems to redirect all the input biomass towards product formation, with consequent increase in product yields (Erb 2019, Bowie et al. 2020).

An example of such a *SivMeN* is the CETCH cycle, designed for biochemical carbon fixation. This cycle can convert two molecules of CO₂ to one molecule of either glyoxylate or malate, both of which are two-carbon metabolites (Schwander et al. 2016). The introduction of engineered enzymes was among the many improvements carried out during the course of its development. Addition of methylsuccinyl-CoA oxidase (Mco) in the CETCH cycle allowed oxidation of methylsuccinyl-CoA by molecular oxygen instead of relying on artificial electron acceptors such as ferrocenium, while the engineered

proofreading enzyme propionyl-CoA oxidase (Pco) prevented metabolite loss due to the build-up of dead-end metabolites such as malyl-CoA.

As reaction kinetics govern the behaviour of *Siv*MeNs, kinetic modelling is essential for pathway design and further optimization. Metabolic network reconstruction and kinetic parameter identification from experimental data may prove useful for gaining additional insights into pathway performance improvement. For example, nodes engaged in unwanted and previously unknown side reactivities can be pinpointed through network structure identification.

Several methods have been developed to deduce the connectivity of chemical species participating in a reaction network from time-series concentration data. For example, in the correlation metric construction (CMC) method, a map of connections is determined through analysis of time-lagged concentration correlation functions of two chemical species at a time (Arkin et al. 1995, Arkin et al. 1997). In comparison, the Numerical Matrices Methods (NMM) are based on the analysis of correlations between kinetic complexes to determine not only the network structure, but also the kinetic parameters using linear algebra and statistical methods. The present work aims at evaluating the capabilities and drawbacks of NMM for analysis of cyclic *Siv*MeNs such as the CETCH cycle.

2. Numerical Matrices Methods (NMM)

NMM was developed for network structure identification as well as for the determination of the respective rate constants through least squares parameter estimation (Karnaukhov et al. 2007). The NMM use the formalisms of kinetic complexes, a key aspect of the chemical reaction network theory according to which linear, bilinear and quadratic rate expressions are applied to identify all possible uni- and bimolecular interactions among chemical species, that may occur in a given reaction network. A short survey of the tools included in the NMM is illustrated in Figure 1.

The first tool, called the Kinetic Matrix Method (KMM), expresses the rate of change of the concentrations $X_j(t_i)$ of the chemical species as a product between a kinetic matrix \tilde{A} and a vector of kinetic complexes \tilde{F} , as shown in Figure 1b. The vector \tilde{F} is calculated from the exponentiation of the species vector $\tilde{X}(t_i)$ with the matrix \tilde{Y} formed from the set of complex vectors \hat{y}_k , as stated in equation (1). The species vector $\tilde{X}(t_i)$ has a number of ns rows, equal to the number of chemical species involved, where its elements represent the concentration X_j of the chemical species j at time point t_i .

$$\tilde{F}(t_i) = [\tilde{X}(t_i)]^{\tilde{Y}} = \prod_{j=1}^{ns} [X_j(t_i)]^{y_{jk}} \quad (1)$$

The KMM initially reconstructs a primary kinetic matrix with null columns, each of which contains elements with values below a certain threshold ε_k . Such columns represent complexes that do not contribute to the network dynamics. A second KMM with more accurate values of the matrix \tilde{A} can be executed through the removal of the non-relevant complexes from the initial matrix \tilde{Y} of complex vectors. Afterwards, the Representation Matrix Method (RMM) is used to decompose the kinetic matrix \tilde{A} into a set of representation matrices \tilde{G}_p and a vector of nonzero parameters $\hat{k} = k_p$ that contains

np nonzero elements of \tilde{A} with absolute values above a certain threshold ε_k , as shown in Figure 1c.

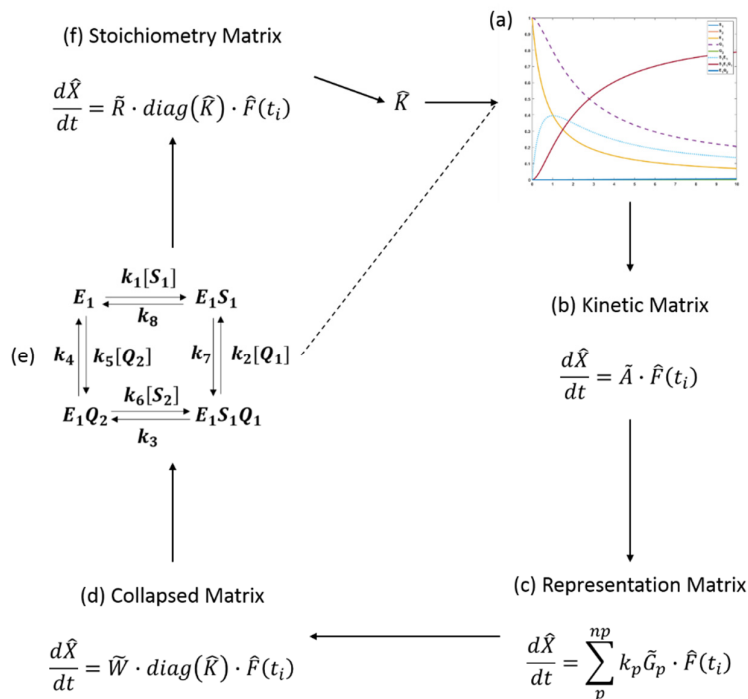


Figure 1. Overview of the tools of the Numerical Matrices Methods. Starting from time series of concentration data for all the chemical species participating in a given reaction network (a), a sequence of matrix-based numerical steps (b-f) is executed that finally leads to the reconstruction of both the network stoichiometric matrix \tilde{R} and the vector of rate constants \tilde{K} .

Then the stoichiometric matrix \tilde{R} is reconstructed from the set of representation matrices \tilde{G}_p through an intermediate “collapsed” kinetic matrix \tilde{W} (see Figure 1d). The columns of \tilde{W} corresponding to multivariant reactions (or reactions that share the same reactant complex, but with different product complexes) are then decomposed into reaction vectors of the stoichiometric matrix \tilde{R} . In the complete absence of multivariant reactions in the chemical reaction network of interest, \tilde{W} will be equal to \tilde{R} .

Once the stoichiometric matrix \tilde{R} is fully known (see Figure 1e), the last method called the Stoichiometric Matrix Method (SMM) is used to calculate the vector \tilde{K} of rate constants. The SMM mathematically expresses the rate of change of $X_j(t_i)$ as a function of the matrix \tilde{R} and the vector \tilde{K} as shown in Figure 1f.

3. Numerical Experiments

MATLAB-based numerical experiments were conducted to test the capabilities of NMM in reconstructing the kinetic mechanisms governing enzyme-catalyzed reaction networks.

In particular, the NMM were tested for its ability to reconstruct the reaction network structure and determine the values of the rate constants starting from time series of concentration data. To achieve this goal, *in silico* data sets describing the concentration profiles over time of three models of enzyme-catalyzed reactions were generated and used as input information for the NMM-based reconstruction approach. The first model represents a single-enzyme mechanism, the second model represents a two-enzyme cascade, and the third model represents a three-enzyme cycle that resembles a strongly simplified CETCH cycle. In all of these cases, the enzymes were modeled to follow the compulsory-ordered ternary complex mechanism. Figure 2 shows the schematic representation of the three enzymes that participate in the designed models.

The aforementioned *in silico* data sets were generated through numerical integration of the differential equations representing the mass balances of the chemical species participating in the three enzymatic networks. Due to the nature of these equations, the generated datasets also contained complete information about the concentration dynamics of the transitional enzyme-substrate complexes. Such information is usually not present in concentration profiles gathered under laboratory conditions. During numerical integration, the maximum chosen duration was $t_{max}=10$, with concentration values known at a resolution of 0.01 mol/l.

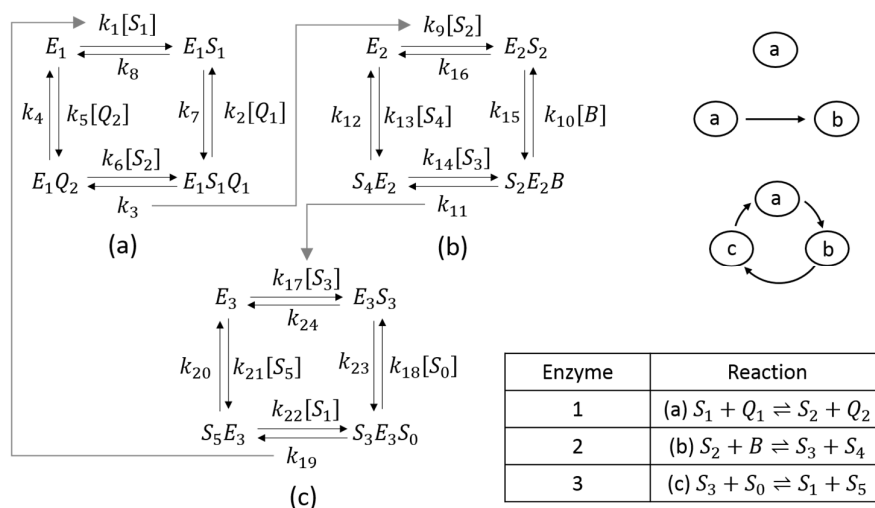


Figure 2. Diagrams that represent the compulsory-ordered ternary complex mechanism that governs the kinetics of the enzymes considered in the network models. The single-enzyme model contains the rate expressions of (a). The two-enzyme model contains the rate expressions of both (a) and (b). The cyclic three-enzyme model contains the rate expressions of (a), (b) and (c).

The values of the rate constants used during synthetic data set generation were chosen to ensure that the enzyme-substrate complex formation steps are fast reactions, while the product formation steps were considered as the rate-determining steps. The difference between the rate constant values of the slowest and the fastest steps of both forward and reverse reactions are chosen to reach up to three or more orders of magnitude. Such a range of values mimics biological reality, as similarly huge differences between rate constants have been observed in enzymes such as the yeast hexokinase (Noat et al. 1968)

and bovine heart mitochondrial malate dehydrogenase (Heyde et al. 1968). The chosen values for each of the rate constants are given in Table 1 and were later varied, as explained in section 4.

To enable faster and more accurate determination of the kinetic matrix \tilde{A} during the execution of the KMM, a preliminary elimination of the complex vectors was carried out. Thereby, we excluded species' interactions that usually do not occur in enzyme catalysis. Examples of such kinetic complexes include the ones that correspond to interactions between small molecules that are not bound to their respective enzymes or enzyme-substrate complexes, interactions between unbound enzymes, interactions between enzyme-substrate intermediates, and interactions between free and bound enzymes. It was also assumed that no dimer formation occurs during the operation of the three network models. For this reason, the complex vectors representing such interactions were eliminated beforehand. We noted that these complex interactions may need to be reconsidered in the future, in order to better reflect the biological reality, for instance with respect to signaling pathways in which interactions between different enzymes are present.

In enzymatic reactions and networks, the linear dependencies among the chemical species due to the conservation of mass in the system of reactions lead to singularities of the kinetic matrix \tilde{A} . To break these *holonomic conditions* intrinsic to stoichiometric matrices of reaction networks, a sufficient number of time-dependent concentration profiles with different initial concentration conditions were collected. The number of time-series of concentration data nm , and the initial concentration values used, were chosen according to a fractional factorial design. Two-level designs were chosen, with the normalized lower level assigned with the value of 0.05 and the normalized higher level with the value of 0.5. These values are given only to the concentrations of the unbound enzymes, substrates, cofactors and products, while to the rest of the chemical species we assigned the initial concentration values of zero during dataset generation. For the single-enzyme kinetic model, a 2^{3-1} design was chosen. For the two-enzyme kinetic model, a 2^{9-4} design was sufficient to break the holonomic constraints. For the three-enzyme kinetic model, a 2^{11-4} design is sufficient, as long as a twelfth factor has non-zero initial concentration values (either 0.05 or 0.5 in any of the 32 numerical experiments).

4. Results

Karnaikhov et al. (2007) tested the NMM by using a network model representing a hypothetical assembly mechanism of a quaternary complex between ribosomal RNA and three RNA-binding proteins. The used rate constant values were in the range of 0.001–1.3. Using the same range of rate constant values for our three network models during numerical dataset generation and setting the threshold $\epsilon_K = 0.0005$, the NMM managed to properly reconstruct the stoichiometric matrices and the values of all rate constants.

But in the case of the three-enzyme network, the SMM gave incorrect stoichiometric matrices when the rate constant values of the first step of the reverse reactions catalyzed by the three enzymes (i.e. $k_{5,13,21}$) were one order of magnitude smaller than the values of the next step involving the binding of substrate. If at least one of these rate constants had a value that was in the same order of magnitude as that of the next step, then the NMM properly determined the correct entries of the stoichiometric matrix. If the rate constant values were changed to get a difference of more than three orders of magnitude between the slowest and fastest steps (e.g. setting either $k_1 \geq 10$ or $k_3 < 10^{-3}$), the SMM

calculations became very slow, as the kinetic matrix \tilde{A} at this point assumed a very high rank and condition number. A similar issue occurred in the cases of the simpler single-enzyme and two-enzyme cascade models, where differences of more than three orders of magnitude between the slowest and fastest steps resulted in incorrect dimensions and incorrect entries of the reconstructed stoichiometric matrix.

Table 1. The non-dimensional rate constants chosen during numerical dataset generation.

Rate constants	Values used	Rate constants	Values used
k_1	1.3	$k_{8,16,24}$	0.011
k_2	0.75	k_9	1.03
k_3	0.001	k_{10}	0.6
$k_{4,12,20}$	0.009	k_{11}	0.006
$k_{5,13,21}$	0.5	k_{17}	1.1
$k_{6,14,22}$	0.85	k_{18}	0.95
$k_{7,15,23}$	0.009	k_{19}	0.005

5. Conclusion

The Numerical Matrices Methods (NMM) are useful tools for kinetic network reconstruction from time-dependent concentration data. However, problems arise when the rate constants in a given network differ by more than three orders of magnitude between fast and slow steps (i.e. the rate-determining steps). The findings suggest treating the network reconstruction task better as a dynamic optimization problem, constrained by a differential-algebraic equation (DAE) system, where the fastest steps are modelled as chemical equilibrium constraints. As rate constants are often not known *a priori*, one should create a family of kinetic network alternatives, from which the best network could be identified by systematic model discrimination. In this way, the model-based design and optimization of synthetic *in vitro* metabolic networks (SivMeNs) can be addressed.

References

- Dudley, Q. M., A. S. Karim and M. C. Jewett (2015). "Cell-free metabolic engineering: Biomanufacturing beyond the cell." *Biotechnology Journal* **10**(1): 69-82.
- Erb, T. J. (2019). "Structural organization of biocatalytic systems: the next dimension of synthetic metabolism." *Emerging Topics in Life Sciences* **3**(5): 579-586.
- Bowie, J. U., S. Sherkhanov, T. P. Korman, M. A. Valliere, P. H. Opgenorth and H. Liu (2020). "Synthetic Biochemistry: The Bio-inspired Cell-Free Approach to Commodity Chemical Production." *Trends in Biotechnology* **38**(7): 766-778.
- Schwander, T., L. S. von Borzyskowski, S. Burgener, N. S. Cortina and T. J. Erb (2016). "A synthetic pathway for the fixation of carbon dioxide *in vitro*." *Science* **354**(6314): 900-904.
- Arkin, A. and J. Ross (1995). "Statistical Construction Of Chemical-Reaction Mechanisms From Measured Time-Series." *Journal of Physical Chemistry* **99**(3): 970-979.
- Arkin, A., P. D. Shen and J. Ross (1997). "A test case of correlation metric construction of a reaction pathway from measurements." *Science* **277**(5330): 1275-1279.
- Karnaikhov, A. V., E. V. Karnaikhova and J. R. Williamson (2007). "Numerical Matrices Method for nonlinear system identification and description of dynamics of biochemical reaction networks." *Biophysical Journal* **92**(10): 3459-3473.
- Noat, G., J. Ricard, M. Borel and C. Got (1968). "Kinetic Study of Yeast Hexokinase. I. Steady-State Kinetics." *European Journal of Biochemistry* **5**(1): 55.
- Heyde, E. and S. Ainswort (1968). "Kinetic Studies On Mechanism Of Malate Dehydrogenase Reaction." *Journal of Biological Chemistry* **243**(9): 2413.

Assessment of intermediate storage and distribution nodes in personalised medicine

Andrea Bernardi^a, Maria Papathanasiou^{a,*}, Matthew Lakelin^b, Nilay Shah^a

^a*Center for Process System Engineering, Imperial College London, SW72AZ, UK*

^b*TrakCel Limited, 10/11 Raleigh Walk, Cardiff, CF104LN, UK*

**maria.papathanasiou1@imperial.ac.uk*

Abstract

Chimeric Antigen Receptor (CAR)-T cell therapies are a type of patient-specific cell immunotherapy demonstrating promising results in the treatment of aggressive blood cancer types. CAR-T cells follow a 1:1 business model, translating into manufacturing lines and distribution nodes being exclusive to the production of a single therapy, hindering volumetric scale up. In this work, we address manufacturing capacity bottlenecks via a Mixed Integer Linear Programming (MILP) model. The proposed formulation focuses on the design of candidate supply chain network configurations under different demand scenarios. We investigate the effect of an intermediate storage option upstream of the network as means of: (a) debottlenecking manufacturing lines and (b) increasing facility utilisation. In this setting, we assess cost-effectiveness and flexibility of a decentralised supply chain and we evaluate network performance with respect to two key performance indicators (KPIs): (a) average production cost and (b) average response treatment time. The trade-off between cost-efficiency and responsiveness is examined and discussed.

Keywords: personalised medicine, supply chain, optimisation.

1. Introduction

Chimeric antigen receptor T cell (CAR-T cells) therapy is a type of adoptive cell immunotherapy (Zhang *et al.* 2017; Guthrie, 2018). T cells are a sub-type of white blood cells that assist cell-mediated immunity. In the case of CAR-T cell manufacturing, T cells are genetically modified and enabled to recognise and kill target cancer cells. Their manufacturing process starts with the extraction of T cells from the donor's (allogeneic) or patient's (autologous) blood stream through a specialised procedure (leukapheresis). The leukapheresis sample is then transferred to the manufacturing facility for further processing that involves cell expansion, genetic modification and Quality Control (QC). The approved therapy is then sent to the hospital and administered to the patient (Levine, 2015). Autologous CAR-T cell therapies have received landmark approvals by the U.S. Food and Drug Administration (FDA) (2017) and the European Medicines Agency (EMA) (2018) and are currently offered for the treatment of recurring, aggressive B-cell lymphoma, under a controlled scheme (EMA, 2018; Novartis, 2018; FDA, 2017). At the moment, autologous CAR-T cells seem to be leading the race in personalised cell therapies, while the allogeneic counterpart is also gaining attention.

A unique feature of autologous CAR-T therapies is that the patient's cells are used as raw material and therefore the therapy manufacturing and distribution that follow are exclusively dedicated to the production and delivery of a single therapy. In such cases, volumetric scale-up is not applicable and is replaced by scale-out solutions, where parallel manufacturing lines are available. In addition, CAR-T cell life cycle involves the

distribution and handling of sensitive materials with short shelf lives. This is translated into tight time constraints that challenge the supply chain network even more. Given the autologous and sensible nature of CAR-T therapies, novel decentralized manufacturing models are considered promising alternatives to a traditional centralized approach, since small scale local manufacturing facilities offer greater flexibility and can significantly reduce transportation costs (Harrison *et al.*, 2018). On the other hand, intermediate storage of the therapies might be necessary to serve as buffer when the demand temporarily exceeds the manufacturing facilities' capacity or to optimise the facility utilisation and is crucial to efficiently coordinate manufacturing and distribution tasks.

In this work, we present a Mixed Integer Linear Programming (MILP) formulation that provides candidate solutions with respect to the location, number and capacity of manufacturing sites, storage duration, and the most suitable mode of transport. In particular, the MILP model is used to assess the supply chain network performance under three different time constraint scenarios (19, 22 and 25 days) total return time. The latter refers to the total duration of the therapy life cycle, starting from the leukapheresis procedure and ending with the delivery of the therapy at the hospital. All the scenarios are assuming a demand of 500 patients per year.

2. Materials and methods

The general supply chain network considered in this work is represented in Figure 1. It consists of 5 nodes in total, namely: leukapheresis site, storage site, manufacturing site, Quality Control and hospital. It should be noted that storage in this case is optional, thus the model is allowed to by-pass it and proceed directly into the manufacturing. The location of both storage and QC facilities is assumed to be co-located with the manufacturing facilities in this work. Nevertheless, the model formulation has the capability of evaluating outsourced storage and QC, whereby the decision becomes a design variable of the supply chain network.

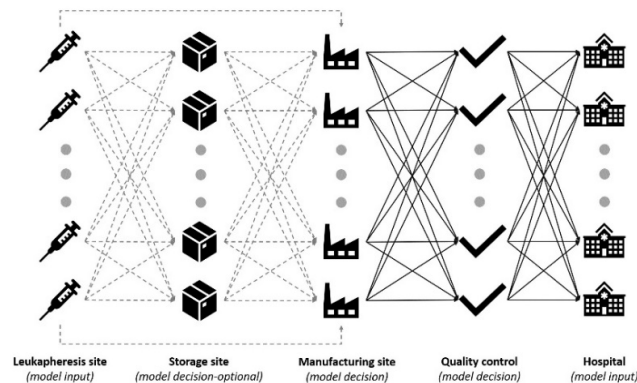


Figure 1: Supply chain network considering 5 nodes: (a) leukapheresis site, (b) storage site, (c) manufacturing site, (d) Quality Control and (e) hospital. Storage of leukapheresis samples is optional allowing the model to by-pass this node and move directly into manufacturing.

Manufacturing facilities of three different capacities are considered. These include facilities of 4, 10 and 31 parallel production lines, corresponding to a maximum capacity of approximately 200, 500 and 1500 therapies per year respectively. Manufacturing costs are based on the information available from Spink and Steinsapir (2018), while QC costs

are proprietary to Trakcel Ltd. The unit transport costs across the various locations are estimated costs from a white glove courier with cell therapy processes. Storage costs consider preservation of samples at -150°C.

Based on economies-of-scale, the variable manufacturing cost per therapy is assumed to be a decreasing function of the percentage at which the respective facility is utilised. This is represented by an originally non-linear formulation that is then linearized using a McCormick envelope (McCormick, 1976). Constraints representing the minimum flow of samples are adapted from Tsiakis *et al.* (2001).

The model considers that manufacturing facilities are placed close to transport hubs, such as airports, ports and train stations to facilitate distribution. In addition, the locations align with existing cell and gene therapy facilities in Europe, US and the UK. In this case, we assume that QC is co-located with the manufacturing facility. We consider that QC facilities are of 20-fold higher capacity when compared to the manufacturing facilities, relaxing therefore QC capacity constraints. Following standard European and UK practices of public and public and/or semi-public healthcare systems, we assume that the choice of collaborating hospitals and specialist centres is not under the manufacturer’s sole control. Therefore, these two nodes are considered as model inputs and not as decision variables. For the leukapheresis sites we consider a capacity of 8 patients per day, while we assume that capacity is not a bottleneck for the administration of the therapy at the hospital site and thus no upper limit is provided. Following the Advanced Therapy Treatment Centre (ATTC) model as discussed in Papathanasiou *et al.* (2020), we assume that leukapheresis sites and hospitals are different facilities, located in the same region. We assume that transportation can happen either within 24 or 48 hours, irrespective of the mode of transport (i.e. car, rail etc). This is to align the model assumptions with standard practice in the cell therapy space where courier contracts are costed based on the estimated time of delivery. An overview of the model formulation is presented in Table 1.

Table 1: Overview of model formulation.

	Index	Mathematical formulation	Description
Obj function	(1)	$C_{tot} = C_m + C_{tr} + C_{QC} + C_{st}$	Total cost
Constraints	(2)	$TRT_p = t_{delivery} - t_{start} \leq U^t$	Total return time of therapy
	(3)	$FCAP_m \cdot E_m - \sum_p INM_{p,m,t} \geq 0$	Manufacturing capacity constraint
	(4)	$\sum_m E_m \leq U^M$	Total number of manufacturing sites
	(5)	$INC_{p,c,t} = OUTC_{p,c,t+TLS}$ $INS_{p,m,t} = OUTS_{p,m,t+Tsd}$ $INM_{p,m,t} = OUTM_{p,m,t+TME}$ $INQ_{p,q,t} = OUTQ_{p,q,t+TQC}$	Sample balance at each node (leukapheresis site, storage, manufacturing and quality control)

We consider a base case where the samples are processed immediately after they reach the manufacturing facility (Scenario A), and four alternative scenarios where the samples can be stored at the manufacturing facility site for up to six days and different time constraints are enforced. Table 2 summarizes the constraints in each scenario. The assumed duration of the manufacturing is 7 days, representing a forward-looking scenario of technology developments that can lead to decreased culture times with respect to current industrial average (~19 days).

Table 2: Summary of the five scenarios considered in this paper.

	Storage	# of manufacturing facilities	U^t
Scenario A	No storage	2	19 days
Scenario B	✓ (up to 6 days)	2	19 days
Scenario C	✓ (up to 6 days)	2	22 days
Scenario D	✓ (up to 6 days)	2	25 days
Scenario E	✓ (up to 6 days)	2	No constraint

3. Results and discussion

The model described above is used for the design and assessment of a supply chain network to deliver 500 therapies per year. The demand distribution is assumed to be homogeneous among each year quarter, for which results are demonstrated here. In all examined scenarios the model is allowed to invest into maximum two manufacturing facilities to be built.

The model has 10,338,522 variables and 2,502,554 discrete variables and all the scenarios have been executed in a 24-Core Xeon E5-2697 machine with 96GB Ram. The scenario with the highest CPU time is Scenario E and it required a CPU time of ~98 hours to reach an optimality gap lower than 3%.

Figure 2a illustrates the results of all five scenarios in terms of average cost and return time of a therapy, Figure 2b focus on Scenarios A, B and C, and Figure 3 reports the utilisation of the manufacturing facilities in Scenario A and C.

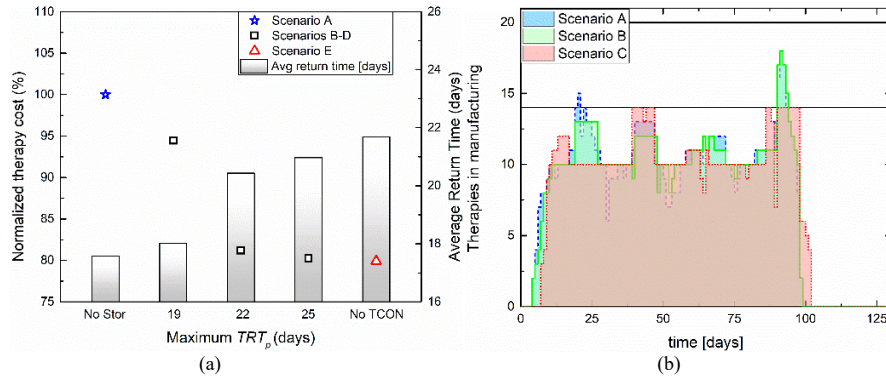


Figure 2: (a) Average therapy cost in Scenario A (blue star), B-D (black squares) and E (red triangle). The bars represent the average return time of a therapy in each scenario. (b) Number of therapies in manufacturing in Scenario A and C. The black horizontal lines represent the thresholds of 14 and 20 therapies.

We can observe that the average therapy cost decreases when storage is chosen as an option, as it allows for improved manufacturing planning. This is achieved by scheduling to manufacture more therapies in parallel, therefore decreasing the operating cost. When the same upper bound for the total return time of a single therapy is considered then a 5% decrease of the cost is achieved with an incremental increase in the average return time (~1%). A sharper decrease of 20% of the average therapy cost can be achieved by increasing the upper bound for the total return time of a therapy to 22 days. Relaxing further this constraint leads to negligible improvements in the cost.

The reason of the decrease of the average therapy cost is explained by the choice and utilisation of the manufacturing facilities. Given the sizes of the manufacturing facilities considered in this work, a different combination of small (4 parallel manufacturing lines) and medium-sized (10 parallel manufacturing lines) is required for scenarios A and C. As we can observe in Figure 2b, in Scenario A the threshold of 14 therapies in manufacturing, which can be satisfied by a small and a medium facility, is temporarily exceeded in a small subset of days. On the other hand, in Scenario C the same total demand can be satisfied by one medium and one small facility, as the threshold of 14 therapies in manufacturing is never exceeded. It is important to note that in Scenario B two medium-sized facilities are still needed in order to satisfy the demand.

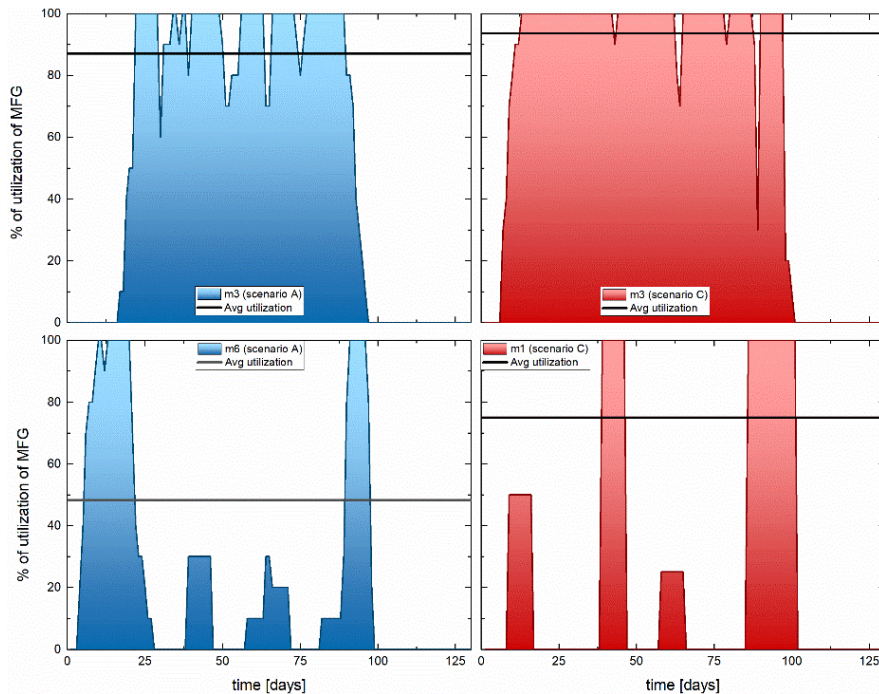


Figure 3: The graphs on the left represents the utilisation of the manufacturing facilities $m3$ (top) and $m6$ (bottom) in Scenario A, the graphs on the right represents the utilisation of manufacturing facilities $m3$ (top) and $m1$ (bottom) in Scenario C. The black horizontal line in each plot represents the average utilisation.

From Figure 3 we can observe that the facility $m6$ in Scenario A is significantly underutilized, resulting into a higher operating cost per therapy. In Scenario C the main facility $m3$ is almost always operating at maximum capacity, while the smaller facility $m1$ is used only when the demand cannot be satisfied by $m3$ and on average the utilisation is higher than the utilisation of $m6$ in Scenario A, due to its smaller size. It should be mentioned that the average utilisation is calculated excluding the days when the facilities are not in operation.

4. Conclusions

In this work we present a Mixed Integer Linear Programming (MILP) formulation for the design and optimisation of supply chain networks in personalised therapies. The latter present the unique characteristic of a 1:1 business model, where manufacturing and distribution lines are occupied for the production and delivery of a single therapy. This is translated into increased per-therapy costs, as well as manufacturing bottlenecks, considering that volumetric scale up is not applicable in such cases. The MILP model investigates the potential of decentralised manufacturing as well as storage in an attempt to create a flexible, cost-efficient network that remains responsive to the patient schedule. We consider different life cycle duration scenarios and an annual demand of 500 patients throughout. Overall, the model indicates that cost decreases as a function of the return time, while the use of multiple facilities may lead to lower operating costs. The latter is related to the variable cost being calculated as an inverse function of the facility utilisation. Lastly, results from this work demonstrate that storing leukapheresis samples upstream of the supply chain can allow for improved manufacturing planning, as well as debottlenecking. This can be attributed to the flexibility added in the overall network that allows samples to be stored until a critical mass is ready to be manufactured in parallel, leading therefore to a higher utilisation percentage of the facility and lower operating cost.

Acknowledgements: The authors would like to acknowledge expert opinion received through multiple conversations with the User Steering Committee of the Future Targeted Healthcare Manufacturing Hub. Funding from the UK Engineering & Physical Sciences Research Council (EPSRC) for the Future Targeted Healthcare Manufacturing Hub hosted at University College London with UK university partners is gratefully acknowledged (Grant Reference: EP/P006485/1).

References

- EMA, 2018. First two CAR-T cell medicines recommended for approval in the European Union. <https://www.ema.europa.eu/en/news/first-two-car-t-cell-medicines-recommended-approval-european-union> [Accessed: 10 Nov 2020].
- FDA, 2017. FDA approves CAR-T cell therapy to treat adults with certain types of large B-cell lymphoma. <https://www.fda.gov/news-events/press-announcements/fda-approves-car-t-cell-therapy-treatadults-certain-types-large-b-cell-lymphoma> [accessed: 10 Nov 2020].
- G. Guthrie, 2018. CAR T-Cell Immunotherapy: The 2018 Advance of the Year. <https://www.cancer.net/blog/2018-01/car-t-cell-immunotherapy-2018-advance-year> [accessed: 10 Nov 2020].
- R. P. Harrison, S. Ruck, Q. A. Rafiq, N. Medcalf, 2018. Decentralised manufacturing of cell and gene therapy products: Learning from other healthcare sectors. *Biotechnology advances* 36 (2), 345–357.
- B. Levine, 2015. Performance-enhancing drugs: design and production of redirected chimeric antigen receptor (CAR) T cells. *Cancer gene therapy* 22 (2), 79–84.
- G. P. McCormick, 1976. Computability of global solutions to factorable nonconvex programs: Part I - Convex underestimating problems. *Mathematical programming* 10 (1), 147–175.
- Novartis, 2018. Kymriah® (tisagenlecleucel), first-in-class CAR-T therapy from Novartis, receives second FDA approval to treat appropriate r/r patients with large B-cell lymphoma. <https://tinyurl.com/y4cgc5dq> [Accessed: 10 Nov 2020].
- M. M. Papathanasiou, C. Stamatis, M. Lakelin, S. Farid, N. Titchener-Hooker, N. Shah, 2020. Autologous CAR T-cell therapies supply chain: challenges and opportunities? *Cancer Gene Therapy*, 1–11.
- K. Spink, A. Steinsapir, 2018. The long road to affordability: a cost of goods analysis for an autologous CAR-T process. *Cell and Gene Therapy Insights* 4, 1105–1116.
- P. Tsiakis, N. Shah, C. C. Pantelides, 2001. Design of multi-echelon supply chain networks under demand uncertainty. *Industrial & Engineering Chemistry Research* 40 (16), 3585–3604.
- C. Zhang, J. Liu, J. F. Zhong, X. Zhang, 2017. Engineering CAR-T cells. *Biomarker research* 5 (1), 1–6.

Thermal treatment in dairy processes: Validation of protein deposition models

Wilfred Kwabena Darko^a, Federico Lozano Santamaria^a, Laurent Bouvier^b, Guillaume Delaplace^b and Sandro Macchietto^{a,*}

^a*Imperial College London, South Kensington Campus, London SW7 2AZ, UK*

^b*Univ. Lille, CNRS, INRAE, Centrale Lille, UMR 8207 - UMET - Unité Matériaux et Transformations, F-59000 Lille, France*

s.macchietto@imperial.ac.uk

Abstract

The optimisation of heat treatment processes in the dairy industry is of high practical interest. Models are key to understanding and optimising equipment design, operating conditions, energy and water utilization, heating and cleaning cycles and economics. Decades of research have led to the development of many models, however, there is still a lack of understanding about the role and mechanism of denaturation and deposition of proteins from whey solutions. Extensive validation of the more recent thermal and fouling models against experimental data is also incomplete.

Here, five sets of detailed data from dynamic experiments in typical Plate Heat Exchangers (PHEs) are used to validate a 2D-distributed, dynamic model under fouling. The model accounts for the exchanger geometry, operating conditions and prevailing local conditions. Three reaction mechanisms for β -lactoglobulin (β -LG) protein unfolding, aggregation and deposition are considered, one of which is novel.

First, a validation was carried out of the thermal PHE model alone (i.e. under clean conditions, with no fouling) against a water-water dynamic experiment, with excellent results. The exit temperatures predictions match the experimental observations within $<1^\circ\text{C}$ on average. The fouling model was then validated against four dynamic water-whey protein solutions (WPS) experiments (hot water heating a cold WPS stream) covering different conditions. Some of the model deposition constants were obtained through parameter estimation. Comparison of predicted and measured quantities (exit temperatures and amount of deposit in each channel at the end of the experiments) show that neither the aggregate protein deposition model nor the unfolded deposition model on their own can describe the observations. It is shown that a new modification that combines both mechanisms results in an improved fitting of the data in all experiments. The combined deposition scheme results indicate that both mechanisms are important in different parts of the exchanger, and even along individual plates.

Keywords: food thermal treatment process, pasteurisation, fouling, experimental model validation, unfolded and aggregate proteins, dynamic simulation, parameter estimation.

1. INTRODUCTION

Thermal treatments such as pasteurisation and ultra-high-temperature (UHT) processes, widely used in the dairy and food industries to insure product hygienic safety and long

shelf life. They typically use plate heat exchangers (PHEs). Fouling reduces PHEs thermal efficiency, requires frequent cleaning that generates large volumes of wastewater and increases costs. The optimisation of such processes is therefore of high practical interest. Reliable, predictive models of heat transfer and fouling are key to optimise equipment design, operating conditions, energy and water utilization, heating and cleaning cycles, productivity and economics. Decades of experimental and modelling research have led to the development of many models to characterise, in particular, the role and mechanism of denaturation and deposition of whey proteins (for example, but not limited to, β -lactoglobulin, β -LG) in dairy derivatives (Loveday, 2016, Blanpain-Avet, et al., 2016, Khaldi, 2016, Sharma and Macchietto, 2019). However, there is still uncertainty about the denaturation/deposition mechanisms, and an extensive validation of the most recent models against experimental data is still missing. Here, the objective is to assess a recently developed dynamic model of PHEs subject to fouling, developed at Imperial College London, against some high quality experimental data collected at UMET/INRAE, France. The work is part of a collaboration aimed at the development and validation of improved fouling and cleaning technology.

1.1. Fouling reaction model

Above a certain temperature β -LG undergoes denaturation whereby the native protein (N) changes to an unfolded form (U) with highly reactive thiol groups exposed. The unfolded proteins irreversibly react to form insoluble aggregates (A). All the reactions are assumed to begin at 70°C, and the unfolding and aggregation reactions to occur in both bulk fluid and boundary layers with the same kinetic parameters. The unfolding reaction is reported to be limiting below 80°C and the aggregation reaction limiting above 80°C. Controversy still exist in literature concerning the denaturated protein species (unfolded and/or aggregated) which are supposed to form the deposit (Bansal and Chen, 2006; Sadeghinezhad et al., 2013, Blanpain-Avet et al., 2016). Consequently, in addition to deposition arising from Aggregates (Type A model) and Unfolded (Type B model) proteins alone, a new model that combines them (Type C model) is proposed (Figure 1). The kinetics parameters of the denaturation reactions are taken from (Khaldi, 2016) and those of the deposition reactions from (De Jong, 1996). The deposition proportionality parameters, β_U and/or β_A , were estimated from the channel mass deposit data. Full details and parameters values are given in Darko, 2020.

1.2. Plate Heat Exchanger model

The PHE dynamic model considers a whey protein fouling fluid being heated as flowing in a channel delimited by two hot plates (Guan and Macchietto 2018). The model is distributed in 2 directions, longitudinally (coordinate x) and across (coordinate y), considering five domains (Figure 2, from left to right): Ω_W^L : Left plate wall domain; Ω_L^L : Left plate deposit layer domain; Ω_F : Channel fluid flow domain; Ω_L^R : Right plate deposit layer domain; Ω_W^R : Right plate wall domain. A uniform distribution is assumed along the plate width, W . Differential mass and energy balances in the 5 domains plus the above reaction models define 2D-distributed temperature profiles, heat fluxes, and deposition rates at each location of each plate, according to the local thermal and hydraulic conditions. A similar model describes the hot fluid channel (for a non-fouling fluid, the fouling reactions are omitted). The models of all channels are then integrated into an overall model (Figure 3, left) through appropriate boundary conditions, according to the PHE design and flow configuration. The resulting partial differential and algebraic equation (PDAE) system was implemented in a commercial simulator (gPROMS, 2018).

The full dynamic model is used both in simulation mode (to calculate all outputs given all parameters and inputs) and in parameter estimation mode (to estimate the best parameters given all inputs and selected output measurements). Details of model and solution method are given in Sharma and Macchietto, 2019 for a single PHE and in Zhu et al., 2020 for application to a complete heat treatment process.

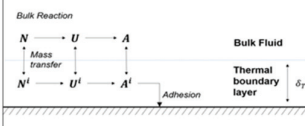
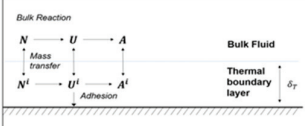
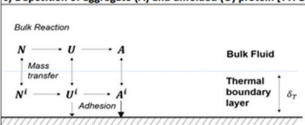
<p>a) Deposition of aggregated (A) protein [TYPE A]</p>  <p> h_f Convective heat transfer coefficient λ_d Deposit thermal conductivity σ_T Thermal boundary layer thickness </p>	<p>Reactions in the bulk of the fluid</p> $r_{fN} = k_{oN} \exp\left(-\frac{E_N}{RT_{fluid}}\right) C_N$ $r_{fU} = \begin{cases} k_{oU1} \exp\left(-\frac{E_{U1}}{RT_{fluid}}\right) C_U^2, & T_{fluid} \leq T^* \\ k_{oU2} \exp\left(-\frac{E_{U2}}{RT_{fluid}}\right) C_U^2, & T_{fluid} > T^* \end{cases}$ <p>Deposition model</p> $\frac{\rho_d}{\sigma_T} \frac{d\delta_A}{dt} = \begin{cases} r_{dA} - r_{eA} = \frac{\rho_d}{\sigma_T} \left[\beta_A \frac{\lambda_d}{h_f} \left(k_{dA} \exp\left(-\frac{E_{dA}}{RT_{film}}\right) C_A^i - \kappa_A \delta_A \right) \right], & T_{fluid} \geq T_{min} \\ 0, & T_{fluid} < T_{min} \end{cases}$	<p>Reactions in the boundary layer (BL)</p> $r_{bN} = k_{oN}^i \exp\left(-\frac{E_N^i}{RT_{BL}}\right) C_N^i$ $r_{bU} = \begin{cases} k_{oU1}^i \exp\left(-\frac{E_{U1}^i}{RT_{BL}}\right) (C_U^i)^2, & T_{fluid} \leq T^* \\ k_{oU2}^i \exp\left(-\frac{E_{U2}^i}{RT_{BL}}\right) (C_U^i)^2, & T_{fluid} > T^* \end{cases}$
<p>b) Deposition of unfolded (U) protein [TYPE B]</p>  <p> h_f Convective heat transfer coefficient λ_d Deposit thermal conductivity σ_T Thermal boundary layer thickness </p>	<p>Reactions in the bulk of the fluid</p> $r_{fN} = k_{oN} \exp\left(-\frac{E_N}{RT_{fluid}}\right) C_N$ $r_{fU} = \begin{cases} k_{oU1} \exp\left(-\frac{E_{U1}}{RT_{fluid}}\right) C_U^2, & T_{fluid} \leq T^* \\ k_{oU2} \exp\left(-\frac{E_{U2}}{RT_{fluid}}\right) C_U^2, & T_{fluid} > T^* \end{cases}$ <p>Deposition model</p> $\frac{\rho_d}{\sigma_T} \frac{d\delta_U}{dt} = \begin{cases} r_{dU} - r_{eU} = \frac{\rho_d}{\sigma_T} \left[\beta_U \frac{\lambda_d}{h_f} \left(k_{dU} \exp\left(-\frac{E_{dU}}{RT_{film}}\right) C_U^i - \kappa_U \delta_U \right) \right], & T_{fluid} \geq T_{min} \\ 0, & T_{fluid} < T_{min} \end{cases}$	<p>Reactions in the boundary layer (BL)</p> $r_{bN} = k_{oN}^i \exp\left(-\frac{E_N^i}{RT_{BL}}\right) C_N^i$ $r_{bU} = \begin{cases} k_{oU1}^i \exp\left(-\frac{E_{U1}^i}{RT_{BL}}\right) (C_U^i)^2, & T_{fluid} \leq T^* \\ k_{oU2}^i \exp\left(-\frac{E_{U2}^i}{RT_{BL}}\right) (C_U^i)^2, & T_{fluid} > T^* \end{cases}$
<p>c) Deposition of aggregate (A) and unfolded (U) protein [TYPE C]</p>  <p> h_f Convective heat transfer coefficient λ_d Deposit thermal conductivity σ_T Thermal boundary layer thickness </p>	<p>Reactions in the bulk of the fluid</p> $r_{fN} = k_{oN} \exp\left(-\frac{E_N}{RT_{fluid}}\right) C_N$ $r_{fU} = \begin{cases} k_{oU1} \exp\left(-\frac{E_{U1}}{RT_{fluid}}\right) C_U^2, & T_{fluid} \leq T^* \\ k_{oU2} \exp\left(-\frac{E_{U2}}{RT_{fluid}}\right) C_U^2, & T_{fluid} > T^* \end{cases}$ <p>Deposition model</p> $\frac{\rho_d}{\sigma_T} \frac{d\delta_A}{dt} = \begin{cases} r_{dA} - r_{eA} = \frac{\rho_d}{\sigma_T} \left[\beta_A \frac{\lambda_d}{h_f} \left(k_{dA} \exp\left(-\frac{E_{dA}}{RT_{film}}\right) C_A^i - \kappa_A \delta_A \right) \right], & T_{fluid} \geq T_{min} \\ 0, & T_{fluid} < T_{min} \end{cases}$ $\frac{\rho_d}{\sigma_T} \frac{d\delta_U}{dt} = \begin{cases} r_{dU} - r_{eU} = \frac{\rho_d}{\sigma_T} \left[\beta_U \frac{\lambda_d}{h_f} \left(k_{dU} \exp\left(-\frac{E_{dU}}{RT_{film}}\right) C_U^i - \kappa_U \delta_U \right) \right], & T_{fluid} \geq T_{min} \\ 0, & T_{fluid} < T_{min} \end{cases}$	<p>Reactions in the boundary layer (BL)</p> $r_{bN} = k_{oN}^i \exp\left(-\frac{E_N^i}{RT_{BL}}\right) C_N^i$ $r_{bU} = \begin{cases} k_{oU1}^i \exp\left(-\frac{E_{U1}^i}{RT_{BL}}\right) (C_U^i)^2, & T_{fluid} \leq T^* \\ k_{oU2}^i \exp\left(-\frac{E_{U2}^i}{RT_{BL}}\right) (C_U^i)^2, & T_{fluid} > T^* \end{cases}$

Figure 1: Protein denaturation and deposition mechanisms and reactions (adapted from Georgiadis and Macchietto, 2000)

1.3. Experimental setup and data

The PHE apparatus in Figure 3 was used in five experiments at different operating conditions (Table 1) (Khaldi, 2016). Flowrates, inlet and outlet temperatures of both hot and cold streams were measured every 15 seconds. Experiment No. 0 used water in both streams (no fouling). This corresponds to a preheat phase, where the PHE is brought to a desired steady state temperature. Fouling experiments 1-4 involved a specially prepared whey protein solution (WPS), heated by a hot water stream. After an initial preheat phase as in Exp. 0, the cold stream is switched from water to the fouling fluid. Initial conditions at this point are given in Table 1. In these fouling runs, the outlet WPS temperature was manually maintained constant by manipulating the inlet hot water temperature (and in some cases, its flowrate). At the end of each experiment the deposit in each channel was collected and weighed (Table 1). Measurement variances are 0.1 g for mass deposition and 0.1°C (at calibration, rising gradually to 0.3°C) for temperature.

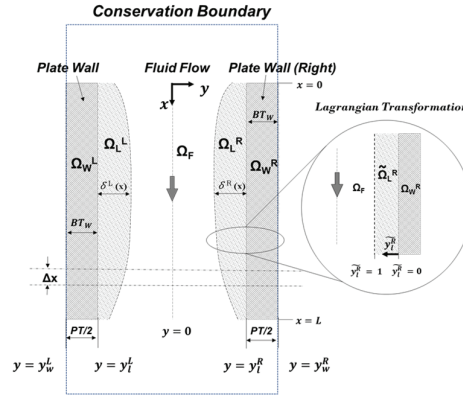
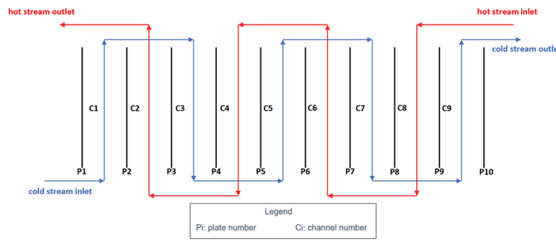


Figure 2 Schematic for a single channel delimited by two plates (Guan and Macchietto, 2018)



Material of construction	Stainless steel 316 Ti
Corrugation angle	90°
Plate length	0.495 m
Plate width	0.15 m
Plate thickness	0.8 mm
Distance between plates	3.93 mm

Figure 3: PHE configuration and design characteristics

Table 1: Initial conditions, and deposit mass in channels C1 to C9 at the end of the experiments

Exp. No.	Whey Protein Solution (WPS)				Hot Water			Deposit in channels [g]					End time [s]
	$\beta - lg$ [kg/m ³]	T _{in} [°C]	T _{out} [°C]	Flowrate [L/min]	T _{in} [°C]	T _{out} [°C]	Flowrate [L/min]	C1	C3	C5	C7	C9	
0	0	75	84	5.00	86	78	4.37	0	0	0	0	0	2520
1	6.58	65	85	5.25	96	70	4.82	0.3	4.2	18.8	38.6	40.4	6752
2	6.58	64	85	5.00	87	79	14.99	3.8	24.1	38.1	43.4	43.5	6716
3	6.58	64	85	5.13	115	66	2.43	0.2	1.8	14.1	46.6	58.4	6712
4	6.58	60	74	5.05	81	64	5.18	0.2	0.9	1.2	1.3	3.4	6692

2. Model Validation

The measured inlet temperature and mass flowrate profiles of both streams were reproduced using linear and quadratic polynomial splines. As an example for Exp. No. 0,

Figure 4 (left) shows that the inputs to the simulations are very well captured. Experiment No. 0 (with no fouling), allows validating the PHE thermal model. Excellent agreement is observed between experimental and predicted exit temperatures of both cold and hot water streams (Figure 4, right), with average absolute deviations of $0.5 \pm 0.6^\circ\text{C}$ and $0.8 \pm 0.5^\circ\text{C}$, respectively. The model has the structural adequacy and predictive ability to describe heat transfer dynamics for the clean PHE. Then, for each fouling experiment 1-4, the channel mass deposit data at the end of the run (Table 1) were used to estimate the remaining constants (β_A and/or β_U) in the three deposition models (A to C) described in Figure 1. Results (e.g. Exp. No. 1 in Figure 5) show that the Type C model (which combines aggregated and unfolded deposition) gives a better fit in all experiments. Post treatments of the numerical data obtained by Type C model seems to indicate that the ratio of aggregated and unfolded proteins in the deposit varies both along the length of the channels, and between channels (Figure 6). Hence both reaction schemes are important in different locations and temperature conditions.

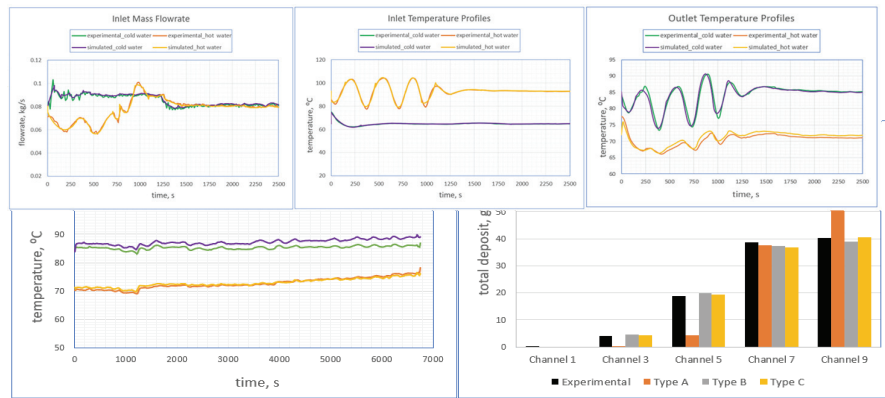


Figure 5 Experiment No. 1: Exit temperature profiles with model Type C (left); Experimental and simulated deposit mass in the fouling channels with all deposition models (right)

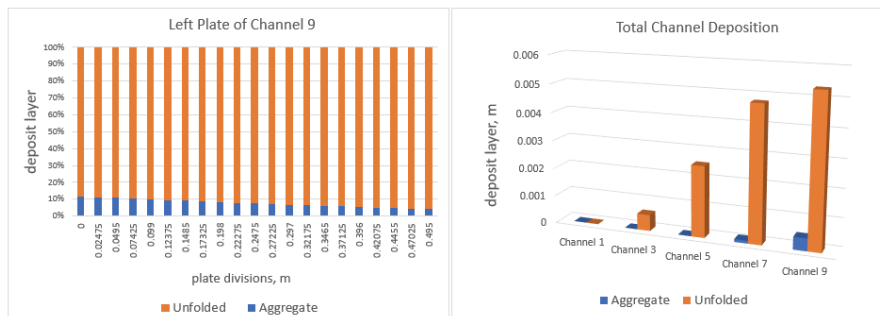


Figure 6 Experiment No. 1: Distribution of aggregate and unfolded proteins in the deposit layer along the length of a plate (left); total protein deposited in the fouling channels, C1-C9 (right)

3. Discussion and Conclusions

The dynamic, 2D distributed PHE model was tested on experimental data from five well documented experiments with three protein reaction/deposition models. Validation of the

PHE thermal model with a highly dynamic water-water run showed an excellent agreement between data and model, without parameter adjustments. The fouling runs (here and Darko, 2020) show that the aggregate proteins (Type A) fouling scheme does not adequately describe the final channel mass deposit data. The unfolded proteins (Type B) scheme performs better, however, neither scheme alone is adequate in all experiments. A linear combination of the two (Type C) gives a better agreement under all conditions. The predicted spatial distribution in the channels of unfolded and aggregate proteins in the deposit shows that both mechanisms are locally important and must be considered together. Protein ratios in the deposit were not measured, so remain to be validated. The combination of thermal and fouling models with quality experimental data enabled a systematic assessment of two traditional protein deposition models and a proposed new one, for dairy heat treatment in PHEs. Additional results and a sensitivity analysis will be reported elsewhere. This work could be extended to whey protein solutions containing calcium, which has a significant influence on denaturation and is known to promote binding of unfolded protein (hence aggregation and fouling). The approach, models and results presented will improve the optimal monitoring, operations and design of heat treatment in dairy and related food processes.

References

- B. Bansal and X. D. Chen, 2006, A critical review of milk fouling in heat exchangers, *Comprehensive Reviews in Food Science and Food Safety*, 5, 27–33.
- P. Blanpain-Avet, C. André, M. Khaldi, L. Bouvier, J. Petit, T. Six, R. Jeantet, T. Croguennec, G. Delaplace, 2016. Predicting the distribution of whey protein fouling in a plate heat exchanger using the kinetic parameters of the thermal denaturation reaction of β -lactoglobulin and the bulk temperature profiles, *Journal of Dairy Science*, 99 (12),
- W. K. Darko, Modelling and optimisation of thermal treatment in dairy and food processes, Imperial College London, 2020
- P. De Jong, 1996, Modelling and optimisation of thermal processes in the dairy industry, NIZO Verslag Ede V341, NIZO Food Research, The Netherlands
- Georgiadis, M. C. and Macchietto, S., 2000. Dynamic modelling and simulation of plate heat exchangers under milk fouling. *Chemical Engineering Science*. 55 (9), 1605-1619.
- gPROMS, 2018. Process Systems Enterprise Limited, www.psenderprise.com.
- S. Guan and S. Macchietto, 2018. A Novel Dynamic Model of Plate Heat Exchangers Subject to Fouling. *Computer Aided Chemical Engineering*, Vol 43 Part B, 1679-1684.
- M. Khaldi, 2016, Etude du Lien entre la Physico-chimie de Dérivés Laitiers et leur Aptitude à l'Encrassement Lors du Traitement Thermomécanique en Echangeur de Chaleur, Université des Sciences et Technologies de Lille 1, France and Université de Carthage
- S. M. Loveday, 2016, β -Lactoglobulin heat denaturation: A critical assessment of kinetic modelling, *Int. Dairy J.*, vol. 52, pp. 92–100
- E. Sadeghinezhad, S. N. Kazi, A. Badarudin, M. N. M. Zubair, B. L. Dehkordi, and C. S. Oon, 2013, A review of milk fouling on heat exchanger surfaces, *Rev. Chem. Eng.*, 29, 169–188.
- A. Sharma and S. Macchietto, 2019, Fouling and cleaning of Plate Heat Exchangers for Milk Pasteurisation: a moving boundary model, *Computer Aided Chemical Engineering* 46, Part B, 1483-1488, Elsevier, Amsterdam.
- M. Zhu, F. L. Santamaria, and S. Macchietto, 2020. A general dynamic model of a complete milk pasteuriser unit subject to fouling, *Computer Aided Chemical Engineering* 48, Part A , 247-252, Elsevier, Amsterdam.

Optimal efficiency of biomass conversion from bio-based byproducts to biofuel production in the Ethiopian sugar industry: A case study in Wonji-Shoa sugar factory, Ethiopia

Shumet Sharew^{a,b,*}, Ludovic Montastruc^b, Abubeker Yimam^a, Stephane Negny^b, Jean-Henry Ferrasse^c

^aAddis Ababa University Institute of Technology, School of Chemical and Bioengineering, Addis Ababa, P.O.Box 385, Ethiopia

^bINP-ENSILACET, allée Emile Monso-BP 44362, 31432 TOULOUSE Cedex 4, Laboratoire de Genie Chimique UMP 5503, France

^cAix Marseille Univ, CNRS, Centrale Marseille, M2P2, Marseille, France

*shumet.sharew@inp-toulouse.fr and/or shumet.sendek@aait.edu.et

Abstract

Energy is one of the essential inputs for socio-economic, and geo-political strategies in nation building. Among potential interests, lignocellulosic biomasses constitute a prime source of biofuel and other value-added products. The thermochemical approach resolve to the degradation of cellulosic biomass into fragments of lignin, hemicellulose and cellulose. From the fragmented components, extraction of fermentable and hydrolysable sugars could be obtained through various combined process steps and finally converted into desired products (alcohols, alkanes, CO, H₂). In this research, an average of wet bagasse (C₃₀H_{120.11}O_{53.84}) with 49.28% moisture content; dry bagasse (C₃₀H_{46.65}O_{17.11}); dry molasses (C₃₀H₅₆O₂₈) and dry filter cake (C₃₀H_{51.14}O_{22.7}), were taken for the conversion analysis, with sulphur and nitrogen remained neglected. Adopting gasification as a proposed approach on all of the incoming wet bagasse (16.53 kg/s), mobilizes 6907 kg of water or 5.03% of total water in it. Since, the factory is characterized by low energy efficiency, using a large fraction of the bagasse produced as fuel in the cogeneration system to supply the process energy requirements. The possibility of selling surplus electricity to the grid and utilizing surplus byproducts as an input for the other processes has motivated to enhance more efficient cogeneration systems and process thermal integration.

Keywords: Energy, Lignocellulose, Bagasse, Hydrolysis, Biorefinery.

1. Introduction

1.1. Background

Energy is one of the essential inputs for the socio-economic development (Brew-Hammond, 2010), and geo-political strategies in nation building. Nations set the ball rolling by putting forward their strategic development vision with clear objectives to meet the energy demands. Among the potential interests to extending alternatives bioenergy resources, lignocellulosic biomass has been identified as the prime source of biofuels and other value-added products (International Finance Corporation, 2017), and geared towards the production of cost-effective biofuel (Boateng & Lee, 2013), from inputs which are not food competitive by nature (Viikari et al., 2012). A new scenario, with the possibility of biomass conversion technologies to the production of alkanes (C_nH_{2n+2}O) from hydrolyzed bagasse with dodecane (C₁₂H₂₆) as the longest carbon chain produced, the conversion of bagasse and molasses into alcohols (C_nH_{2n+2}O). Eventually, conversion of biomass into hydrogen(H₂) and carbon monoxide (CO) are considered through conversion mechanisms at a higher level to sustainable economy.

In Ethiopia, the Wonji-Shoa sugar factory crushes about 906,089 tons of cane per year and produces about 111,014 tons of sugar. Its energy demand depends on the cogeneration plant designed to produce 31 MW power of electricity from surplus bagasse and consumes 11 MW power of electricity to self-sufficiency. However, the factory has been historically characterized by low energy efficiency due to the lack of process energy integration in the sugar mills.

2. Biofuel processing routes from biomass

2.1. Biofuel processing routes from sugarcane bagasse

Essentially, there are two different approaches (i.e., thermochemical and biochemical) for the bioethanol production from biomass(bagasse) (Cesaro & Belgiorno, 2015). Both approaches resolve to the degradation of cellulosic biomass into fragments of lignin, hemicellulose and cellulose. The biochemical pathway includes physical and biological pretreatments of the lignocellulosic biomass. The physical pretreatment is commonly used methods subject to high pressure saturated steam at temperatures of 160–260°C and pressure of 0.7 to

4.8MPa (Kucharska et al., 2018). Biological pretreatment (enzymatic hydrolysis) is an ideal approach for degrading cellulosic biomass (bagasse) into fermentable sugars and generally it can be accomplished by the role of at least three major class of enzyme synergetic action such as endoglucanases, exoglucanases, and β -glucosidase to extract fermentable sugar. later on, the sugars are converted into bioethanol *via* fermentation and anhydrous ethanol is produced *via* distillation. Thermochemical biomass conversion includes a number of possible routes to produce useful fuels and chemicals from the initial biomass feedstock. Biomass can be used as a solid fuel, or converted into liquid or gaseous forms for the production of electric power, heat, chemicals, or gaseous and liquid fuels (Kundu et al., 2018). The thermochemical approach under process consideration includes incineration, combustion, sophisticated oxidation such as HydroThermal oxidation (SHTO) and Wet Air Oxidation (WAO) (Goffè & Ferrasse, 2019). This technology requires high level of heat and results in syngas (synthesis gas) such as CO, H₂, and CO₂. These gases are chemically converted into a mixture of alcohol using metal catalysts. The process can occur through direct combustion, pyrolysis, gasification, torrefaction and liquefaction. The process to biomass conversion use three ways. It can be burned to produce heat and electricity, changed to gaseous fuels such as methane, hydrogen, and carbon monoxide or changed to liquid fuels (Cesaro & Belgiorno, 2015).

2.2. Pretreatment

This process has an essential effect on the overall process of biofuel from lignocellulose, which makes the cellulose accessible to hydrolysis for conversion to biofuel. Lignocellulosic biomass is pretreated by diverse methods including physical, chemical, physicochemical, biological pretreatment or combined pretreatment (Asgar et al., 2020). In this step, it is applied to modify the macroscopic as well as the microscopic size and structure of lignocellulosic feedstock (Kazemi Shariat Panahi et al., 2020). Steam explosion taken as the sole pretreatment method for sugar yield improvement is been considered for the biofuel production (Boateng & Lee, 2013). One of the most important evidences of implementing pretreatment methods for the alcohol and alkanes production is the removal of lignin and hemicellulose, through cost-effective processes.

2.3. Hydrolysis

The goal of this process is to generate fermentable monomeric sugars from hemicellulose (C₅) and cellulose (C₆) content of lignocellulosic biomass in conversion scheme. Its performance is highly associated with the selected pretreatment process (Achinas & Euverink, 2016). This step can be accomplished by two different processes steps, namely, acid hydrolysis and enzymatic hydrolysis. In acid hydrolysis either sulphuric acid or hydrochloric acid-based hydrolysis process is operated under two different conditions; (i) a process that uses concentrated acids operates at a lower temperature and, (ii) a process that uses dilute acid concentration and operates at a higher temperature. On the other hand enzymatic hydrolysis, which is an ideal approach for degrading cellulosic biomass into reducing sugars, generally can be accomplished by the role of at least three major classes of enzyme synergetic action (section 2.1). However, enzymatic hydrolysis depends on certain critical optimal conditions for maximal efficiency such as sample moisture content can cause an irreversible collapse of the pore cellulosic structure, decreasing the hydrolyzability (Taherzadeh & Karimi, 2007) and the cellulose access to the enzymatic reaction and limit the number of hydrolyzed carbon atom within the reaction stoichiometry (Da Silva & Chandel., 2014). The hydrolysis reactions of biomass specifically cellulose and hemicellulose components under go monomeric generation shown in reaction equation (1 & 2) respectively. The lignin component resistant to hydrolysis could be a source of heat and electricity in cogeneration plant during gasification.



Where, n is the stoichiometric coefficients of reactants consumed and products generated from the utilized biomass for reaction (1 & 2). Taking into account reaction (1) each glucose unit in the long chain combines with a water molecule, and 180 mass units of glucose are released from 162 mass units of glucan and 18 mass units of water, an 11.1% substantial mass gain and from hemicellulose hydrolysis over 13.6% mass gain is reported from the theoretical weight percent ratio. For this research the mass of cellulose and hemicellulose from 141285.7334 ton/year of moisture free bagasse calculated accordingly to the weight percent value reported on table (1).

Table 1A: biochemical composition of bagasse

Biochemical composition	Weight %
Cellulose	44
Hemicellulose	27
Lignin	25
Sucrose loss	4

Table 1B: biochemical composition of bagasse

Elemental composition	Weight %
Carbon	44
Hydrogen	27
Oxygen	25

3. Materials and methods

The methodology of the considered system includes the conversion reaction and combustion reaction that provides energy to the system for reactants and products in their standard states and the results presented from the MATLAB code in which coded by **J.-H. Ferrasse**. In the case of material selection with respect to the desired products (alcohol, alkane, carbon monoxide, hydrogen), thermochemical path way is chosen for this research sugarcane bagasse whose characteristics have been extracted from the database for biomass and waste of the energy research center of the Netherlands (**Goffé & Ferrasse, 2019**). An average of wet bagasse ($C_{30}H_{120.11}O_{53.84}$) with 49.28% moisture content (factory records); dry bagasse ($C_{30}H_{46.65}O_{17.11}$); dry molasses ($C_{30}H_{56}O_{28}$); dry filter cake ($C_{30}H_{51.14}O_{22.7}$); taken for conversion analysis, where sulphur and nitrogen were neglected. The formula values used for the study assumed on a carbon basis number of 30, sulphur and nitrogen are neglected. To construct comparability among results one kilogram per second of biomass is considered as an input of the designed reaction system. Biomass to which shall put the subject of biomass conversion in the proper energy perspective to variables in the conversion process (reaction stoichiometry), technology, and the desired end product (**Detroy & St Julian, 1982**). **Table (2)** list the values used for the study on a carbon basis number of 30, sulphur and nitrogen are neglected.

Table (2)Biomass composition used for the study, ash and moisture-free basis.

Biomass	Formula	C (%)	H (%)	O (%)	H/C	O/C
Sugarcane bagasse	$C_{30}H_{46.65}O_{17.11}$	52.90	6.86	40.24	1.56	0.57
Molasses	$C_{30}H_{56}O_{28}$	41.56	6.50	51.94	1.87	0.93
Filter cake	$C_{30}H_{51.14}O_{22.7}$	46.50	6.6	46.90	1.70	0.76

3.1. Conversion products

The first biomass conversion is the production of alkanes ($C_nH_{2n+2}O$) from hydrolized bagasse with dodecane ($C_{12}H_{26}$) as the longest carbon chain produced, to conventional corresponding products found in liquid hydrocarbons. The second product with the longest carbon chain family dodecanol ($C_{12}H_{26}O$), is the conversion of biomass into alcohols ($C_nH_{2n+2}O$). The third product with the stright carbon chain family n-nonane ($C_9H_{20}O$), is the convesion of biomass (molasses) into alcohols (C_nH_{2n+2}). Eventually, conversion of biomass into hydrogen(H_2) and carbon monoxide (CO) are considered through conversion mechanisms. The biomass and product (s) considered in this research paper by taking into account inputs bagasse, molasses and filter cake mixed together to simultaneous fermentation after individual components pretreatment and hydrolysis, if biochemical pathway is favoured. Since input components differ from each other in their sugar content and number of atoms in their chemical formula resulting to differ their lower heating value (LHV) defined in the Boie equation (Eq. (3)) as thermochemical biomass conversion is the sole process prototype.

$$LHV_{Boie} = 348.35C + 938.7H - 1080 + 62.8N + 104.65S \quad Eq. (3)$$

Each capital letter (eq. 3 and 4) is the mass percent (wt%) of corresponding atoms in the element and the LHV is expressed in kJ/kg. The LHV of input components is subject to its moisture content. In the case of additional drying, the energy (if we wanted to dry) represents 2%, assuming that the sun brings us to a dryness of 85% (which remains to be verified). Therefore, the LHV of inputs can be specified as wet and dry to remark the impact of moisture content on biomass to energy conversion scenario. For wet bagasse with its expression ($C_{30}H_{120.11}O_{53.84}$) the LHV is: $LHV_{wet, bagasse} = 10.816$ MJ/kg and for dry bagasse with its expression ($C_{30}H_{46.65}O_{17.11}$) the LHV is: $LHV_{dry, bagasse} = 20.519$ MJ/kg. For wet molasses with its expression ($C_{30}H_{94}O_{47}$) the LHV is: $LHV_{wet, molasses} = 10.997$ MJ/kg and for dry molasses with its expression ($C_{30}H_{56}O_{28}$) the LHV is: $LHV_{dry, molasses} = 14.964$ MJ/kg. For wet filter cake with its expression ($C_{30}H_{51.14}O_{22.7}$) the LHV is: $LHV_{wet, filter\ cake} = 4.906$ MJ/kg and for dry filter cake with its expression ($C_{30}H_{51.14}O_{22.7}$) the LHV is: $LHV_{dry, filter\ cake} = 17.331$ MJ/kg. In the case of alkanes and alcohol, the modified version of Boie equation is used to liquid and gaseous fuel (Eq. 4).

$$LHV_{modified} = 349.1C + 958.3H - 103.4O - 15.1N - 100.5S \quad Eq. (4)$$

Biomass and products representation helps to structured elements depend on the two atomic ratio on the basis of carbon atom, H/C and O/C. By increasing H/C and decreasing O/C it is possible to visualize the conversion impact. The biomass conversion performance can be seen from material and energy ratio indicators in which inputs calculated from the biomass and energies comes from biomass and their products. Depend on the energy supply to the system either biomass or products become a role play to material ratio indicator calculation. Here, if only biomass is involved, equation (5) used to calculate material ratio.

$$\eta_{\text{material}} = \frac{n_{\text{product obtained}} * M_{\text{product}}}{(n_{\text{biomass conversion}} + n_{\text{biomass burned}}) * M_{\text{biomass}}} \quad \text{Eq. (5)}$$

on the other hand , if the combustion of products become an energy supply source to the system equation (6) is used to determine kilogram of bomass converted.

$$\eta_{\text{material}} = \frac{(n_{\text{product obtained}} - n_{\text{product burned}}) * M_{\text{product}}}{n_{\text{biomass conversion}} * M_{\text{biomass}}} \quad \text{Eq. (6)}$$

where, n_i and M_i are refer to the mole number and molar mass of component “i” respectively. The energy ratio indicator considered as a measure of energy value contained by the biomass or the product(s) from its lower heating value (LHV) and the source of energy supply to the system. If the biomass is used to supply energy to the system, it will follow equation (7) to determine biomass conversion.

$$\eta_{\text{energy}} = \frac{n_{\text{product obtained}} * LHV_{\text{product}}}{(n_{\text{biomass conversion}} + n_{\text{biomass burned}}) * LHV_{\text{biomass}}} \quad \text{Eq. (7)}$$

when the energy supply to the system is originated from the product conversion, equation (8) is used to determine energy efficiency of the system.

$$\eta_{\text{energy}} = \frac{(n_{\text{product obtained}} - n_{\text{product burned}}) * LHV_{\text{product}}}{n_{\text{biomass conversion}} * LHV_{\text{biomass}}} \quad \text{Eq. (8)}$$

Parameters for the system can be considered as conversion and thermo-chemical for biomass conversion and energy analysis. In thermo-chemical parameters the energy analysis viewed from the first law of thermodynamics in which the calculation of reagents and products done in their standard states ($T = 298 \text{ K}$ and $P = 1.01325 * 10^5 \text{ Pa}$) and the LHV described in equation (3 and 4) from simulis[®] thermodynamics database. In this stance the amount of energy necessary to supply to the system can be calculated by the standard enthalpy of reaction described in equation (9). In conversion parameters the biomass could be a limiting reagent or reaction rate determining agent by taking into account some assumptions according to biomass elemental composition. Upon the behavior of limiting reagents and converted products biomass perceived carbon and hydrogen rules that marked the conversion ratio of carbon and hydrogen atoms in biomass conversion.

$$\Delta_r H_{298}^{\circ} = \sum_i v_i \Delta_f H_{298}^{\circ} \quad \text{Eq. (9)}$$

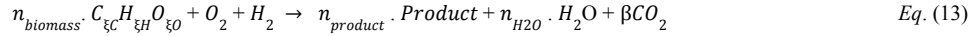
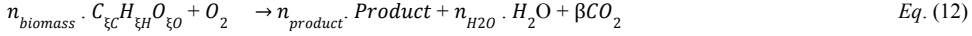
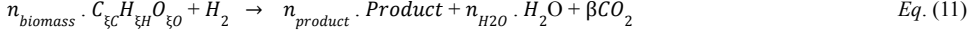
3.1.1 Carbon as limiting reagent

Carbon rule details about the amount of atomic carbon conversion present in the biomass (numbered as carbon 30) into valuable products such as alkanes, alcohol, or carbon monoxide with range of conversion from 0 to 100% in proportion to the production of carbon dioxide from the reaction. The proportional carbon conversion ration given by equation (10).

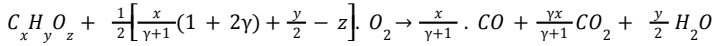
$$C\% = \frac{n_{\text{product}} * \xi_{\text{Cproduct}}}{n_{\text{biomass}} * \xi_{\text{Cbiomass}}} \quad \text{Eq. (10)}$$

Where, n_{biomass} , n_{product} refers to stoichiometric number of moles of biomass and products in equilibrium reaction respectively. The formation of carbon dioxide in the reaction measures the carbon conversion ratio and

the supply of oxygen into the reaction. As well the formation of carbon monoxide basically appreciated by the partial oxidation of biomass described by equation (14) . The general conversion reaction of the biomass equation (11 - 13) considered sulphur and nitrogen are negligible. The reactions were conducted with providing only oxygen and only hydrogen equation (12 and 13), and with simultaneous provision of oxygen and hydrogen, equation (13).

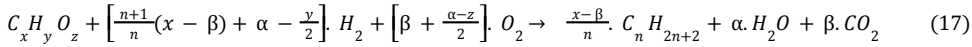
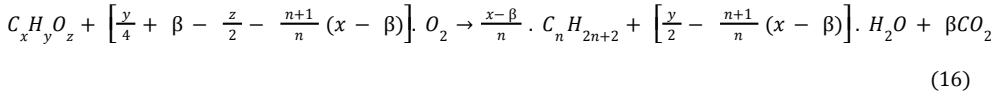
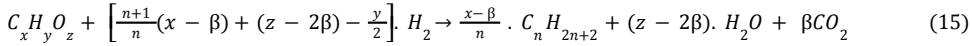


where, the parameter (β) used to adjust the CO_2 production and carbon conversion ratio (C%).



Eq. (14)

Hence, the general biomass conversion, $C_x H_y O_z$ into alkanes can be described as equation (14 - 16) with provided only oxygen, only hydrogen, and with simultaneous provision of both oxygen and hydrogen in the conversion reaction.

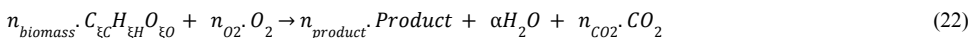
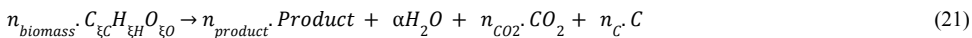
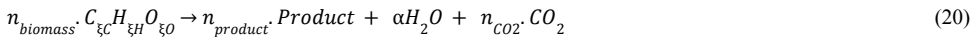
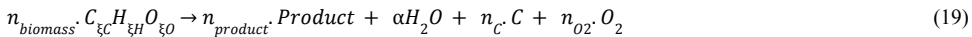


3.1.2 Hydrogen as limiting reagent

In this perspective the atomic hydrogen in biomass taking a role play career as a limiting reagent in the biomass conversion reaction. The amount of hydrogen contain in the biomass converted into useful products such as alkanes, alcohol, or hydrogen. on the same analogy of carbon the hydrogen atom in the biomass conversion performance described by hydrogen conversion ratio equation (18).

$$H\% = \frac{n_{product} \cdot \xi_{Hproduct}}{n_{biomass} \cdot \xi_{Hbiomass}} \quad (18)$$

Like the carbon conversion ration (C%), hydrogen conversion ratio (H%) used a defined adjustable parameter (α) for hydrogen conversion and water production. The potential of hydrogen conversion differ accordingly with the biomass composition and H/O ratio. Based on the unified generalized molecular formula of dry bagasse ($C_{30} H_{46.65} O_{17.11}$), dry molasses ($C_{30} H_{56} O_{28}$), and dry filter cake ($C_{30} H_{51.14} O_{22.7}$), the H/O ratio accounts 2.73, 2, and 2.25 respectively considered to the capture or supply of carbon and oxygen for possible reactions. The possible reactions to decide the product either alkanes or alcohols shown equation (19 - 23).



$$n_{biomass} \cdot C_{\xi_C} H_{\xi_H} O_{\xi_O} \rightarrow n_{product} \cdot Product + \alpha H_2O + n_{CO_2} \cdot CO_2 + n_{O_2} \cdot O_2 \quad (23)$$

The production of water and hydrogen conversion appreciated by exothermic reaction of biomass without the supply of external heat described by equation (24).

$$C_x H_y O_z + \frac{1}{2}(2x - z + \alpha) \cdot O_2 \rightarrow \frac{1}{2}(y - 2\alpha) \cdot H_2 + \alpha H_2O + x \cdot CO_2 \quad (24)$$

Similarly considering hydrogen as a limiting reagent to the biomass conversion to the product formation of alkanes given equation (25 - 28).

$$C_x H_y O_z \rightarrow \frac{y-2\alpha}{2n+2} \cdot C_n H_{2n+2} + \alpha H_2O + \left[x - \frac{n(y-2\alpha)}{2n+2} \right] \cdot C + \frac{z-\alpha}{2} \cdot CO_2 \quad (25)$$

$$C_x H_y O_z \rightarrow \frac{y-2\alpha}{2n+2} \cdot C_n H_{2n+2} + \alpha H_2O + \left[x - \frac{n(y-2\alpha)}{2n+2} - \frac{z-\alpha}{2} \right] \cdot C + \frac{z-\alpha}{2} \cdot C \quad (26)$$

$$C_x H_y O_z + \left[x - \frac{n(y-2\alpha)}{2n+2} - \frac{z-\alpha}{2} \right] \cdot O_2 \rightarrow \frac{y-2\alpha}{2n+2} \cdot C_n H_{2n+2} + \alpha H_2O + \left[x - \frac{n(y-2\alpha)}{2n+2} \right] \cdot CO_2 \quad (27)$$

$$C_x H_y O_z \rightarrow \frac{y-2\alpha}{2n+2} \cdot C_n H_{2n+2} + \alpha H_2O + \left[x - \frac{n(y-2\alpha)}{2n+2} \right] \cdot CO_2 + \left[\frac{z-\alpha}{2} - x + \frac{n(y-2\alpha)}{2n+2} \right] \quad (28)$$

In the case of alcohol the $C_n H_{2n+2}$ terms of equation (25 - 28) replaced by $C_n H_{2n+2}O$. The parameter β varies accordingly with the amount of carbon, ξ_C . The variation can be between $\beta = 0$, and $\beta = 30$. When $\beta = 0$, the carbon conversion is the same as the amount of carbon contained in the biomass. When $\beta = 30$, all the carbon atom in the biomass converted to CO_2 . Similarly, the parameter α varies with the conversion of hydrogen in the biomass into water and hydrogen.

4. Results and Discussion

4.1. Hydrolysis of bagasse

To achieve the conversion of bagasse to biofuel, since it is unrealistic to convert water (moisture content), the assumption is that water can hydrolyze cellulose and hemicellulose to sugar. The sugar obtained will have for molar composition the same composition as that already contained (presented) in the bagasse. Thus, the hydrolyzed biomass has a composition of $C_{30} H_{50.34} O_{18.18.96}$ and a LHV of 19.60 MJ / kg if it is considered to be dry. Virtually this consists of using 4.66% wet biomass (kg water / kg bagasse). In this specific case all, the water is consumed. If this operation is done on all of the incoming bagasse (16.53 kg/s), it will mobilize 6907 kg of water or 5.03% of the total water that constitutes the moisture in the bagasse. Hence, wet biomass: $C_{30} H_{120.11} O_{53.84}$ (LHV = 10.81MJ/kg) flow rate of 16.53 kg / s; dry biomass: $C_{30} H_{46.65} O_{17.11}$ (LHV = 20.51 MJ/kg) flow rate of 8.38 kg / s and hydrolyzed biomass: $C_{30} H_{50.34} O_{18.18.19}$ (LHV of 19.60 MJ / kg) flow rate of 8.79 kg / s. There is a slight drop in LHV, but a change in composition and a slight increase in throughput. The two offset each other. The gain is therefore not energy, but it gives a "realistic upper limit to the water that can be taken into account". In this stance 100% dry bagasse amount come into effect hydrolysis and of which 1 mole of cellulose gain 1 mole of water to release simple sugar. Since cellulose is a polymer and to define exact number of moles of hydrolyzed product assumed that $1 \cdot 10^6$ as a constant factor. Therefore, the number of moles of cellulose generated from the given mass can be calculated as equation (29).

$$n_{cellulose} = \frac{1E6 \cdot \dot{m}_{cellulose}}{M_{cellulose}} \quad (29)$$

Thus, from 62,165.7227 ton/year cellulose, $n_{cellulose}$ accounts 383739029.1 moles. The gain of water which is the same molar value assumption with the hydrolyzed cellulose and can be calculated as equation (30).

$$\dot{m}_{H_2O} = n_{cellulose} \cdot \frac{M_{H_2O}}{1E6} \quad (30)$$

The mass of water gained during hydrolysis goes to 6907.302524 ton/year, which is only 5.032% of the water contained in the original bagasse it accounts 49.28% moisture content from the total 278560.20 ton/year of bagasse production. Consider that the calculated dryness of 50.7% moisture free bagasse. In the case of additional drying, the energy (if necessary to dry) represents 2%, assuming that the sun brings us to a dryness of 85% (which remains to be verified). This question of drying will be central if design to make a biorefinery scheme, less important if this is satisfied with a theoretical analysis and relying on the sun. The amount of water evaporated is considered as the difference of amount of water in the original biomass and the amount of water required for hydrolysis process equation (31), which is about 130367.164 ton/year.

$$\dot{m}_{H_2O, \text{ evaporated}} = \dot{m}_{H_2O, \text{ biomass}} - \dot{m}_{H_2O, \text{ hydrolysis}} \quad (31)$$

To reach to complete dryness of bagasse it requires energy for evaporation depend on the latent heat of evaporation and the amount of water evaporated as described in equation (32).

$$E_v = \dot{m}_{H_2O} * L_v \quad (32)$$

Where, E_v , and L_v are heat energy required for evaporation and specific latent heat of vaporization (2250 kJ/kg) of water respectively. Therefore, energy required to evaporated the specified amount of water is $2.93326 * 10^{11}$ kJ/year. The amount of energy contained in hydrolyzed biomass according to its lower heating value equation (3) from the molecular formula ($C_{30}H_{50.34}O_{18.96}$) is calculated as equion (33).

$$E_{\text{hydrolyzed biomass}} = \dot{m}_{\text{hydrolyzed biomass}} * LHV_{\text{hydrolyzed biomass}} \quad (33)$$

Therefore, the hydrolyzed bagasse energy content refers to $2.90482 * 10^{12}$ kJ/year. Taking into account the percent ratio of energy required to evaporation to the energy contained in the hydrolyzed biomass ($E_v/E_{\text{hydrolyzed biomass}}$) about 10. Thus to evaporate the water gained for hydrolysis, in order to obtain a dry product, one can consider a complete or partial thermal action in addition to solar drying. In the case of complete drying, without the integration of heat, the energy required to represents 10% of the energy contained in the hydrolyzed bagasse.

4.2. Site power requirements

The power requirement to sugar production assumed be the unit's overall needs are covered by bagasse. The calculations presented (Table 1), are for dry bagasse. But the same is found for the other two forms of biomass (molasses and filter cake). The designed power requirement of the factory is 11 MW of electricity generated from its back pressure turbine with an average steam to cane ratio of 0.56 from 906089.5 ton/year crushed cane. The annual steam production could reach 507410.12 ton/year at 180°C. Therefore, the steam power can be calculated as given equation (34).

$$P_{st} = \frac{\dot{m}_{\text{steam}} * H_{v, H_2O}}{t_{\text{operation}}} \quad (34)$$

Where, P_{st} and $t_{\text{operation}}$ are steam power potential and annual operation time (4680 hr/year) of the case factory. Hence, the steam power calculated is 83.785313 MW for bagasse 49.28% moisture. The heat requirement for electricity generation can be determined from the electricity consumption to bagasse feed ratio ($\frac{P_{\text{consumption}}}{\dot{m}_{\text{bagasse}}}$) or (11 MW/16.5337251 kg/s), of which 0.6653 MW/kg bagasse. On the same way the heat requirement for steam production follow steam to bagasse ratio ($\frac{\dot{m}_{\text{steam}}}{\dot{m}_{\text{bagasse}}}$), about 1.821546 with enthalpy of the steam at specified temperature (180°C). Table (3) describes the heat requirement values for electricity generation and steam production for a flowrate of 1 kg/s of bagasse.

Table (3) heat requirement values for electricity generation and steam production for 1 kg/s biomass.

Heat for electricity generation	1.869661953 MW/kg bagasse	$(\frac{P_{consumption}}{m_{bagasse}}) / (1 - (\sqrt{\frac{T_{cold}}{T_{hot}}}))$, $T_{cold} = 300k$ and $T_{hot} = 723k$
Heat for steam production	5.067539993 MW	$(\frac{m_{steam}}{m_{bagasse}}) * \Delta H_{steam, T=180^{\circ}C}$

Therefore, it becomes 1.31 MW/kg of total dry bagasse. Electricity is produced from steam at 300 °C and it consider an efficiency to be closer to finite time production $(1 - (\sqrt{\frac{T_{cold}}{T_{hot}}}))$, with $T_{cold} = 300K$, $T_{hot} = 573 K$.

Table 4. Biomass to energy conversion result on moisture free feedstocks and hydrolysed bagasse

Feedstock	Alcohol			Alkane			CO+H ₂		Flow (kg/s)	LHV (MJ/kg)
	Energy [kJ]	Mass [kg]	C	Energy [kJ]	Mass [kg]	C	Energy [kJ]	Mass [kg]		
Hydrolyzed bagasse	93.30	56.20	12	93.30	41.40	12	94.00	123	2.37 (3.30*)	19.60
Dry bagasse	93.30	54.60	12	93.30	43.40	12	93.40	129	2.36 (3.17*)	20.51
Dry molasses	93.30	46.70	9	93.20	31.00	10	94.60	1	1.20	14.96
Dry filter cake	93.30	52.60	10-12	93.10	36.50	12	94.30	1,12	1.97	17.33

(C refers to carbon atom number), (* refers to hydrolysed biomass flowrate)

For steam, the power found is 83 MW. The ratio of 0.56 (steam to cane ratio), for steam becomes in MW/kg dry bagasse is 3.59. To fix ideas it is necessary to articulate the energy equivalent of 1kg/s of wet bagasse is the production of 4.74 MW electricity and 9.99 MW steam. However, with a dry bagasse flow rate of 8.38 Kg/s and a LHV 20.51 MJ/kg to cover the plant's needs, it will therefore be necessary to burn 1.93 kg/s for electricity and 4.08 kg/s for steam generation.

In this case Steam/electricity production is not optimized in the sense of cogeneration. But, in view of the ratio of vapor taken, this is a point to be discussed and it leaves a flow rate of 2.36 kg/s of dry biomass. This calculation done for the wet biomass, results 2.59 kg/s and 2.37 kg/s for the hydrolyzed biomass. It's preferred to retain the two values of 2.36 kg/s and 2.37 kg/s to conversion for LHV of wet biomass. Hence, the power requirement can be seen with cogeneration and extraction, and without cogeneration schemes. The cogeneration and extraction scheme considered that the steam cycle at a hot temperature of 450°C to 180°C withdrawn the turbinéd horse power (HP) with no stretch to 27°C. This is substantial gain of 3.1 kg/s (Table 4) of dry biomass which is almost 107 theoretical MW power and 98,000 tons of biofuel. Where as, without cogeneration integration scheme the production of biofuel (alcohol and alkanes), goes to 93 MW of power from 5.1 kg/s biomass and 86,000 tons of possible biofuel during the whole operation time in a year.

5. Conclusion

Biomass is a versatile and abundant resource, which can be used to produce energy via different routes. Several kinds of biomass can be used either to produce bioethanol or alkanes, CO and H₂. To this end, some boundary conditions, including the area of application and the presence of existing infrastructure to process integration, have to be identified. It is suspected that from the complete balance sheet of the biomass composition, water has significant role in the final energy content (LHV). Although biochemical pathway is reliable techniques, already common at full-scale, the integrated thermochemical process feasibility for industrial application should be more deeply investigated. Nevertheless, the strongest integration is the water treatment unit. Water is three times more important than biofuel. This treated water (and separated in fact if it is concentrate the material in sugar + biofuel) also has a significant value. Since, 92% of Ethiopia's policy is based on the biomass resource; biorefinery integration into the sugar factories is of prime importance.

References

- A.Brew-Hammond, 2010, Energy access in Africa: Challenges ahead, *Energy Policy*, 38 (5) 2291–2301.
- IFC, 2017, International Finance Corporation, *Converting Biomass to Energy : A Guide for Developers and Investors*. Washington, DC.
- O.Boateng, K.T.Lee, 2013, Comparative thermodynamic sustainability assessment of lignocellulosic pretreatment methods for bioethanol production via exergy analysis, *Chemical Engineering Journal*, 228, 162–171.
- L.Viikari, J.Vehmaanperä, A.Koivula, 2012, Lignocellulosic ethanol: From science to industry, *Biomass and Bioenergy*, 46, 13–24.
- A.Cesaro, V. Belgiorno, 2015, Combined biogas and bioethanol production: Opportunities and challenges for industrial application. *Energies* 2015, 8(8), 8121-8144.
- K.Kucharska, P.Rybarczyk, I.Hołowacz, R.Lukajtis, M.Glinka, M.Kamiński, 2018, Pretreatment of lignocellulosic materials as substrates for fermentation processes. *Molecules*. 23(11) 2937.
- K.Kundu, A.Chatterjee, T.Bhattacharyya, M.Roy, A.Kaur, 2018, Thermochemical Conversion of Biomass to Bioenergy: A Review in Prospects of Alternative Transportation Fuels, Springer Singapore, 235–268.
- J.Goffé, J.H. Ferrasse, 2019, Stoichiometry impact on the optimum efficiency of biomass conversion to biofuels. *Energy*, 170, 438–458.
- A.A.Rastegari, A.N.Yadav, A.Gupta, 2020, *Prospects of Renewable Bioprocessing in Future Energy Systems: Production by Cyanobacteria*, Springer International Publishing, Springer Nature Switzerland AG, 10, 518
- H.Kazemi Shariat Panahi, M.Dehghani, M.Aghbashlo, K.Karimi, M.Tabatabaei, 2020, Conversion of residues from agro-food industry into bioethanol in Iran: An under-valued biofuel additive to phase out MTBE in gasoline. *Renewable Energy*, 145, 699–710.
- S. Achinas, G.J.W. Euverink, 2016, Consolidated briefing of biochemical ethanol production from lignocellulosic biomass, *Electronic Journal of Biotechnology*, 23,44-53.
- M.J.Taherzadeh, K.Karimi, 2007, Enzyme-based hydrolysis processes for ethanol from lignocellulosic materials: A review. *BioResources*, 2(4)707–738.
- S.S.DaSilva, A.K.Chandel, 2014, *Biofuels in Brazil: Fundamental aspects, recent developments, and future perspectives*. *Biofuels in Brazil: Fundamental Aspects, Recent Developments, and Future Perspectives* Springer International Publishing Switzerland, pp. 435
- R.W. Detroy, G.St Julian, 1982, Biomass conversion: Fermentation chemicals and fuels. *Critical Reviews in Microbiology*, 10(3)203–228.

Data-driven prediction of antiviral peptides based on periodicities of amino acid properties

Chris A. Kieslich*, Fatemeh Alimirzaei, Hyeju Song, Matthew Do, Paige Hall

**Auburn University, Auburn, AL 36849 USA*

kieslich@auburn.edu

Abstract

With the emergence of new pathogens, e.g., methicillin-resistant *Staphylococcus aureus* (MRSA), and the recent novel coronavirus pandemic, there has been an ever-increasing need for novel antimicrobial therapeutics. In this work, we have developed support vector machine (SVM) models to predict antiviral peptide sequences. Oscillations in physicochemical properties in protein sequences have been shown to be predictive of protein structure and function, and in the presented work we have taken advantage of these known periodicities to develop models that predict antiviral peptide sequences. In developing the presented models, we first generated property factors by applying principal component analysis (PCA) to the AAindex dataset of 544 amino acid properties. We next converted peptide sequences into physicochemical vectors using 18 property factors resulting from the PCA. Fourier transforms were applied to the property factor vectors to measure the amplitude of the physicochemical oscillations, which served as the features to train our SVM models. To train and test the developed models we have used a publicly available database of antiviral peptides (<http://crdd.osdd.net/servers/avppred/>), and we have used cross-validation to train and tune models based on multiple training and testing sets. To further understand the physicochemical properties of antiviral peptides we have also applied a previously developed feature selection algorithm. Future work will be aimed at computationally designing novel antiviral therapeutics based on the developed machine learning models.

Keywords: Computational biology, Machine learning, Support vector machines, Feature selection, Antiviral peptides

1. Introduction

With the increasing threat of viruses on human populations around the world, as evidenced by the recent COVID-19 pandemic, there is significant need for approaches for rapid development of treatments for novel viral outbreaks. If and when a new viral outbreak poses an eminent threat, the availability of tools for therapeutic design could enable the fast and efficient development of novel antiviral treatments. One promising class of antiviral treatments are anti-viral peptides (AVPs), which can act in a variety of ways, such as inhibiting replication, preventing binding to host cells, and interrupting virus-induced host signalling. Rational approaches have been previously used to successfully design AVPs, and more recent efforts have been aimed at using computational methods to predict their function based on the peptide sequence. One challenging aspect of developing machine learning models is identifying how to best encode or represent a peptide's sequence or properties. Most datasets of peptide function include peptides of varied length and to train a machine learning model one

must ensure that every peptide is represented by the same number of features. Past efforts for predicting AVPs have used peptide features that include the number of positively/negatively-charged amino acids, the charge of the peptide, the frequency of each amino acid in the sequence, the amount of possible H-bonds, molecular weight, and average hydrophobicity. One major limitation of these features is they do not maintain information regarding the ordering of amino acids or properties along the peptide structure, which is known to be crucial to protein structure and function. An alternative approach for developing physicochemical descriptors of protein sequences has been previously proposed that takes advantage of underlying periodicities in protein/peptide physicochemical properties (Eisenberg et al. 1984; Rackovsky 1998). Rackovsky (1998) has shown that periodicities of physicochemical properties along the sequence of a protein can be used to categorize families of protein structure/function. By using Fourier transforms and numerical tricks, it is possible to encode peptide sequences of varying lengths in terms of the same number of features based on the oscillation of amino acid properties.

In this work, we have used data analysis (i.e., PCA) and machine learning (i.e., support vector machines) to develop accurate models for predicting AVP sequences based on periodicities of amino acid properties. Additionally, by ranking the importance of the developed Fourier-based features, we were able to train SVM models with improved accuracy and generalizability, while also beginning to gain some insights into the importance of oscillations in physicochemical properties for AVP function.

2. Methods

In this work, we have used the R statistical language to perform all steps of our analysis including the generation of amino acid property factors using PCA, Fourier-based feature extraction, training/validating support vector machines, and feature selection. Below are more detailed descriptions of how these elements of our approach were implemented.

2.1. AVP Dataset

To develop data-driven classification models we need to have access to sufficiently large datasets, which contain both amino acid sequences and function labels. At present, there are multiple publicly available databases that hold the identities of some known antiviral peptides, including AVPpred (Thakur et al. 2012), APD3 (Wang et al. 2016), and CAMPR3 (Waghu et al. 2016). In the current study, we have focused on the AVPpred dataset, which contains 544 experimentally validated antiviral peptide sequences along with two sets of negative AVP: i) 407 experimentally validated nonactive peptides; and ii) 544 randomly selected non-secretory peptides. To eliminate the possibility of potential bias, we filtered the AVPpred dataset to eliminate any sequences with greater than 40% sequence identity. This was performed by first using the Clustal Omega webserver (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) to align all sequences of the AVPpred dataset, and then applying hierarchical clustering in R to identify clusters of sequences sharing more than 40% sequence identity. The medoid of each cluster was selected as the representative sequence, resulting in 195 AVP sequences, 259 nonactive AVP sequences, and 492 randomly selected non-secretory peptides. The filtered dataset is what was used for all of the model training and validation in presented work.

2.2. Principal component analysis

To generate physicochemical Fourier-based features, we first need to convert amino acid sequences into numerical vectors based on amino acid properties. The AAindex dataset, found in the *protr* R package, is a collection of 544 amino acid properties from the literature that include various physicochemical descriptors. All of the 544 amino acid properties could be used to convert the amino acid sequences into property vectors; however, this would result in thousands of potential features once the property vectors were converted into Fourier coefficients. Alternatively, we can perform dimensionality reduction to reduce the number of amino acid property vectors prior to conversion to Fourier coefficients. In the current work, we have applied principal component analysis to generate amino acid property factors, as has been proposed previously. We used the *prcomp* function in R to extract principal components and to select a subset of the principal components based on contributions to the overall variance in the data (Figure 1).

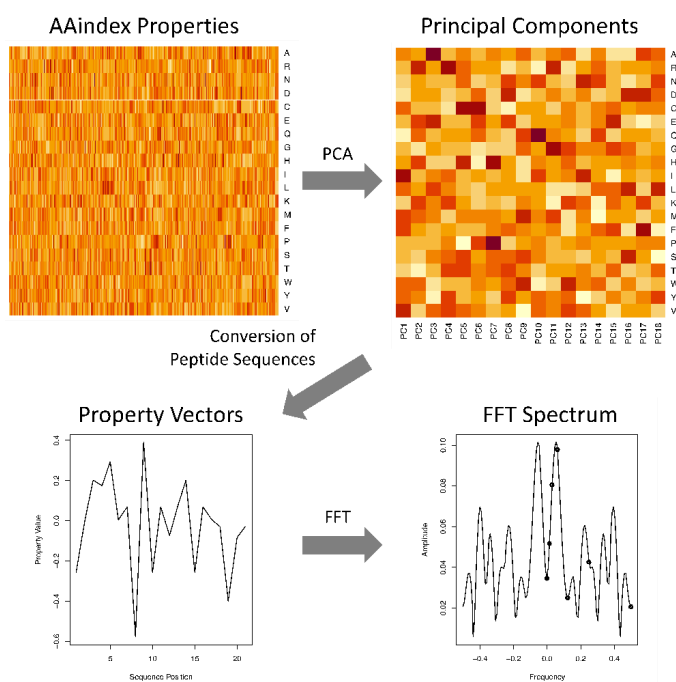


Figure 1. Schematic of feature extraction procedure based on amino acid property periodicities.

2.3. Feature extraction

Based on the generated amino acid property factors, we then converted the amino acid sequence of each peptide into physicochemical vectors (Figure 1). Fourier transforms, using the *fft* function in R, were then applied to each of the property factor vectors. To ensure that the same number of frequency (Fourier) components were generated for each property vector we used zero padding and an assumed maximum sequence length of 128 amino acids (maximum AVP sequence length in the dataset is 107 amino acids). The moduli of the complex Fourier coefficients for frequency values of 0, 0.015625, 0.03125, 0.0625, 0.125, 0.25, and 0.5 were selected as the features for training models (Figure 1). The frequency components (features) corresponding to periods that are

longer than a given peptide were set to zero. We eliminated features (columns) from the full set of features if more than 70% of peptides had a value of zero.

2.4. Support vector machines

All support vector machines were trained using the *svm* function of *e1071* R package based on the radial basis function nonlinear kernel. The cost and gamma hyperparameters of the SVM models were tuned using a grid-search with cost and gamma values based on powers of two, $2^n \forall n \in \{-9, \dots, 8\}$, where n is an integer. Five-fold cross-validation, based on balanced training sets containing 435 AVP and 435 non-AVP sequences, was used to tune and validate the models based on first sorting the peptide sequences according to length and then select five training and testing sets with an equal number of samples for each peptide class. Model performance was measured based on classification accuracy and is reported as the fraction of classes (AVP or non-AVP) that was predicted correctly in the testing sets (Figure 3). The reported cross validation accuracies are the average of the classification accuracies for the five training and testing sets.

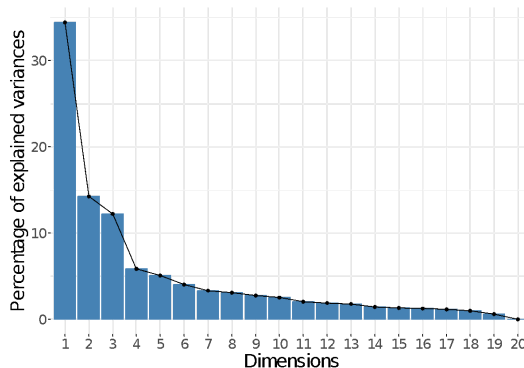


Figure 2. Variance explained by the 20 principal components from the PCA of the 544 AAindex amino acid properties.

2.5. Feature selection

Feature selection is a crucial aspect of data science as it can enable the identification of an essential set of predictive descriptors (features), as well as it can increase the robustness of models to prevent overfitting. Previously, we have developed a feature selection algorithm based on non-linear SVMs, which is general in nature and has been applied to predicting fault detection in chemical plants (Onel et al. 2018; Onel et al. 2019) and HIV-1 viral entry (Kieslich et al. 2016). The algorithm is model-based and requires first training a SVM model prior to computing a criterion that quantifies the contribution of each feature to the SVM objective function to determine which features to remove. The criterion (Eq. 1) is derived based on sensitivity analysis of the dual formulation of SVM models.

$$crit_k = - \frac{1}{2} \sum_i \sum_j \alpha_i^* \alpha_j^* y_i y_j \frac{\partial K(x_i^{\circ z}, x_j^{\circ z})}{\partial z_k} \Big|_{z=1} \quad (1)$$

The algorithm uses a greedy approach to rank the features, where we start with a training model based on all of the features, compute the criteria for all features, and remove a fraction of the features with the largest criteria values. In the presented work,

we removed 25% of the remaining features after each iteration of the algorithm and returned the hyperparameters after each iteration of the algorithm. The feature ranking procedure was applied to each of the five training sets and a consensus ranking was generated based on the average rank of each feature across the five training sets.

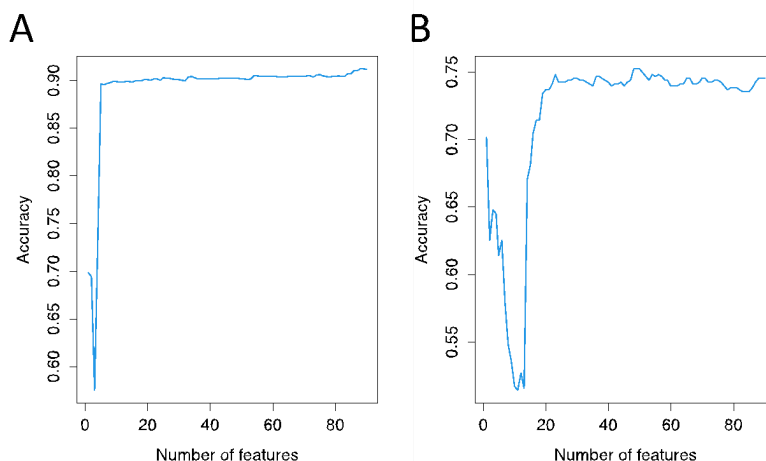


Figure 3. Feature contributions to model accuracy based on feature selection ranking. A) Feature selection results for classifying AVPs vs. random peptides; B) Feature selection results for classifying AVPs vs. non-AVPs.

3. Results

The principal component analysis of the amino acid properties from AAindex dataset was based on a data matrix consisting of 20 instances (amino acids) of 544 variables (properties). The PCA analysis generated 20 principal components that describe the more than 500 amino acid properties, and the contribution of each principal component is visualized in Figure 2. In choosing which principal components to use as amino acid property factors, we selected the principal components which contribute more than 1% of overall variance in the data. The first 18 principal components met our criteria (>1% variance) (Figure 2), and together describe over 99% of the total variance. The analysis resulted in 18 amino acid property factors that were used to generate 18 property vectors for each of the peptide sequences. For each peptide, the FFT spectrum of each property vector was computed and the frequency components corresponding to the sequence average and the oscillations with periods of 2, 4, 8, 16, 32, or 64 amino acids were extracted. This resulted in 126 features based on 7 frequency components from 18 property vectors for each sequence, which was filtered to 90 features by removing features with at least 70% of the values being zero.

Based on the generated features, we developed two SVM models, one to distinguish the AVP peptides from each of the types of non-AVP peptides (nonactive and random non-secretory peptides). For both classification tasks, we performed feature selection to rank the physicochemical features. To measure the contribution of each feature to model accuracy we performed five-fold cross validation after adding each feature one at a time starting with the highest ranked feature. As can be seen in Figure 3, distinguishing AVPs from random non-secretory peptides is an easier classification task

than distinguishing AVPs from nonactive AVPs, since the maximum accuracy when using random peptides is 0.912 and only 0.752 for the nonactive peptides. Only 5 of the 90 features are necessary to achieve the majority of the accuracy of the AVP-vs-random model, while about 4 times as many features are necessary to achieve the maximum accuracy of the AVP-vs-nonactive model, which is further evidence of the difficulty distinguishing between active and nonactive AVPs.

4. Conclusions

In this study, we have developed support vector machine models that distinguish between antiviral peptide sequences and two classes of nonAVP sequences. To develop these models, we first generated amino acid property factors by applying principal component analysis to a dataset on amino acid properties from the literature. We then used the property factors to convert AVP sequences into property vectors that served as the input for Fourier analysis to extract the features used in training our models. The proposed approach for feature extraction and model development, including the incorporation of the feature selection algorithm, have potentially applications in prediction of peptide properties and function. Future work will be aimed at improving the Fourier-based encoding of peptide sequences and applying the approach to predicting various peptide functions/properties, as well as further development of approaches for SVM-based feature selection. The models developed in this study could have potential use in designing novel antiviral peptides but given the remaining challenges in distinguishing between active and nonactive AVPs further investigation is necessary, which may need to include both computational and experimental studies.

References

- C.A. Kieslich, P. Tamamis, Y.A. Guzman, M. Onel, C.A. Floudas, 2016, Highly accurate structure-based prediction of HIV-1 coreceptor usage suggests intermolecular interactions driving tropism. *PLOS ONE*, 11(2), e0148974.
- M. Onel, C.A. Kieslich, E.N. Pistikopoulos, 2019, A nonlinear support vector machine-based feature selection approach for fault detection and diagnosis: Application to the Tennessee Eastman process. *AIChE J.*, 65 (3), 992-1005.
- M. Onel, C.A. Kieslich, Y.A. Guzman, C.A. Floudas, E.N. Pistikopoulos, 2018, Big Data Approach to Batch Process Monitoring: Simultaneous Fault Detection and Identification Using Nonlinear Support Vector Machine-based Feature Selection. *Comput. Chem. Eng.*, 115, 46-63.
- S. Rackovsky, 1998, Hidden sequence periodicities and protein architecture, *P. Natl. Acad. Sci. USA*, 95, 8580-8584.
- N. Thakur, A. Qureshi, M. Kumar, 2012, AVPpred: collection and prediction of highly effective antiviral peptides, *Nucleic Acids Res.*, 40, W199- W204.
- D. Eisenberg, R.M. Weiss, T.C. Terwilliger, 1984, The Hydrophobic Moment Detects Periodicity in Protein Hydrophobicity, *P. Natl. Acad. Sci. USA*, 81(1), 140-44.
- F.H. Wagh, R.S. Barai, P. Gurung, S. Idicula-Thomas, 2016, CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nuc. Acids Res.*, 4, 44,1094-1097.
- G. Wang, X. Li, Z. Wang, 2016, APD3: the antimicrobial peptide database as a tool for research and education, *Nuc. Acids Res.*, 4, 44,1087-93.

Allosteric Regulation of CRYs in Mammalian Circadian Clock

Onur Ozcan^a, Seref Gul^{b,c} and Ibrahim Halil Kavakli^{a,b*}

^a*Koc University, Molecular Biology and Genetics, Rumelifeneri Yolu, Istanbul, 34450, Turkey*

^b*Koc University, Chemical and Biological Engineering, Rumelifeneri Yolu, Istanbul, 34450, Turkey*

^c*Current address: Istanbul University, Faculty of Sciences, Department of Biology, Biotechnology Division, Istanbul, 34134, Turkey*

hkavakli@ku.edu.tr

Abstract

Mammalian circadian clocks are composed of transcriptional-translational feedback loops. Transcriptional activators (BMAL1/CLOCK) form positive arm and transcriptional repressors (CRYPTOCHROMEs (CRYs) and PERIODs (PERs)) form the negative arm of the clock mechanism in mammals. CRYs have conserved primary (FAD binding) and secondary pockets critical to interact with different proteins. Despite high structural similarities between CRY1 and CRY2, studies suggest each CRY plays a different role in the circadian clock. For example, the binding of the CRY1 to CLOCK is regulated by a dynamic serine-rich loop (Ser-loop) around the secondary pocket. A recent report showed that a distant residue Arg-293 allosterically regulates the Ser-loop in CRY1. Here, using molecular dynamics simulations, we showed that Arg-311 to His mutation in CRY2 (homolog of Arg-293 of CRY1) had a minor effect on the allosteric path to Ser-loop.

Keywords: Circadian Clock, Cryptochromes, Allosterism, Molecular dynamics simulation

1. Introduction

The circadian clock regulates several behavioral and physiological processes such as alertness, memory, heart rate, blood pressure, and immune responses through rhythmic transcriptional regulation (Cermakian *et al*, 2013). Several diseases e.g. diabetes, obesity and some sleep disorders are associated with circadian rhythm abnormalities (Sahar & Sassone-Corsi, 2009).

At a molecular level, BMAL1 and CLOCK form heterodimer and bind to E-box (CACGTG) in the promoter region of clock-controlled genes (CCG) including *Period* (*Per*) and *Cryptochrome* (*Cry*) to initiate their transcription in mammals (Aryal *et al*, 2017). After accumulation and translocation of PER and CRY along with casein kinase I ϵ (CKI ϵ) into the nucleus, they bind BMAL1/CLOCK dimer and inhibit transcription (Aryal *et al.*, 2017; Kavakli *et al*, 2017; Kavakli & Sancar, 2002). The degradation of

CRY and PER proteins relieves the repression of BMAL1/CLOCK, restarting the cycle (Takahashi, 2016).

There are two *Cry* genes in mammals and studies suggested that CRYs have different roles in the circadian clock as discussed in (Cal-Kayitmazbatir *et al.*, 2021; Fribourgh *et al.*, 2020; Gul *et al.*, 2020). CRYs contain a photolyase homology region (PHR) comprising primary (FAD binding) and secondary pockets and an extended C-terminal region (Kavakli *et al.*, 2017; Ozber *et al.*, 2010; Partch *et al.*, 2005). PHR and C-terminal regions of the CRYs are necessary to maintain rhythmicity and amplitude of circadian rhythm, respectively (Gao *et al.*, 2013; Khan *et al.*, 2012). CRY1 interacts directly with CLOCK and BMAL1 through two distinct regions in the PHR domain (Michael *et al.*, 2017). C-terminal regions of CRYs interact with the transactivation domain (TAD) of BMAL1 (Czarna *et al.*, 2011; Xu *et al.*, 2015). Mutational analysis of CRYs showed that residues around the secondary pocket of CRY1 and CRY2 are critical for the CLOCK binding and their repression activity (Rosensweig *et al.*, 2018). A human gain-of-function CRY1 variant (exon 11 skipping mutation in C-tail of CRY1), exhibiting high affinity to BMAL1/CLOCK, and another CRY1 mutant (exon 6 skipping mutation in PHR of CRY1), causing an arrhythmic phenotype in a cell line, were found in people suffering from familial delayed sleep phase disorder and attention deficit/hyperactivity disorder (Onat *et al.*, 2020; Patke *et al.*, 2017).

Our recent study revealed that the Arg-293 of CRY1, within the FAD binding pocket is important for the allosteric regulation of Ser-loop and has affected the molecular clock (Gul *et al.*, 2020). Sequence alignment showed that the Arg-293 is a conserved amino acid residue in both CRY1 and CRY2. Thus, in this study, we performed the molecular dynamics simulation to understand the role of Arg-311 in CRY2 (the homolog of Arg-293 in CRY1) in the allosteric regulation of the Ser-loop.

2. Materials and Methods

Complete photolyase homology region of mouse CRY2 with accession number of NP_034093.1 was obtained from NCBI database. Sequence was submitted to the RaptorX web-server to get complete PHR structures for CRY2 (Kallberg *et al.*, 2012). CRY2-Arg-311His (homolog of Arg-293 of CRY1) mutant protein structure was generated using VMD psfgen plugin. Protonation states of ionizable residues of all four models calculated using PDB2PQR webserver that uses PROBKA. The CHARMM36m forcefield and NAMD software were used for parametrization and to run simulations respectively. Systems were solvated in a rectangular box that extends 10 Å from the closest protein surface by TIP3P water. Periodic boundary conditions and Particle Mesh Ewald (PME) used for full electrostatic interactions with a 12 Å cutoff. All covalent hydrogen bond lengths were constrained with SHAKE algorithm with a time step of 2 fs. Systems were, then, neutralized with the appropriate amount of sodium and chloride ions. Energy minimization was performed for 100,000 steps. After, systems were heated and equilibrated for 140ps. The temperature increased gradually, 5K per step, until 310K using a Langevin thermostat. NPT equilibration performed at 310K and under 1 atm, pressure using Langevin barostat. During equilibration, harmonic restraints of 3 kcal·mol⁻¹·Å⁻² on Ca atoms were gradually removed. Production simulations were run for CRY2, and CRY2-Arg-293His for 300ns and 150ns, respectively. Temperature and pressure were maintained by the Langevin thermostat and the Langevin barostat, respectively. NAMD software and CHARMM force fields were used to run all MD simulations (Phillips *et al.*, 2005).

For analysis of simulations, all trajectories were aligned to the initial structure, and we discarded first 10 ns of all simulations. RMSD, H-bonds and RMSF calculations were performed by using VMD. Each of the refined trajectories was saved as two multi-frame pdb files with 1 ns and 100 ps intervals for analysis to be used in POVME and WISP, respectively. Secondary pocket volume was calculated using POVME 3.0 software (Wagner *et al.*, 2017), with a 1 ns intervals. The Wilcoxon rank-sum test was used to determine the significance of the difference between the pocket volumes implemented in R (Wickham, 2009). WISP 1.1 software was used to analyze the allosteric pathways (Van Wart *et al.*, 2014).

3. Results

Our previous work showed that p.Arg293His CRY1 variant exhibits CRY2-like *in vitro* properties (Gul *et al.*, 2020). Our further molecular dynamic simulation revealed that the Arg-293 is responsible for allosteric regulation of the Ser-loop, which plays an important role in CLOCK-CRY1 binding. The multiple sequence alignment with other CRYs from different organisms showed the conservation of this residue, including mammalian CRY2. To understand the role of Arg-311 (homolog of Arg-293 of CRY1) in allosteric regulation of CRY2 we employed molecular dynamics (MD) simulations using CRY2 and CRY2-R311H structures. For comparison, previously reported simulation data of CRY1 and p.Arg293His CRY1 were used (Gul *et al.*, 2020). Root-mean-square deviation (RMSD) values for each structure was calculated to show that simulations reached equilibrium (**Figure 1A**). Next, the dynamics of CRY1 and CRY2 were compared by plotting root-mean-square fluctuation (RMSF) of each amino acid residue in CRYs. Overall, the dynamicity of both CRYs is different, especially in α/β domains including Ser-loop and P-loop (**Figure 1B**). Ser-loop in CRY1 was found to be highly dynamic compared to Ser-loop of CRY2. We then compared RMSF plot of CRY2-R311H with CRY2 to evaluate change in the dynamicity of both structures. Results demonstrated that Ser-loop dynamicity are very similar in both structures. However, the P-loop behavior was greatly affected by the mutation in CRY2 (**Figure 1B, C, D**). The same mutation increased the dynamicity of C-lid that is critical for CRY-PER2 and CRY-FBXL3 interactions (Czarna *et al.*, 2013), in both CRYs (**Figure 1C, D**). Collectively, our data suggested that the highly conserved Arg amino acid residue in CRY1 and CRY2 has differential effects on the CRYs.

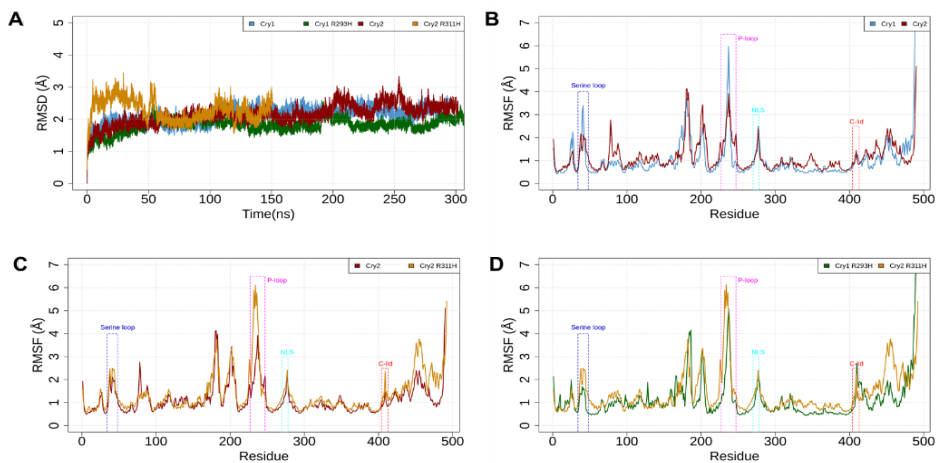


Figure 1: **A)** Root mean square deviation (RMSD) of all simulations (CRY1, CRY1-R293H, CRY2, CRY2-R311H). Root mean square fluctuation (RMSF) of **B)** CRY1 and CRY2; **C)** CRY2 and CRY2-R311H; **D)** CRY1-R293H and CRY2-R311H. For the sake of figure clarity, residue numbering in CRY1 was followed. The critical and functional parts of CRYs were shown between dashed lines. NLS: Nuclear localization signal.

We then calculated the secondary pocket volumes (SPV) of CRY2 and CRY2-R311H and compared them previously reported CRY1 and CRY1-R293H results (Gul *et al.*, 2020) to analyze the effect of mutations, where CLOCK binds. The SPV of the CRY2-R311H was quite larger than both CRY1 and CRY1-R293H and comparable to that of CRY2 (**Figure 2A**). Unlike Arg-293 mutation to His in CRY1, Arg-311 to His mutation in CRY2 did not affect the SPV. The distance between the Ser-loop and secondary pocket was not affected in CRY2-R311H, however, the same distance was affected in CRY1-R293H (**Figure 2B**). Finally, the number of H-bonds was calculated between the Ser-loop and the rest of the protein. The decreased distance between Ser-loop and secondary pocket in CRY1-R293H can be explained by more hydrogen bond formation throughout the simulation compared to CRY1. Amount of hydrogen bond per frame was similar in CRY2 and CRY2-R311H. H-bond analysis suggests that the effect of Arg-311 mutation in CRY2 has a different effect on CRY2 than that of its homolog mutation in CRY1 (**Figure 2C**). To address the effect of R311H mutation on the allosteric regulation of Ser-loop in CRY2 we investigated the allosteric paths between Ser-loop and Arg-311 using WISP. The shortest lengths of the allosteric paths (between Ser-loop and residue 311) in CRY2 and CRY2-R311H were 3.4 Å and 2.2 Å, respectively. Other suboptimal paths to Ser-loop were quite different in CRYs (**Figure 2D**). WISP analysis for CRY2 structures in this study and our previous CRY1 results (Gul *et al.*, 2020) showed that: 1) CRY1 and CRY2 path lengths are changing in different magnitudes, implying that allosteric regulations between the Arg293 and the Arg-311 and Ser-loop in CRY1 and CRY2 are different, respectively; 2) while Arg293His mutation changed the shortest path length dramatically in CRY1, Arg-311 to His mutation in CRY2 had less effect on the path length. To note, SPV mainly affected by the position of highly dynamic unstructured Ser-loop in CRY1, however, Ser-loop in CRY2 has a turn of an alpha helix that increases its rigidity. Because of the rigidity of

Ser-loop in CRY2, the effect of mutation might not alter its dynamics as drastically as in the CRY1.

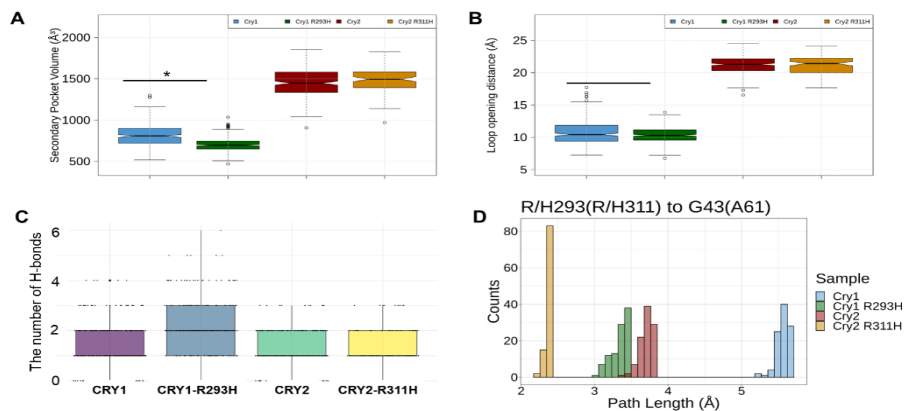


Figure 2: Box-plot representation of the **A**) secondary pocket volume (Wilcoxon rank sum test *: $p < 0.0001$); **B**) loop distance between Ser-loop (Gly43/Ala61) and Phe 105/123 (first residue number is for CRY1, second for CRY2), **C**) number of H-bonds between Ser-loop and rest of the protein per frame. The same number of pdb files were used for each calculation. **D**) Histogram representation of pathways' length calculated by using WISP.

Conclusions

Despite high primary sequence and tertiary structure similarities between CRY1 and CRY2, their different functions in the mammalian circadian clock remained elusive. One of the most notable differences between CRY1 and CRY2 is their differential affinity to CLOCK. Recent reports shed light on the critical amino acids and loop dynamics for the CRY-CLOCK binding (Fribourgh *et al.*, 2020; Rosensweig *et al.*, 2018). Our recent paper identified unique allosteric regulation between the Arg-293 in the primary pocket and Ser-loop in CRY1 (Gul *et al.*, 2020). In this report, we analyzed the role of homolog Arg residue (Arg-311) of CRY2 in allosteric regulation using MD simulations. Full-length PHR domains of CRY2 and CRY2-R311H were modeled and simulated and compared with previously reported CRY1 simulations. RMSF analysis showed that CRY1 and CRY2 have different dynamics. Mutations on CRY1 and CRY2 differentially affected the overall dynamicity. While dynamics of Ser-loop were significantly reduced in CRY1-R293H mutant, it didn't get affected in CRY2-R311H mutant. H-bond analysis suggested that the reduced Ser-loop dynamicity may stem from an increase in H-bond interaction between the loop and the rest of the protein in CRY1-R293H. Unlike CRY1-R293H, the secondary pocket volume of the CRY2-R311H remained unchanged. The dynamics of CRY1 and CRY2 were drastically different in primarily two regions, Ser-loop and P-loop. While CRY1-R293H had reduced the dynamics of both regions, CRY2-R311H had a subtle effect on the dynamics of Ser-loop but increased the dynamics of the P loop. Consistently both mutations increased the dynamicity of C-lid. Experimental studies need to be carried out to show the physiological significance of such dynamical changes in the CRY2.

References

- Aryal RP, Kwak PB, Tamayo AG, Gebert M, Chiu PL, Walz T, Weitz CJ (2017) Macromolecular Assemblies of the Mammalian Circadian Clock. *Mol Cell* 67: 770-782 e776
- Cal-Kayitmazbatir S, Kulkoyluoglu-Cotul E, Growe J, Selby CP, Rhoades SD, Malik D, Oner H, Asimgil H, Francey LJ, Sancar A, Kruger WD, Hogenesch JB, Weljie A, Anafi RC & Kavakli IH (2021) CRY1-CBS binding regulates circadian clock function and metabolism. *FEBS J* 288: 614-639
- Cermakian N, Lange T, Golombek D, Sankar D, Nakao A, Shibata S, Mazzocchi G (2013) Crosstalk between the circadian clock circuitry and the immune system. *Chronobiol Int* 30: 870-888
- Czarna A, Berndt A, Singh HR, Grudziecki A, Ladurner AG, Timinszky G, Kramer A, Wolf E (2013) Structures of Drosophila Cryptochrome and Mouse Cryptochrome1 Provide Insight into Circadian Function. *Cell* 153: 1394-1405
- Czarna A, Breitkreuz H, Mahrenholz CC, Arens J, Strauss HM, Wolf E (2011) Quantitative Analyses of Cryptochrome-mBMAL1 Interactions mechanistic insight into the transcriptional regulation of the mammalian circadian clock *J Biol Chem* 286: 22414-22425
- Fribourgh JL, Srivastava A, Sandate CR, Michael AK, Hsu PL, Rakers C, Nguyen LT, Torgrimson MR, Parico GCG, Tripathi S, Zheng N, Lander GC, Hirota T, Tama F & Partch CL (2020) Dynamics at the serine loop underlie differential affinity of cryptochromes for CLOCK:BMAL1 to control circadian timing. *Elife* 9
- Gao P, Yoo SH, Lee KJ, Rosensweig C, Takahashi JS, Chen BP, Green CB (2013) Phosphorylation of the cryptochrome 1 C-terminal tail regulates circadian period length. *J Biol Chem* 288: 35277-35286
- Gul S, Aydin C, Ozcan O, Gurkan B, Surme S, Baris I, Kavakli IH (2020) The Arg-293 of Cryptochrome1 is responsible for the allosteric regulation of CLOCK-CRY1 binding in circadian rhythm. *J Biol Chem* 295: 17187-17199
- Kallberg M, Wang HP, Wang S, Peng J, Wang ZY, Lu H, Xu JB (2012) Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7: 1511-1522
- Kavakli IH, Baris I, Tardu M, Gul S, Oner H, Cal S, Bulut S, Yarpavar D, Berkel C, Ustaoglu P, & Aydin C (2017) The Photolyase/Cryptochrome Family of Proteins as DNA Repair Enzymes and Transcriptional Repressors. *Photochem Photobiol* 93: 93-103
- Kavakli IH, Sancar A (2002) Circadian photoreception in humans and mice. *Mol Interv* 2: 484-492
- Khan SK, Xu H, Ukai-Tadenuma M, Burton B, Wang Y, Ueda HR, Liu AC (2012) Identification of a novel cryptochrome differentiating domain required for feedback repression in circadian clock function. *J Biol Chem* 287: 25917-25926
- Michael AK, Fribourgh JL, Chelliah Y, Sandate CR, Hura GL, Schneidman-Duhovny D, Tripathi SM, Takahashi JS, Partch CL (2017) Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of CRY1. *P Natl Acad Sci USA* 114: 1560-1565
- Onat OE, Kars ME, Gul S, Bilguvar K, Wu Y, Ozhan A, Aydin C, Basak AN, Trusso MA, Goracci A, Fallerini C, Renieri A, Casanova JL, Itan Y, Atbasoglu, CE, Saka MC, Kavakli IH & Ozelcik T (2020) Human CRY1 variants associate with attention deficit/hyperactivity disorder. *J Clin Invest* 130: 3885-3900
- Ozber N, Baris I, Tatlici G, Gur I, Kilinc S, Unal EB, Kavakli IH (2010) Identification of two amino acids in the C-terminal domain of mouse CRY2 essential for PER2 interaction. *BMC Mol Biol* 11: 69
- Partch CL, Clarkson MW, Ozgur S, Lee AL, Sancar A (2005) Role of structural plasticity in signal transduction by the cryptochrome blue-light photoreceptor. *Biochemistry* 44: 3795-3805

- Patke A, Murphy PJ, Onat OE, Krieger AC, Ozcelik T, Campbell SS, Young MW (2017) Mutation of the Human Circadian Clock Gene CRY1 in Familial Delayed Sleep Phase Disorder. *Cell* 169: 203-215
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781-1802
- Rosensweig C, Reynolds KA, Gao P, Laothamatas I, Shan YL, Ranganathan R, Takahashi JS, Green CB (2018) An evolutionary hotspot defines functional differences between CRYPTOCHROMES. *Nat Commun* 9
- Sahar S, Sassone-Corsi P (2009) Metabolism and cancer: the circadian clock connection. *Nat Rev Cancer* 9: 886-896
- Takahashi JS (2016) Molecular Architecture of the Circadian Clock in Mammals. In: *A Time for Metabolism and Hormones*, Sassone-Corsi P., Christen Y. (eds.) Cham (CH)
- Van Wart AT, Durrant J, Votapka L, Amaro RE (2014) Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis. *Journal of Chemical Theory and Computation* 10: 511-517
- Wagner JR, Sorensen J, Hensley N, Wong C, Zhu C, Perison T, Amaro RE (2017) POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J Chem Theory Comput* 13: 4584-4592
- Wickham H (2009) ggplot2 Elegant Graphics for Data Analysis Introduction. *Use R!* 1-+
- Xu HY, Gustafson CL, Sammons PJ, Khan SK, Parsley NC, Ramanathan C, Lee HW, Liu AC, Partch CL (2015) Cryptochrome 1 regulates the circadian clock through dynamic interactions with the BMAL1 C terminus. *Nat Struct Mol Biol* 22: 476-U470

Reflections on the development of scenario and problem-based chemical engineering projects

A. Tsatse, E. Sorensen*

Department of Chemical Engineering, University College London, Torrington Place, WC1E 7JE London, UK

e.sorensen@ucl.ac.uk

Abstract

This work reflects on the use of scenario- and problem-based learning as a way of conveying not only fundamental knowledge, but also to provide training in the use of computational Process Systems Engineering (PSE) tools applied to open-ended real world problems. The teaching framework also has a strong emphasis on the development of professional skills and to evaluate the recommended design solutions considering multiple perspectives such as economics, safety, environment and societal context. The framework is implemented through week-long group projects called Scenarios, taking place mainly in the first two years of study, and examples are given of different variations of Scenarios. This teaching approach has multiple benefits, including but not limited to, students' understanding of PSE tools and the development of their critical engineering thinking.

Keywords: problem-based learning, Scenarios, Process Systems Engineering

1. Introduction

University College London (UCL) has long been considered a global leader in engineering education, and has also recently been identified amongst one of the strongest emerging trailblazers delivering cutting edge engineering education (Graham, 2018). The latter is partly due to the introduction of UCL's Integrated Engineering Programme (IEP) which was introduced in 2014/15. The IEP is a teaching framework that allows both specialist and interdisciplinary engineering education to be delivered across different engineering disciplines and in programmes that teach fundamental technical knowledge through interdisciplinary, research-based projects and professional skills. The programme aims to produce independent and self-directed engineers based on innovative teaching practices which includes the use of scenario- and problem-based learning using real world examples (Sorensen, 2016).

One of the main aspects of the IEP is a strong emphasis on modelling and design, starting with two compulsory modules dedicated to mathematical modelling & analysis for all engineering students in Years 1 and 2 (Molaei Chalchooghi and Sorensen, 2018), which for chemical engineering is supplemented by a third module on computational modelling and analysis. The use of computational tools is embedded within the curriculum and within most individual modules, and the students become proficient in the use of Matlab, GAMS, gPROMS and Aspen Plus. More importantly, they are taught

how to critically evaluate the results based on the assumptions made in the definition of the problem.

Most of the curriculum is still dedicated to core engineering disciplinary content, mainly for accreditation reasons, which is delivered through lectures, tutorials and labs, however, the IEP also introduced a series of project-based learning opportunities, in particular, during the first two years of studies. Students in their very first term take on two five-week engineering design challenges, where traditional lectures are replaced by smaller design workshops. During these workshops, they participate in modelling, testing and prototype-making of their engineering solutions for global challenges such as sustainability and global health (Sorensen, 2016). Similarly, within each department, including UCL's Department of Chemical Engineering, core modules have been re-configured into four-week blocks such that they offer a theoretical context for a one-week team-based project, within the IEP called Scenarios, following the block. Scenarios are one week-long projects where students working in groups are asked to tackle an open-ended real world problem that is related to the modules considered in that teaching term. The undergraduate curriculum currently delivers six such Scenarios across Years 1 and 2, all of which have a strong element of modelling and design (Molaei Chalchooghi and Sorensen, 2018), as well as of safety (Pollock and Sorensen, 2018), the latter not considered in this work but is also an integral part of each Scenario. In the following, the structure of the Scenarios will be presented, and examples of Scenario problems will be provided with particular emphasis on Process Systems Engineering (PSE), followed by a discussion of the benefits of this innovative approach on students' learning.

2. Problem-based learning through Scenarios

The Scenarios are research-based projects, firmly embedding the six dimensions of UCL's Connected Curriculum which aims to ensure that all UCL students are able to learn through participating in research and enquiry at all levels of their programme of study (Fung, 2017). The Scenarios teach fundamental engineering knowledge, but also have a strong focus on professional skills, ethics and safety. Of particular focus is the Statement of Ethical Principles set out by UK's Engineering Council and the Royal Academy of Engineering (Royal Academy of Engineering, 2017), an engineering concern that is not normally addressed in technical modules but is highly relevant to the discipline.

Each of the six Scenarios focuses on a different problem, the subject of which is of relevance to the topics that have been covered during the teaching term in the core modules, for instance, transport phenomena, thermodynamics or mass transfer. The Scenarios are usually related to the chemical industries, thus giving the students some insight into the types of problems they will face as practicing engineers during their career. The students are given five days to come up with a solution to the problem, attending progress meetings and facilitation sessions throughout the week, before submitting and presenting their results in a report and through a poster, presentation or video. Whilst the technical topics of the Scenarios are linked to core technical modules, the marks contribute to IEP modules on Design and Professional Skills, covering general learning outcomes related to these themes rather than technical aspects.

The department has used gPROMS ModelBuilder (Process Systems Enterprise, 2020) in undergraduate teaching for over 15 years, within the IEP introduced in the second term of Year 1. Process flowsheeting is introduced at the start of Year 2, with particular emphasis on Aspen Plus (Aspentech, 2017), which together with gPROMS are the main simulation tools used across the curriculum. Both tools are used within the Scenarios, which therefore also serve to introduce students to the importance of PSE and to how they can use their theoretical knowledge and modelling skills in order to model a chemical process, interpret its behaviour and think critically about how the process could be further improved from a design and/or operational perspective.

Team working is a key aspect of the IEP and for each Scenario teams are allocated on a different basis, for instance, based on accumulated average mark or even allocated randomly. A discussion of the rationale behind the different team allocations methods is unfortunately beyond the scope of this paper, so it will therefore suffice to say that, at the end of the sixth and last Scenario, and before the students start their capstone design project in Year 3, students are generally very confident that they can perform in various teams towards a common goal, ensuring that all team members have an equal workload and are respecting each other.

The Scenarios are normally launched on a Monday morning during a plenary introductory session where the Scenario topic and deliverables are presented by the teaching team and the expectations are stated to make sure the students understand the goals for this Scenario. Next, students are informed about the two client (progress) meetings during the week, the helpdesk sessions for any questions they might have and for software support, the deadlines they need to meet, as well as suggested literature. In this way, the teaching team ensures that the students are sufficiently guided, but at the same time have freedom to organize their work within their groups of five to six students (Wankat and Oreovicz, 1993). The teaching team usually consists of four members of teaching/academic staff and two Postgraduate Teaching Assistants (PGTAs) supporting a cohort of around 120 students. Staff members participate in the progress meetings and in the final presentations, whilst PGTAs mainly support students in the helpdesk sessions.

3. Scenario examples

In the following, three different Scenarios will be discussed. The focus will be on a Scenario normally delivered as the fourth Scenario, taking place at the end of the first term in Year 2, to illustrate different ways of delivering the same learning outcomes. The Scenario considered is linked to an introductory module on Separation Processes which covers fundamental aspects of distillation, absorption and extraction as well as basic column design.

Different Scenarios problems are considered each year. This is quite challenging for the teaching staff but ensures that students are not able to plagiarise a previous problem, or even seek assistance from older students. As the module lecturer is always one of the Scenario staff members, they can focus the delivery of the module around the Scenario problem to ensure that all the learning outcomes for the module are met regardless, hence the variation in the Scenario topic is not a concern.

a) Reactive distillation for the production of MTBE

The design of a reactive distillation process was considered in the 2017/2018 academic year. As reactive distillation is not exhaustively covered in the module curriculum, students were introduced to the concept of process intensification in depth for the first time in this Scenario. The first deliverables were therefore focused around reviewing the process; understanding its working principle and why it is considered a prime process intensification example; discussing the advantages and potential limitations of the industrially relevant case study, i.e. the production of MTBE, etc. Then, students were asked to simulate the case study using Aspen Plus (Aspentech, 2017), but were given a template input file from which to start. After the fundamental process modelling step, the next deliverable included suggestions from each team with regards to potential process improvements, demonstrating the benefits of their suggested design revisions on process performance and economics, and grounded in the simulation work.

b) Carbon Capture and Sequestration

Another example includes consideration of Carbon Capture and Sequestration (CCS), introduced in 2019/2020, where students were asked to design a carbon capture absorption unit based on a post-combustion method. The deliverables included reviewing the social context of the method based on recent European greenhouse gases regulations. In addition, teams were required to evaluate alternative process development options, and to design the absorption column for post-combustion using Aspen Plus (Aspentech, 2017) whilst evaluating the associated process safety aspects by conducting a Failure Mode and Effect Analysis (FMEA). The final deliverable also focused on suggestion of process alternatives to encourage students to think about how their process can be further improved, and to justify their choices based on the current (and future?) regulations, process design limitations and simulation results. The students were also asked to consider the concept of sequestration, and to recommend whether or not this should be considered by future engineers. A number of student teams reported concerns with sequestration and suggested other ways of combating climate change should instead be considered.

c) Fermentation bio-butanol process

In a slightly different case study, introduced in 2018/2019, student teams were asked to design a simple batch biofuel separation process for the production of biofuel for use by anthropologists studying indigenous Indian tribes in the Amazon delta. Students first explored current bio-butanol processes, in particular, identify design limitations of the process such as potential azeotropes based on VLE diagrams generated using simulation tools. The next deliverable was to provide suggestions for process developments related to raw materials, resources, biological aspects etc. and to demonstrate the associated benefits based on their simulations. Finally, each team discussed the impact of their suggested process design with regards to local land use, water contamination etc., concluding on whether their solution is acceptable or not, from a sustainability and social perspective, for use in a vulnerable area such as the Amazon. In this particular Scenario, the tutors therefore encouraged the teams to make their decisions not only in terms of process feasibility and performance, but also in terms of its social and environmental impact. The majority of student teams recommended that their design should *not* be implemented due to concerns for the impact on the indigenous tribes, thus clearly demonstrating that the IEP focus on professional skills and social responsibility is taken very seriously by the students.

All the Scenarios mentioned above make use of simulation tools as an integrated part of the Scenario. The aim is very much to make the students proficient in the use of computational tools but for them to see them precisely just as tools to achieve the engineering calculations they require rather than a means to an end.

Prior to the helpdesk sessions, a short video is issued which explains how the process considered works, and how it can be simulated using the appropriate software to ensure that the students have understood the fundamental concepts and how they could proceed with setting up their simulation. During the helpdesk sessions, the students can ask questions, addressing them not only to the tutors (staff and PGTAs) but also to the rest of the class as peer discussion has been found to be particularly effective as a teaching and learning practice (U.S. Department of Education, 1986).

In all Scenarios, progress meetings are held twice during the Scenario week. During these meetings, each student team is given limited time to critically present their findings so far, demonstrating that they are able to meet all deliverables. In cases when teams have difficulty addressing some of the deliverables, the tutors can direct them without directly providing the answer. With regards to the design of the processes considered, all teams are encouraged to properly justify their design decisions based on their simulations, and using proper model analysis tools such as sensitivity analysis.

In the final presentations, each team presents their findings and all team members are expected to participate, including answering questions from the tutoring team. From a teaching perspective, presenting their work in teams is beneficial for students' learning as they learn how to cooperate and obtain knowledge both within the group and individually, and to defend a team decision. For their presentations, time is deliberately very limited, forcing the students to think critically about which findings to present and how to justify their final decisions, leaving the rest for their written report.

4. Discussion

The example Scenarios presented above have hopefully illustrated how different engineering concepts can be considered using project-based learning with a strong focus on team working and professional skills. Each Scenario considers a different real world problem, for which students must make use of computation tools and to use their critical engineering thinking in order to reach a solution. For all the Scenarios presented, the concepts considered have been carefully selected in terms of the knowledge and professional skills students should ultimately obtain. Firstly, students learn how to apply fundamental engineering knowledge (mathematic equations, process principles etc.) in order to tackle an open-ended real world problem. Secondly, understanding how computational tools can be used in order to investigate a process, as well as learning how to take into consideration the impact of their design decisions on a number of aspects (e.g. economic, environmental, societal etc.) is another key aim of these Scenarios. Through scenario-based learning students learn how to convert their theoretical knowledge into not only modelling equations, but also how to investigate the process using sensitivity analyses and optimisation making therefore, design choices which can be properly justified.

It is truly impressive to see how much students achieve during a Scenario given that they were first introduced to the problem at the beginning of the week, having therefore very limited, if any previous, background on the topic. The use of project-based learning undoubtedly has a significant impact on student learning, as has the use of simulation tools required to meet the deliverables. From a teacher perspective, it is highly satisfying to see how the students develop from one Scenario to the next; in particular, how they feel more confident about their suggested solutions and how their confidence in arguing their decisions increases. This is, in our opinion, in part due to the fact that their decisions are based on proper investigations within a computer-aided process engineering framework.

5. Conclusions

This work has presented the rationale behind the use of project-based learning in an integrated engineering programme, implemented through six week-long Scenarios, including the use of computational tools in considering open-ended real world problems. Several example Scenarios have been presented and discussed, highlighting how this teaching approach has multiple benefits, including but not limited to, students' understanding of PSE tools and the development of their critical engineering thinking in their use.

References

- Aspentech (2017). Aspen Plus USA, Aspen Technology. Inc., <https://www.aspentech.com/en/products/engineering/aspen-plus> [Accessed 05/11/2020]
- D. Fung (2017). *A Connected Curriculum for Higher Education*, UCL Press.
- R. H. Graham (2018). *The global state of the art in engineering education*. Cambridge, MA: Massachusetts Institute of Technology.
- M. Molaei Chalchooghi and E. Sorensen (2018). *Supporting the use of PSE computational tools across a chemical engineering program*, Computer Aided Chemical Engineering, 44, 1651-1656.
- M. Pollock and E. Sorensen (2018). *Reflections on embedding safety throughout the process engineering program*, Computer Aided Chemical Engineering, 44, 1633-1638.
- Process Systems Enterprise (2020), gPROMS ModelBuilder, <https://www.psenderprise.com/products/gproms/modelbuilder>. [Accessed 05/11/2020]
- Royal Academy of Engineering (2017). <https://www.raeng.org.uk/publications/reports/statement-of-ethical-principles> [Accessed 05/11/2020]
- E. Sorensen (2016). *Changing the World*. The Chemical Engineer, 27-30.
- U.S. Department of Education (1987). *What Works. Research about teaching and learning* (2nd ed.), Washington, D.C.
- P. C. Wankat and F. S. Oreovicz (1993). *Teaching engineering*, McGraw-Hill: New York.

Programming skills across the (bio)engineering curriculum – a students' perspective

Simoneta Caño de las Heras^a, Carina L. Gargalo^a, Krist V. Gernaey^a, Ulrich Krühne^{a*}

^a *Process and Systems Engineering Centre (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søtofts Plads, Building 228 A, 2800 Kgs. Lyngby, Denmark*

ulkr@kt.dtu.dk

Abstract

Industry 4.0 is moving towards the digitalization of traditional manufacturing processes, and thus, the educational system must adapt and meet the demand of engineers able to work in this digital paradigm. In this context, students, and engineers, are now more than ever required to possess programming skills. Therefore, in this work, we collected and evaluated the students' perceptions about the need of acquiring programming skills in (bio)chemical engineering education. The data has been collected through a quantitative survey distributed to undergraduate and graduate students in the chemical and biochemical engineering study programs at the Technical University of Denmark. In general, the students pointed towards an insufficient presence of programming content in their study plan, and they favor Python as the preferred programming language. Moreover, the students selected actions to improve the integration of programming in their curriculum, focusing on increasing the content of programming applied to the (bio)chemical engineering field and using only one programming language. In addition, the students pointed out modeling and optimization as the most useful topics for their future career. Finally, a learning design has been developed to fulfill these needs and it has been implemented in an educational computer-aided tool, called BioVL (**Bioprocess Virtual Laboratory** - www.biovl.com).

Keywords: Computer-aided tool, Education, Programming, (Bio)chemical engineering.

1. Introduction

Industry 4.0 is a current trend in the bio-manufacturing industry, with the focus to introduce and use digital solutions inside traditional manufacturing. In this context, the digital realm has an important part to play, with examples such as the Internet of Things (IoT) or Digital twins (DT). Accordingly, educational systems must become active agents for this ongoing transformation (Gargalo *et al.*, 2020; Narayanan *et al.*, 2020); universities are engaged in integrating the current and future needs of the digital transition into their curriculum. This digital dimension relies on computational tools and methods to implement the interactions among the operation units, share information between different processes and systems, and adopt decisions without human intervention. Therefore, although there is an increasing need for engineers with programming skills, there is no data on the students' perception and opinion about the

programming content in their education. Meanwhile, students' perception of an activity's value is strongly linked to performance and success in the task. This corresponds to the Expectancy-Value Theory (Wigfield and Eccles, 2000), which is the educational theoretical framework used in this study. In this context, we have collected and evaluated the students' interests and opinions regarding the programming content and implementation level in their curriculum. Moreover, these insights are then used to design a topic-related pedagogical description, which is then implemented in an educational computer-aided tool; called BioVL (**Bi**oprocess **V**irtual **L**aboratory).

2. Methods

A quantitative assessment of the students' perceptions about the need of programming skills in the curriculum was performed. The data was obtained by using a multiple-choice questionnaire. This survey was distributed to students through a link in the following courses: 1) Chemical reaction engineering (the third and final year of their BEng education) (DTU, 2020a); and 2) Process adaptation in Fermentation Based Biomanufacturing (first year of their MSc) (DTU, 2020b). Both courses belong to the chemical and biochemical engineering study programs at the Technical University of Denmark.

3. Results and Discussion

In 2020, 52 students completed the survey: 39% were BEng students, while 61% were MSc students. The questions asked were related to previous knowledge about programming and how they perceived programming to be implemented in their education. This assessment aims to provide a complete view of the students' education gaps and requests; and deliver a road map to better integrate, if necessary, programming in their studies. Therefore, the main aim of this work is to propose a learning design to bridge the identified gaps as well as include, as much as possible, wishes brought up by the students.

3.1. Students' perception: a quantitative approach

Initially, the students were asked if they perceived their programming education as sufficient. The collected data is presented in Figure 1.

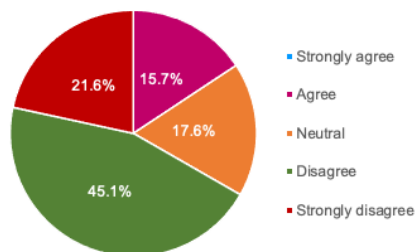


Figure 1. Survey responses to: do my studies include enough programming?

As it can be observed in Figure 1, 66.7% of the participants are convinced that their studies do not include enough programming, and none of the participants strongly agrees on the extent of their current programming education. Hence, this data clearly

identifies that the students feel that the extent of programming in their curriculum is not enough.

The second line of research has to do with gathering the students' views on including the teaching and practical implementation of programming skills into their curriculum.

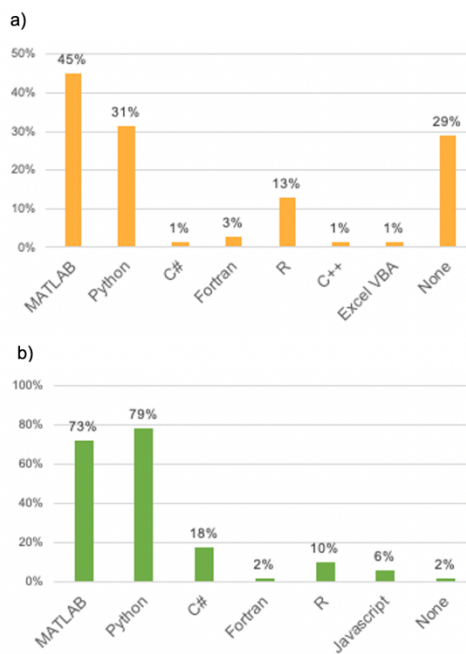
Therefore, as presented in Figure 2, we have started by identifying the students' preferred programming language. It shows that most of the students have previous knowledge on MATLAB or/and Python. Notably, approximately 30% of students declared that they do not have knowledge or experience in any programming language. Considering that the students are in the last year of their BEng or the first year of their MSc, we believe this to be a concerning reality.

On the other hand, Figure 2-b) shows that students prefer Python (79%) and MATLAB (73%) over other programming languages, such as R. However, when asked to select only one programming language, 60% of the participants choose Python, while 30% voted for MATLAB. The remaining students selected R (5%) or were indecisive (5%). The data showed a difference between the programming language that the students had previous knowledge about and the language that they see as most useful for their future careers. To understand the inbuilt motivations behind the students' choices, they were questioned about the characteristics of the different languages that impacted their decision (see Table 1).

Table 1. The most important characteristics identified by the students for the selection of a programming language.

Statement	Percentage
Easy to read and easy to code	66.7%
Open source / Free license	52.9%
Can be used for artificial intelligence and machine learning	11.8%
Several available packages	17.6%
A great online community and support	68.8%
Multidisciplinary	29.4%
Availability of models in chemical and biochemical engineering	39.2%

Table 1 shows a strong agreement in three characteristics: 1) great online community; 2) easy to read and to code; and 3) open source. By selecting these attributes, the students are again supporting their choice for Python, a very popular and well-documented open-source language, built on a philosophy emphasizing code readability. Moreover, the students used the 'extra comments' section in the survey to highlight their interest in favoring the programming languages currently being used in the industry.



Once the students have evaluated their programming skills, they had the opportunity to assess ways to improve them in their education. The survey statements and the collected data are presented in Table 2. Here, a high number of the students advocated for more practice on programming through an applied programming course in the latest part of their studies (65.4%) as well as more applied exercises (59.6%). Besides, most students also agree on their interest in using only one programming language (57.5%). These strategies have already been implemented in universities, such as Lund University that systematically uses MATLAB in their engineering education, or (dos Santos, Vianna and Le Roux, 2018) who implemented a programming course in Fortran in the final year of chemical engineering education with positive but inconclusive results at the University of São Paulo.

Table 2. Students' perspectives on how to improve the teaching of programming in their education.

Statement	Percentage
Only one programming language for all subjects in my studies (e.g. MATLAB, Python)	57.5%
More exercises on programming related to a specific topic	59.6%
Introductory sessions on the programming language before using it in the classroom	55.8%
A teacher assistant for programming in the classroom	36.5%
Hands-on programming course for chemical and biochemical engineering in the latest part of my study plan	65.4%
None of the above	0%

Furthermore, the students added suggestions in the comments section, coming up with proposals such as inviting experts from the industry.

Finally, the students were asked about the expected purpose of the use of their programming skills in their future careers; the collected data is presented in Figure 3.

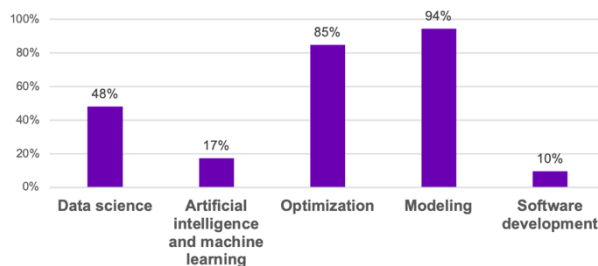


Figure 3. Survey responses related to the expected use of programming in the students' future careers.

These results show that almost all students (94%) selected "modelling" as the main application of programming in their future, closely followed by "optimization" (85%). On the other hand, only 17% have chosen "artificial intelligence and machine learning", even though it currently is a hot topic in (bio)chemical engineering.

To sum up, students were critical about the content of programming in their studies, highlighting a gap in their educational curriculum.

3.2. Learning Design

A specific learning design should be developed to overcome the identified pedagogical need, and later on implemented and validated. This work will mainly focus on the definition of a learning design and its implementation. The developed learning design aims to establish a pedagogical frame to teach applied modelling in Python for chemical and biochemical engineers in the final years of their education mostly through programming exercises.

In general, a learning design consists of the specification of the following points: 1) prerequisites of learning; 2) settings; 3) learning goals; 4) content; 5) learning process; and 6) assessment. This didactic frame was first proposed by (Hiim and Hippe, 1997) and later expanded by (Weitze, 2016).

Table 3 shows the defined characteristics of the developed and implemented learning design in this study. It is important to highlight that the learning goals in Table 3 intent to include and integrate the higher levels of the Bloom taxonomy (Abdulwahed and Nagy, 2009) associated with meta-cognitive knowledge.

To conclude, this learning design has been implemented into a computer-aided tool - BioVL. A prototype of BioVL is available at www.biovl.com. This software focuses on explaining, solving, and enabling the modification of conditions and parameters in mechanistic models for bioprocesses focused on undergraduate education. Currently, BioVL has been designed as a complementary material for the students' education to provide more programming exercises. It includes among its functionalities, a Python library with the models, questionnaires, mini-games, a chatbot, and the possibility to confront realistic operational problems (de Las Heras *et al.*, 2019).

Table 3. Learning design for the teaching of programming to (bio)engineering students.

Element	Definition
Pre-requisite of learning	Students with previous knowledge on process design and a basic knowledge of modelling and programming.
Setting	A computer-aided tool, called BioVL.
Learning goals	It aims to: 1) apply Python to solve (bio)chemical engineering problems; and 2) create (and implement) process models in Python.
Content	It covers: (i) theoretical and practical knowledge of Python and object-oriented modelling; (ii) theoretical knowledge on how to create a model.
Learning process	Collaborative learning between students, a chatbot and the Kolb's experiential cycle.
Assessment	Use of variable feedback and self-evaluation as the students test their hypothesis and develop and test abstract conceptualization (Hattie and Timperley, 2007).

4. Conclusions

Industry 4.0 has the potential to greatly improve (bio)chemical processing. However, to shift paradigms and bring traditional manufacturing to the digital era, engineering students need to be able to communicate and connect the physical and digital world. This is done through programming, and therefore, programming skills must be seen as an essential tool and integrated into the university's curriculum. In addition, the students' involvement in the development of a learning design for the teaching of programming can have significant benefits, such as increased motivation and performance. Therefore, in this work, we quantitatively assessed the students' perceptions about their knowledge of programming, as well as their main interests. After the students have indicated that their curriculum's programming content was insufficient, the survey asked how to improve this based on the students' perception. Furthermore, the students indicated their preferred programming language as being Python, to be applied for modelling and optimization. These needs identified by the students were used as guidelines to define a learning design. Finally, the appropriate learning design is integrated into a computer-aided tool. This computer-aided tool is called BioVL (www.biovl.com), and it is currently in its prototype stage, focusing on modelling of bioprocesses.

References

- Abdulwahed, M. and Nagy, Z. K. (2009) 'Applying Kolb's Experiential Learning Cycle for Laboratory Education', *Journal of Engineering Education*, 98(3), pp. 283–294. doi: 10.1002/j.2168-9830.2009.tb01025.x.
- DTU (2020a) 28342 *Chemical Reaction Engineering*. Available at: <https://kurser.dtu.dk/course/28342> (Accessed: 12 November 2020).
- DTU (2020b) 28455 *Process adaptation in Fermentation Based Biomanufacturing*. Available at: <https://kurser.dtu.dk/course/28455>.
- Gargalo, C. L. et al. (2020) *Towards the Development of Digital Twins for the Bio-manufacturing Industry*. Adv Biochem Eng Biotechnol. doi: https://doi.org/10.1007/10_2020_142.
- Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Review of Educational Research*, 77(1), pp. 81–112. doi: 10.3102/003465430298487.
- Hiim, H and Hippe, E. (1997) 'Learning through experience, understanding and action', *Gyldendal Uddannelse*.
- de Las Heras, S. C. et al. (2019) *Facilitating learning by failure through a pedagogical model-based tool for bioprocesses, 29th European Symposium on Computer Aided Process Engineering*. Elsevier Masson SAS. doi: 10.1016/b978-0-12-818634-3.50305-2.
- Narayanan, H. et al. (2020) 'Bioprocessing in the Digital Age: The Role of Process Models', *Biotechnology Journal*, 15(1), pp. 1–10. doi: 10.1002/biot.201900172.
- dos Santos, M., Vianna, A. S. and Le Roux, G. A. C. (2018) 'Programming skills in the industry 4.0: are chemical engineering students able to face new problems?', *Education for Chemical Engineers*. Elsevier B.V., 22, pp. 69–76. doi: 10.1016/j.ece.2018.01.002.
- Weitze, C. L. (2016) 'Designing for learning and play: The smiley model as a framework', *Interaction Design and Architecture(s)*, 29(1), pp. 52–75.
- Wigfield, A. and Eccles, J. S. (2000) 'Expectancy–value theory of achievement motivation', *Contemporary educational psychology*. Elsevier, 25(1), pp. 68–81.

Engineering Social Responsibility: Rising to the Challenge of Sustainable Development

Iain Duncan Stalker^a, Rinkal K. Desai^b

^a*Institute of Management, University of Bolton, Bolton BS1 1SW, United Kingdom*

^b*WMG, University of Warwick, Coventry CV4 7AL, United Kingdom*

Abstract

To successfully rise to the challenges of contemporary physical and social contexts, and thus secure a more sustainable world, (process) engineering (under)graduates must learn to engage effectively with other professionals. This demands that we structure learning opportunities that make explicit with whom engineers need to engage and promote softer skills that form an essential foundation of effective professional engineering practice. As a step towards this, we present a workspace that provides a holistic perspective on the intersections of the various value chains that obtain in manufacturing and process industries; this framework helps to identify key partners from professions and stakeholder groups. We also show how this framework informs the discussion around employability and professional competence.

Keywords: Engineering Education, Ethical Practice, Sustainable Development, Employability, Circular Economy.

1. Background

Engineers shape the world around us, changing both physical and social contexts. Ingenious materials and innovative industrial processes yield technologies and business models that improve standards of living. Yet, this increased quality of life is not universally enjoyed and in the longer-term the procurement of profits threatens to exact a heavy price on people and the planet; cf. Elkington (1997) and Savitz and Weber (2006). Growing awareness of the cumulative environmental impact of extraction, synthesis, manufacturing, and transport provokes increasing demands for accountability and corporate responsibility (Horrigan 2010). A groundswell of opinion challenges industries to be more transparent in their practices and proactive in replacing the traditional “take-make-waste” model with a so-called “Circular Economy” (MacArthur Foundation 2019).

Engineers are precisely the professionals to take a lead in rectifying such imbalances; indeed, resolving contradictions is a fundamental aspect of an engineer’s mindset; recall for example, TRIZ (Altshuller 2006). Moreover, responsibility in processing, manufacturing, distribution, and use has long been a key imperative in Engineering, cf. Life-Cycle Assessment (e.g., Ashby 2005) and initiatives for sustainable industrial systems (Pozo et al, 2020). Redress demands the “creative destruction” of established processes (Schumpeter 1934) and the innovation of new, more sustainable practices; and this cannot be done in isolation. Engineering systems must be recognised as systems within larger systems; thus, practitioners must work together with stakeholders and

other professionals to accommodate the needs of all those directly or indirectly affected by the outputs—concepts, systems, products, etc.—of their industries; see Freeman (1994).

We must ensure that engineering graduates know with whom to speak, how to engage, why and about what. We must equip them with appropriate skills to debate; and we must ensure that they are familiar with ethical perspectives to promote appreciation of different viewpoints in the pursuit of rapport. As a step towards this, we present a ‘workspace’ that provides a holistic perspective on the intersections of the various value chains that obtain in process engineering, manufacturing and product development; this makes explicit the many paths through which products and services are created and developed (“innovation trajectories”) and the (artificial and natural) contexts from which these draw; this in turn helps to identify key partners from professions and stakeholder groups. Knowing with whom one must engage, how and why directly supports softer skills that are an essential foundation of an effective professional engineering practice. Our key contribution here is a framework to structure appropriate learning opportunities, ensuring that engineering education fulfils these essentials. We also consider this framework a useful mechanism to inform the discussion around employability and professional competence; cf. Dacre Pool and Sewell (2007).

The structure of the article is as follows: in Section 2, we clarify sustainable development; in Section 3, we present our workspace, outlining its construction and briefly illustrating its application; we close in Section 4 with an indication of its relevance to employability

2. Sustainable Development

There are many uses of the terms ‘sustainability’ and ‘sustainable development’; perhaps the most oft-cited definition of sustainable development is that of the Brundtland report (WCED 1987): “development which meets today’s needs without compromising the ability of future generations to meet theirs” (p. 41).

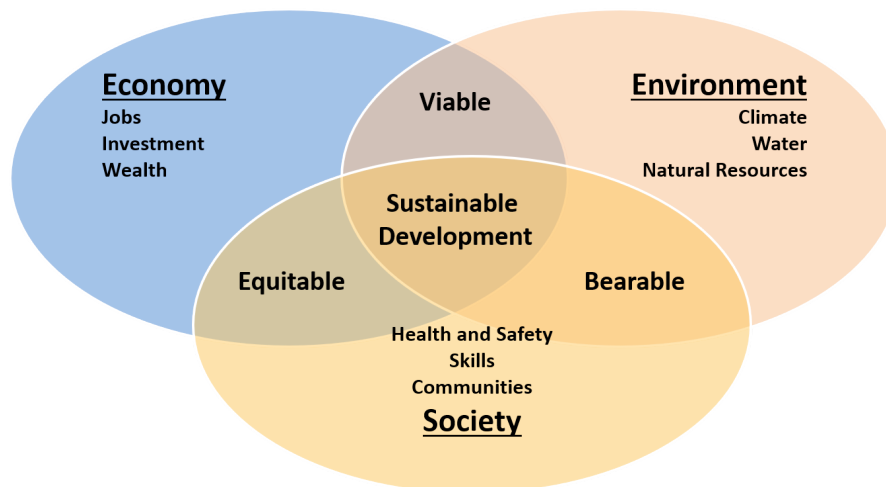


Figure 1: Sustainable Development

We prefer a more wholistic conceptualisation that recognises the three pillars of Elkington (1997), see Figure 1: unless there is explicit consideration of the needs of people (society) and economic viability in conjunction with responsible curation of the environment, we will not have a truly sustainable approach; cf. the observations of *inter alia* Savitz and Weber (2006) that sustainable development arises where “business interests and the interests of the environment and society intersect”.

3. The SAMITE Workspace

3.1. Framework and Trajectories

The SAMITE workspace of Figure 2 (Stalker *et al* 2011) comprises: a framework that coordinates elements essential to a full characterisation and understanding of a product-service bundle, its composition and interaction with users and markets; and ‘innovation trajectories’ that elucidate routes through the framework

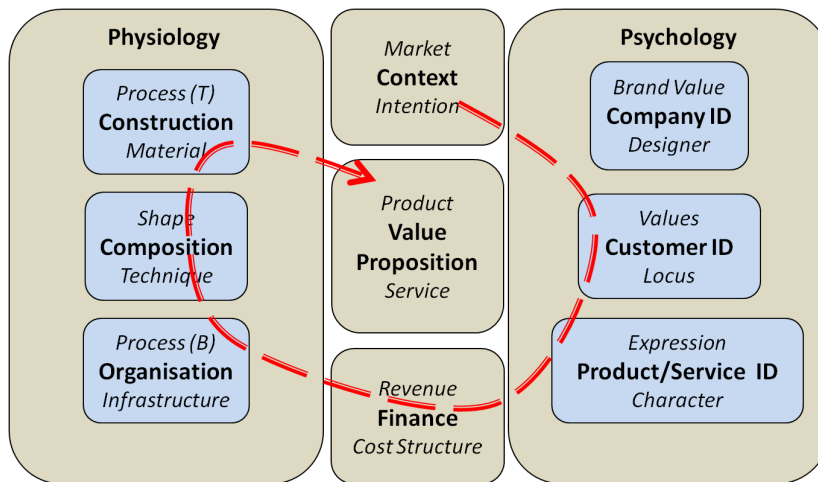


Figure 2: SAMITE Workspace with a Market-Driven Innovation Trajectory

Value proposition denotes a bundle of products and services offered to a specific customer. It acts as the focal point of the framework.

Physiology denotes the structural aspects of the value proposition and comprises

- **Construction** which treats the physical form, material(s), and the technical process(es) employed.
- **Composition** which addresses design approach(es), specific shape(s) and technique(s) applied.
- **Organisation** which addresses the means through which the value proposition is brought to market; it also addresses the intangible analogue(s) of physical form for services; it consists in business processes and the organisational infrastructure or ecosystem.

Market captures the commercial context, identifying target customer(s), i.e. market(s), and design intention.

Finance captures the revenue stream(s) through which the value created will be appropriated and makes explicit the cost structure(s) involved in creating and delivering the value proposition.

Psychology denotes the intangible aspects of the value proposition; it follows the product expression model of Gotzsch (2006) and comprises Company ID, User ID¹ and Product/Service ID.

An ‘(innovation) trajectory’ is a path through the framework; it begins at an ‘epicentre’ and visits each main area; it is descriptive rather than prescriptive. Tracing possible trajectories clarifies how to bring products, services, or bundles to market, e.g. what to consider and whom to consult. This reveals the fundamental nature of the framework as a clarifying structure: it elucidates aspects to be evaluated and considered; it identifies potential roles where teams coordinate their efforts; and it gives suggestions of whom, i.e. what categories of stakeholders, to approach to realise products, services or bundles. Figure 2 illustrates a trajectory with an epicentre in ‘context’: we refer to this as a ‘market-driven innovation trajectory’. The driver for the innovation is a new market opportunity. From this, a profile of the target customer is developed and used to define a product-service bundle to meet the demand. If the market opportunity can be clarified in terms of price, then the product-service bundle definition is augmented with strategic pricing and an appropriate (target) cost structure. This cost structure is likely to influence the choice of business processes, composition techniques and the physical construction of the final offering. To refine the choices and decisions that will ultimately lead to the value proposition, the trajectory will be traced a number of times; each tracing provides opportunities to explore the economic viability of the value proposition, the social contribution and costs of developing the value proposition and the environmental/ecological impact of delivering it.

Naturally, other epicentres are possible and give rise to complementary trajectories; for example, an epicentre in ‘physiology’ deriving from (say) the development of a new (technical) process that is more environmentally friendly would be referred to as a ‘process driven innovation trajectory’ (Stalker *et al* 2011). Often, an innovation will unite and unify a number of developments; for example, a new market opportunity may coincide with the development of a new process, e.g., the more ethically minded consumer demands more sustainable fabrics that can be realised through the use of ‘newer’ fibres such as bamboo. In this case, we consider multiple trajectories from multiple epicentres and refer to the coordination of the individual trajectories into a single trajectory as ‘(innovation) trajectory alignment’ (Stalker *et al* 2011). These invite the coordination of efforts from different disciplines and motivate the education of engineers and other graduates of the future in the context of product-service developments.

3.2. *Summary Case Study: Synthetic Biology*

Synthetic biology is an exciting subfield of bioengineering that promises disruptive innovation in the development and production of drugs and vaccines and therapeutics (Freemont 2019). It comprises *in silico* design informed by real biological data and was identified by the European Parliament as an emerging technology with the potential to combat the COVID-19 pandemic (Kritikos 2020). By its very nature, being

¹ Here, we use ‘locus’ to signify expressions of time, place, status, and culture.

laboratory-informed and computation-based, it promises sustainable approaches to product design and process development. We believe that it makes an interesting and topical case study for the use of the SAMITE workspace in the education of (process) engineers to foster sustainable development and stakeholder engagement: what we refer to as “Engineering Corporate Responsibility”. For example, Bruynseels (2020) examines responsible innovation through synthetic biology and data positionality, exploring several ethical issues on the way. We can view this as a market-driven approach to vaccine development that, when scaled up to production, would trace an innovation trajectory like that of Figure 2. The trajectory makes clear the stakeholders whose engagement is essential to the realisation of commercially viable sustainable process (for the final vaccine) at industrial scale. Moreover, that the preliminary product design is computational—as indeed could be the process synthesis, e.g., cf. Fraga *et al* (2000)—means that the translation of the designs into physical and business processes affords ample opportunity to convene implementation groups of specific stakeholders and professionals that not only have necessary expertise, but would have a vested interest working to ensure the most sustainable process is delivered owing to their engagement.

4. Conclusion

Sustainable development and practices cannot be realised in isolation. To ensure that the challenges of physical and social contexts are confidently met and surmounted, and thus a sustainable world secured, future (process) engineers must convene effective teams of appropriate professionals and relevant stakeholders. They must be educated in what such teams look like. We believe that SAMITE is a constructive step towards this. The notions of epicentre and trajectory foster systematic engagement: making explicit the many paths through which products and services are created and developed and the (artificial and natural) contexts from which these draw helps to identify key partners from professions and stakeholder groups. Moreover, this provides a fundamental platform for communication: it offers an explicit recognition of the interfaces at which the vocabularies of disparate areas must be aligned.

The SAMITE workspace is also a useful mechanism to inform the broader discussions around employability, contributing directly to development of a fuller understanding of the context of ones subject, confidence in knowing how to assemble an appropriate team to address challenges, and reflection on the need to engage with stakeholders; cf. Understanding, Efficacy and Metacognition in the USEM model of Knight and Yorke (2004). Knowing with whom one must coordinate and why directly supports softer skills that are an essential foundation of an effective professional engineering practice; for example, it explicitly supports the ‘personal and professional skills’ components of the CDIO framework (www.cdio.org); the structures can be used to structure interdisciplinary project teams, which supports the ‘interpersonal skills’ component of the CDIO Framework; and of course other professional frameworks.

References

- Altshuller, G (2006). *And Suddenly the Inventor Appeared. TRIZ, the Theory of Inventive Problem Solving. Eighth Printing.* Worcester, MA: Technical Innovation Center, Inc.
- Ashby, M F (2005). *Materials Selection in Mechanical Design (Third Edition).* Oxford, UK: Elsevier.

- Bruynseels, K (2020). Responsible innovation in synthetic biology in response to COVID-19: the role of data positionality. *Ethics Inf Technol.* <https://doi.org/10.1007/s10676-020-09565-9>
- Dacre Pool, L and Sewell, P (2007). The key to employability: developing a practical model of graduate employability. *Education and Training.* Vol. 49 No. 4, pp. 277-89.
- Elkington, J (1997). *Cannibals with Forks: the Triple Bottom Line of 21st Century Business.* Oxford, UK: Capstone.
- Fraga, ES, Steffens, MA, Bogle, IDL & Hind, AK (2000). *An object oriented framework for process synthesis and simulation*, in "Foundations of Computer-Aided Process Design," M F Malone, J A Trainham & B Carnahan, editors (AIChE Symposium Series) 96(323):446-449.
- Freemont, P S (2019). Synthetic biology industry: Data-driven design is creating new opportunities in biotechnology. *Emerging Topics in Life Sciences/* Vol. 3, pp. 651–657.
- Horrigan, B (2010). *Corporate Social Responsibility in the 21st Century: Debates, Models and Practices Across Government, Law and Business.* Cheltenham, Glos, UK: Edward Elgar Publishing.
- Kritikos, M. (2020). *Ten technologies to fight coronavirus.* Brussels: European Parliamentary Research Service (EPRS) - Scientific Foresight Unit STOA.
- MacArthur Foundation (2019). Concept. What is a Circular Economy? MacArthur Foundation. <https://www.ellenmacarthurfoundation.org/circular-economy/concept>. Accessed 7/09/20.
- Pozo, C., A. Galán-Martín, D. Cortés-Borda, M. Sales-Pardo, A. Azapagic, R. Guimerà, G. Guillén-Gosálbez (2020). Global environmental inequality: burdens distribution, dynamic evolution and roadmap towards equality underpinned by systems optimisation. *Journal of Cleaner Production* 270, 121828, 1-12.
- Savitz, AW and Weber, K (2006). *The Triple Bottom Line.* San Francisco, CA: Jossey-Bass/Wiley.
- Schumpeter, J (1934). *The Theory of Economic Development.* Cambridge, MA: Harvard University Press.

Virtualization of Pressure Control Demonstration

Unit – A Step Towards Virtual Laboratory

Miloš Bogataj^{a,*}, Nejc Arh^a, Zdravko Kravanja^a

^a*Faculty of Chemistry and Chemical Engineering, University of Maribor, Smetanova ulica 17, SI-2000 Maribor, Slovenia*

milos.bogataj@um.si

Abstract

In this paper, we present virtualization of the pressure control demonstration unit. The two main objectives were: to determine a dynamic mathematical model of the unit, and to use the mathematical model to virtualize the process that enables open and closed loop simulations. The dynamic model of the unit, which was developed based on experimental input-output data, shows Fit to Working Data greater than 95 %. Finally, the unit was virtualized in a form of a graphical user interface that hides all the modeling components from the user. The virtual unit is thus designed to enable students with limited or no prior knowledge of control theory and modeling of dynamic systems to study and analyze the dynamics of the system and to observe the effects of feedback control mechanisms.

Keywords: Education, Modeling, Process Dynamics, Process Control, Virtual Laboratory.

1. Introduction

The pandemic of 2020 has changed the way knowledge is passed on from teachers to students. Classroom lectures have been replaced by online lectures practically overnight. Due to restrictions, some laboratory exercises, which are essential for undergraduate chemical engineering students to acquire practical competences and skills, had to be modified, condensed, recorded by technical staff and presented online to students (Pintarič and Kravanja, 2020), leaving them with no or only limited practical experience. Process control is regarded as one of core topics of most undergraduate chemical engineering curricula. The course not only exposes students to the rigors of process dynamics and control, but also teaches them how to tackle complex problems and apply good engineering judgment to obtain results of engineering usefulness (Byrne, 2006). Putting theory in practice is crucial in this respect. In a recent study, Zender and Greiner (2020) conclude that learning by computer simulation performs similarly to learning by experimental method. However, to achieve the best learning outcome, the two methods should be considered complementary.

If laboratory or demonstration equipment is not available, there are many simulation environments that can provide virtual learning experience. Some of them are commercial (e.g. *Matlab & Smulink* (www.mathworks.com), *SimulationX* (www.simulationx.com), *PiDISTILL*, *PITOPS-TFI*, *SIMCET* (www.picontrolsolutions.com)) and some are opensource (e.g. *OpenModelica* (www.openmodelica.org), *Scilab & Xcos* (www.scilab.org)).

In the remainder of the manuscript, we present a study on virtualization of the pressure control demonstration unit RT-634 (GUNT GmbH, Hamburg) – a step towards making the laboratory equipment virtual so that it can be made available to students to enhance their learning experience.

2. Case Study and Methods

The pressure control demonstration unit RT-634 is shown in Fig. 1. Its main components are hardware controller (a), control valve (b), pressure tanks (c, d), pressure sensor (e), and needle valve for air tapping (f). The air pressure at the inlet of the controller valve (p^{in}) is assumed constant at 4.5 bar. The actuator driving the control valve is connected to a standard modulated signal (0 V – 10 V). At 0 V the valve is fully closed and at 10 V the valve is fully open. The needle valve can be adjusted to allow for a specific maximal pressure (p^{max}) to be attained in the system. The equipment can be operated either by the hardware controller or by a software controller within the provided instrumentation and control software. In addition to the control functionality, the software also enables data logging.



Figure 1: Pressure control demonstration unit RT-634 (Source: GUNT GmbH).

2.1 Experimental

Experiments were conducted on unit RT-634 to obtain three sets of input-output data needed for data-driven model development at three different values of maximal pressure ($p^{\text{max}} = 3.3$ bar, 3.6 bar, and 4.0 bar). An identical set of input data (i.e. modulated voltage applied to the actuator) was used in all experiments. The input data set was designed by a random number generator (Fig. 2). In this way, most of the important dynamics of the system were excited. The system output (pressure (p)) was recorded in 0.5 s intervals over the duration of the experiments.

2.2 Model identification

The identification of the RT-634 demonstration unit model was performed in Matlab (Matlab, R2019a) using experimental input-output data and System Identification Toolbox.

2.3 Modeling and virtualization

Virtualization of the RT-634 unit was performed in Simulink (Matlab, R2019a) using the identified dynamic input-output model and graphical modeling, simulation, and visualization capabilities of Simulink.

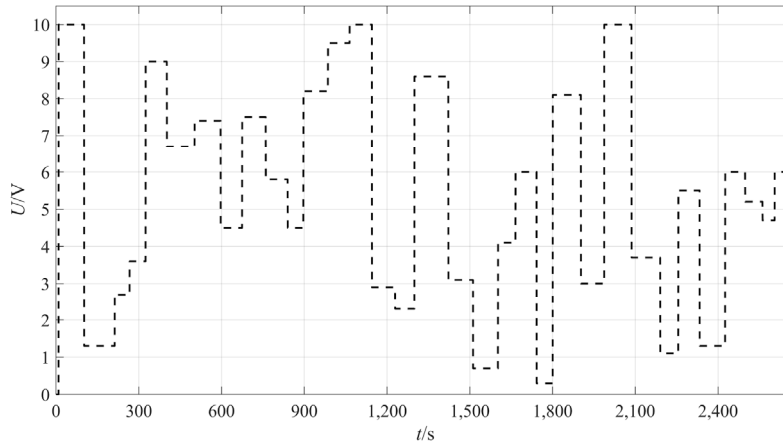


Figure 2: Piecewise-constant input data set (modulated voltage) with variable time step.

3. Results

In this section we present the results of the model identification, and modeling and virtualization of the RT-634 unit.

3.1 Model identification results

Among the many models available in the System Identification Toolbox, the nonlinear Hammerstein-Wiener (NLHW) model was found to produce the best agreement between experimental and simulated data. The NLHW model was set up with a 10-segment piecewise linear function as input nonlinearity, and a 4th order one-dimensional polynomial as output nonlinearity. The metrics related to the quality of the identified models – Fit to Working Data (FWD), Final Prediction Error (FPE) and Loss Function (LS) – are listed in Table 1.

Table 1: Model quality metrics.

Mode	p^{\max} (bar)	FWD (%)	FPE (/)	LF (/)
l				
M3.3	3.3	96.74	0.001	0.001
M3.6	3.6	95.51	0.002	0.002
M4.0	4.0	95.02	0.004	0.004

The results given in Table 1 show that the models provide a good fit to the experimental data. This is indicated by FWD values (normalized root mean squared error as %), which are greater than 95 %. In addition, the values of complementary metrics (FPE and LF), are less than 0.004 in all the cases. The plot representing simulated response to changes in voltage applied to the actuator versus the experimental response to those changes for model M4.0 is presented in Fig. 3. The plot shows that even at FWD = 95 %, the simulated

response is in good agreement with the experimental data over the entire range of tested input voltage values.

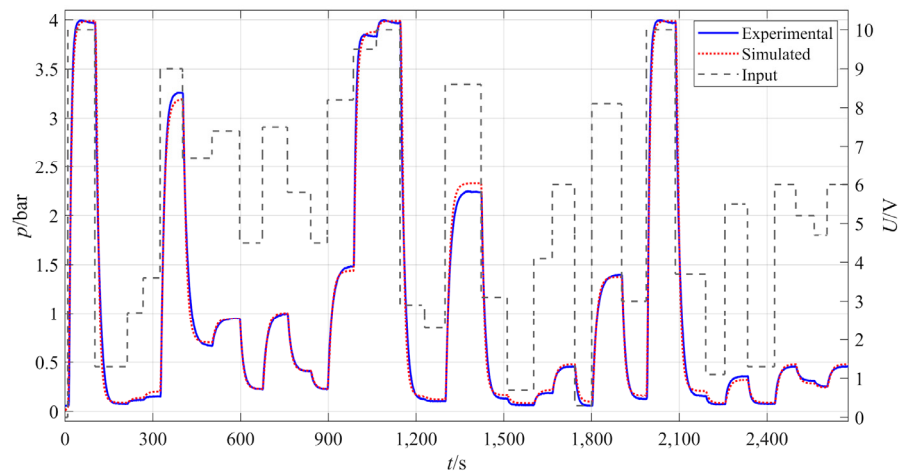


Figure 3: Experimental and simulated response to modulated voltage input (Model: M4.0).

3.2 Modeling open and closed loop response

Fig. 4 shows the Simulink model of the demonstration unit RT-634. The model can be used to simulate open and closed loop response. The developed dynamic models (M3.3–M4.0) are embedded in subsystem NLHWs and can be switched on and off during simulation as required. To avoid unnecessary discontinuities in the model, switching between open and closed loop operation is accomplished through gain K , which can take a value of 1 or 0. The model operates in an open loop mode if $K = 1$, on the other hand, if $K = 0$, the model operates in closed loop mode. For consistent and realistic simulation results and to prevent numerical difficulties, a saturation block is put downstream of the NLHWs subsystem, with the lower saturation limit set to 10^{-5} and the upper saturation limit to 4.5. In addition, the PID output is limited between the values 0 and 10.

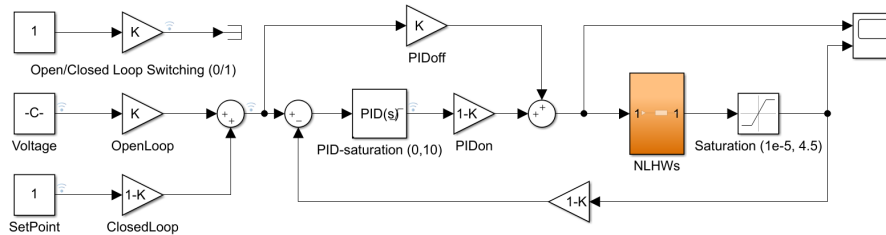


Figure 4: Simulink model.

The model shown in Fig. 4 is fully functional in terms of simulation capabilities, but a novice unfamiliar with the Simulink environment might find the user experience overwhelming. A graphical user interface that hides the model and presents it in a more familiar look, revealing only the necessary features, can be built using elements from Simulink block libraries such as subsystems, dashboard switches, sliders, gauges, scopes etc.

3.3 Graphical virtualization

The virtual RT-634 graphical user interface is shown in Fig. 5. The block model (Fig. 4) is embedded in a subsystem and masked by the image of the demonstration unit. The control panel is equipped with a rotary switch to select the maximal attainable pressure (p^{\max}). A toggle switch enables the user to switch between open and closed loop mode. In open loop mode, the voltage applied to the actuator can be set using the slider. On the other hand, in closed loop mode, the user can adjust the set-point value and set the values of PID controller constants. All the inputs available on the control panel can be adjusted during the simulation run.

The simulation results are shown in two plots. In open loop simulation, the results shown in the plots are the value of the process variable (p) and the voltage applied to the actuator. In loop simulation, the results shown are the value of process variable (p) and the set-point value.

Using Simulink Simulation Pacing, the simulation time can be synchronized with the wall clock time, making the learning experience more realistic. All data generated during the simulation is saved to Matlab Workspace and can be used for additional analyses (e.g. system identification, model linearization, controller tuning).

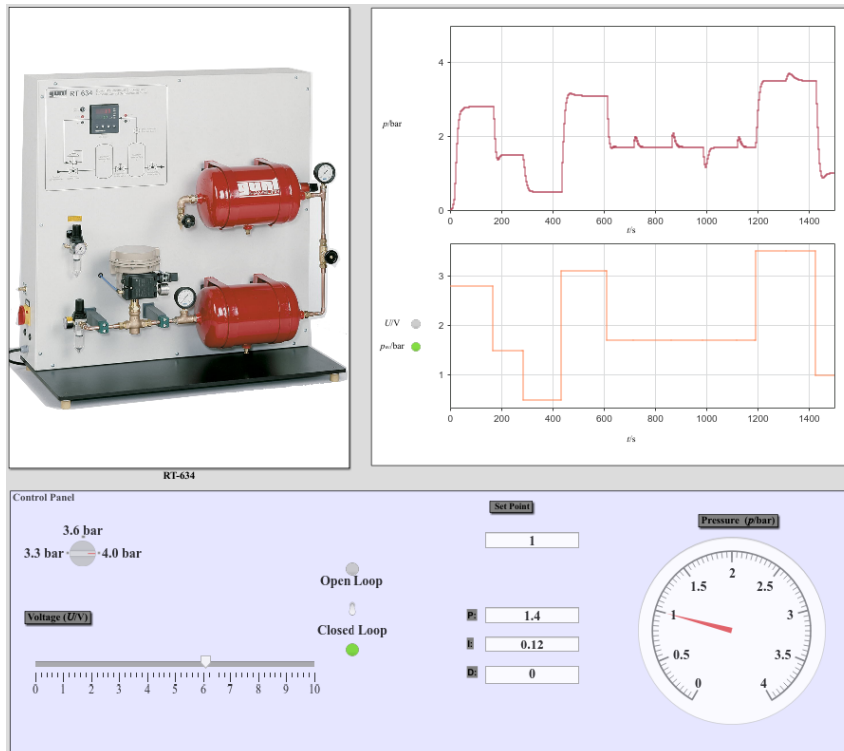


Figure 5: Virtual pressure control demonstration unit – graphical user interface.

4. Conclusions

A dynamic model of the RT-634 unit was developed based on experimental input-output data using MATLAB's System Identification Toolbox. The Hammerstein-Wiener model

with piecewise linear input nonlinearity and polynomial output nonlinearity provided the best fit among the tested models. Fit to Working Data was greater than 95 % indicating that the model captures the experimental behavior of the unit. The RT-634 was virtualized in Simulink in a form of graphical user interface.

The virtual unit is designed as to enable students with limited or even without previous knowledge of control theory and modelling of dynamic systems to study and analyze system dynamics and to observe the effects of P, PI and PID feedback control mechanisms in the context of servo and process control problems in real time.

Although, practical training with real demonstration unit and control hardware, even if computer-controlled, allows for more realistic and memorable experiences, the virtual counterpart is a valuable alternative. One of the main advantages is its continuous availability. Finally, future work will focus on virtualization of flow, level, pH, and temperature control demonstration units, so that the students will have virtual laboratory at their disposal.

Acknowledgements

The authors would like to acknowledge the Slovenian Research Agency (Research Program P2-0032 and Project J7-1816) for financial support.

References

- A. Zandler, H. Greiner, 2020, The effect of two instructional methods on learning outcome in chemistry education: The experiment method and computer simulation, *Education for Chemical Engineers*, 30, 9–19.
- E. P. Byrne. 2006, The Role of Specialization in the Chemical Engineering Curriculum, *Education for Chemical Engineers*, 1(1), 3–15;
- GUNT GmbH, Available online: <https://gunt.de/en/>, (Accessed on : 5. November, 2020).
- MATLAB, 2019, Version: 9.6.0.1072779 (R2019a), Natick, Massachusetts: The MathWorks Inc.
- OpenModelica, Available online: <https://www.openmodelica.org/>, (Accessed on : 5. November, 2020).
- PiDISTILL*, Available online: <https://www.picontrolsolutions.com/>, (Accessed on : 2. February, 2021).
- PITOPS-TFI*, Available online: <https://www.picontrolsolutions.com/>, (Accessed on : 2. February, 2021).
- SIMCET*, Available online: <https://www.picontrolsolutions.com/>, (Accessed on : 2. February, 2021).
- SimulationX, Available online: <https://www.simulationx.com/>, (Accessed on : 5. November, 2020).
- Scilab, Available online: <https://www.scilab.org/>, (Accessed on : 5. November, 2020).
- Z. Novak Pintarič, Z. Krvanja, 2020, The Impact of the COVID-19 Pandemic in 2020 on the Quality of STEM Higher Education, *Chemical Engineering Transactions*, 81, 1315–1320.

Complementarities of part-time and full-time postgraduates in process systems engineering (PSE): when industry and academia merges

Brenno C. Menezes,^{a,*} Robert E. Franzoi,^b Mohammed Yaqot,^a Mohamed E. H. Sawaly,^{a,c} Adnan A. M. A. Al-Banna,^{a,d} Salman Ashkanani,^{a,e} Jeffrey D. Kelly^f

^a*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar Foundation, Qatar*

^b*Department of Chemical Engineering, University of São Paulo, Av. Prof. Lineu Prestes 580, São Paulo 05508-000, Brazil*

^c*Division of Supply Chain, Hamad Medical Corporation, Doha, Qatar*

^d*Division of Supply Chain, Milaha, Doha, Qatar*

^e*Execution Planning and Services, Qatargas Operating Company Ltd., Doha, Qatar*

^f*Industrial Algorithms Ltd., 15 St. Andrews Road, Toronto M1P 4C3, Canada*

bmenezes@hbku.edu.qa

Abstract

Academia is an arena where practitioners from industry are integrated to theoreticians. Such alliance has been intensified by the industry 4.0 (I4) age from which these counterparts are seeking to merge efforts towards society 5.0 (S5), enabling next generations to easily accept novelties and changes in well-established operations, process-of-work, behaviours, etc. In educational centres, such a pace into the I4-S5 state pushes new ways of adopting (or adapting current) sharing of work among peers since this may potentially become a tool for an efficient process-of-research. Thus, we particularly cover postgraduate centres with part-time (PT) and full-time (FT) students in the fields of process system engineering (PSE) and we are widely relying on computer aided process engineering (CAPE) tools, algorithms, software, packages, etc. A collaborative research and development of PSE-CAPE systems may a) involve PT and FT postgraduates in multi-disciplinary fields of science and engineering and b) go across physics, math, and technologies to include social sciences, public policies, and beyond. The proposition is to analyse PT-FT synergies considering their experiences, accessibility of data to validate models, viability to handle CAPE tools, etc. An example of collaboration between PT and FT students, involving a university, a research center, a consulting company, and a medical corporation, is highlighted to optimise healthcare treatment systems for social progress and sustainable development amid COVID-19.

Keywords: Education in PSE, part-time postgraduates, shared working, teamwork.

1. Introduction

The increasing number of part-time students in postgraduate centres worldwide has brought concerns about their evolution as researchers and developers of technology and related impacts in communities. How supervision tasks should adapt to this and which mechanisms must be adopted to address such situations are the points to be answered in this paper. This study particularly focuses on postgraduates from the fields of process system engineering (PSE) or widely counting on computer aided process engineering (CAPE) tools, algorithms, software, packages, etc. In the argumentation, complementarities of part-time (PT) and full-time (FT) postgraduates of PSE/CAPE are analysed from the point of view of the industry (organisations, companies, ministries, etc.) aiming to identify gaps and drawbacks of PT postgraduates that work in parallel with the FT postgraduates. Considering companies' investments in employees, release time and PT's extreme efforts given their job workloads and academic duties, challenges to facilitate such a scenario need to be discussed to improve the synergies in the relationship between industry and academia.

Lacking data on postgraduation numbers in terms of distribution and effectiveness of PT and FT students, an analogy with postsecondary education in the United States is introduced. Data from the Integrated Postsecondary Education Data System (IPEDS) published for the first time in 2017 shows the alarming results on postsecondary part-time students' efficiency, which demonstrates the need to improve the path of their academic endeavour. According to IPEDS (2017), among the postsecondary students, which represents 37% of the PT students, only 17% received a degree or certificate within six years (considering five years as the effective time of graduation as a bachelor). In comparison, around 56% of FT postsecondary students graduated within six years of enrolment. The poll covered about 1.2 million PT students from 2008 to 2017. Several other details on the distribution of PT per sector and data differentiating private and public and profit and non-profit institutes are given in the IPEDS survey.

For several years, sharing of information among students, despite being PT or FT, postsecondary or postgraduates, has been established as a framework to equalise the quality of the knowledge given that bridges are built to allow students with weaknesses to succeed towards standards of learning. Nevertheless, as we sustain in this work, a certain level of intrusive management on behalf of the PT and FT learning success must emerge in supervision activities. It is clearer when PT students are allowed in PSE education centres and it gets stronger when most of the students are PT. By doing a postgraduation in parallel while working in public and private sectors, the available time of PT students to research and to develop a PSE postgraduate study is reduced. It is intensified when CAPE tools are involved in the research and development of a product to be deployed further in industrial, ministerial, etc., sectors that have followed the evolution of their employees and allowed the investment in their released time as PT students in academia (Kelly and Menezes, 2020).

From such study, complementarities of PT and FT students are evaluated by comparing pros and cons in terms of practitioners' and theoreticians' capabilities and efforts to be conjugated or combined into a teamwork process of making research and development for a complete deployment of the postgraduate topic. The coordination of such cooperation among PT and FT students resembles the hierarchical decomposition heuristic (Kelly and Zyngier, 2008) that focuses on solving the overall problem when managing a decentralized and distributed system integration of competing targets and

directions of multiple sub-systems. These must be defined systematically every semester and re-tuned whenever needed. As potential assistants of supervisors, FT students have more time to spend in the postgraduation studies covering theories, configuration and codification of algorithms in mathematical and computing languages, extra paper writing (e.g., review papers), etc. However, challenges exist such as a) tools to allow sharing of work without compromise learning; b) exposition of weaknesses of students who are underperforming; c) balances of number of PT and FT students for teamwork supervision; and d) extra work for advisory coordination and control of effectiveness of the PT learning and FT assistance.

2. Previous literature on PSE/CAPE tools for teamwork applications

Into the applicability of process system engineering in industry to facilitate PT and FT students in a staged fashion within an easy pathway, Menezes and Kelly (2018a) presented open-use and community-based tools for education in PSE with several examples in industrial applications from decision automation (or decision-making or prescriptive analytics) to data analytics (or predictive analytics). Menezes and Kelly (2018b) extended their previous work describing prescriptive and predictive analytics examples found in batch and continuous processes in industry by using the mathematical modeling platform IMPL (Industrial Modeling and Programming Language) that embeds the unit-operation-port-state superstructure (UOPSS) and the quantity-logic-quality phenomena (QLQP) from Kelly (2005). In their work, the so-called industrial modeling frameworks (IMF) are given as an initial phase for the development of PSE applications by using the built-in UOPSS flowsheet network. Within this stage of the learning process, equations are not coded, instead, the problems are configured using the input data in specific meaningful frameworks with respect to keywords or semantics to be learned as in any modeling and programming language. This type of modeling platform is known as structure-based (as opposed to the set-based such as GAMS and scalar-based such as MATLAB), whereby the equations are constructed automatically by the configuration of the problems using the UOPSS constructs and QLQP concepts (Kelly and Menezes, 2019).

Kelly and Menezes (2020) presented a path to evolve towards advanced modeling and programming skills by using a blend of IMPL's IML (Industrial Modeling Language), IPL (Industrial Programming Language), and IMPC (Industrial Modeling and Programming Code). This hybrid modeling allows sharing work between PT and FT students (configuration in IML and codification in IPL and IMPC). In such a way, the PT students can drive the achievements of the project by using the IML configuration considering the vast number of parameters, variables, and constraints and the UOPSS-QLQP elements without coding. For more advanced steps to programmatically model the problems integrated with other systems connecting graphical user interface and database, the use of an application program interface (API) may be accomplished by FT students using IMPL's IPL codifications via computing languages as Python, Julia, R, C, C++, C#, Java, Visual Basic, etc. The development of ad hoc libraries for PSE packages (thermodynamics, kinetics, etc.) can be developed in IMPL's IMPC using Intel-Fortran. IML, IPL, and IMPC can coexist simultaneously in a model in a hybrid PT-FT work. More recently, IMPL-DATA has been released which has always been a part of IMPL to pre- and post-process problem data, but is exposed to support the user in the creation of simpler and smaller analytics, estimation and optimization deployments. Therefore, the PT students can progress from a low-skilled modeler or

mere user of off-the-shelf tools to a high-skilled developer that drives the design and implementation of solutions to be deployed in the postgraduation and their companies.

3. Postgraduates from industry and academia

For those postgraduation centres with part-time students for research and development in process system engineering topics, some particularities must be addressed. In analysing potential synergies and trade-offs or pros and cons in the combination of PT and FT experiences, time spent for researching, accessibility of data to validate models, viability to handle CAPE tools, supervisors learn from such study everywhere this applies.

In this way, we initialise the argumentation with a digression. In such comparative analyses, the uniqueness aspects of industry and academia aim to be merged in favour of the researching, development, and deployment (RD&D) in the engineering management and decision sciences on the process of making decisions to be managed the executions. In the digressive example, both the trial-and-error procedures based on previous experiences and the automated decision-making via exact methods of mathematical programming are found. One would be inclined to say that the latter would be more efficient than the former. This may be true if the automated engine based on decision sciences is representative enough to be accepted as the result to be followed in the execution of the decisions. However, industrial practitioners may act as validators of the theoretical decisions. Joly et al. (2017) discussed lessons learned from developing and implementing refinery production scheduling technologies pointing out the weaknesses and strengths of both development and implementation or deployment part of the whole, when involving researchers and developers in research centres and end-users (the validators) in industrial production sites.

3.1. Part-time (PT) students from industry, organisations, companies, ministries, etc.

Campbell and Bombardieri (2017) address postsecondary learning and fulfilment steps of PT and FT students. In the postsecondary environment, the majority of PT are older and financially support themselves and their families. Approximately 75% of them are currently working and 40% of them work 40 hours per week. Obviously, this situation is even worse in postgraduation since the students are within a further stage of their career than in the postsecondary stage. As an extrapolation, the complicated state for PT postgraduates of PSE centres must be considered by the advisers in the coordination of their challenges to succeed.

Despite the barriers of PT students, their strengths related to previous experiences on the topics of the postgraduation's researching and development may allow them to act as validators of the theories and proofs from the academic environment. Furthermore, the PT student's access to data must be considered as a powerful attribute, however such data mostly needs to be sanitized to hide confidential information. These PT qualities are shown in the example of the healthcare treatment system to be seen.

3.2. Full-time (FT) students: the additional arms of PT supervision

Postgraduate FT students need some essential core skills if they are going to assist their supervisors in advising PT students. Some core skills are required for a better supervision, such as leadership, time management, technical expertise, communication

skills, and ability to judge. One of the pillars of growth in both business and academia is to have good leadership and supervision skills over team members. Therefore, improving the FT student expertise in leadership and supervision is required to interact with other peers, which also supports the faculty supervision process through the enrolment period. The core skills are following. First, being a good leader is critical for FT, it is the first step towards managing tasks delegated by faculty supervisors and collaborate with team members (FT or PT students) as they would rely on their leader for guidance and mentorship. Second, time management skill is considered the seed to thriving in any task or activity as it is implemented in every workplace and for any position. Third, strong technical skills are required when a peer asks for technical support and knowledge that FT should be able to deliver as close as their supervisor. Fourth, communication skills are needed since the FT should be prepared as it helps to professionally transfer the message or tasks to others. Fifth, judgmental skills are required in certain situations to overcome unexpected outcomes, although it will improve based on previous advising experiences.

4. Example of a merge of industry and academia

The problem in Figure 1 is addressed by a collaboration between PT and FT students within a coalition of a university, a research center, a consulting company, and a medical corporation to develop predictive and prescriptive analytics based on machine learning and optimisation for the optimal healthcare systems in the planning, scheduling, and coordination of the treatment networks amid the COVID-19 pandemic.

This study aims to better predict the spread of the disease in order to prepare for the avalanche of the number of patients in the healthcare treatment systems. In measuring or simulating the selected inputs (social distance, lockdown policies, out-of-stock of personal protective equipment (PPE), climate, etc.), the spread of the new number of cases per day (epidemiology curve) can be modeled through a data analytics machine learning approach. In a second step, using epidemiology curve predictions, an optimal design, operation, and control of the healthcare treatment systems can be determined. Then, the sensitivity analysis for an assessment on social progress and sustainable economy can be performed. In the first and second steps of the project, FT students with machine learning and optimisation skills develop the configuration, codification, interfaces, etc., needed to predict the epidemiology curves of COVID-19 positive patients in the near future and this will be an input of the optimisation and control of the healthcare treatment systems. The data, modeling and results of such complex problems are validated by the PT students that work in the medical corporation.

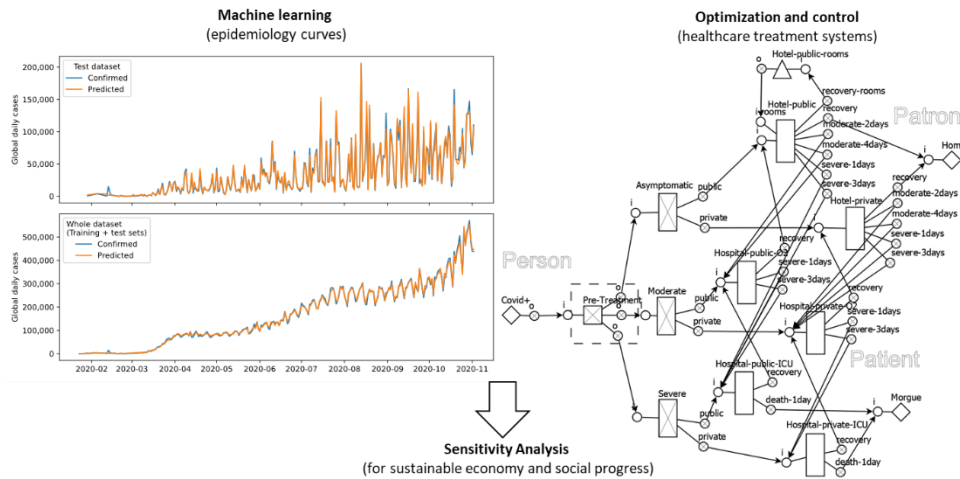


Figure 1. Advanced analytics for healthcare systems: merge of industry and academia.

For a medical resource supply chain resilience against the COVID-19 pandemic, a generalised performance model can be developed considering minimisation of deviations from targets, standards, policies, etc., to be gathered by the PT student. A target can be defined as the safety level of PPE stocks in the location for use and in the warehouse for distribution. However, one of the biggest challenges is to determine the forthcoming number of patients, considering models of the contagious curve and how this can be used in a performance model in agreement with a) standards (soft bounds or targets); b) limits in the treatment capacities; and c) correlated processes by using lower and upper bounds of variables. Under certain circumstances, the use of penalties can be considered to avoid infeasible problems (out of stock on the PPE as an example).

5. Conclusions

The complementarities of part-time and full-time postgraduates in process system engineering centres are a topic to be addressed by policymakers, institutional leaders, and researchers to improve part-time student efficiency. Education in PSE/CAPE tools highlights the importance of merging industry and academia to validate case studies, find synergies, coordinate the cooperation of postgraduates, etc. It also participates in building a future society that balances economic development, addresses social issues and incorporate technology to solve sustainably real-world problems. The example of optimisation and control of healthcare systems illustrates how industry, represented by the medical corporation, must be in consonance with the knowledge provided by academia. The case was developed based on a synergy between a part-time student exposed to the needed tools from academia and real data gathered from the industry.

References

- C. Campbell, M. Bombardieri, 2017, Accessed in 23/11/2020 at <https://www.americanprogress.org/issues/education-postsecondary/news/2017/10/18/440997>.
 IPEDS, 2017, Accessed in 23/11/2020 at <https://nces.ed.gov/ipeds/use-the-data>.
 M. Joly, M.Y. Miyake, 2017, Lessons learned from developing and implementing refinery production scheduling technologies, *Frontiers of Engineering Management*, 4(3), 325-337.

- J.D. Kelly, 2005, The Unit-Operation-Stock Superstructure (UOSS) and the Quantity-Logic-Quality Paradigm (QLQP) for production scheduling in the process industries, In Multidisciplinary International Scheduling Conference Proceedings: New York, United States, 327-333.
- J.D. Kelly, D. Zyngier, 2008, Hierarchical decomposition heuristic for scheduling: coordinated reasoning for decentralized and distributed decision-making problems, *Computer and Chemical Engineering*, 80, 140-154.
- J.D. Kelly, B.C. Menezes, 2019, Industrial Modeling and Programming Language (IMPL) for off- and on-line optimization and estimation applications, In: Fathi M., Khakifirooz M., Pardalos P. (eds) *Optimization in Large Scale Problems*, Springer Optimization and Its Applications, 152, 75-96.
- J.D. Kelly, B.C. Menezes, 2020, Industrial software for computer aided process engineering (CAPE) modeling and programming skills development, *Computer Aided Process Engineering*, 48, 2035-2040.
- B.C. Menezes, J.D Kelly, 2018a, Education in process systems engineering for prescriptive and predictive analytics in industry, *Blucher Chemical Engineering Proceedings*, 1, 4767.
- B.C. Menezes, J.D Kelly, 2018b, Open-use and community-based tools for education in process system engineering: industrial applications from decision automation to data analytics, *Computer Aided Process Engineering*, 43, 1705-1706.

Building Knowledge Capacity for Quantum Computing in Engineering Education

Deborah Carberry,^{a,*} Amirhossein Nourbakhsh,^d Jay Karon,^b Mark N. Jones,^{a,c} Mojgan Jadidi,^d Kyarash Shahriari,^d Christian Beenfeldt,^e Martin Peter Andersson,^a Seyed Soheil Mansouri,^{a,*}

^a*Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søtofts Plads, Building 228A, 2800 Kgs. Lyngby, Denmark*

^b*Intelligent Data Analytics, 613 - 7191 Yonge Street, Thornhill, ON, L3T 0C4, Canada*

^c*Molecular Quantum Solutions ApS, Krusågade 19, 3.th., 1719 Copenhagen V, Denmark*

^d*Lassonde School of Engineering, York University, 4700 Keele Street, Toronto ON M3J 1P3, Canada*

^e*Knowledge Hub Zealand, Biotech City Kalundborg, 4400 Kalundborg, Denmark*
debca@kt.dtu.dk, seso@kt.dtu.dk

Abstract

The Quantum industry is currently at an embryonic stage. If it is to grow, it will require new markets, and, a workforce with the requisite skills and knowledge to support it. Anticipating this potential growth, this paper will explore capacity building within engineering education on the subject of Quantum Computing (QC). This work has two aims. On the one hand, it seeks to illustrate the need for developing education on the subject, inferred by trends in open literature. On the other hand, it seeks to suggest a starting point for quantum computing education in higher education. Since 2018, a sharp incline can be observed in the number of publications on topics related to QC. These publications are arising within several fields related to engineering including, but not limited to, material science, chemistry and computer science. In response to this trend, this paper will evaluate several third party educational approaches to teaching emergent technologies with a view to developing a model for teaching QC. Due to a lack of precedent in a wide range of industry applications and the current limitations in the state-of-the-art of this technology, the educational model proposed will be one that exploits imagination, as opposed to knowledge acquisition, in the pursuit of new knowledge building.

Keywords: Education, Emergent Technologies, Quantum Computing, Engineering.

1. Introduction

This paper highlights the current progress of a project for which the objective is to build capacity within engineering education at the Department of Chemical and Biochemical Engineering at the Technical University of Denmark and at the Lassonde School of Engineering at York University in Canada. The goal is that when Quantum Computing technology is ready for widespread adoption, we will be ready to support its growth.

1.1. Quantum Computing

Quantum Computing (QC) can be described as an emergent technology. Whilst its scientific principles (e.g. superposition and entanglement) are well understood, it currently possesses several limitations in relation to its useful application in the real

world. Despite these limitations, several early use cases have been shown to be competitive compared to classical systems whilst many areas have been identified as having great potential for quantum advantage in the future. For example, in the field of Process Systems Engineering (PSE), Ajagekar and You (2020) published experimental results where a proposed QC-based fault diagnosis model outperformed both state-of-the-art data-driven approaches and deep neural network based models in almost all cases. Activities such as process optimization and process data analytics (Pistikopoulos et al., 2021), and, areas such as drug discovery (Hassanzadeh, 2020), and, energy systems and supply chains (Ajagekar, You, 2019), are projected to benefit from QC.

1.2. Quantum Computing Education

Despite the potential for the future application of quantum computing in engineering, no unified curriculum exists to bring these approaches to engineering education across its various disciplines. This may be because teaching and learning an emergent technology presents many unique challenges. More precisely and unlike an established technology, it may not be possible to present a full body of science. Further, adequate tools and techniques may not be developed, and real-world examples will be sparse. Day and Shoemaker (2000) proposed several unique characteristics that can be ascribed to emergent technologies that result in significant challenges in the sphere of teaching and learning. These characteristics include ‘uncertainty, an evolving understanding of functional benefits, formative regulatory standards, speculative use patterns, incomplete market knowledge, embryonic industry structure, new market players and emergent and fast changing rules of the game’. Quantum computing can be qualified as possessing many of these characteristics and therefore, we will approach the challenge of designing education on this subject in the broader context of designing education for an emergent technology. Whilst adopting this broader approach will make it easier to conceptualise the problem, it does not make it easy to solve. Due to the breadth of characteristics that apply to emerging technologies, and the variation that may be inherent within each, a bespoke design is often required.

2. The Growing Interest in Quantum Computing

To explore the growing investment in quantum computing within academia, a quantitative analysis was undertaken with a view to identifying trends across time. ‘Count of publications’ was selected as the quantifying parameter whilst a list of keywords served as the qualifying parameter. The keywords were as follows: (i) Quantum Computing, (ii) Quantum Computation and (iii) Quantum Computer.

The keywords were applied to journal ‘Titles’ in the first instance and journal ‘Topics’ in the second, resulting in the production of a list of journal entries for each search respectively. ‘Title counts’ were intended to illustrate the growth in focus on the subject of quantum computing whilst ‘Topic counts’ were intended to illustrate a more general growth in interest. The results of the analysis are illustrated in Figure 1.

As Figure 1 demonstrates, the annual count of ‘Titles’ containing a keyword was reasonably uniform between 2001 and 2017. A significant upturn occurred in 2018, which continued into 2019 where an increase in excess of 100 % is illustrated in the number of ‘Titles’ containing a keyword compared to just two years previous.

Turning to the count of ‘Topics’ containing a keyword, a more gradual increase in interest can be observed in quantum computing over the same period. In agreement with the trend in ‘Title’ counts, ‘Topic’ counts also illustrate a sharp upturn in growth in both 2018 and 2019 offering further support that a transition is on the horizon. With a view to

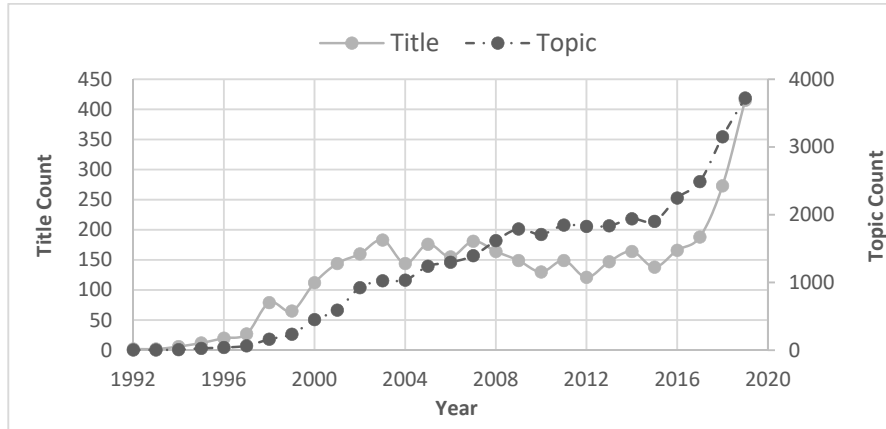


Figure 1: Count of ‘Titles’ and ‘Topics’ containing the search keywords

developing a more granular overview of the growing interest in QC, the topics were then broken down by ‘field’, as illustrated in Figure 2.

Figure 2 illustrates that physics and mathematics are the fields that are producing the greatest output in terms of the number of publications. Relatively speaking, mechanics, computing, optics, and chemistry are also making a respectable number of contributions, whilst material science is close behind. Whilst the trend line for mechanics, and to a lesser extent optics and chemistry, appears to be well correlated with that of physics and mathematics, the computing trend line is not. Between 2009 and 2015, computing appears to have experienced a trough in interest for reasons we are unable to identify here.

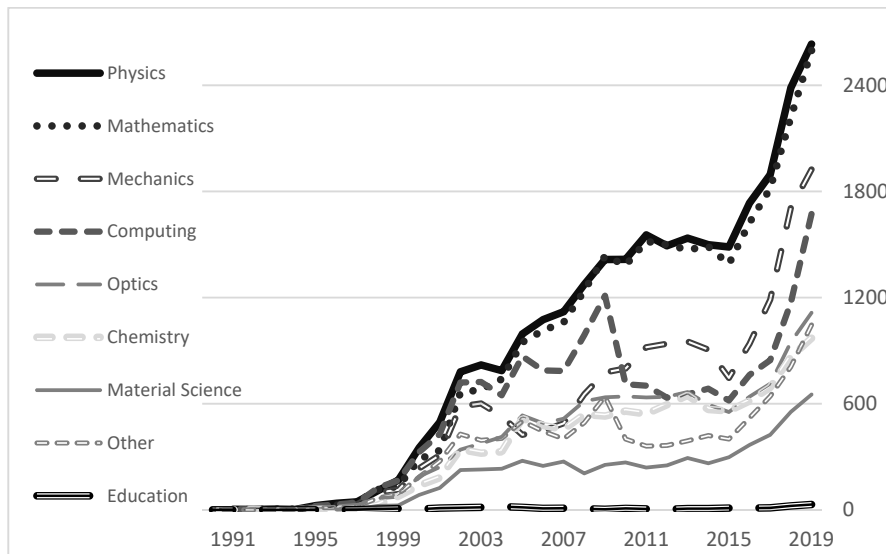


Figure 2: Count of ‘Topics’ containing the search keywords, broken down by Field

However, what is apparent is that several fields are experiencing accelerated growth in interest in the area of quantum computing, and with this, one can expect an emerging demand for a growing talent pool with appropriate knowledge, skills and creativity. And, whilst the education field has also experienced a relative increase in the number of publications on quantum computing over the last couple of years (at 22 and 30 publications for 2018 and 2019 respectively) we might have reasons to be concerned that the level of interest expressed here is simply not enough to match demand elsewhere. Validation for this concern is supported by a recent Black and White report commissioned by D-Wave where 59% of the survey respondents who are not using or considering using quantum computing in the near future are not doing so because of skills issues (451 Research, 2020).

3. State-of-the-Art: Teaching Emerging Technologies

Consulting literature on the topic of Education for emergent technologies, several methodologies have been proposed and/or implemented. Three methods will be outlined below.

Nilsen & Puroo (2005) outline how a trio of pedagogical traditions, objectivist, constructivist and social-cultural could be mapped to a trio of learning objectives, namely the learning of tools, the understanding of principles and the application of these tools and principles in problem solving respectively. Employing a case study research methodology, they were able to identify several challenges to teaching emergent technologies within their own field (Information Systems) and hence, proposed some recommendations to overcome same. These recommendations included the use of flexibility in structuring courses, the use of practical elements in content building, and, a need to engineer roles that can evolve over time, thus permitting the use of multiple pedagogies in the delivery of a single course.

The Pritzker School of Molecular Engineering (PME) at the University of Chicago (IL, USA) created an elective titled 'Introduction to Emerging Technologies' with a view to attracting new students to diversify their student body. In their recent paper, Fowler, et al., propose how systems thinking can be integrated into their current course by adding a number of systems thinking techniques to a series of case studies (Fowler et al., 2019). The subject of the case studies range from tissue engineering to quantum information technologies whilst the proposed systems thinking techniques, many of which can equally be applied to a range of other pedagogies, include think-pair-share, relationship identification between actors within a system, and, questioning (carefully selected in advance). Fowler, et al., correctly identify that systems thinking has applications beyond the sphere of technology in assessing and addressing problems in the wider societal landscape.

A few years earlier, Bishop (2009) opted to use a socio-technological development model for the design of an elective course involving the study and analysis of emerging technologies for the systems engineering department at the United States Naval Academy. The foundation of the course is to address three fundamental questions: 1) What is possible? 2) What is achievable? 3) What is valuable? A number of models including the McKenzie certainty trough and the Disruptive Innovation model are woven into a series of case studies that allow the course participants to expand their skills in recognising *the factors that affect and effect technological change* (Bishop, 2009).



Figure 3: Visual Illustration of Bishop's three fundamental questions

4. Analysis

Nilsen & Purao's (2005) course attempted to address the full range of knowledge and skills required to be competent in a technology. The year was 2004 and their emergent technology was XML. XML and Quantum Computing are not comparable technologies, hence, the challenges they encountered were quite different from the challenges we are encountering with quantum computing today. Using their trio of pedagogies and corresponding learning outcomes, it was the principles and the potential use cases of XML where the ambiguity was most prominent, and hence most challenging for education. Whilst Quantum Computing also suffers from ambiguity in the region of potential use cases, it differs significantly from XML in that it is the lack of tools, and not principles, that pose the greater challenge. To cut through the ambiguity residing with XML technologies in 2004, explanations, guidelines and standards were required. To cut through the ambiguity residing in quantum computing today, a technological leap is required.

Working independently, both Fowler et al. (2019) and Bishop (2009) approached the topic of emerging technologies from a non-technical angle. Unlike Nilsen & Purao (2005), they focused their programs on the socio-cultural aspect of the technology, relieving themselves of the need to employ multiple pedagogies within a single course. Whilst it is essential to understand something about the scientific principles and state-of-the-art that underpin, or occasion, the technologies that they present, there is no requirement for the participants to be expertly versed or skilled in them.

At the same time, there are a few departures between these two socio-technological approaches. As stated above, Fowler, et al., are proposing a systems thinking methodology whilst Bishop is leveraging a system of probing. Systems thinking champions joined up thinking, which is a valuable way to assess the viability of a new technology, however, seeing beyond what already exists is not a core function of this pedagogical style. Put another way, it may be that systems thinking is better at telling us what is likely or unlikely to succeed as opposed to assisting us in developing new ideas. Alternatively, Bishop's more open-ended approach of questioning draws us closer to the human cognitive capacity for imagination.

5. Proposed Model

The educational model that this paper proposes to develop is one that separates the teaching of the core technology from that of exploring its socio-cultural applications. Tailoring a single course for engineering students where the objective of the course is to design new use cases, participants will be encouraged to investigate how quantum computing can be harnessed where other methods are currently utilized across industry. Excluding the more practical implementations of quantum computing, and according to McKenzie's certainty trough, we anticipate that the participants of the course will be more optimistic about the possibilities that the technology presents, given their distance from the technology itself. This course will also lean towards Bishop's system of probing.

6. Conclusion

Designing education for emergent technologies presents unique challenges as knowledge, skills and use cases pertaining to the technologies are not well established and subject to change. Quantum computing is one such technology and whilst this paper does not predict when this technology will become established, our quantitative analysis of published papers on the subject indicates that we are drawing closer. To build capacity within engineering education we must confront these educational challenges. Otherwise, we may not be ready to support the market for quantum computing as the technology becomes more frequent and accessible. We have proposed to implement a learning design that will exercise the cognitive muscle for imagination as well as joined up thinking. Ultimately, we hope that our pedagogical approach will permit more freedom of the imagination, thus widening the circle of what is deemed possible within several engineering industries, and, in quantum computing more generally.

References

- A. Ajagekar, F. You, 2020, Quantum computing assisted deep learning for fault detection and diagnosis in industrial process systems, *Computers & Chemical Engineering*, Vol 143
- A. Ajagekar, F. You, 2019, Quantum computing for energy systems optimization: Challenges and opportunities, *Energy*, Vol 179, Pages 76-89
- B. Bishop, 2009, Teaching Emerging Technologies using a socio-technological development model, *American Society for Engineering Education*
- 451 Research, 2020, Enterprises Are Preparing to Quantum Leap Today, Black & White, D-Wave. Accessed November 23, 2020.
https://www.dwavesys.com/sites/default/files/10962_Advisory_BW_D-wave%20R2.pdf
- B. Bishop, 2009, Teaching Emerging Technologies using a socio-technological development model, *American Society for Engineering Education*
- E. N. Pistikopoulos, A. Barbosa-Povoa, J. H. Lee, R. Misener, A. Mitsos, G. V. Reklaitis, V. Venkatasubramanian, F. You, R. Gani, 2021, Process systems engineering – The generation next?, *Computers & Chemical Engineering*, Vol. 147, DOI: 10.1016/j.compchemeng.2021.107252
- G. Day and P. Shoemaker, 2000, *Whaton on Managing Emerging Technologies*, Wiley, Hoboken, United States
- W. Fowler, J. Ting, S Meng, L. Li, M. Tirrell, 2019, Integrating Systems Thinking into Teaching Emerging Technologies, *Journal of Chemical Education*, 96, pp. 2805-2813
- H. Nilsen, S. Puroo, 2005, Balancing Objectivist and Constructivist Pedagogies for Teaching Emerging Technologies: Evidence from a Scandinavian Case Study, *Journal of Information Systems Education*, 16, 3, pp 281-292
- P. Hassanzadeh, 2020, Towards the quantum-enabled technologies for development of drugs or delivery systems, *Journal of Controlled Release*, Vol 324, Pages 260-279

Teaching PSE Mastery During, and After, the COVID-19 Pandemic

Daniel R. Lewin

Department of Chemical Engineering, Technion. I. I. T., Haifa 32000, Israel

Email for correspondence: dlewin@technion.ac.il

Abstract

The author has been teaching process systems engineering (PSE) courses to undergraduates at the Technion for more than 30 years, evolving his teaching to active learning methods, and in the last five years, to the “flipped class” model. In the spring of 2020, teaching became particularly challenging, since it was taught on-line for the first time, with students having to collaborate remotely with each other also for the detailed design work. This contribution presents the experiences and conclusions resulting from the first COVID-19 semester (spring 2020). At the time of writing, the recommendations are being implemented on two flipped courses, on process design and process control, which are being taught completely on-line in winter 2020 to the same class of 53 students. The presentation will thus summarize the lessons learned over a complete year of practice

Keywords: Process design instruction, process control instruction, project-based learning, active learning, flipped classroom, on-line learning.

1. Introduction – typical instructional objectives for PSE mastery

All of us in the PSE community will agree about the importance of taking a systems approach in chemical engineering design and analysis instruction (Silverstein et al, 2013). Within the framework of PSE, this instruction would include at least courses in the central expertise areas of numerical methods, process control and process design, composed of curricula like those listed below. A helpful way of teaching these materials is by making use of concept maps, which facilitate explaining the connection between the course components. An example of a concept map for a course on numerical methods is presented in Figure 1. Representative learning outcomes for these three key courses are outlined next.

1.1 Numerical Methods: This course ideally instructs the students in the understanding of the basic building blocks of numerical methods, before continuing to providing tools for their practical application. On the completion of such a course, students are expected to select the appropriate numerical method for a given problem, implement it, and interpret the obtained result. Typical course outcomes are as follows:

Building blocks:

- Efficient solution of linear systems
- Finite difference approximations (derivatives, interpolation, integration)
- Efficient solution of nonlinear systems
- Mastery in unconstrained and constrained minimization (Linear Programming)

Applications:

- Linear and nonlinear regression capabilities
- Efficient solution of ODEs, IVPDEs and BVPs
- Integrated problem-solving capabilities

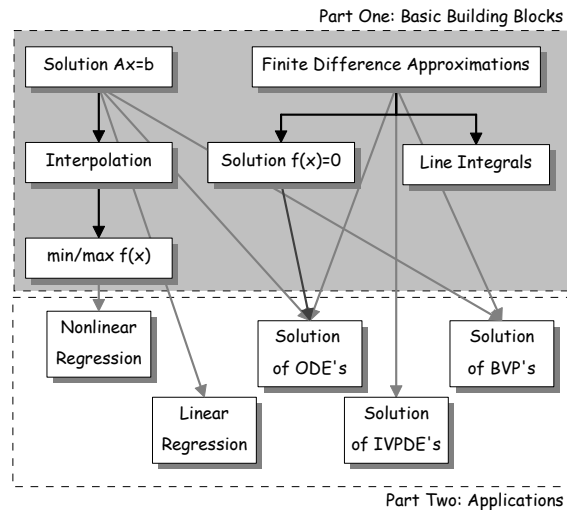


Figure 1. Concept map for a typical course on numerical methods.

1.2 Process Control: This course provides the tools to develop first principles and empirical process models, and then using the derived models, to design simple control systems to meet desired closed-loop performance. Typical course outcomes (Seborg et al, 2004) are as follows:

Process modelling:

- First-principles modelling capability
- Ability to generate state-space and transfer function models
- Block diagram manipulation capability
- Ability to analyse the transient response of linear systems

Process control synthesis:

- Frequency domain analysis capabilities
- Stability analysis capability
- Capability to synthesize control systems to meet response specifications using the root locus method
- Knowing how to tune PID controllers effectively
- Capability to design cascade and feedforward control systems

1.3 Process Design: The capstone design course represents the acid test of a student's ability to apply the engineering tools he/she has acquired, with typical desired outcomes (Seider et al, 2017) being as follows:

- Capability to carry out plant costing and profitability analysis
- Separation sequence synthesis capability for both zeotropic and azeotropic systems
- Capability to perform MER targeting and heat exchanger network synthesis
- Plant-wide control system configuration capability
- Capability to perform a HAZOP and to carry out a HAZAN
- Proven cooperative design project capability, demonstrating both team and individual skills

2. Student-centered vs Teacher-centered teaching

As postulated by Bloom (1968), the degree to which students achieve mastery depends on four conditions: (1) Clear definition of what constitutes mastery; (2) Systematic, well-organized instruction, focused on student needs; (3) Assistance for students when and where they experience difficulties; (4) Provision of sufficient time for students to achieve mastery. Bloom reports the modes of learning that improve outcomes, with the most significant obtained by personal tutoring, which increases the degree of mastery as exhibited by exam grades up to two standard deviations higher than for students taught conventionally by a lecture-based approach.

In a course that is taught in a teacher-centred approach, the contact time between the teacher and the students is mostly utilized for lectures by the teacher, often with modest involvement of the students. In recitations, the assistant will often take the same approach. This means that in a teacher-centred approach, students are largely passive in most of the contact time available, with the students expected to take an active role mostly when tackling homework sets on their own. These deficiencies reduce the degree to which students acquire mastery in higher-level design and evaluation capabilities.

In contrast, in a student-centred approach, the contact time is focused on giving opportunities for students to become involved in class activities, with the teaching staff acting as mentors. Amongst the activities are class quizzes leading to discussions, brainstorming, cooperative problem-solving, and student presentations. By nurturing student involvement, the teacher will be able to better assess the degree of mastery being built up by the students. Student involvement is even more critical in the recitations, where the focus should be on giving students time to work problems for themselves. For students to learn, they need to be given opportunities to make mistakes, understand the reasons for the mistakes, and correct them. This takes time, and the more recitation time taken up by the TA explaining his/her problem-solving strategies, the less time the students will have for their own efforts. Mentoring students' work, should fill most of the recitation time, enabling staff to mentor and assess student capabilities.

This formative assessment can only be ascertained if the teachers and assistants reduce the amount of time that they are lecturing in favour of providing time for active learning by the students (Velegol et al, 2015). One of the best ways to make this happen is to transition to flipped classes, which move a large part of the information transfer to on-line, pre-recorded lectures, which the students need to complete as their homework in advance of class activities. The *flipped classroom*, involves the following sequence of activities, repeated in every week of each course:

- a. *On-line Materials* – Produced by converting lectures to pre-prepared, on-line lessons composed of 5-15 min video clips interspersed by on-line quizzes. Students are expected to cover these materials on their own as homework in advance of each week of activity and are given course credit for it. **Benefits:** Students learn the basic materials covered in each week at their own pace, and their learning is reinforced by addressing the on-line activities as they follow the materials. The on-line activities can be tailored to achieve specific objectives in each stage of the course. These can be: (a) Regular quizzes: Quiz questions posed as multiple-choice, matching, or numerical computations; (b) “Your turn” extended calculations and small-scale designs: A problem for the student to tackle independently is defined at the end of a video clip, which is followed by a movie in which a possible solution to the problem solved is presented, which students can compare to their solutions; (c) Preparing for brainstorming: A video clip can present a problem that requires group effort to address, for which students are requested to collect information, write down ideas,

and bring their results to class for groups discussion. Note that all these activities increase the students' stake in their learning and will prepare them to make better use of the next resource – the Class Meeting.

- b. *Class Meetings* – Moving from *teacher-cantered lecturing* to *student-cantered meetings in the classroom*. A typical class meeting combines quizzes, class discussions and open-ended problem solving, with the focus being to keep the students active. **Benefits:** Giving students the opportunity to prepare ahead increases their effective participation in class and impacts positively on the degree to which they learn and master the application of what they have learned. The specific benefits of each type of activity that could be utilized are as follows: (a) Quizzes for comprehension: These could be clicker questions, to test comprehension of concepts learned at home, or to reinforce previous, related materials. The lecturer can check the level of understanding exhibited by all of the students, instantly; (b) Quizzes to generate discussion: When the questions raised may have more than one solution, it pays dividends to use them to generate class discussion. Learning from incorrect answers is often more valuable than focusing only on correct ones; (c) Open-ended problem solving: This is one of the main reasons for having class meetings. The focus should be on getting students to participate in the development of solutions. For particularly complex problems, dividing the class in separate workgroups may have benefit.
- c. *Active Tutorials* – For students to master course content, they need to apply themselves to independently work problem sets covering the curriculum. The job of the teaching assistant in this setting is to be the enabler for student efforts rather than a demonstrator of solutions. **Benefits:** In active tutorials, students working in teams solve the classwork (previously referred to as homework) in class time. This ensures that: (a) All students who participate in the sessions are actively involved in working problems; (b) Assistance can be provided by staff and from students, helping each other; (c) Students, assistants and the lecturer all receive feedback in a timely fashion (in real time).

3. On-line challenges and how to address them

The spring of 2020, with the resulting COVID-19 lockdowns, introduced additional challenges to effective teaching. Several problems surfaced, associated with a need for social distancing and on-line lessons. Here is an itemized list of problems together with the ways that have been found helpful in overcoming them:

- a. Undesirable on-line behaviour of students, such as students turning off cameras and microphones or passive and/or low student attendance. **Fixes:** (a) Request that students turn on cameras with microphones on mute, turning on microphones to participate. A bright and positive attitude by the lecturer will go far in securing cooperation of the students. (b) What worked outstandingly well was to invite all of the students to an on-line “BYOB Party” before the start of classes, to get to know them and to use the informal meeting as a chance to share expectations. After that, the ice was broken and most of the students were cooperative in the on-line Zoom sessions. Attendance was high (usually over 80% of the students), with many students participating in class discussion.
- b. Undesirable on-line behaviour of teachers, such as the teacher talking most of the class time, or teachers demonstrating solutions of problems, with little involvement of students, or allowing a few students to dominate the in-class discussions. **Fixes:** (a) Pause in presentation to give students a chance to ask questions. Respond to the

- questions and check that the response fully-addresses them; (b) In-class problem solving should involve the students. Do not provide full solutions up-front but get students to contribute suggestions and partial solution steps by brainstorming with student involvement; (c) Use on-line quizzes to promote class discussion, with all students participating. Use polling software to involve the whole class in this, and use the class answers, especially the wrong ones, to generate discussion in the class.
- c. Too many students (typically 15-25%) not preparing for the synchronous meetings by studying the on-line lessons in advance. **Fix: You cannot afford to lose 15-25% of the class!** Not taking steps to bring these non-performers back into the fold can mean a large proportion of underperformers who do not even pass a course. Efforts need to be made to track the non-collaborators, reaching out to them from Week 1 of the course, and bring them back in, and it is surprising how easy this is to do. Many of them will take kindly to your outreach, especially if your communication is positive and focused on how much you care about their success. If the percentage of students truly on-board is maintained high during the entire course, the whole class will benefit, and the outcomes at the end of the course will reflect this.
- Most of these suggested fixes will work in a regular, face-to-face (F2F) setting also.

4. A flipped roadmap for the future

The author has had a long and successful experience with the effective implementation of the flipped classroom to the teaching of both process control and process design, now for seven consecutive years. There is evidence for improved outcomes in process design instruction resulting from the implementation of active methods (Lewin and Barzilai, 2020a). In the year of the pandemic, and the consequently imposed lockdowns, the flipped classroom was relatively easily adapted to on-line learning (Lewin, 2020b). The experiences gained in the second semester of the pandemic to a relatively large group of students who simultaneously participated in courses in process design and process control have led to a clear conclusion that a correctly implemented flipped paradigm is highly effective. This implementation involves the following eight key components:

1. *Have a game plan.* Balance expectations of the lecturers, teaching assistants and students, as all three stakeholder groups need to be on board. It is recommended that a lecturer with no previous experience in flipping try the paradigm first on a single week of class, selecting the week that is the most challenging to fully-cover using a conventional approach. In addition to preparing the on-line lesson as homework, the class meeting and the active tutorial should be included in this trial.
2. *Preparation of on-line lessons.* Define instructional objectives for each lesson. Divide the lecture into video segments of between 5–15-minute duration, ensuring that the content is complete (e.g., cover all steps in a mathematical development). Write and use a script when recording the video segments and practice the delivery before recording. Audio quality is more critical than video quality.
3. *Preparation of effective quiz questions.* Follow each video segment with a quiz question/cluster/activity to test students' understanding. Write useful explanations of all answers (especially important for the wrong ones) and allow students to retry the questions that they get wrong. This is not a test – it is part of their learning!
4. *Lesson assembly and testing.* Upload questions and videos and generate a Moodle Lesson (or similar). The teacher should test the flow and system response first, and have an assistant perform an independent check.

5. *Require students to complete the lessons before Class Meetings.* Students should be given credit for this crucial preparatory step. Continuously follow up on students who do not do this, starting from the first week of the semester.
6. *Plan for a useful Class Meeting.* Prepare additional materials and do not repeat what the students have already learned on-line. The following is a partial list of activities that have been found to be useful: (a) Short quiz questions – to be used to foster class-discussion; (b) Open-ended exam-style questions to be solved with class participation; (c) Project/design work, executed in “break out rooms”; (d) Short student presentations.
7. *Schedule an Active Tutorial.* Schedule sufficient time as this activity largely replaces what used to be “homework.” Allow time to discuss solution strategies in class. Divide the class into small work groups, using breakout rooms if on-line, or by ensuring appropriate seating arrangements if F2F. Make sure question levels in each week’s problem set span from easy to difficult (exam level), and make solutions available on-line. It is unreasonable to expect students to handle exam-level questions well in the final exam without giving them the opportunity to practice solving similar questions for themselves in the Active Tutorials during the semester.
8. *Follow up on every component.* All three steps of the flipping paradigm are critical to success and all of them can be continuously improved. For the On-line Lesson, were there any problematic video segments, and were there any problematic or particularly useful quiz questions, and should more questions be added? For the Class Meeting, were there enough students active, and how many attended? Were the planned activities suitable? For the Active Tutorial, how many students attended, and how many of them were actively engaged and completed the assignments?

5. Conclusions

Experience with the flipped-class approach indicates that engagement with the materials throughout the semester improved the students’ level of confidence in their mastery of the subjects. These observations could explain the improved performance in the final exams in both the process design and the process control courses since adopting active learning and flipping in both. The encouraging outcomes obtained in both courses suggest that this format can be taught to good effect in more than one course at a time, and equally well both on-line and in F2F teaching. Hopefully, these findings and recommendations will encourage others in the PSE community to move to active learning methods.

References

- Bloom BS (1968). "Learning for Mastery," *UCLA - CSEIP - Evaluation Comment*. 1(2):1-12.
- Lewin, D. R. and A. Barzilai (2020a). "Teaching Process Design to Chemical Engineering Undergraduates – an Evolution," *Chemical Engineering Education*, in press.
- Lewin, D. R. (2020b). "Teaching Plant Design in the Year of COVID-19," submitted to *Chemical Engineering Education*.
- Seborg, D. E., T. F. Edgar and D. A. Mellichamp (2004). *Process Dynamics and Control*, 2nd Ed., John Wiley and Sons, New York.
- Seider, W. D., D. R. Lewin, J. D. Seader, S. Widagdo, R. Gani and K. M. Ng (2017). *Product and Process Design Principles: Synthesis, Analysis and Evaluation*, 4th Ed, John Wiley, NY.
- Silverstein, D. L., L. G. Bullard, W. D. Seider, and M. A. Vigeant (2013). "How We Teach: Capstone Design," presented at 120th ASEE Annual Conference & Exposition, Atlanta, Georgia.
- Velegol, S. B., Zappe, S. E. and Mahoney, E. (2015). "The Evolution of a Flipped Classroom: Evidence-based Recommendations," *Advances in Engineering Education*, 4(3):1-35.

Real-Time Chemical Process Monitoring with UMAP

Zachary Webb^a and José Romagnoli^{a,*}

^a*Louisiana State University, 3307 Patrick F. Taylor Hall, Baton Rouge, LA 70803, USA*
jose@lsu.edu

Abstract

In the modern chemical world, engineers have access to millions of data points at their fingertips. Using this data properly can help us to recognize vast improvements in the way plants are run, especially in process monitoring. In this paper three aspects of process monitoring are considered: visualization, fault identification, and fault diagnosis. For each aspect, the use of modern machine learning techniques for addressing these issues is discussed, and improvements over outdated methods are illustrated. This proposed approach is tested using the Tennessee Eastman Process (TEP) for several types of faults.

Keywords: Process Monitoring, Dimensionality Reduction, Clustering, Fault Detection, Fault Diagnosis

1. Introduction

The chemical industry has always had a need for a system to monitor the process and ensure the quality of chemical product. Such a monitoring system has four goals: fault detection, fault identification, fault diagnosis, and process recovery. In the past, these goals could be achieved to a degree for any individual process with a model developed carefully with specific process knowledge, using tools like univariate error charts and alarms. Methods have been developed to bring about more sophisticated process monitoring that takes advantage of multivariate data information. Common methods like Principal Components Analysis (PCA) and recursive PCA (RPCA) are useful in some cases, but are deficient in dealing with data containing nonlinearities, which is especially problematic in the chemical industry. The other drawback is that a PCA model is unable to trace such normal process drifting, while most of the industrial processes are time-varying due to catalyst deactivation, equipment aging, and tube coking.

Recent breakthroughs in dimension reduction (DR) techniques and topology have popularized the Uniform Manifold Approximation and Projection (UMAP) algorithm for projecting high dimensional data into 2 or 3 dimensions. With UMAP, we can reduce the dimensionality of process data and apply modern clustering algorithms to quickly and effectively categorize it. Once a model has been created using the historical information, incoming data points can be passed through the same UMAP model and a fault detection program (FDP). The FDP uses an adaptive k-Nearest Neighbor (AkNN) method to determine the belongingness of each incoming data point to the clusters of data in the historical model. Data which correspond to known clusters are added to a repository of historical information, and data which do not are classified as “faulty”. We can diagnose

the cause of faulty points using a subspace greedy algorithm (SGS), which determines the variables that contribute to their differences from known clusters. After diagnosis is accomplished, the route to process recovery will be much clearer for the responsible engineer. This proposed method is tested using the Tennessee Eastman Process (TEP) for several types of faults. The proposed method of process monitoring outperforms traditional methods. The online monitoring tool provides fast and useful information to engineers or operators in the case of a fault, leading to speedy recovery of the process.

2. Methods

2.1. Uniform Manifold Approximation and Projection (UMAP)

UMAP is a dimension reduction technique developed by McInnes et al. in 2018. It is a manifold learning technique based on Riemannian geometry and algebraic topology. UMAP works by first discovering the relationships of the input data in the high-dimensional space, using fuzzy simplicial complexes, which are a topological representation of a local neighborhood graph. Once a graph has been constructed, UMAP uses stochastic gradient descent optimization to find a low-dimensional representation that is most similar to the high-dimensional input. UMAP improves upon the previous state-of-the-art DR algorithm, t-SNE, by better preserving the global structure of the input data in the low-dimensional space. For low-dimensional t-SNE representations, the within-cluster distances were meaningful for determining similarity of data, but the between-cluster distances were not guaranteed to be important. Because the real-time process monitoring workflow hinges on the separation of faulty data in the low-dimensional space, this improvement is critical.

2.2. K-Nearest Neighbors Fault Detection

The k-Nearest Neighbor (kNN) algorithm is a popular classification method for unknown data. By using the Euclidean distance, we can classify a data point based on information known about its k-nearest neighbors. If we set $k=1$, then each point will be classified just by its single nearest neighbor, and if we set k to a large integer value, it will be classified by a vote of the aggregate classes of its k nearest neighbors. Using the output distances from the kNN, we can determine the likelihood that a single point belongs to a known group of points. In a plant setting, we can use kNN to scan incoming data points in real-time, compare them to a historical database of pre-classified points, and determine if the incoming point characterizes an abnormal event based on a threshold distance.

2.3. Adaptive kNN

In order to run a kNN fault detection program for any amount of time, the trained kNN graph must have the ability to adapt over time. The behavior of a chemical plant is subject to time-varying behavior and process drift. Because the pairwise kNN graph construction is computationally expensive, we can reduce the workload of the computation engine by maintaining a database of only data which are prototypical for each class we are interested in. For a data point x , the number of its neighbors, in data set D laying within the k th threshold distance d_a , denoted as N_x , is defined:

$$N_x = |\{q \in D | \text{dist}(x, q) \leq d_a\}| \quad (1)$$

Thus, we can define a prototype by naïve criteria. A point x is considered a prototype if the N_x is equal to k . This definition is too strict, so we can expand the scope of the prototype by using an updating factor, p , where $p \in (0,1)$. Now we can define the extended prototype:

$$k(1 - p) < N_x < k(1 + p) \quad (2)$$

We can now also define a fault:

$$N_x \leq k(1 + p) \quad (3)$$

Using these definitions, we now have an adaptive kNN algorithm that can determine if incoming data points would be redundant in the model or if they otherwise provide new useful information. We can also define a “valley” point within the sorted k-dist graph:

$$x = \min \{x \in [a, b] | \frac{f(x + \frac{s}{2}) - f(x - \frac{s}{2})}{s} \geq \frac{f(b) - f(0)}{b}\} \quad (4)$$

Where $f(x)$ is the sorted k-dist curve, and the searching step is defined as s where $s > 1$.

2.4. Variable Contribution using Subspace Greedy Search Algorithm

Now that the monitoring framework can use AkNN to detect faulty data, the next stage is to diagnose those faults as soon as they occur. If we can determine which of the real input variables contribute to the differences in a faulty data point, recovering from the abnormal event will be much easier. The goal of fault identification is to find the combination of the most contributing group of variables (subspace) which causes the discriminant between normal and faulty data. A naïve approach which tests all possible subspaces for a difference in disparity between normal and faulty data will fail for two reasons. First, the curse of dimensionality will dull this effect on the score as the data grows in dimension. Second, with a number of possible subspaces given by 2^d , where d is the number of input variables, it is impractical to calculate the score for each subspace due to computational complexity. In order to address these problems, subspace greedy search (SGS) is implemented. The testing score is calculated by comparing the k-distance between the normal and faulty data for each target subspace. Use of a greedy algorithm beginning in low-dimensional subspaces allows the search to terminate in a reasonable number of step and return a list of input variables ranked by their contribution. This addition will allow users to react appropriately to data found to be faulty.

3. Results

3.1. The Tennessee Eastman Process

The Tennessee Eastman Process (TEP) is a simulation developed by Downs and Vogel, which is used as a benchmark for studying process control and monitoring. It consists of

two liquid products along with an unwanted byproduct, which are generated by four gaseous reactants and one inert. The simulation offers 20 process faults among 53 measurements. We studied TEP0 (normal data), TEP2 (step change in the inert composition), and TEP10 (random variation in the feed temperature). For each of these faulty conditions, the goal is to create a good visualization in 2D space using UMAP, identify the faults using AkNN, and to diagnose the root cause of the event using SGS.

3.2. Projection into 2D space

For the step change introduced in TEP2, all three methods of visualization accomplish the goal of separating the TEP0 data from the TEP2 data in 2D space. For this task, UMAP was run with $k=50$ neighbors and $\text{min_dist}=0.5$. These projections are depicted in Figure 1.

For the random variation introduced in TEP10, the three methods give us drastically different results. PCA struggles to separate the TEP0 data from the TEP10 data, so in 2D space, all of the data points fall into one group. The t-SNE results are much better, separating out two groups of TEP10 data from the larger group of TEP0 data at the top left. UMAP is able to separate the two groups of TEP10 and show the separation clearly in 2D space, outperforming the other two DR algorithms. For this task, UMAP was run with $k=10$ neighbors and $\text{min_dist}=0.05$. These projections are depicted in Figure 2.

3.3. Error Testing

The AkNN fault detection system was able to characterize the faulty data at over 99% accuracy for both faults TEP2 and TEP10 compared to the normal data TEP0 (shown in Figures 3 and 4). With faulty data showing a knn-dist measurement above the given threshold, AkNN can identify faults and alert personnel quickly when faults occur.

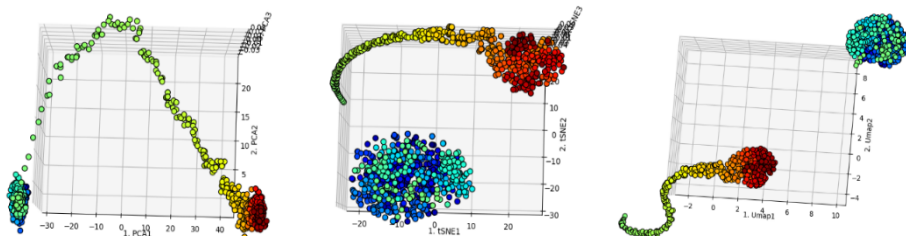


Figure 1: TEP0 (Normal) and TEP2 (step change) visualized with PCA, t-SNE, and UMAP (from left to right)

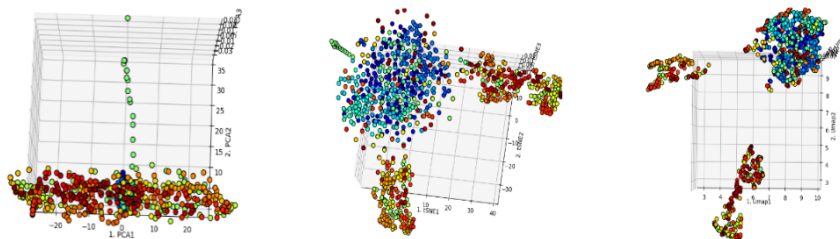


Figure 2: TEP0 (Normal) and TEP10 (random variations) visualized with PCA, t-SNE, and UMAP (from left to right)

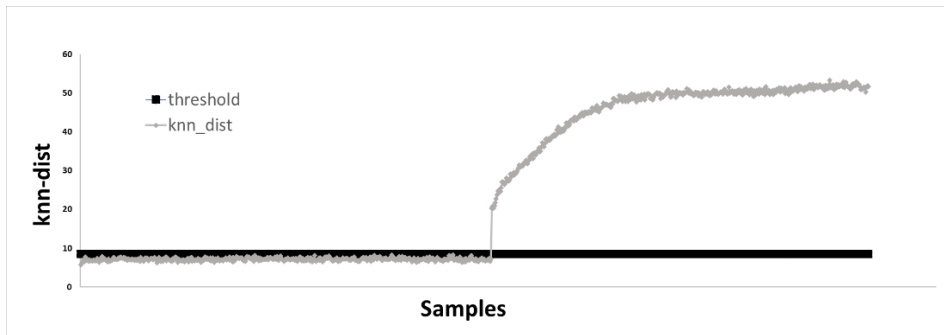


Figure 3: TEP0 (Normal) and TEP2 (step change) error measured by knn-dist

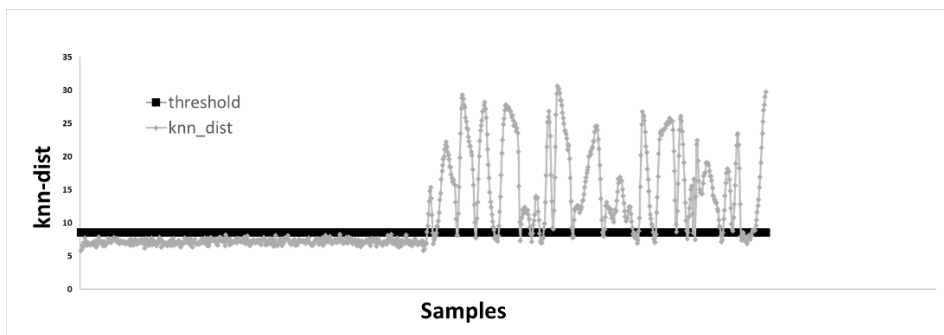


Figure 4: TEP0 (Normal) and TEP10 (random variations) error measured by knn-dist

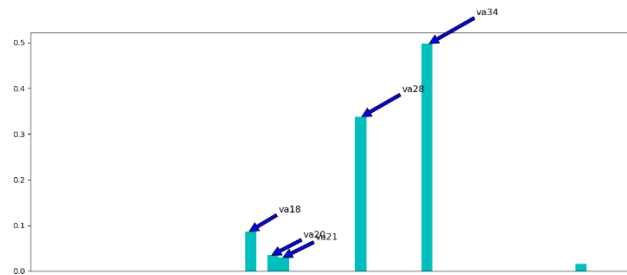


Figure 5: TEP0 (Normal) and TEP2 (step change) contribution chart

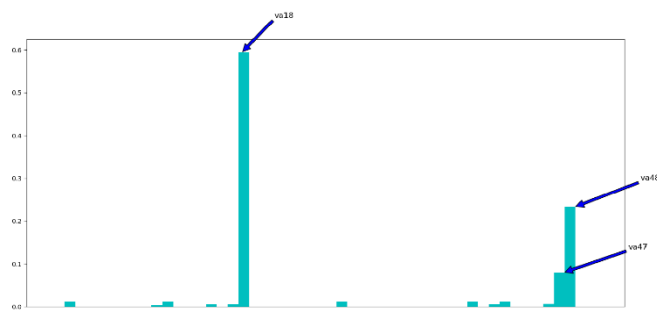


Figure 6: TEP0 (Normal) and TEP10 (random variations) contribution chart

3.4. Variable Contribution

The subspace greedy search algorithm was able to identify the primary source of the fault TEP2 (step change in the inert composition) and TEP10 (random variation in the feed temperature). When compared to the variable contributions given by the explained variance of PCA, this SGS algorithm is clearly preferable. SGS often returns only one or two contributing input variables, which speeds up the recovery process by pointing to the most responsible contributors, while PCA can give a long list of partially contributing variables, leaving personnel with not much assistance.

4. Conclusions

In this paper, we propose three tools: dimension reduction for improved process visualization with UMAP, real-time process monitoring including fault identification based on AkNN, and variable contribution fault diagnosis with SGS to form a full process monitoring environment for detecting and recovering from abnormal events, especially in the chemical processing world. We have shown that these modern tools are competitive with traditional techniques and in many cases improve upon them. An environment which uses all of these methods in conjunction with a user-friendly interface will be the introduction of many in the chemical industry to the world of machine learning and computer science. These advanced methods improve the popularity real-time process monitoring, which would improve the quality of chemical products and better the safety of those nearby.

References

- [1] Zhu, W., Sun, W., & Romagnoli, J. (2018). Adaptive k-nearest-neighbor method for process monitoring. *Industrial & Engineering Chemistry Research*, 57(7), 2574-2586.
- [2] McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- [3] Russell, E. L., Chiang, L. H., & Braatz, R. D. (2012). *Data-driven methods for fault detection and diagnosis in chemical processes*. Springer Science & Business Media.
- [4] Chiang, L. H.; Braatz, R. D.; Russell, E. L. *Fault detection and diagnosis in industrial systems*; 2001; pp 30 – 54.
- [5] Downs, J. J.; Vogel, E. F. A plant-wide industrial process control problem. *Computers & chemical engineering* 1993, 17, 245–255.
- [6] He, Q. P.; Wang, J. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing* 2007, 20, 345–354.
- [7] Mícenková, B.; Ng, R. T.; Dang, X.-H.; Assent, I. Explaining outliers by subspace separability. *2013 IEEE 13th International Conference on Data Mining 2013*, 518–527.

Efficient Energy Integration and Design of Distillation Separation Sequences

José A. Caballero*, Juan Javaloyes

Institute of Chemical Process Engineering, University of Alicante, Carretera de San Vicente s.n. 03690, Alicante, Spain.

caballer@ua.es

Abstract

In this paper we show that for a zeotropic mixture the simultaneous implementation of different heat integration alternatives like direct reboiler-condenser heat exchange, Direct or reverse vapor recompression cycles, multi-effect distillation, partially or totally breaking of thermal couples and substitution of thermal couples by the equivalent only liquid transfer alternative, when possible, considerably increase the thermal efficiency of a sequence of distillation columns and its total cost. We illustrate the procedure with the separation of a four-component mixture. Results show that we can get impressive reductions in total annual cost.

Keywords: Thermally Coupled Distillation, Heat Integration, Process Integration, Vapor Recompression Cycle.

1. Introduction

In the chemical industry engineers have a large number of unit operations to separate components (e.g. adsorption, extraction, membranes, absorption, etc.). However, distillation accounts around 90-95% of all the liquid separations and purifications in industry, and this situation is not likely to change in the near future (Humphrey and Keller, 1997). In spite of the predominance of distillation, it is not uncommon the claim that distillation is an inefficient separation technology. However, Agrawal and Tumbalam Gooty, (2020) have shown that this belief is a misconception likely based on the fact that distillation involves vaporization of one or more components and therefore is more energy intensive than other alternatives that do not require phase change. However, the energy required in distillation is not necessarily proportional to the reboiler duty and distillation can be much more efficient than intuition seems to suggest. Conversely, for many applications, the high efficiency of distillation makes it challenging for alternatives to even achieve the same energy as that of distillation (Agrawal and Tumbalam Gooty, 2020). In this work we will show how the systematic implementation of heat integration alternatives can substantially increase the energy efficiency in distillation.

The characteristics of the search space for the sharp split of M components in N fractions, were established by different authors around 10-15 years ago by Caballero and Grossmann, (2006) and following a different approach by Giridhar and Agrawal, (2010). Shah and Agrawal, (2010) defined a regular configuration as a sequence of distillation columns formed by exactly $N-1$ columns (where N is the number of key components to be separated). If each one of the columns of the sequence has a reboiler and a condenser, then the sequence is a basic sequence. Basic sequences can be generated only by relations

between separation tasks and therefore it is possible to separate the structural considerations associated with the separations from the heat requirements of each distillation sequence. Alternatively, Caballero and Grossmann, (2006) used the concept of «structurally different configurations» to refer to a given sequence with a specific separation task without specifying any particular assignment of heat exchangers. The extension from structural different separations (or Basic) to the regular ones only requires to assign a heat exchanger or a thermal couple to each one of the streams connecting two columns in the sequence. Those authors showed that if we do not consider further integration or intensification, the optimal sequence is a regular sequence.

While there have been important advances in theory and algorithms for the generation of distillation-based separation sequences, and a good number of alternatives have been proposed for improving the energy efficiency (i.e. heat integration, vapor recompression, multi-effect distillation, partially thermally coupled columns, heat pumps, etc.) there are no a systematic framework for combining all those possibilities (and of course the new ones that could eventually appear) to generate efficient distillation sequences. In this work we present a systematic approach for designing separation sequences of zeotropic mixtures that simultaneously consider thermally coupled distillation (TCD) sequences, together with different alternatives for heat integration.

2. Algorithm for the efficient energy integration in thermally coupled distillation.

A good number of alternatives have been proposed to increase the energy efficiency in distillation. Direct heat integration between condensers and reboilers. Multi-effect distillation. Thermally coupled distillation, VRCs (or RVRC) are also a common approach to save energy when the temperature difference is not too large (typically no more than 30°C) and it is usually implemented between the condenser (or the rectifying section) and the reboiler (or the stripping section) of a column. Navarro-Amorós et al. (2013) showed that in thermally coupled distillation is possible to take advantage of the inherent inefficiency of TCD systems, by withdrawing the excess vapor (liquid) inside certain column sections and using it in a VRC (RVRC). The benefit is twofold; on one hand, the optimal operating conditions of the TCD system are recovered and therefore the diameter of some column sections and the utilities consumption are reduced. On the other, the VR or RVR allow further reductions in utilities consumption. Other alternatives, like partially thermally coupling; or the only liquid transfer alternative (Jiang and Agrawal, 2019; Skiborowski, 2020), are structural modifications that can be used to improve the heat integration.

The simultaneous design of the distillation sequence and all previous heat integration alternatives (and others that could eventually appear) is likely to produce very large and complex models that could be, at least, very hard to solve. Instead, we propose a sequential approach:

1. Select the best sequence of distillation columns in the space of regular configurations. In this work we use the disjunctive approach proposed by Caballero and Grossmann, (2006). This model, that relies on the Underwood-Fenske-Gilliland shortcut model, was validated using the commercial process simulator Aspen-Hysys. The next steps are developed using rigorous simulations.
2. For the selected configuration, identify:

- 2.1. Inefficiencies in the connection points between separation tasks that belong to the same column. The excess of vapor/liquid can be removed by heat exchanger. Those heat exchangers use utilities at intermediate temperatures, which in some cases allows the use of cheaper utilities. Besides, the internal flows decrease in some of the decoupled sections and therefore also the column diameter.
If the unbalance of flows in the connection points is large then consider to implement a VRC or RVRC to further reduce the energy costs.
- 2.2. Identify the possibility of heat integration between reboilers and condensers. Consider also the possibility of changing the pressure of some columns, breaking totally or partially some thermal couples if needed to get direct reboiler-condenser heat integration or indirect through a VRC/RVRC cycles.
- 2.3. For difficult separation task, consider the possibility of a multi-effect column integration. If that separation task is thermally coupled with other separation task, there are different alternatives: a) consider the possibility of partially (or totally) breaking the thermal couple introducing a new heat exchanger. b) Use the 'liquid transfer only' proposed by Agrawal. This last approach requires adding a new column section and a heat exchanger. With this approach, at least one of the columns that form the multi-effect arrangement can be maintained thermally coupled with the rest.
3. Repeat step 2 until no further improvement be possible.
4. Consider the possibility of intensification. This last point is not included in this work, but a review of different alternatives can be found in (Jiang and Agrawal, 2019)

3. Example

We illustrate the algorithm by designing the separation of a equimolar mixture of propane (A), isobutane (B), n-butane (C) and pentane (D). The objective is to obtain each component with a purity of at least 0.99 mol fraction. The best distillation sequence is shown in Figure 1a. The result, including the sizing of the columns was validated using ASPEN-HYSYS. Table 1 shows the data for the example.

Table 1. Data for the example.

Components	Composition (mol fraction)		
A: Propane	0.25	Feed Flow	200 kmol/h
B: i-Butane	0.25	Nominal Pressure	1000 kPa
C: n-Butane	0.25		
D: Pentane	0.25	Thermodynamics:	Peng Robinson
Cold Utilities	Cost (\$/kW·y)	Hot Utilities	Cost (\$/kW·y)
Refrigerated water (5-15 °C)	127.6	LP Steam (5 bar 160 °C)	404.67
Cold Water (20 – 35°C)	10.2	HP Steam (10 bar 180 °C)	427.14
Electricity 0.06 \$/kWh equivalent to 480\$/kW·y based on 8000 h/y of operation			
Annualization factor = 0.162 (interest = 10% in 10 years)			
Cost estimation based on correlations by (Turton et al., 2013)			

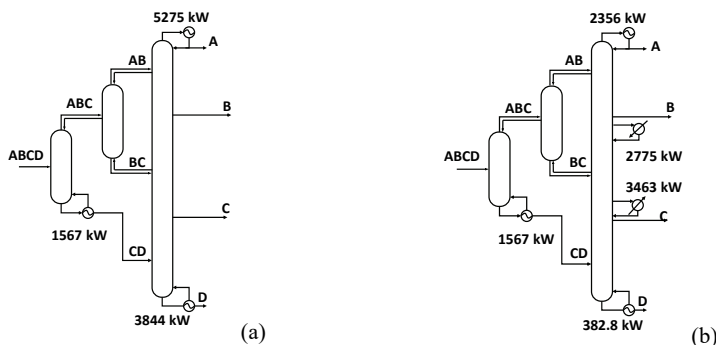


Figure 1. a) Optimal thermally coupled initial configuration, b) configuration after adding intermediate heat exchangers.

The temperature difference between the component B (i-C4) and C(n-C4) at the operational pressure is around 13-14 °C, so the integration of a VRC could eventually be of interest (Figure 2.a). On the other side, it would be possible to increase the pressure in the first column to try a heat integration with the reboiler C (Figure 2.b). However, to that end we need to remove the thermal couple in the top of the first column and substitute it by a condenser. Of course, this change is not for free, (remember that the optimal solution does not include this heat exchanger): it affects the rest of the system, adds a new heat exchanger, and the increase in pressure increases the hot utilities. With only the VRC we can reduce the TAC up to 3255 k\$/y, and with only the direct heat integration up to 3144 k\$/y, around a 29 and 31% compared to the base case. In both cases, there is still an important duty that is not satisfied in reboiler C. But with the simultaneous heat integration and the VRC we completely remove the duties in the reboiler C (See Figure 2.c), and the TAC is reduced in a 36.5% (2924.4 k\$/y).

Going back to the initial configuration it is possible to implement a different strategy for heat integration. Due to the fact that separation of butanes (B/C) is much more difficult than the rest, we can consider the possibility of a multi-effect integration. To that end, it is necessary to brake, at least partially, the thermal couple associated to the stream 'BC'. There are at least two possibilities: one is to add an extra column section and a new reboiler to convert the thermal couple BC in an equivalent 'only liquid transfer stream.'

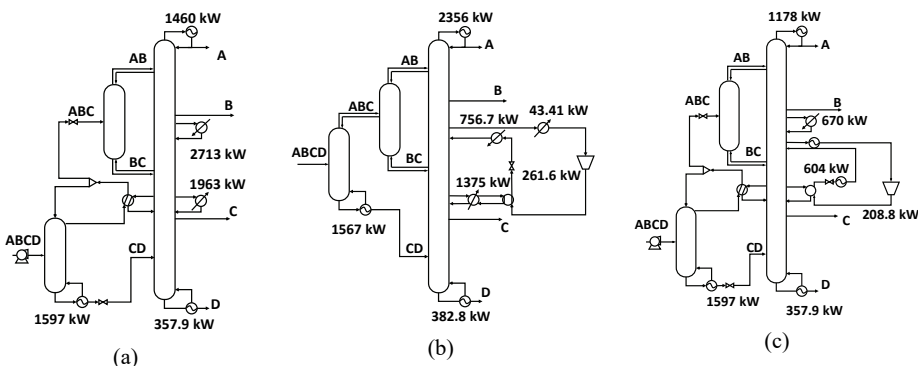


Figure 2. a) Direct heat integration between condenser of the first column and reboiler in C, b) Vapor recompression cycle between states B and C, c) Direct heat integration + VRC.

Another consists of adding a reboiler in stream BC, but at the same time partially maintain the thermal couple BC. With the second approach we avoid adding an extra column section, and therefore, it is the selected approach. The flows in the thermal couple and the liquid stream that form the new set of streams BC are adjusted to allow the heat integration between the two columns that form the thermal couple. Note that in this case, we can reduce the pressure of one of the columns that separate B from C to around 550 kPa, and maintain the thermal couple in the other separator B/C, and at the same time the integration of this last column B/C with C/D in the same column. In this way, we have halved the reboiler duty associated to stream C, but at the prize of adding two small reboilers (one in the stream BC and another in stream B) and a condenser (in stream B of the low-pressure B/C separator). Even though with this configuration the TAC is 3152.6 k\$/y a 31.6 % lower than the base case -Figure 3.1-.

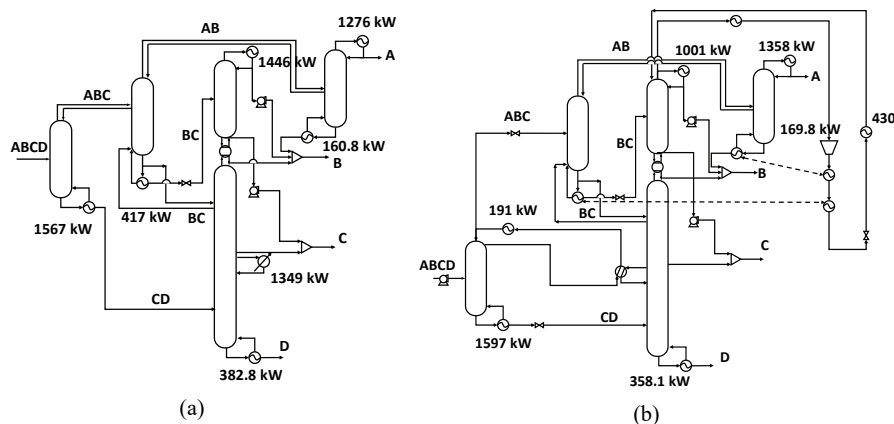


Figure 3. a) Partial thermal couple in state BC, and multi-effect integration by dividing the separation B/C in two columns at different pressure. b) Same that in (a) adding direct heat integration with first column and a VRC.

Following the same reasoning that in first configuration, we can break the thermal couple in ABC, and increase the pressure in the first column to integrate the condenser of this column with the reboiler associated to stream C. Adding this direct heat integration the total cost reduces up to 2349.6 k\$/y a 49% lower than the base case. Finally, the temperature in the condenser of the low-pressure B/C separator is around 44°C and the temperatures in the reboilers of separators A/B and AB/BC are around 66 and 72 °C respectively. Then, a VRC from the upper of low-pressure B/C separation can provide duty enough to satisfy those two heat exchangers. The cost of this last configuration is reduced to 2163 k\$/y a 53% lower than the base case -Figure 3.b.

4. Conclusions

In this paper we have shown that starting from the best thermally coupled sequence of distillation columns obtained from the set of regular configurations it is possible to increase the efficiency of distillation columns by sequentially and iteratively adding different alternatives of heat integration. While no one of those alternatives is new, the

simultaneous consideration of all/some of them can produce an impressive increase in energy efficiency that is reflected in considerable reduction in total costs.

We have illustrated the procedure with the separation of a four-component mixture that includes a difficult separation. The identification of points of inefficiency and the systematic inclusion of intermediate heat exchangers, vapor recompression cycles, breaking totally or partially some thermal couples to implement direct heat integration of multi-effect heat integration have produced two configurations one with a reduction of 36% in TAC and the second one over a 50%

5. Acknowledgements

The authors acknowledge financial support to the Spanish «Ministerio de Economía, Industria y Competitividad» under project CTQ2016-77968-C3-2-P (AEI/FEDER, UE) and to the «Generalitat Valenciana» under project PROMETEO/2020/064.

6. References.

1. Agrawal, R., Tumbalam Gooty, R., 2020. Misconceptions about efficiency and maturity of distillation. *AIChE J.* 66, 2–4. <https://doi.org/10.1002/aic.16294>
2. Caballero, J.A., Grossmann, I.E., 2006. Structural Considerations and Modeling in the Synthesis of Heat-Integrated–Thermally Coupled Distillation Sequences. *Ind. Eng. Chem. Res.* 45, 8454–8474. <https://doi.org/10.1021/ie060030w>
3. Giridhar, A., Agrawal, R., 2010. Synthesis of distillation configurations: I. Characteristics of a good search space. *Comput. Chem. Eng.* 34, 73.
4. Humphrey, J., Keller, G., 1997. *Separation Process Technology*. McGraw-Hill Education.
5. Jiang, Z., Agrawal, R., 2019. Process intensification in multicomponent distillation: A review of recent advancements. *Chem. Eng. Res. Des.* 147, 122–145. <https://doi.org/10.1016/J.CHERD.2019.04.023>
6. Navaro-Amorós, M.A., Ruiz-Femenia, R., Caballero, J.A., 2013. A new technique for recovering energy in thermally coupled distillation using vapor recompression cycles. *Aiche J.* 59, 3767–3781.
7. Shah, V.H., Agrawal, R., 2010. A matrix method for multicomponent distillation sequences. *AIChE J.* 56, 1759–1775. <https://doi.org/10.1002/aic.12118>
8. Skiborowski, M., 2020. Energy Efficient Distillation by Combination of Thermal Coupling and Heat Integration, in: Pierucci, S., Manenti, F., Bozzano, G., Manca, D. (Eds.), *Proceedings of the 30 European Symposium on Computer Aided Process Engineering (ESCAPE30)*. Elsevier B.V., pp. 991–996. <https://doi.org/10.1016/B978-0-12-823377-1.50146-4>
9. Turton, R., Bailei, R.C., Whiting, W.B., Shaeiwitz, J. A., & Bhattacharyya, D., 2013. *Analysis, Synthesis and Design of Chemical Processes.*, 4th editio. ed. Pearson Education, Inc., Upper Saddle River, NJ. USA.

Enhancing operability during early stage of process synthesis considering flexibility and inherent safety simultaneously

Andreja Nemet*, Klavdija Zirngast, Zdravko Kravanja, Zorka Novak Pintarič

Faculty of Chemistry and Chemical Engineering, Smetanova ulica 17, 2000 Maribor Slovenia

andreja.nemet@um.si

Abstract

Obtaining optimal solutions in process synthesis has great potential to improve the future profitability and sustainability of the process. Nowadays, most processes are either over- or undersized, resulting in suboptimal designs or designs with low operability performance. The latter means processes that in most cases do not meet production requirements or do not operate optimally under varying process parameters. There are several aspects of operability, all of which should be considered to achieve process designs that operate successfully in the real world. Moreover, the operability analysis should be included in the early stages of the process design. Once the process design is selected, there are limited opportunities for improvement and in most cases no substantial enhancement can be achieved. In this study, process synthesis is upgraded to include an operability view, that simultaneously considers flexibility and inherent safety to achieve economically optimal, flexible and safe process designs. The results of the case study show that an increase in flexibility of 3417 % and reduction in risk of 43 % can be achieved with a negligible loss in economic viability of 0.85 %.

Keywords: MINLP process synthesis, operability, flexibility, inherent safety, simultaneous approach

1. Introduction

Heat Integration remains to be a core task in the process synthesis due to its high potential for energy and emission reduction. The methods for obtaining Heat Exchanger Network (HEN) are getting increasingly developed resulting in highly integrated networks. Although at nominal data these network designs seem to represent an optimal solution, the operability of such networks might be highly uncertain. One should be aware that the more the systems are integrated the lower is the degree of freedom, which narrows the operating window (Tian et al., 2018). Therefore, when obtaining HEN designs the operability should not be neglected, since HENs are inevitably highly integrated in process systems, connecting different parts of processes. The traditional hierarchical approach considers the operability of the energy systems at the third stage of the development, after synthesizing network and obtaining utility system. This approach has its disadvantages as the system configuration is already selected, which can lead to “topological trap”, when the operability of the energy system cannot be substantially

improved (Andiappan et al., 2017). In previous work Escobar et al. (2013) considered the flexibility and controllability of HEN design using two-step approach.

In this work two aspects of operability have been considered, flexibility and inherent safety. Flexibility is enhanced, when heat exchangers are optimally oversized, allowing larger number of degrees of freedom (Novak Pintarič and Kravanja, 2004). On the other hand, decreased risk (increased safety) is obtained when smaller and more compact heat exchangers are used (Nemet et al., 2017). Considering both, flexibility and inherent safety simultaneously, leads to trade-off solutions among different aspects of operability. Objective of this study is to simultaneously achieve flexible and inherently safer HEN designs.

2. Methodology

A multi-scenario mixed-integer nonlinear programming (MINLP) model was developed to simultaneously consider the inherent safety and flexibility in process synthesis. The initial HEN synthesis model by Yee and Grossmann (1990) was extended by i) considering different types of heat exchangers (Soršak and Kravanja, 2004), ii) calculations of the inherent safety index (Nemet et al., 2017), and iii) strategies for achieving flexible HEN under varying input parameters (Zirngast et al., 2019).

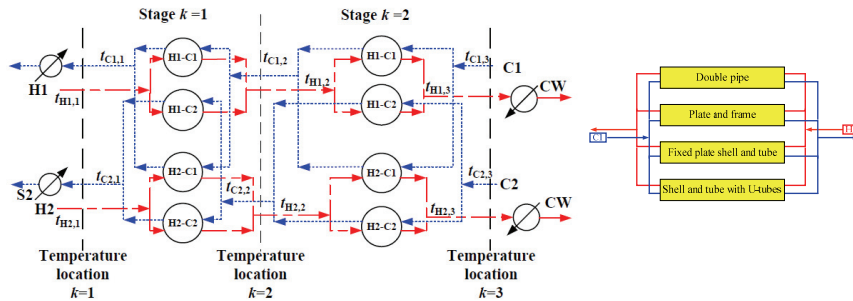


Figure 1: HEN superstructure.

Both sequential and simultaneous approaches were performed. With the sequential approach, a Pareto solutions of flexible HEN designs at various risk limits were obtained. The economic objective was the expected net present value (NPV) of the savings NPV^{savings} . Savings were defined as the difference between the solution where no Heat Integration was considered, $NPV^{\text{no_HI}}$, and the case with Heat Integration, NPV^{HI} (Eq. 1). The NPV consisted of investment (I), discounted annual cash flow (F_c) terms considering the discount rate r^{disc} and the lifetime t_{LT} of HEN.

$$NPV^{\text{savings}} = NPV^{\text{no_HI}} - NPV^{\text{HI}} = -\Delta I + F_c \cdot \frac{(1 + r^{\text{disc}})^{t_{\text{LT}}} - 1}{r^{\text{disc}} \cdot (1 + r^{\text{disc}})^{t_{\text{LT}}}} \quad (1)$$

Eq. 2 presents the discounted annual cash flow (F_c), which takes into account the tax rate (t), revenue (R), expenses (E) and depreciation (D).

$$F_c = (1 - t) \cdot (R - E) + t \cdot D \quad (2)$$

In the simultaneous approach, the safety/risk was also considered in the objective function. The flexible HEN with no safety consideration was taken as a reference case.

The objective function Z was defined as a difference between NPV of savings normalized with NPV obtained at flexible design $NPV_{flexible}^{savings}$ and risk index RI normalized with risk index determined at flexible design $RI_{flexible}$ (Eq.3). The aim was to maximize Z . In this way, the economic performance tends to be as high as possible and the risk as low as possible.

$$Z = \frac{NPV_{flexible}^{savings}}{NPV_{flexible}^{savings}} - \frac{RI}{RI_{flexible}} \quad (3)$$

The RI (Eq. 4) is determined for each substance on the hot hp and cold cp side (process) separately, for matches in each temperature stage k , for each type of heat exchanger hx . The risk index depends on the failure rate of the heat exchanger f_{hx}^{fail} , density of the substance in heat exchanger ρ , heat exchanger area $A_{hp,cp,k,hx}$, and area density of the heat exchanger β_{hx} . Additionally, the risk index depends on risk factors accounting for i) process installation versus storage f_{1hx} , ii) indoor/outdoor positioning f_{2hx} and and iii) process conditions f_{hp}/f_{cp}^3 divided by limiting value G . The limiting value G for toxic substances depends on lethal concentration LC50 (rat, inhalation, 1 h) at 298 K and phase of the substance as presented in Nemet et al (2017).

$$RI_{hp,cp,k,hx} = \sum_{hx \in HX} \left(f_{hx}^{fail} \cdot \frac{\rho \cdot A_{hp,cp,k,hx}}{\beta_{hx}} \cdot \frac{f_{hx}^1 \cdot f_{hx}^2 \cdot f_{hp}^3}{G_{hp}^h} \right) + \sum_{hx \in HX} \left(f_{hx}^{fail} \cdot \frac{\rho \cdot A_{hp,cp,k,hx}}{\beta_{hx}} \cdot \frac{f_{hx}^1 \cdot f_{hx}^2 \cdot f_{hp}^3}{G_{cp}^c} \right) \quad (4)$$

$\forall hp \in HP, cp \in CP, hx \in HX, k \in K$

The overall risk RI^{tot} is determined (Eq. 5) by summation of individual risk $RI_{hp,cp,k,hx}$ over substances in hot and cold process streams and all types of heat exchangers in each stage.

$$RI^{tot} = \sum_{hp} \sum_{cp} \sum_k \sum_{hx} RI_{hp,cp,k,hx} \quad (5)$$

3. Case study

The synthesis of a HEN was carried out as an illustrative example to test the operability (Table 1). The uncertain parameters were the heat capacity flowrates (CF) of the streams H1 and C1 as presented in Table 1. In this example, flexibility is understood as the minimum oversizing of the HEN required to cope with disturbances in the heat capacity flowrates. This type of variations corresponds to exogenous uncertainty as it does not depend on the decisions made. A lifetime of 15 years, a discount rate of 7 % and a tax rate of 20 % were assumed.

Table 1: Input data for process streams and utilities

Stream	T_s /K	T_t /K	CF /kW K ⁻¹	h / kW m ⁻² K ⁻¹	LC ₅₀ (rat,1h, inh) mg ⁻¹
H1	500	400	150±30	0.6	100
H2	450	390	390	1	600
C1	330	390	240±24	0.7	10
C2	380	380	50	1	500
Hot utility	510	510		5	-
Cold utility	300	321		1	-

Figure 2 shows the solution obtained at nominal values. As can be seen, the types of heat exchangers selected were the less compact ones, shell and tube, U-tube, and double pipe.

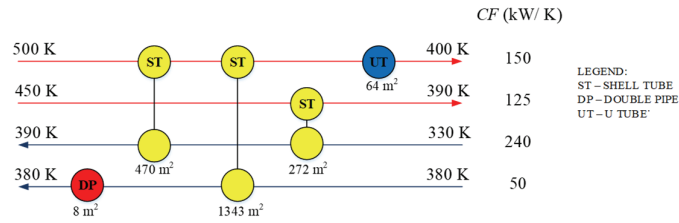


Figure 2: Solution at nominal input values

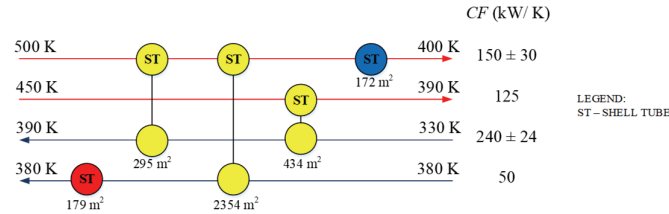


Figure 3: Design considering flexibility

Figure 3 shows the HEN design obtained using the strategy for flexible designs. Note that the total heat exchanger area installed was now higher by almost 60 % 3,434 m², compared to design obtained at nominal input data (3,434 m² vs. 2,157 m²); however, in both cases mostly the less compact shell and tube heat exchangers were selected ($\beta = 720$ m²/m³). Using a sequential approach, a bi-criteria optimization of NPV vs. risk index was performed. The optimization was performed firstly at nominal conditions and secondly by the strategy to obtain flexible designs. Figure 4a presents both Pareto curves for nominal and the flexible cases.

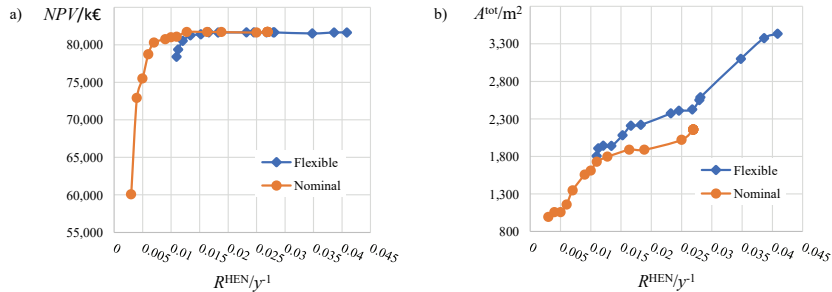


Figure 4: Pareto curve of a) NPV and b) total area versus risk index for nominal and flexible HENs

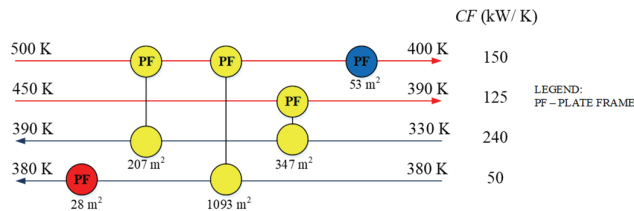


Figure 5: The solution at nominal input values and risk index at 0.011

The NPV curves remain relatively flat until a certain point while still decreasing the risk. In general, flexible designs have a higher risk compared to the solutions obtained at nominal input data. Improving safety in flexible HEN designs is more limited as the lowest risk index achieved was 0.010952, while at nominal values it was 0.003. Figure 4b presents the Pareto curve of total HEN area installed versus the risk index. One can see that flexible design does require more extensive areas. A probable space of optimal solutions would be in the range of risk index 0.01-0.015 and the NPV of savings around 80,000 k€, looking at the Pareto solution obtained for flexible designs. Figure 5 presents the safest HEN design obtained at nominal value having NPV 81,062 k€. All the selected heat exchangers are of compact plate and frame type ($\beta = 1300 \text{ m}^2/\text{m}^3$). More compact heat exchangers result in a lower mass of substances in the heat exchangers, which pose a lower risk. The total installed area was 1729 m². Figure 6 illustrates the safest flexible HEN design. The total area of the safest flexible HEN design was 1,806 m² and the NPV obtained was 78,401 k€. The safety was enhanced at the expense of the economic performance while keeping the flexibility of the HEN design.

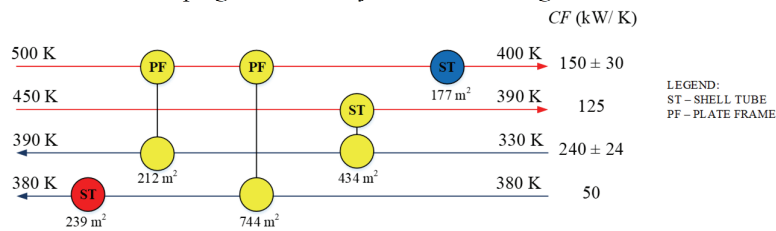


Figure 6: Flexible solution with decreased risk having risk index value 0.0109

Using the sequential approach, the possible space of optimal design could be determined around kink of Pareto curves; however, not in the unique way. For this purpose, the simultaneous approach, considering the economic viability and the risk in the objective function simultaneously was applied.

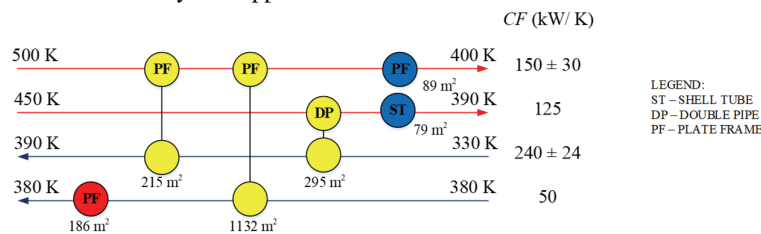


Figure 7: Design obtained with multi-objective optimization considering normalized NPV and risk index

The result obtained with the simultaneous approach was within the optimal region determined with the sequential approach. The objective function of the simultaneous solution was $Z = 0.9548$, NPV 81,013 k€ and the risk index 0.015. Comparison between this solution and the most economically efficient solution reveals that the NPV decreased from 81,642 k€, by 629 k€ (0.7 %) and the risk index decreased from 0.04086 by 0.02586 (63.3 % decrease). The obtained solution design is presented in Figure 7, showing that four of six heat exchangers are the compact plate and frame type. Compared to the most

economically viable flexible HEN design (Figure 3), there is an additional cooler for stream H2. The HEN's total area was 1,996 m², which is a significant decrease from 3,434 m² (Figure 3).

Comparison between the simultaneous solution and the nominal solution leads to a very similar conclusion. The NPV of the solution at nominal input data was 81 713 k€, while for the simultaneous solution, it was 81,013 k€, which is a 700 k€ (0.85 %) decrease. This decrease is negligible when considering the operability performance. The flexibility increased from 0.029 to 1.02 representing 0.991 (3,417 %) increase, while the risk index decreased from 0.0269 to 0.0153 representing a 0.0116 (43.1 %) decrease.

4. Conclusions

A mixed-integer nonlinear programming (MINLP) model for HEN synthesis considering simultaneously two aspects of operability, flexibility and inherent safety, was developed. The results show that enhancing flexibility leads to an increase in the total area of HEN to cope with variations in the uncertain parameters. On the contrary, inherent safety is enhanced by selecting smaller heat exchangers with fewer potentially harmful substances. By considering flexibility and inherent safety simultaneously, appropriate trade-offs between these effects are established, leading to optimal compromise solutions. Sequential and simultaneous approaches were used to obtain results. The sequential approach showed that significant improvement in safety and, hence, risk, can be obtained, while the economics of HEN designs were not significantly reduced. It also revealed that flexible designs are associated with higher risks. The result obtained with the simultaneous approach shows that the NPV decreased by less than 1 % compared to the value obtained at the nominal conditions. At the same time, the flexibility was more than 30 times higher and the safety almost twice as high. Therefore, even with a negligible reduction in economic performance, the operability of the HEN design was enormously improved.

References

1. V. Andiappan, D.K.S. Ng, R. R. Tan, 2017, Design Operability and Retrofit Analysis (DORA) framework for energy systems, *Energy*, 134,1038-1052.
2. M. Escobar, J.O. Trierweiler, I.E. Grossmann, 2013, *Computers and Chemical Engineering*, 55, 158-180.
3. A. Nemet, J. J. Klemeš, I. Moon, Z. Kravanja, 2017, Safety Analysis Embedded in Heat Exchanger Network Synthesis, *Computer & Chemical Engineering*, 107, 357-380.
4. Novak Pintarič Z., Kravanja Z, 2004, A strategy for MINLP synthesis of flexible and operable processes, *Computers and Chemical Engineering*, 28 (6-7), 1105-1119.
5. Y. Tian, S.E. Demirel, M.M.F. Hasan, E.N. Pistikopoulos, 2018, An overview of process system engineering approached for process intensification: State of the art, *Chemical Engineering & Processing: Process Intensification* 133, 160–210.
6. A. Soršak, Z. Kravanja, 2004, MINLP retrofit of heat exchangers networks comprising different exchanger types, *Computer & Chemical Engineering*, 28,1-2,235-251.
7. T. F. Yee, I. E. Grossmann, 1990, Simultaneous optimization models for heat integration – II. Heat exchanger network synthesis, *Computers & Chemical Engineering*, 14, 10, 1165-1184.
8. K. Zirnigast, L. Čuček, Ž. Zore, Z. Kravanja, Z. Novak Pintarič, 2019, Synthesis of flexible supply networks under uncertainty applied to biogas production, *Computers & Chemical Engineering*, 129, 106503, 1-12.