ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes

Gramado, Brazil, April 2-5, 2006

Preprints Volume II

Papers of the 3rd (Tuesday) and 4th (Wednesday) days

EDITORS

Francis J. Doyle III
University of California, Santa Barbara, CA, USA

Jorge O. Trierweiler
Federal University of Rio Grande do Sul, Brazil

Argimiro R. Secchi
Federal University of Rio Grande do Sul, Brazil

ASSISTANT EDITORS

Mehmet Mercangoz
University of California, Santa Barbara, CA, USA

Luciane S. Ferreira
Federal University of Rio Grande do Sul, Brazil

The ADCHEM Organizing Committees
gratefully acknowledges the support of:

Diamond Sponsor

PETROBRAS
www.petrobras.com.br

Gold Sponsors

Braskem
Petroquímica Brasileira de Classe Mundial
www.braskem.com.br

REFAP
Alberto Pasqualini - Refap S/A
www.refapsa.com.br

Silver Sponsors

ARACRUZ
www.aracruz.com.br

chemtech
A Siemens Company
PAVILION TECHNOLOGIES
www.chemtech.com.br

COPESUL
COMPANHIA PETROQUÍMICA DO SUL
www.copesul.com.br

Bronze Sponsors

aspentech
www.aspentech.com

Ipiranga Petroquímica
www.ipq.com.br

Support

UCSB
doyle.chemengr.ucsb.edu

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
www.eng.ufrgs.br/gimscop

Brazilian National Funding Agencies

CAPES
www.capes.gov.br

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico
www.cnpq.br

INTERNATIONAL PROGRAM COMITTEE

Francis J. Doyle III, USA—Chairman

Program Area Chairmen

Optimization and Scheduling
Stratos Pistokopoulos, UK

Batch Process Modeling and Control
Richard Braatz, USA

Process Control Applications
Robert Parker, USA

Model Based Control
Martin Guay, Canada

Modeling and Identification
Mayuresh V. Kothare, USA

Process and Control Monitoring
Dale Seborg, USA

MEMBERS

Allgöwer, Frank (Germany)
Alvarez, Jesus (Mexico)
Araujo, Ofélia (Brazil)
Arkun, Yaman (Turkey)
Bandoni, José A. (Argentina)
Barolo, Max (Italy)
Bequette, Wayne (USA)
Bonvin, Dominique (Switzerland)
Brambilla, Alessandro (Italy)
Christophides, Panagiotis (USA)
Cinar, Ali (USA)
Daoutidis, Prodromos (USA)
De Souza Jr., Mauricio (Brazil)
Edgar, Tom (USA)
Engell, Sebastian (Germany)
Forbes, Fraser (Canada)
Foss, Bjarne (Norway)
Gao, Furong (Hong Kong)
Gatzke, Edward P. (USA)
Georgakis, Christos (USA)
Giudici, Reinaldo (Brazil)
Hahn, Jürgen (USA)
Hasebe, Shinji (Japan)
Henson, Mike (USA)
Hoo, Karlene (USA)
Huang, Biao (Canada)
Jacobsen, Elling (Sweden)
Jorgensen, Sten Bay (Denmark)
King, Rudibert (Germany)
Kravaris, Costas (Greece)
Lee, Jay (USA)
Lee, Peter (Australia)
Lewin, Danny (Israel)
Lima, Enrique (Brazil)
Maciel Filho, Rubens (Brazil)
Marchetti, Jacinto (Argentina)
Marlin, Thomas E. (Canada)
Marquardt, Wolfgang (Germany)
McAvoy, Thomas (USA)
Morari, Manfred (Switzerland)
Moro, Lincoln F. (Brazil)
Nikolaou, Mike (USA)
Ogunnaike, Babatunde A. (USA)
Ohshima, Masahiro (Japan)
Nascimento, Claudio O. (Brazil)
Palazoglu, Ahmet (USA)
Park, Sunwon (Korea)
Perrier, Michel (Canada)
Preisig, Heinz (Norway)
Qin, Joe (USA)
Rawlings, Jim (USA)
Romagnoli , Jose (Australia)
Scali, Claudio (Italy)
Shah, Shirish L. (Canada)
Skogestad, Sigurd (Norway)
Soroush, Masoud (USA)
Swartz, Chris (Canada)
Thornhill, Nina (U.K.)
Ydstie, Erik (USA)
Yoon, En Sup (Korea)
Young, Robert E. (USA)
Yu, Cheng-Ching (Taiwan)

NATIONAL ORGANIZING COMITTEE

Jorge Otávio Trierweiler, Brazil — Chairman

Argimiro R. Secchi (UFRGS)
Darci Odloak (USP)
Luciane S. Ferreira (UFRGS)
Marcelo Farenzena (UFRGS)

ADCHEM 2006

The ADCHEM 2006 (International Symposium on Advanced Control of Chemical Processes) was held in Gramado, Brazil from April 2 to 5, 2006 and was organized under the auspices of International Federation of Automatic Control (IFAC) and Brazilian Society for Automation (SBA). The ADCHEM is a continuing series of international symposia held most recently in Hong Kong, China (2003/2004), in Pisa, Italy (2000), Banff, Canada (1997), Kyoto, Japan (1994), and Toulouse, France (1991). These meetings focus on advances in methods for modeling and control for all types of chemical processes. They are part of a three year rotation of IFAC meetings, which also include the IFAC DYCOPS Symposium Series (recently held in Boston, USA) and the IFAC World Congress.

Promoter Organizations:

http://www.ifac-control.org/

http://www.sba.org.br/

Contents of Volume II – Tuesday and Wednesday Program

Plenary Session 3

Keynotes 5 and 6

Session 4.1 - Biomedical Systems Modeling, Analysis and Control

Session 4.2 - Bioprocess Modeling and Identification

Session 4.3 - Estimation and Adaptive Control

Keynotes 7 and 8

Session 5.1 - Analysis and Control of Separation Processes

Session 5.2 - Modeling of Particulate Systems

Session 5.3 - Process Monitoring

Session 6.1 - Modeling and Identification

Session 6.2 - Optimization and Scheduling

Session 6.3 -Process Monitoring

Keynotes 9 and 10

Session 7.1 -Optimization and Design Applications

Session 7.2 -Control of Complex Systems

Session 7.3 - Process Control

Keynotes 11 and 12

Session 8.1 - Optimization and Control of Petrochemical Systems

Session 8.2 - Practical Applications of Modeling and Identification

Session 8.3 - Performance Assessment of Closed-Loop Systems

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PARAMETER IDENTIFICATION VIA THE ADJOINT METHOD: APPLICATION TO PROTEIN REGULATORY NETWORKS

Robin L. Raffard * **Keith Amonlirdviman** **
Jeffrey D. Axelrod *** **Claire J. Tomlin** ****

* *Department of Aeronautics and Astronautics, Stanford University, CA 94305-4035 USA.*
rraffard@stanford.edu
** *Department of Aeronautics and Astronautics, Stanford University, CA 94305-4035 USA.* amon@stanford.edu
*** *Department of Pathology, Stanford University School of Medicine, Stanford CA 94305-5324 USA.*
jaxelrod@stanford.edu
**** *Department of Aeronautics and Astronautics, Stanford University, CA 94305-4035 USA.* tomlin@stanford.edu

Abstract: An adjoint-based algorithm for performing automatic parameter identification on differential equation based models of biological systems is presented. The algorithm solves an optimization problem, in which the cost reflects the deviation between the observed data and the output of the parameterized mathematical model, and the constraints reflect the governing parameterized equations themselves. Preliminary results of the application of this algorithm to a previously presented mathematical model of planar cell polarity signaling in the wings of *Drosophila melanogaster* are presented.

1. INTRODUCTION

A key problem in systems biology is the identification of parameters in the mathematical models that describe biological systems. This problem is generally difficult due to both the number of state variables and parameters, and the fact that the governing equations are usually nonlinear functions of these states and parameters. It is also advantageous to perform the parameter identification problem relatively quickly, since this allows one to efficiently test the feasibility of different mathematical models.

In this paper, we present an algorithm for performing automatic parameter identification on differential equation based models of biological systems. The algorithm attempts to minimize an objective function which encodes the deviation between the observed data of the system and the output of the parameterized model, with the governing parameterized equations forming the constraints of this optimization problem. The algorithm relies on the adjoint method, which calculates the gradient of the objective function with respect to the unknown parameters, essentially describing analytically how to minimize the objective by varying the parameters. We augment this gradient based method by using additional information provided by the derivative of the gradient to give well-conditioned optimization even when the optimal parameter values are several

[1] This research was supported by DARPA under the BioComp program, by NIH under grant R01 GM075311-01, and by a Stanford Bio-X IIP Award.

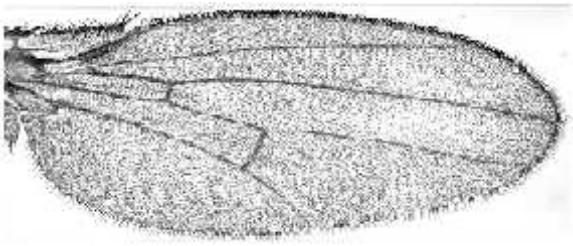


Fig. 1. *Drosophila* adult wing epithelium. Proximal edge is to the left, distal edge is to the right.

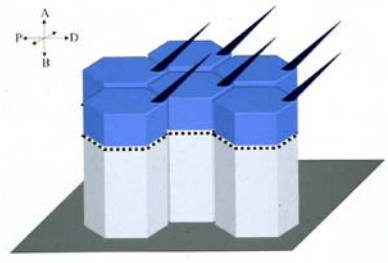


Fig. 2. Diagram shows that each epithielial cell constructs a hair that protrudes from its distal vertex and points distally, creating a virtually error free parallel array.

orders of magnitude different from each other. We present preliminary results of this algorithm on a previously described mathematical model (Amonlirdviman et al. (2005)) of the signaling network regulating the planar cell polarity of *Drosophila* wing epithelial cells orthogonal to their apical-basal axes. This network is termed planar cell polarity (PCP).

2. PLANAR CELL POLARITY (PCP)

In adult *Drosophila*, each epithelial cell on the wing produces a single hair, or trichome. The hairs grow from the distal edge (edge of the cell closest to the wing tip) of each cell and all point in the same direction, towards the wing tip, as shown in Figures 1 and 2 (note that all images in this paper follow the convention that the proximal side of the cell/wing is to the left of the image, distal is to the right). Genetic analyses have identified a group of proteins that are required to correctly polarize these arrays (Adler (2002), Strutt (2002)), and the regular array of hairs is caused by spatially asymmetric distributions of these proteins in the plane of the epithelium. The process by which the proteins controlling hair polarization localize to different areas within each cell during the development of the fly is called planar cell polarity (PCP) signaling. The wing epithelial cells aggregate in a hexagonal close-packed array (Figure 2).

In the presence of cell clones mutant for some PCP genes, the hair polarity in neighboring wild-type cells is disrupted, a phenomenon termed domineering non-autonomy. Domineering non-autonomy reverses hair orientation on either the proximal or distal side of the clone in a manner characteristic to the particular mutant protein. Based on the available biological data, a feedback loop mechanism describing the interaction of a group of PCP molecules was proposed to mediate PCP signaling (Axelrod (2001), and Tree et al. (2002)). The signaling diagram is drawn schematically in Figure 3, in which an arrow indicates a positive influence, and a line indicates a negative influence. The diagram describes the following: Frizzled (Fz), a membrane protein, promotes the localization of Disheveled (Dsh), a cytoplasmic protein, to a membrane; Dsh stabilizes Fz location; Fz promotes the localization of Van Gogh (Vang), a membrane protein, and Prickle (Pk), a cytoplasmic protein, on the membrane of a neighboring cell; Pk and Vang inhibit the recruitment of Dsh to a membrane. Experimentally, it has been observed that, in steady state, Dsh and Fz proteins localize to the distal edge and Pk and Vang to the proximal edge of all cells in the array, thus the large font indicates that the wild type protein localizes at this location. It is believed that the hair grows at the site of the highest concentration of Dsh protein.

A mathematical model based on the feedback loop model (Tree et al. (2002)) and a global directional cue that biases the direction toward which the feedback loop orients (Yang et al. (2002), Ma et al. (2003)) was used to demonstrate, through simulation, the feasibility of the model to reproduce all of the most characteristic PCP phenotypes (Amonlirdviman et al. (2005)). The logic of the feedback loop is encoded in the mathematical model by representing interactions as binding to form protein complexes. For example, the interaction between Fz and Dsh is represented as a reaction forming the complex DshFz, which can interact with other proteins and complexes, and it can undergo a backward reaction that separates it back into its components Fz and Dsh. The mathematical model includes the four original proteins, as well as six complexes, the last four of which form across the cell boundary with the adjacent cell: DshFz, VangPk, FzVang, DshFzVang, FzVangPk, DshFzVangPk. While positive influences are encoded by complex formation, negative influences are through terms that aid the reverse reaction. The state variables of the mathematical model are the local concentrations of these proteins (for example [Fz] represents the concentration of Fz) which are assumed to be continuous. The mathematical model assumes that protein molecules move by diffusion: Dsh and Pk diffuse within the

cell interior, while Vang, Fz, and all the complexes diffuse only in the membrane (or shared membranes).

The mathematical model is represented by ten reaction-diffusion partial differential equations (PDEs). All of the model parameters, including reaction rates, diffusion constants and initial protein concentrations were not directly observable from the available data, so parameter values were identified by being constrained to result in the desired qualitative features of the hair pattern phenotypes. The Nelder-Mead simplex method (Nelder and Mead (1965)) was used to attempt to minimize an objective function composed of quadratic penalty functions corresponding to these feature constraints to produce a feasible solution set of parameters. The model includes 37 parameters, and each evaluation of the objective function required 13 runs of the model simulation corresponding to each of the experimental cases that the model was meant to reproduce. The complete development of the model and results of this analysis are available in Amonlirdviman et al. (2005).

3. THE PARAMETER IDENTIFICATION PROBLEM

In the current work, we strive to replace the simplex method with a more efficient optimization method. As we have described above, the governing equations for PCP consist of ten reaction-diffusion equations describing the time and space evolution of the concentrations of the four PCP proteins and six of their complexes. If $x(t,s) =$ ([Dsh], [Pk], [Fz], [Vang], [DshFz], [VangPk], [FzVang], [DshFzVang], [FzVangPk], [DshFzVangPk]) $\in \mathbb{R}^{10}$ represents the vector of all protein concentrations, and if $s \in \Omega = \mathcal{V} \times \partial\mathcal{V}$ represents all space variables (covering both the cytoplasm $\mathcal{V}$ and the membrane $\partial\mathcal{V}$), the governing equations can be written in the following compact form:

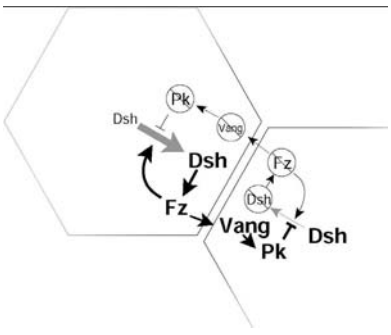


Fig. 3. Four protein PCP signaling network.

$$\frac{\partial x(t,s)}{\partial t} = P(s, x(t,s), \theta) + \mu(\theta)\Delta x(t,s) \ , \quad \forall s \in \Omega \tag{1}$$

which means that the rate of change of each protein concentration is equal to its net rate of production $P(s, x(t,s), \theta)$, plus its rate of diffusion $\mu(\theta)\Delta x(t,s)$.
If protein i reacts with protein j to form complex k, P_i is a function of the type $R_i x_i x_j - \lambda_i x_k$ – it includes more reaction terms if protein i is present in more than one reaction. The forward rates of reaction R and the backward rates of reaction λ are stored in the parameter $\theta \in \mathbb{R}^{37}$, which has to be estimated. Finally, $\mu(\theta)$ is the constant of diffusion of each protein and Δ represents the Laplacian operator.
Eight of the ten proteins diffuse in the membrane. Therefore their reaction-diffusion equations are specified on a periodic domain and do not require boundary conditions. However, Dsh and Pk diffuse in the interior of the cell, which is a finite domain, and therefore require boundary conditions noted in compact form

$$\mu(\theta)\nabla_s x(t,s) \cdot n = CP(s, x(t,s), \theta) \ , \quad \forall s \in \partial\Omega \tag{2}$$

in which n represents the unit normal vector to the membrane. The matrix C is a 10×10 matrix with all zero entries, except the first two diagonal elements (corresponding to Dsh and Pk) which are equal to one. C filters the last eight proteins, for which the diffusion across the membrane is zero. For Dsh and Pk, the rate of diffusion across the membrane is equal to their rate of production.

Experimental data consist of pictures of hair polarity which are provided at final time T, taken to be at the end of the signaling process. In our PCP model, hair polarity is predicted based on the Dsh concentration in the cells and is stored in a vector Y^{model} comprising as many entries as simulated cells, and calculated by

$$Y^{\text{model}} = \int_{\Omega} h(x(s,T), s)\, ds \,, \tag{3}$$

in which h is a differentiable function, which gives a score of 1 to a cell with Dsh localization on the distal side, -1 to a cell with Dsh localization on the proximal side and 0 with no Dsh localization. Similarly the data Y^{obs} is a vector with entries ranging from -1 to 1; -1 for cells with reverse polarity and 1 for cells with polarity. The problem of identifying the unknown parameters is the one of finding, among our parametrized set of PCP models, the model which best explains the experimental data. Therefore, it consists of minimizing the prediction error, *i.e.,* the deviation between the observed

data and the output of the parametrized model. Mathematically, it reads

$$\begin{aligned} \text{minimize} \quad & J(\theta) = ||Y^{\text{model}} - Y^{\text{obs}}|| \\ \text{subject to} \quad & \frac{\partial x(t,s)}{\partial t} = P(s, x(t,s), \theta) \\ & \qquad + \mu(\theta)\Delta x(t,s) \end{aligned} \tag{4}$$

Usually, the norm $||.||$ is chosen as a quadratic norm. Besides its mathematical convenience, such a norm is often chosen because it recovers the maximum likelihood criterion. Indeed, suppose measurements are stochastic data consisting of the sum of the true model outcome and normally distributed noise:
$Y^{\text{obs}} = \int_\Omega h(x^{\text{true}}(s,T), s) + v$, where v is a normal random variable with mean 0 and covariance Σ. The likelihood of the observations is equal to

$$\begin{aligned} &\text{PDF}(v = Y^{\text{obs}} - Y^{\text{model}}) = \frac{1}{(2\pi)^N \det\Sigma} \\ &\exp\big(-\frac{1}{2}(Y^{\text{obs}} - Y^{\text{model}})\Sigma^{-1}(Y^{\text{obs}} - Y^{\text{model}})\big) \end{aligned} \tag{5}$$

where PDF refers to probability density function and is a Gaussian in the present case. The parameter which maximizes the likelihood of the observations is then given by

$$\begin{aligned} \theta^* &= \arg\max\{\exp(||(Y^{\text{obs}} - Y^{\text{model}})||^2_{\Sigma^{-1}})\} \\ &= \arg\min\{||(Y^{\text{obs}} - Y^{\text{model}})||^2_{\Sigma^{-1}}\} \end{aligned} \tag{6}$$

We will assume, in the remainder of the paper, that Σ is the identity matrix.

4. SOLUTION METHOD VIA OPTIMAL CONTROL THEORY

The parameter identification problem consists of an optimization program in which the variables are constrained by a PDE. In this section, we will show how to efficiently solve such a problem.

4.1 Gradient computation

Many optimization algorithms rely on descent methods, which require the computation of the gradient of the objective function. For the case of PDE optimization programs, calculating the gradient can be efficiently done via a version of the adjoint method, which was developed by Jameson (1998) largely for use in nonlinear aerodynamic optimization problems. We will first review the adjoint method.

4.1.1. Adjoint method Let us consider an objective function J given by

$$J(\theta) = f(x, \theta) \tag{7}$$

where f is a differentiable function and x is the solution of a differential equation (DE), noted

$$D(x, \theta) = 0 \tag{8}$$

Under technical conditions (see Lions (1971) for more details), the function J is differentiable and

$$\begin{aligned} \lim_{h\to 0} \frac{J(\theta + h\tilde{\theta}) - J(\theta)}{h} &= \nabla_x f(x,\theta)\tilde{x} \\ &+ \nabla_\theta f(x,\theta)\tilde{\theta} \end{aligned} \tag{9}$$

in which $\tilde{x}$ is the solution of the linearized form of the original differential equation (8)

$$\nabla_x D(x,\theta)\tilde{x} + \nabla_\theta D(x,\theta)\tilde{\theta} = 0 \tag{10}$$

At this stage, computing the derivative in each direction, $\tilde{\theta}$, requires one to solve the DE (10) for each of these directions and then form the derivative according to (9).
The adjoint method allows one to obtain the derivative in all directions – in other words, the gradient, by computing the solutions to only two DEs. It proceeds as follows: taking the inner product with an arbitrary costate q (lying in the same function space as x), we obtain

$$q \cdot \nabla_x D(x,\theta)\tilde{x} + q \cdot \nabla_\theta D(x,\theta)\tilde{\theta} = 0 \tag{11}$$

Adding this term to the derivative, we obtain,

$$\begin{aligned} &\lim_{h\to 0} \frac{J(\theta + h\tilde{\theta}) - J(\theta)}{h} = (\nabla_x f(x,\theta) + \\ &q \cdot \nabla_x D(x,\theta))\tilde{x} + (\nabla_\theta f(x,\theta) + q \cdot \nabla_\theta D(x,\theta))\tilde{\theta} \end{aligned} \tag{12}$$

Choosing q so as to cancel the effect of the state perturbation

$$\nabla_x f(x,\theta) + \nabla_x D(x,\theta) \cdot q = 0 \ , \tag{13}$$

the derivative in any direction $\tilde{\theta}$ is $\big(\nabla_\theta f(x,\theta) + q.\nabla_\theta D(x,\theta)\big)\tilde{\theta}$ and therefore the gradient is

$$\nabla J(\theta) = \nabla_\theta f(x,\theta) + q \cdot \nabla_\theta D(x,\theta) \tag{14}$$

With the gradient in hand, it is now possible to perform a descent algorithm, called the quasi-Newton method, for which we will see an effective illustration in section 4.2.

4.1.2. Adjoint equations for PCP The method presented in the previous section is systematic and can be followed step by step for the PCP model. The regularity of the PCP partial differential equations (PDEs) provides us with enough technical conditions to compute the derivative of J as follows

$$\lim_{h\to 0}\frac{J(\theta+h\tilde{\theta})-J(\theta)}{h} = 2\Big(\int_\Omega h(x(s,T),s)\,ds - Y^{\text{obs}}\Big)^T \int_\Omega \nabla_x h(x(s,T),s)\tilde{x}(s,T)\,ds \;; \tag{15}$$

in which $\tilde{x}$ is the solution of the following linear PDE

$$\frac{\partial\tilde{x}}{\partial t}(t,s) = \nabla_x P(s,x(t,s),\theta)\tilde{x}(t,s) + \mu(\theta)\Delta\tilde{x}(t,s) + \big(\nabla_\theta P(s,x(t,s),\theta) + \Delta x(t,s)\nabla\mu(\theta)\big)\tilde{\theta} \tag{16}$$

With linear boundary conditions

$$\nabla\mu(\theta)\tilde{\theta}\nabla_s x(t,s)\cdot n + \mu(\theta)\nabla_s\tilde{x}(t,s)\cdot n = C\big(\nabla_x P(s,x(t,s),\theta)\tilde{x}(t,s) + \nabla_\theta P(s,x(t,s),\theta)\tilde{\theta}\big) \tag{17}$$

Taking the inner product of this linear PDE with an arbitrary costate q

$$\int_\Omega\int_0^T q^T\frac{\partial\tilde{x}}{\partial t}(t,s) = \int_\Omega\int_0^T q^T\big(\nabla_x P\tilde{x} + \mu(\theta)\Delta\tilde{x}(t,s)\big) + \int_\Omega\int_0^T q^T(\nabla_\theta P + \Delta x\nabla\mu(\theta))\tilde{\theta} \tag{18}$$

Integrating by parts,

$$\begin{aligned}\int_\Omega q^T\tilde{x}(T) = \int_\Omega\int_0^T \tilde{x}^T(\frac{\partial q}{\partial t} + \nabla_x P^T q + \mu(\theta)\Delta q) \\ + \int_\Omega\int_0^T q^T(\nabla_\theta P + \Delta x\nabla\mu(\theta))\tilde{\theta} \\ + \int_{\partial\Omega}\int_0^T \tilde{x}^T(\nabla_x P^T C^T q - \mu(\theta)\nabla_s q\cdot n) \\ + \int_{\partial\Omega} q^T(\nabla_\theta P - \nabla_s x\cdot n\nabla\mu(\theta))\tilde{\theta}\end{aligned} \tag{19}$$

We are now in a position to extract the gradient of J. Provided that q solves the following linear PDE

$$-\frac{\partial q}{\partial t} = \nabla_x P^T q + \mu(\theta)\Delta q \tag{20}$$

with boundary conditions

$$\mu(\theta)\nabla_s q\cdot n = \nabla_x P^T D^T q \tag{21}$$

and terminal condition

$$q(s,T) = 2\Big(\int_\Omega h(x(s,T),s) - Y^{\text{obs}}\Big)^T \nabla_x h(x(s,T),s) \tag{22}$$

the gradient is

$$\nabla J = \int_\Omega\int_0^T (\nabla_\theta P + \Delta x\nabla\mu(\theta))^T q + \int_{\partial\Omega}(\nabla_\theta P - \nabla_s x\cdot n\nabla\mu(\theta))^T q\tilde{\theta} \tag{23}$$

4.2 Second order method

The gradient algorithm is numerically efficient when the problem is well conditioned, meaning that the derivatives in all the directions have the same order of magnitude. In the case of PCP, the parameters are unknown and may range over several orders of magnitude. Therefore the problem is likely to be poorly conditioned, in which case a second order method is preferable. A second order method, such as the Newton method, rescales the variables so that in the new system of variables the problem is well conditioned and consequently the descent algorithm is fast, yet no tractable method currently exists for executing the Newton method in optimization programs involving general PDEs. However, it is possible to implement a quasi-Newton method (Gill et al. (1999)), in which the second order derivative of the objective function, called the Hessian, is computed via finite differences on the gradient.

By doing so, we can form an approximate Hessian H and the descent direction is taken as the one which minimizes the quadratic approximation of the objective function: $\delta\theta = -H^{-1}\nabla J$.

4.3 Summary of the algorithm

Algorithm 1. (`2nd order adjoint based algo.`). Start with an initial guess for the parameters θ^{guess} and an initial guess for the Hessian H^{guess}.
Repeat

(1) Solve the governing equation (1) for x, using the current parameter vector θ.
(2) Solve the adjoint equation (20) for q, using the current θ and x.
(3) Determine the gradient ∇J according to equation (23).
(4) Update the Hessian H via finite difference between the current gradient and the previous ones.
(5) Form the descent direction $\Delta\theta = -H^{-1}\nabla J$.

(6) Line search: compute $\beta > 0$ so that $J(\theta + \beta\Delta\theta)$ is minimized.
(7) Update $\theta := \theta + \beta\Delta\theta$.

Terminate when $\nabla J^T H \nabla J$ is small
Return $\theta^* = \theta$.

4.4 Computational complexity

The adjoint method drastically reduces the complexity of the gradient computation. It only requires two PDE calculations, whereas calculating the gradient via finite difference would have required at least $d + 1$ PDE computations, in which d is the number of parameters to estimate ($d = 37$ in our PCP model). Each iteration of the algorithm moreover consists of a coarse one-dimensional minimization (line search), which is typically terminated after three to six PDE (1) computations. In total and from a conservative view-point, each iteration requires eight to ten times the computational time of running the governing PDE (1). Finally, in terms of convergence, the algorithm generally terminates after 50 iterations; therefore the algorithm requires on the order of 50×10 objection function evaluations.

5. PRELIMINARY SIMULATION RESULTS

An important validation of the algorithm is to make sure that it efficiently searches the parameter space. For this purpose, we present preliminary simulation results for the wild type case. For this simulation, we used a simplified version of the PDE model presented in Amonlirdviman et al. (2005), which assumes that the diffusion terms in the equations can be replaced by their quasi-steady-state solutions. This assumption permits elimination of the diffusion terms from the original PDEs, reducing the model to a system of ordinary differential equations (ODEs) in which the number of parameters to identify is reduced from 37 to 27. The complete ODE model development is presented in Amonlirdviman (2005) and Ma et al. (2005). The adjoint method and quasi-Newton algorithm applied is that presented in the previous section.

In these simulations, we assume some true values of the parameters, which are believed to generate a phenotype consistent with the characteristic PCP phenotypes (termed "true"). Then, we deviate from these parameters and we verify that the output of the search algorithm recovers the true phenotype, which we want to match. In practice, we set $\theta^{\text{guess}} = \theta^{\text{true}}(1 + \sigma\mathcal{N}(0, 1))$ with $\sigma = 1$ and we first run a simplex algorithm, not taking advantage of any gradient information, and second, our quasi-Newton method. The results are shown in Figure 4. In this Figure, we display concentrations of Dsh protein, with "cool" colors representing relatively low concentrations, and "warm" colors representing relatively high concentrations. This result indicates that given the same amount of computational effort, the adjoint-based quasi-Newton method has almost recovered the true phenotype, whereas the simplex method is still far from converging.

For the quasi-Newton method, the computation involved for 30 function evaluations is 60 ODE calculations. We note that here, we are only matching the wild type phenotype. To match wild type and all mutant phenotypes as in Amonlirdviman et al. (2005), we would parallelize the computation; and the total computational time would be equal to the number of ODE computations multiplied by the time to compute the slowest phenotype.

This is the subject of our current work.

ACKNOWLEDGEMENTS

We would like to thank Professor Jonathan Goodman for his help on the optimization procedure.

REFERENCES

P. N. Adler. Planar signaling and morphogenesis in *drosophila*. *Developmental Cell*, 2(5):525–535, 2002.

K. Amonlirdviman. *Mathematical Modeling of Planar Cell Polarity Signaling in the Drosophila Melanogaster Wing*. PhD thesis, Stanford University, August 2005.

K. Amonlirdviman, N. A. Khare, D. R. P. Tree, W.-S. Chen, J. D. Axelrod, and C. J. Tomlin. Mathematical modeling of planar cell polarity to understand domineering nonautonomy. *Science*, 307(5708):423–426, January 2005.

J. D. Axelrod. Unipolar membrane association of Dishevelled mediates Frizzled planar cell polarity signaling. *Genes Dev.*, 15(10):1182–1187, May 2001.

L. Evans. *Partial Differential Equations*. AMS Press, 2002.

P. E. Gill and M. W. Leonard. Reduced-hessian quasi-newton methods for unconstrained optimization. *SIAM J. Optim. Vol. 12, No 1, pp.209-237*, 2001.

P. E. Gill, W. Murray, and Margaret H. Wright. *Practical Optimization*. Academic Press. Harcourt Brace and Company, 1999.

A. Jameson. Aerodynamic design via control theory. *Princeton University Report MAE 1824, ICASE Report No. 88-64, November 1988, also, J. of Scientific Computing, Vol. 3, 1988, pp. 233-260*, 1998.

J.-L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. translated by S.K. Mitter, Springer Verlag, New York, 1971.

D. Ma, K. Amonlirdviman, C. J. Tomlin, and J. D. Axelrod. Irregularities in cell packing necessitate a robust planar cell polarity signaling mechanism. 2005. Submitted.

D. Ma, C.-H. Yang, H. McNeill, M. A. Simon, and J. D. Axelrod. Fidelity in planar cell polarity signalling. *Nature*, 421:543–547, 2003.

J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7: 308–313, 1965.

D. I. Strutt. The asymmetric subcellular localisation of components of the planar polarity pathway. *Semin. Cell Dev. Biology*, 13(3):225–231, 2002.

D. R. Tree, J. M. Shulman, R. Rousset, M. P. Scott, D. Gubb, and J. D. Axelrod. Prickle mediates feedback amplification to generate asymmetric planar cell polarity signaling. *Cell*, 109 (3):371–381, May 2002.

C.-H. Yang, J. D. Axelrod, and M. A. Simon. Regulation of frizzled by fat-like cadherins during planar polarity signaling in the *drosophila* compound eye. *Cell*, 108(5):675–688, 2002.

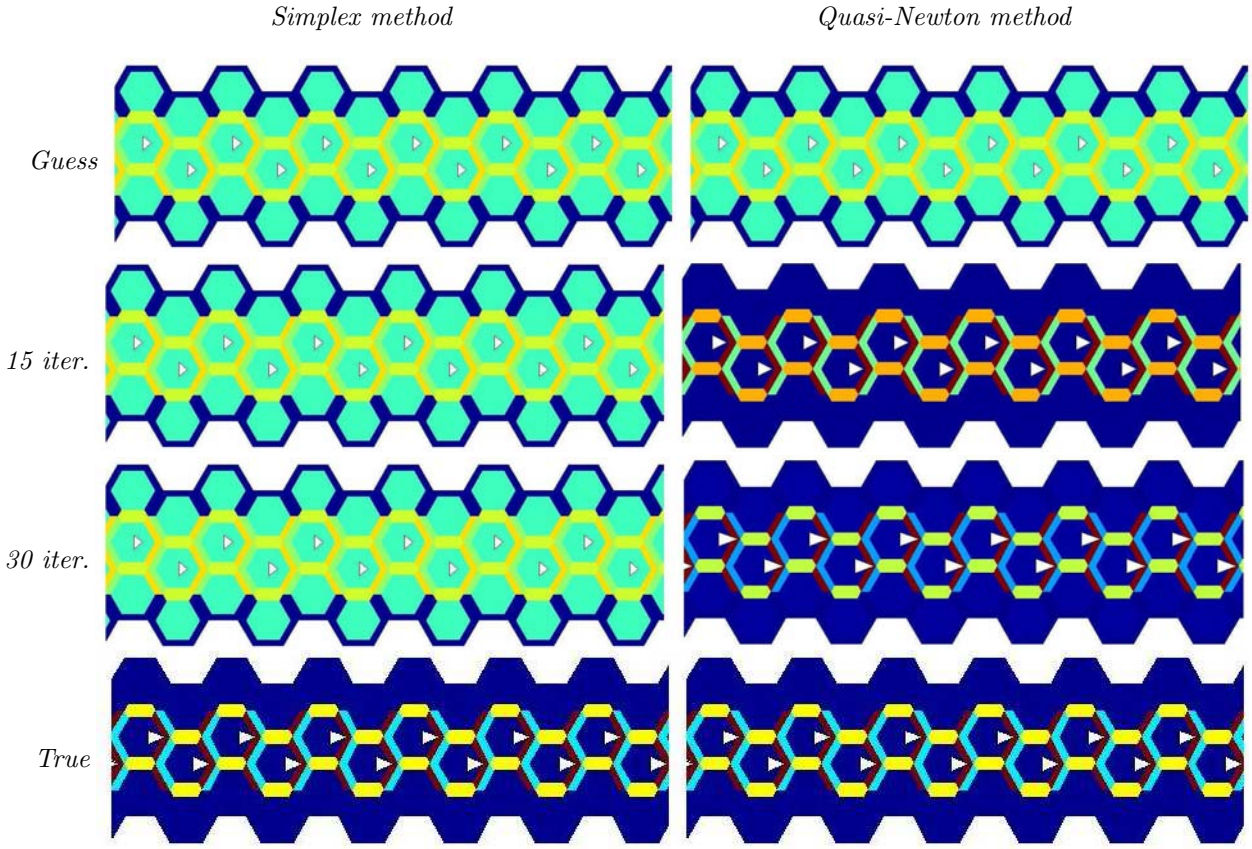


Fig. 4. Comparison between the simplex method and the quasi-Newton method described here for the parameter identification problem. After 30 iterations, the quasi-Newton method has almost recovered the true phenotype.

Keynote 5

Modeling of HIV Infection: Vaccine Readiness, Drug Effectiveness and Therapeutical Failures

X. Xia
University of Pretoria

Keynote 6

Stability and Controllability of Batch Processes

B. Srinivasan and D. Bonvin
Ecole Polytechnique Fédérale de Lausanne

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

MODELLING OF HIV INFECTION: VACCINE READINESS, DRUG EFFECTIVENESS AND THERAPEUTICAL FAILURES

Xiaohua Xia *

* *Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa, Phone: +27 (12) 420 2165, Fax: +27 (12) 362 5000, E-mail: xxia@postino.up.ac.za.*

Abstract:
The paper starts with a review of the research of the pathogenesis of HIV and its interplay with the study of mathematical models. Advances are then highlighted on model identifiability, identification techniques and the applications to current biomedical research and up-to-date clinical practise. An identifiability study is an answer to problems of what quantities and how frequently to measure in blood plasma. Parameter identification methods are chosen and developed for sparse and rough samples. Results are reported on two case studies: vaccine readiness in Southern Africa, drug effectiveness and therapy failures on existing patients in France. Ongoing research programmes and future opportunities are pointed out.

Keywords: biomedical, HIV/AIDS, identifiability, identification, nonlinear systems

1. INTRODUCTION AND A REVIEW OF HIV MODELS

The key markers of the disease progression are the CD4+ T cell and viral levels in the plasma. The typical dynamics of the disease progression, in an untreated individual, for these populations is shown in Figure 1 (Fauci et al, 1996). Highly active antiretroviral therapy (HAART), therapeutic regimens employing drug combinations has shown its ability to cause dramatic and sustained suppression of viral replication and immune system recovery. A typical patient's response is shown in Figure 2 (Ho et al, 1995).

A basic 3D model (Nowak and May, 2000) has been developed to reveal the dynamics in these figures. (Nowak and May, 2000):

$$\begin{cases} \dot{T} = s + rp(T,v) - dT - \beta Tv, \\ \dot{T}^* = \beta Tv - \mu T^*, \\ \dot{v} = kT^* - cv, \end{cases} \tag{1}$$

where T denotes the healthy CD4 cells, T^* denotes the infected CD4 cells, v denotes the free virus particles, and $p(T,v)$ denotes the proliferation of the CD4 cells.

It is estimated that as many as 10^{10} virions are produced and destroyed in an infected individual each day (Perelson et al, 1996). These findings are consistent with a simple steady-state analysis of the model (1) (see (Perelson and Nelson, 1999)). The equilibrium or set point of virus depicted by the model (1) is

$$v^* = \frac{ks}{\mu c} - \frac{d}{\beta}. \tag{2}$$

It can be seen that a model of such a simple nature is able to adequately reflect the disease progression from the initial infection to a stage where the set-point is reached. This is one of the reasons why this model was used in the estimation of set-points in the vaccine programme (Gray et al, 2005; Filter, Xia and Gray, 2005) (see also section 3.1).

One of the interpretations of antiviral drugs, reverse transcriptase inhibitors (RTI) and protease inhibitors (PI) in particular, is that they reduce infection of healthy cells (Nowak and Bonhoeffer, 1995). Under an ideal situation, a 100% effective inhibitor corresponds in the model to setting $\beta = 0$. This understanding has yielded estimates for μ and c in (Perelson et al, 1996) at roughly 0.45/day and 3/day, respectively.

The pool of virus producing cells has been estimated to be very small (3×10^7).If there are no other sources of hidden virus, the eradication of all the virus would take only about 25 days $(3 \times 10^7 \exp(-0.45 \times 25) < 1)$.

Even though the basic model is only valid for a short period of the disease progression, researchers still use it to explain the hallmark long term depletion of CD4+ T cells. The prolonged, high level output of HIV *in vivo* reflects an active, ongoing, process in which CD4 lymphocytes are being infected and killed in large numbers.

Perelson *et al* observed that, after the rapid first phase of decay during the initial 1–2 weeks of antiretroviral treatment, plasma virus levels declined at a considerably slower rate. This second phase of viral decay was attributed to the turnover of a longer-lived virus reservoir of infected cell population, which can be adequately described by a long-lived cell model.

A latently infected cell model was also proposed, because additional cellular reservoirs of virus were found in lymphoid tissues (Haase et al, 1996), particularly on the surface of follicular dendritic cells (FDC). The source underlying the second-phase kinetics might be the release of virions trapped in the lymphoid tissues. It could be linked to infected macrophage, and/or due to the activation of latently infected cells.

The long-lived cells were determined to have a half-life of 1–4 weeks. This means that on average it would take between two and a half to three years of perfectly effective treatment to eradicate the virus. This estimate generated a lot of enthusiasm and optimism in 1996 (Perelson et al, 1996). As it turned out that a third phase of HIV decay was observed from continued follow-up of persons who had remained on HAART for extended periods of time. It suggests that there possibly exists a reservoir of long-lived CD4+ memory T lymphocytes. The kinetics of decay are extremely slow, and the half-life of the memory cell reservoir has been estimated at between 6 and 44 months. As a consequence, the predicted time required for effective antiretroviral therapy to fully eradicate HIV from the body ranges from 9 to 72 years. It implies that conventional antiretroviral regimens are not a true virologic cure.

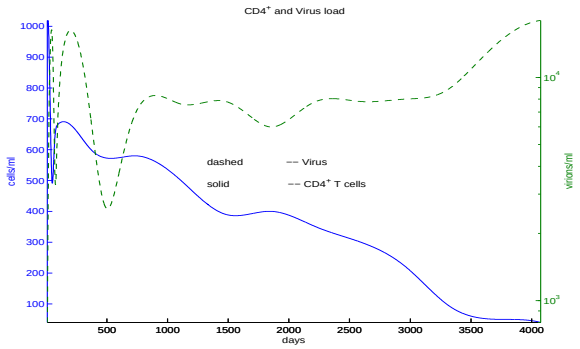


Fig. 1. Typical HIV/AIDS course (Fauci et al, 1996)

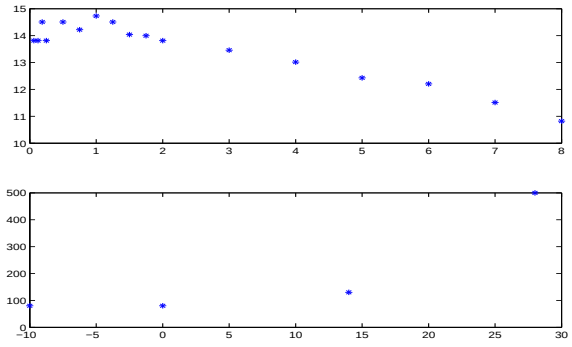


Fig. 2. Typical post-treatment dynamics (Ho et al, 1995)

Table 1. Virus reservoir and life span

Infected Cell	Size	Half-life	Eradication
Active CD4	3×10^7	1 day	25 days
FDC	3×10^8–10^{11}	1–4 wks	0.5–2.8 yrs
Macrophage	n.a	1–4 wks	n.a
Memory CD4	10^5–10^6	6–44 mon	9–72 yrs

Table 1 is a summary of these important findings. Most of information can be found or deduced from (Dewhurst, da Cruz and Whetter, 2000).

The above theories and models are the results in only one direction of the research devoted to address some deficiencies of the basic model (1). Refer to (Covert and Kirschner, 2000) for some of the early models, and (Perelson and Nelson, 1999; Nowak and May, 2000; Callaway and Perelson, 2002) for other important models. Refer to (Wu and Tan, 2005) for some of the newest models.

These developments have brought researchers to face the challenges. It is clear that the modelling

approach has paved the way for theoretical research and has changed the perception of people about the disease. This is achieved by extracting key features from the model parameters. Mathematical modelers are believers of determinism, and accurate models are able to determine the evolution of the infection and the disease. This approach, fundamentally different from the "orthodoxical" medical approach based on statistics over large population, relies on a rapid collection of individual patient's data over a short period of time. Even though there are general observations that can be made from the model and its structure, it is only when the model is tailored to each patient's individual parameters that clear benefits in the treatment strategy arise. More complex models involve more variables, and thus need more data. The records of patients in current clinical practise are typically sparse and rough. There is a need to strike a balance in model complexity and usability in order to use the HIV/AIDS models as a tool for treatment decisions. The following sections show applications of some simple models. Control engineering techniques are brought into the model building.

2. HIV/AIDS PARAMETERS

In this section, it will first be shown how an identifiability study of some simple HIV/AIDS models helps in formulating guidelines for clinical testing and measurement. Some parameter estimation methods will be described.

2.1 Identifiability

For HIV/AIDS parameter identification, the first question to ask is what variables can be measured? Clinically, many variables can be measured, though some of them with less accuracy and high cost. The clinical practise recommended by some medical guidelines (USPHS, 2003) is to measure the viral load and the CD4+ T cell counts in plasma. Since CD4+ T cells are predominantly healthy cells (Janeway and Travers, 1997), also for technical reasons, it is assumed that viral load and healthy CD4+ T cells are measured outputs.

The following questions then arise: what is the minimal number of measurement samples for the CD4+ T cell and the viral counts? when should these measurements be taken? and how? and can one get accurate estimates of the parameters from these measurements?

Identifiability is a basic system property to address these questions. Assume no proliferation is considered in (1), and taking the outputs as

$$\begin{aligned} y_1 &= T, \\ y_2 &= v, \end{aligned} \tag{3}$$

the system (1) is both observable and identifiable, as was shown in (Xia and Moog, 2003). The system (1) has six parameters, denoted by

$$\kappa = (s, d, \beta, c, \mu, k)^T.$$

Identifiability or precisely, algebraic identifiability (Xia and Moog, 2003), means that all parameters κ can be determined from the measured output. To actually find the parameters, higher order differential equations of the output can be calculated,

$$\dot{y}_1 = \theta_1 + \theta_2 y_1 + \theta_3 y_1 y_2, \tag{4}$$

$$\ddot{y}_2 = \theta_4 \dot{y}_2 + \theta_5 y_2 + \theta_6 y_1 y_2, \tag{5}$$

where $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)^T = (s, -d, -\beta, -\mu - c, -\mu c, k\beta)^T$. Θ defines a one-to-one map for $\beta \neq 0$ and $c > \mu$ (Nowak and May, 2000). Therefore, the identification of the original parameters of (1) is equivalent to that of Θ. Thus, all the original parameters are identifiable from the measurement of the viral load and the CD4+ T cell counts in the blood of an HIV patient.

It is necessary to generate a minimum of six equations based on (4) and (5), three from each equation. This can be achieved by differentiating (4) and (5) two more times, resulting in derivatives of y_1 and y_2 up to the order of 3 and 4 respectively. To cope with these order of derivatives, one concludes that at least four measurements of the CD4+ T cell count y_1 and five measurements of the viral load are needed for a complete determination of all the HIV/AIDS parameters in the three dimensional model (1).

Identifiability of other models, and with different measured outputs, is studied in (Xia and Moog, 2003; Jeffrey, Xia and Craig, 2003b; Jeffrey and Xia, 2005).

2.2 Parameter estimation

2.2.1. Least square estimates For simplicity, assume that measurements y_1^0, y_1^1, y_1^2, y_1^3, y_2^0, y_2^1, y_2^2, y_2^3, and y_2^4 are available, the following three equations can be generated based on (4), in which the derivative of y_1 is approximated by $\Delta y_1 / \Delta t$,

$$A \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} 1 & y_1^0 & y_1^0 y_2^0 \\ 1 & y_1^1 & y_1^1 y_2^1 \\ 1 & y_1^2 & y_1^2 y_2^2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} \dfrac{y_1^1 - y_1^0}{d_1} \\ \dfrac{y_1^2 - y_1^1}{d_2} \\ \dfrac{y_1^3 - y_1^2}{d_3} \end{bmatrix}$$

If the matrix A is nonsingular, then there is a unique solution for θ_1, θ_2 and θ_3, and hence estimates for s, d and β. These are essentially least square (LSQ) estimates.

On the other hand, when either y_1 or y_2 is constant, A can never be nonsingular for any choice of measurement interval. In the long asymptomatic stage, the viral load y_2 remains constant, and in the short period after chemotherapy treatment, the CD4+ T cell count does not change much (see the assumptions made in (Ho et al, 1995; Wei et al, 1995)). Therefore during these two time periods, a complete determination of s, d and β is impossible.

Similar conclusions can be drawn from working with (5) for the estimates of μ, c and k.

This analysis also helps to indicate the most likely period for a complete estimation of parameters. An intuitive interpretation of the above analysis is that when the curves of y_1 and y_2 are "bent enough" and "the cumulated strength of virus" (y_2) is bounded, all six parameters can be estimated with confidence of accuracy. Two typical such phases in HIV/AIDS progression are the primary infection stage and the period after chemotherapy treatment when both the viral load and CD4+ T cell counts are changing.

Coincidentally, one notices that all previous estimations of the virus clearance rate (c) and the death rate of infected cell (μ) were made for a post-treatment period of very strong chemotherapy with reverse transcriptor inhibitors and protease inhibitors in (Ho et al, 1995; Wei et al, 1995; Perelson et al, 1996). This choice becomes obvious from the above analysis of parameter convergence.

Of course, this pure LSQ would fail with noisy measurements. The measurement error of the Roche Amplicor assay method could be 0.18 log 10 copies/ml (Schuurman et al, 1996). Another problem with the pure LSQ is that the measurements of viral load and CD4 cells have to coincide in time. Ways to overcome these include adaptive algorithms (Xia, 2003) and improved versions of the LSQ method.

2.2.2. A penalty function approach The penalty function method is in essence LSQ based, but with two important differences: firstly, derivative estimation is only present when a nominal curve is generated by a numerical ordinary differential equation solver and, thus, this estimation is not influenced by measurement noise. Secondly, the cost function is not limited to the LSQ distance, thus, it can be expanded to accommodate a diverse base of knowledge in order to increase the accuracy of parameter estimation.

Table 2. Comparisons of estimates

Patient No.	Method	$\hat{c}$ (day^{-1})	$t_{c\frac{1}{2}}$ (days)	$\hat{\mu}$ (day^{-1})	$t_{\mu\frac{1}{2}}$ (days)
107	published	3.1	0.2	0.5	1.4
	penalty function	2.07	0.33	0.50	1.39
	x_0 *modified*	*3.07*	*0.23*	*0.49*	*1.41*

In addition, since the penalty function method does not require any product terms, there is no constraint on the length of CD4+ T cell and virus data vectors. Together with N measurements of T and K measurements of v, at time $t_1, \ldots, t_N$, and $\tau_1, \ldots, \tau_K$, respectively, the basic cost function is defined as

$$J_w = \sum_{n=1}^{N} \frac{(\hat{T}(t_n) - T_n)^2}{N\text{mean}(T_n)} + \sum_{k=1}^{K} \frac{(\hat{v}(\tau_k) - v_k)^2}{K\text{mean}(v_k)}.$$

It can been seen that the points of the two data vectors need not coincide in time.

Additional refinements of the cost function can be made incorporating outside knowledge of the dataset and the parameters. For example,

$$J_r = J_w + \kappa_1 \max(\frac{d\hat{v}_s}{dt}, 0) + \kappa_2 \max(\hat{\mu} - \hat{c}, 0),$$

where $\hat{v}_s$ is the vector of computed viral load, truncated after a few days. κ_1 and κ_2 are two scaling constants. The first refinement term corresponds to the knowledge that the patient is in steady state before initiation of therapy. The second refinement term corresponds to the statement that the average infected CD4+ T cell lives longer than free virions.

To validate the method, the parameter estimation for the three patients in (Perelson and Nelson, 1999) was repeated in (Filter and Xia, 2003). It was used to extract the same two parameters as in the experiment. Note that all the assumptions described in the experiment were included in the estimate by customizing the penalty function as described above. The results are listed in Table 2 for patient number 107 whose data was plotted in Figure 2.

There is a distinct, and consistent difference in $\hat{c}$ between the published results and the estimation by custom penalty function. Since the estimation of c is dependent on the shoulder region of the virus count (Nowak and May, 2000), a dependence on x_0 is to be expected. For an optimal adjusted x_0, one can estimate the parameters again with the same cost function. The small data window for this experiment does not allow clear information to be found about the initial conditions. From the results it is clear that the estimation of c, and to a lesser extent, of μ, is dependent on outside information about x_0.

2.2.3. A deterministic optimization method To speed up the calculation, (Ouattara et al, 2004; Ouattara, 2005) employed an estimation procedure applied to the discrete-time model of the system (1). The principle of the estimation procedure is explained as the following. Suppose an optimization algorithm (*e. g.*, steepest descent, simplex) is chosen. Running this algorithm to minimize an objective function on large number of initial conditions, one can see that the solutions are distributed in the neighborhood of the real optimum. A median and an interquartile range (IQR) of the calculated solutions offer a desired estimate and the confidence interval of the estimate.

Since a deterministic approach is taken, the procedure is called a deterministic optimization method.

The IQR measures the dispersion of the results. It depends on the chosen tolerance and the convexity of the objective function. The IQR gives an important information on the confidence on the results. This method is extremely useful with sparse data samples. In the HIV/AIDS case, for ethical and financial reasons, it is impossible to collect large amount of data. With the deterministic optimization method, (Ouattara et al, 2004; Ouattara, 2005) were able to compute estimates of all the parameters of the model (1) with minimum number of samples.

3. CASE STUDIES

3.1 HIV vaccine readiness

An interesting application of the estimation procedures, described in section 2.2.2, is the extraction of parameters for patients who took part in an HIV/AIDS vaccine readiness study (Gray et al, 2005). In this study, HIV viral load may be a critical endpoint in vaccine trails by which to judge efficacy. It is important to define viral dynamics in unvaccinated infected individuals, especially in non-B subtype infections where little information is available. The main aim was to determine the set-point for these patients,and find the time from seroconversion for this set-point to be reached.

Fifty-one individuals with recent HIV infection were recruited within 18 months of acquiring HIV infection from four countries in southern Africa (10 from Zimbabwe, 6 from Malawi, 16 from Zambia and 19 from South Africa). Participants were followed at 2, 4, 7 and 9 months after enrolment. At each visit, blood samples were obtained for plasma RNA levels, lymphocyte subset analysis and DNA isolation. Participants were not on antiretroviral treatment. The majority (42/51) were female. The median age was 28 and the median interval from seroconversion to first viral load measurement was 8.9 months (interquartile range of 5.5 – 14.1). Comparison of $\log_{10}$ RNA copies/ml in participants at enrolment between countries showed no significant difference and, based on this, were grouped as one cohort.

Table 3. Data points for two sample patients

days	viral load	cd4+	days	viral load	cd4+
P**15**	*(5+4)*		P**42**	*(4+3)*	
391	46 699		411	67 813	
426	24 463	187	474	11 569	
503	62 364	136	586	39 887	186
573	25 079	193	685		178
636	29 821	143	775	19 359	272

Quite coincidentally, thirty-four of the fifty-one participants had four CD4+ T cell and five viral load counts, thus satisfying the minimum requirement of algebraic identifiability of the basic model. Ten additional patients from the cohort had insufficient data points for a complete evaluation of parameters on their own. For these patients, the assumption was made that their parameter value for c, the death rate constant for virus, did not differ significantly from other patients in the cohort. The rest of the data (seven patients) were useless. Thus, in total, parameters were estimated for forty-four of the fifty-one participants.

Typical data sets are displayed in Table 3 for two patients P15 and P42.

Even though the minimum requirements for parameter estimation were met by most of the patients in this study, the time difference between points, lack of initial CD4+ T cell data and imprecise measurements, required a set of assumptions to be made before the model parameters were estimated: i) Patients were in the early stages of infection; ii) The midpoint between the last negative and first positive sample was taken as an estimate for the time of seroconversion; iii) Patients did reach a steady state in viral load; iv) Initial values for the viral load coincided with the first viral load measurement; v) The order of c and μ was not dictated in this instance of the cost function, and vi) others (for details, see (Filter and Xia, 2003; Filter, Xia and Gray, 2005)).

After the parameters for each patient had been estimated, the set-point was calculated according to (2). In order to find the time to reach the set-point, the fluctuations in viral load were considered. The time from seroconversion to the point where the fluctuations fell within log 0.5 of the set-point was taken as an estimate for the time to reach set-point.

In Figure 3 a detailed view is given for P42 with four data points in the viral load. The markers

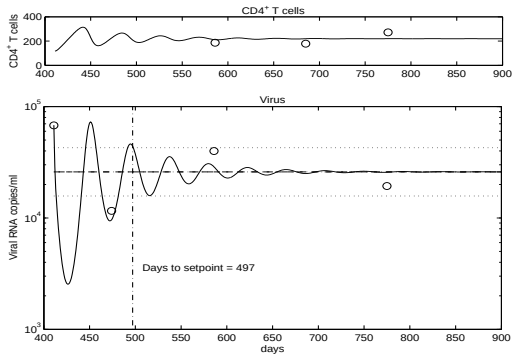


Fig. 3. Model and data example (Filter, Xia and Gray, 2005)

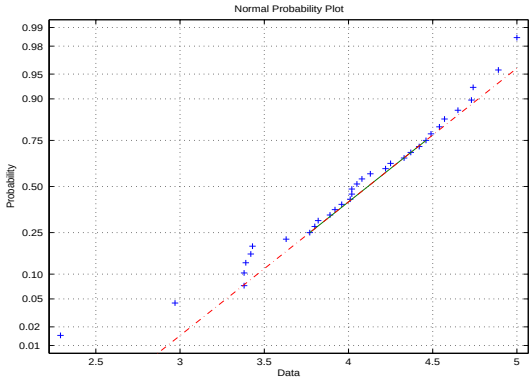


Fig. 4. Normal probability plot of set-point estimations (Filter, Xia and Gray, 2005)

indicate data points and the corresponding model prediction (solid line), derived from parameter estimates. The estimation of the set point is indicated by a dashed line, and the time to set point is taken at the point where the modelled viral load falls between the dotted lines.

The sparsity of the data set did not allow a conclusion about an individual patient, but the results of 44 patients can provide some statistical information about subtype C viral dynamics. The following median estimates of parameters were found

$$\hat{\chi} = (7.48, 0.00085, 1.4 \times 10^{-6}, 1.56, 0.80, 2834)^T.$$

Figure 4 shows the normal probability plot of $\log_{10}$ set-point estimations for all patients. It is clear that the estimates follow a log-normal distribution. The Bera-Jarque parametric hypothesis test of composite normality confirms this, with a significance level of 0.298. The calculated median time to set-point was 16.57 months and the median of the calculated set-point distribution was $4.08 \log_{10}$ (12143 RNA copies/ml). Interestingly, these estimates appear to be no different from reported studies of subtype B HIV infected male cohorts ((Mellors et al, 1996; Mellors et al, 1997; Schacker et al, 1998).

3.2 Therapy effectiveness and therapeutical failures

In another case study, parameter estimates were obtained using the deterministic optimization method for two representative patients from the CHU of Nantes, France (Nantes University Hospital).

The first patient was 43 years of age. He was treated in two consecutive periods. In the first period from day zero to day 272, the patient was treated with two RTIs: zidovudine (AZT) and lamivudine (3TC) and one PI: saquinavir soft gel (SQV). During this period, the viral load drops from 65000 copies/ml to 5000 copies/ml in 113 days, and then increases to 21000 copies/ml. In the second period from day 273 to day 2379, the patient's therapy consists of two RTIs: 3TC and stavudine (d4T) and one PI ritonavir (ABT-538). The viral load drops under the threshold of 50 copies/ml.

The second patient was 42 years old. The first line of antiretroviral therapy is a bi-associated combination of RTI (AZT+3TC): from day zero to day 124, then a tritherapy, AZT +3TC+IDV (indinavir), for the next 769 days. During the second period, the patient received several associations: AZT+3TC+ABT-538+IDV (treatment stopped by the patient's decision); AZT+3TC+NFV (nelfinavir) (treatment stopped due to a virological failure); AZT+3TC+IDV (treatment stopped for toxicity). All these therapies have not been shown to be efficient. In the last period, the patient is treated with AZT +3TC +EFV (efavirenz). No PI was involved.

In the parameter estimation, it was assumed that the proliferation term in (1) takes the following form $p = rTv/(K + v)$.

Two thousand initial conditions were used, and they were assumed to be uniformly distributed on an admissible interval of the underlining parameter. For example, the admissible interval was chosen to be $[10^{-20}, 20]$ assuming that the production of CD4 cells is always positive and smaller than 20 $CD4/mm^3$

The estimates for patient 2 are tabulated in Table 4, in which estimates outside of brackets are for the first period, and estimates in brackets $(\ldots)$ are for the period from day 1280 to day 2388.

From these results, it can be observed that for patient 2 and in the second period of treatment, the parameter β increases (1450 times higher), and k decreases by 9 times lower than that in the first period. This implies that the AZT+3TC+EFV combination is more efficient than the AZT+3TC

Table 4. Estimates of patient 2

	estimates	IQR	$CI_{50\%}$
s	0.46	0.57	0.13,0.17]
	(0.13)	(0.24)	($[5.55 \cdot 10^{-4}$, 0.24])
d	$2.09 \cdot 10^{-3}$	$3.68 \cdot 10^{-3}$	$[8.44 \cdot 10^{-5}$, $3.77 \cdot 10^{-3}]$
	$(2.79 \cdot 10^{-3})$	$(6.78 \cdot 10^{-3})$	($[4.30 \cdot 10^{-4}$, $7.20 \cdot 10^{-3}]$)
μ	0.06	0.037	[0.044, 0.074]
	(0.12)	(0.1)	([0.079, 0.18])
c	0.092	0.068	[0.061, 0.13]
	(0.64)	(0.97)	([0.3, 1.26])
β	$1.02 \cdot 10^{-9}$	$1.79 \cdot 10^{-8}$	$[5.42 \cdot 10^{-14}$, $1.79 \cdot 10^{-8}]$
	$(1.50 \cdot 10^{-6})$	$(2.13 \cdot 10^{-6})$	($[6.19 \cdot 10^{-7}$, $2.75 \cdot 10^{-6}]$)
k	2604	5186.5	[152.04, 5338.5]
	(296.24)	(604.73)	([102.21, 706.94])
r	$1.08 \cdot 10^{-4}$	$8.04 \cdot 10^{-3}$	$[2.79 \cdot 10^{-9}$, 0.008]
	$(3.23 \cdot 10^{-3})$	$(7.75 \cdot 10^{-3})$	($[1.04 \cdot 10^{-5}$, $7.76 \cdot 10^{-3}]$)
K	9.53	1009	$[3.59 \cdot 10^{-5}$, 1009]
	(17.36)	(146.5)	([0.087, 146.59])

combination in the first period. The lack of PI results in higher value of the parameter β. The higher value of r suggests a more active proliferation in the second period. It could also recommended that the RTIs of the second period (AZT +3TC+EFV) combined with the PI of the first period (IDV) could further lower both parameters β and k.

Virological failure is due to persistent replication of the viral load under treatment. Thus virological failure could be interpreted as extreme values of the parameters k and β. Immunological failure is defined when the amount of CD4 cells remains below the level of 200/mm^3 after 6 months of treatment. Immunological failure could be detected by the indicator

$$t_{200} = \frac{1}{\tilde{d}}[-\log_e(1 - 200\frac{\tilde{d}}{s}) + \log_e(1 - \frac{\tilde{d}}{s}T_0)],$$

where $\tilde{d} = d - \frac{r\bar{v}}{K+\bar{v}}$, $\bar{v}$ is the mean of the viral load, and T_0 is the initial CD4 cell level. If $t_{200} > 6$ months, then there is immunological failure.

It could be verified that $t_{200} = 33$ months, indicating an immunological failure of the first period of treatment on patient 2. Clinical data shows that approximately 25 months were necessary for him to reach the 200 CD4/mm^3 level.

A superposition of both virological and immunological failure is referred to as biological failure, and clinical failure is characterized by the clinical manifestation of opportunistic diseases.

4. CONCLUDING REMARKS

The research of the pathogenesis of HIV has reached a point where control system engineering can play a constructive role.

The parameter estimation schemes are also useful for immune prognosis. Anti-retroviral therapies are usually effective in suppressing the viral load in a short period of time. The response of immune systems takes a longer time (usually more than a month) to manifest. It is desirable to predict the immune system response with less than a month's samples. Preliminary results in this regard have been done for six naive patients at CHU of Nantes, France, and further experiments are currently done for more difficult patients.

Given the fact that a large amount of raw data have been accumulated around the world of HIV patients under HAART. The drug effectiveness and therapeutical failure models have to be populated with different HIV subtypes and with different geographic data. For this purpose, an automated computer routine has the advantage, and is also under current investigation.

The system modelling ideas can certainly be borrowed to HIV and TB co-infections and other infectious disease.

And perhaps the biggest scientific open problem is whether a true cure for HIV/AIDS can eventually be found. If there is a cure, then one thing is certain that it comes as a treatment strategy which includes not only combination of drugs but also the manner in which these drugs are administered. From a practical point of view, affordable and effective means to prevent and stop infection and progression is the key to the success of the fight against AIDS. Applying these ideas to create suitable educational platforms for the high-risk groups of people is also deemed necessary (Craig, Xia and Venter, 2004; Craig and Xia, 2005).

Acknowledgement: I would like to express my gratitude to all my collaborators on the research of HIV/AIDS for allowing me to use some materials from our joint publications and research.

REFERENCES

Callaway, D. S., & Perelson, A. S. (2002). HIV-1 infection and low steady state viral loads. *Bulletin of Mathematical Biology*, 64, 29-64.

Covert, D., & Kirschner, D. (2000). Revisiting early models of the host-pathogen interactions in HIV infection. *Comments Theoretical Biology*, 5, 383–411.

Craig, I., & Xia, X. (2005). Can HIV/AIDS be controlled? *IEEE Control Systems Magazine*,25, 80–83.

Craig, I. K., Xia, X. & Venter, J. W. (2004). Introducing HIV/AIDS education into the electrical engineering curriculum at the University of Pretoria. *IEEE Transactions on Education*, 47, 65–73.

Dewhurst, S., da Cruz, R. L. W., & Whetter, L. (2000). Pathogenesis and treatment of HIV-1 infection: recent developments (Y2K update). *Frontiers in Bioscience*, 5, 30–49.

Fauci, A. S., Pantaleo, G., Stanley, S., & Weissman, D. (1996). Immunopathogenic mechanisms of HIV infection. *Annals of Internal Medicine*, 124, 654–663.

Filter, R., & Xia, X. (2003). A penalty function approach to HIV/AIDS model parameter estimation. *13th IFAC Symposium on System Identification*, Rotterdam, Netherlands, 27–29 August 2003.

Filter, R., Xia, X., & Gray, C. (2005). Dynamic HIV/AIDS parameter estimation with application to a vaccine readiness study in Southern Africa. *IEEE Transactions on Biomedical Engineering*, 52, 284–291.

Gray, C. M., Williamson, C., Bredell, H., Puren, A., Xia, X., & et al (2005). Viral dynamics and CD4+ T celll counts in subtype C human immunodeficiency virus type 1-infected individuals from Southern Africa, Aids Research and Human Retroviruses, vol. 21, no. 4, 2005, pp. 285–291.

Haase, A. T., Henry, K., & et al (1996). Quantitative image analysis of HIV-1 infection in lymphoid tissue. *Science*, 274, 985–989.

Ho, D. D., Neumann, A. U., & et al (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*, 273, 123-126.

Janeway, C. A., & Travers, P. (1997). *Immunobiology: The Immune System in Health and Disease*, Garland, New York.

Jeffrey, M., & Xia, X. (2005). Identifiability of HIV/AIDS models, in: DETERMINISTIC AND STOCHASTIC MODELS OF AIDS AND HIVWITH INTERVENTION, Chapter 11, H. Wu and W. Y. Tan (Eds), World Scientific Publications, Singapore.

Jeffrey, M., Xia, X., & Craig, I. K. (2003a). When to initiate HIV therapy: a control theoretic approach. *IEEE Transactions on Biomedical Engineering*, 50(11), 1213–1220.

Jeffrey, M., Xia, X., & Craig, I. K. (2003b). Identifiability of an extended HIV model. *5th IFAC Symposium on Modelling and Control in Biomedical System*, Melbourne, Australia, 21–23 August 2003.

Mellors, J. W., Rinaldo, & et al (1996) Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*, 272, 1167–1170.

Mellors, J. W., Munoz, A., & et al (1997). Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann. Intern. Med.*, 126, 946–954.

Nowak, M. A., & Bonhoeffer, S. (1995). Scientific correspondence. *Science*, 375, p. 193.

Nowak, M. A., & May, R. M. (2000). *Virus Dynamics: mathematical principles of immunology and virology.* New York: Oxford University Press.

Ouattara, D. A. (2005). Mathematical analysis of the HIV-1 infection: parameter estimation, therapies effectiveness and therapeutical failures. *27th IEEE EMBS Annual International Conference*, September 1-4, 2005, Shanghai, China.

Ouattara, D. A., Bugnon, F., Raffi, F., & Moog, C. H. (2004). Parameter identification of an HIV/AIDS model. *13th International Symposium on HIV and Emerging Infectious Diseases*, Toulon, France, September 2004.

Ouattara, D. A., Bugnon, F., Raffi, F., & Moog, C. H. (2005). Therapy effectiveness and therapeutical failures for HIV-1 infection. *IEEE Transactions on Biomedical Engineering*, submitted.

Padhye, N. V., Nelson, R. M., & et al (2002). High-speed amplification of bacillus anthracis DNA using a pressurized helium-CO_2 gas thermocycler. *Genetic Eng. News*, 22, 42–43.

Perelson, A. S., Essunger, P., Markowitz M., & Ho, D. D. (1996) How long should treatment be given if we had an antiretroviral regimen that completely blocked HIV replication? *XIth Intl. Conf. on AIDS*, 1996.

Perelson, A. S., & Nelson, P. W. (1999). Mathematical analysis of HIV-1 dynamics *in vivo*. *SIAM Review*, 41, 3–44.

Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M., & Ho, D. D. (1996). HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time. *Science*, 271, 1582–1586.

Schacker, T. W., Hughes, J. P., & et al (1998). Biological and virologic characteristics of primary HIV infection. *Ann. Intern. Med.*, 128, 613–620.

Schuurman, R., Descamps, D., & et al (1996). Multicenter comparison of three commercial methods for quantification of human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.*, 34, 3016–3022.

USPHS. (2003). Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. HIV/AIDS Treatment Information Service, http://www.aidsinfo.nih.gov.

Wei, X., Ghosh, S. K., & et al (1995). Viral dynamics in HIV-1 infection. *Nature*, 273, 117-112.

Wu, H., & Tan, W. Y. (2005). Deterministic and Stochastic Models of AIDS and HIV with Intervention, World Scientific Publications, Singapore 2005.

Xia, X. (2003). Estimation of HIV/AIDS parameters. *Automatica*, 39, 1983–1988.

Xia, X., & Moog, C. H. (2003). Identifiability of nonlinear systems with application to HIV/AIDS models. *IEEE Trans. Automat. Contr.*, 48, 330–336.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

STABILITY AND CONTROLLABILITY OF BATCH PROCESSES

B. Srinivasan[1] and D. Bonvin[2]

[1] *Department of Chemical Engineering*
École Polytechnique Montreal, Montreal, Canada H3C 3A7
[2] *Laboratoire d'Automatique*
École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne, Switzerland

Abstract: Improving the performance of batch processes requires tools that are tailored to the specificities of batch operations. These include a mathematical representation that explicitly shows the two independent time variables (the run time t and the run index k) as well as the two types of outputs (the run-time and run-end outputs). Furthermore, corrective action can be taken via both on-line and run-to-run control. This paper investigates the important notions of stability and controllability for batch processes, where it is shown that a value rather than a yes-no answer needs to be considered. The tools required for evaluating these properties are readily adapted from the literature. Finally, the various control strategies are illustrated via the simulation of a semi-batch reactor, and references are made to the appropriate tools for evaluating stability and controllability.

Keywords: Batch Processes, Repetitive Processes, On-line Control, Run-to-run Control, Stability, Controllability.

1. INTRODUCTION

The majority of control studies in the literature have dealt with continuous processes operating around around an equilibrium point. In recent years, however, the class of systems where the process terminates in finite time has received increasing attention. An interesting feature is the fact that most of these processes are repeated over time. Many industrial operations, especially in the areas of batch chemical production, mechanical machining, and semiconductor manufacturing do fall under this category.

In a batch process, operations proceed from an initial state to a very different final state. Hence, there exists no single operating point around which the control system can be designed (Bonvin 1998). Also, since batch processing is characterized by the frequent repetition of batch runs, it is appealing to use the results from previous runs to improve the operation of subsequent ones. This has generated the industrially relevant topic of run-to-run control and optimization (Campbell *et al.* 2002, Francois *et al.* 2005). Repetition provides additional degrees of freedom for meeting the control objectives since the work does not necessarily have to be completed in a single run but can be distributed over several runs. This brings into picture an additional type of outputs that need to be controlled, the run-end outputs. The main difficulty is that these outputs are typically only available at the end of the run.

Though a lot of work has been reported recently in the literature on batch process control and optimization (Abel *et al.* 2000, Srinivasan *et al.* 2003, Flores-Cerrillo and MacGregor 2003, Chin

et al. 2004), there is still a lack of understanding of their system-theoretical properties. Due to the finite-time nature of batch processes, the standard definitions of properties such as stability, controllability and observability cannot be used.

This paper presents definitions and analysis tools for the two important properties of stability and controllability for batch processes. It is important to emphasize that the contribution of this paper is in discussing the various notions of stability and controllability and choosing the right notions for the analysis of batch processes. The analysis tools are then readily adapted from those existing in the literature.

The paper is organized as follows. Section 2 introduces a brief mathematical description of batch processes and discusses the implications of two time scales and two types of output for control. Stability and controllability are analyzed in Sections 3 and 4, respectively. An illustrative example is presented in Section 5, and conclusions are drawn in Section 6.

2. CONTROL OF BATCH PROCESSES

A batch process can be seen as a repetitive dynamical process that is characterized by the presence of a finite terminal time and thus the possibility of having several sequential runs, with each run being dynamic. Batch processes have the following main characteristics: (i) There are two time scales, i.e. the continuous time t within the run and the discrete run index k, (ii) the time of a run is limited (finite), (iii) there is no steady-state operating point with respect to t, i.e. the analysis has to be performed around trajectories rather than an equilibrium point, and (iv) two types of measurements are available, i.e. during the run and at the end of the run.

2.1 Terminology and notations

Let $\mathbb{R}$ be used for the space of real numbers and $\mathbb{L}$ for that of functions, and let $\mathbb{Z}_+$ represent the set of positive integers excluding zero. The various elements of a batch process can be defined as follows:

(1) *Run:* One realization of a repetitive process.
(2) *Run time:* The time within a run, $t \in [0, T] \subset \mathbb{R}_+$, where T is the finite terminal time.
(3) *Run index:* The number of a run, $k \in \mathbb{Z}_+$.
(4) *Inputs:* The inputs, $u_k(t) \in \mathcal{U} \subset \mathbb{R}^m$, evolve with t during run k. The input trajectories for run k are denoted by $u_k[0, T] \in \mathbb{L}^m$.
(5) *States:* The states, $x_k(t) \in \mathcal{X} \subset \mathbb{R}^n$, evolve with t during run k. x_k^{ic} are the initial conditions at time $t = 0$. The corresponding state trajectories are denoted by $x_k[0, T] \in \mathbb{L}^n$.
(6) *Outputs:* The outputs are of two types: (i) The run-time outputs, $y_k(t) \in \mathbb{R}^p$, correspond to the on-line measurements during run k; (ii) the run-end outputs, $z_k \in \mathbb{R}^q$, include the measurements that become available at the end of run k. The latter might also depend on the state evolution during the entire run, e.g. the average value of a state.
(7) *System dynamics:* They describe the state and output evolutions for a single run. For example, the nonlinear time-invariant model describing the process behavior during run k reads:

$$\dot{x}_k(t) = F(x_k(t), u_k(t)), \quad x_k(0) = x_k^{ic} \tag{1}$$
$$y_k(t) = H(x_k(t), u_k(t)) \tag{2}$$
$$z_k = \mathcal{H}(x_k[0, T], u_k[0, T]) \tag{3}$$

The dynamics over several runs stem from the possibility to update the initial conditions and the inputs on a run-to-run basis.

The system properties will be analyzed around selected reference trajectories, for which the accent $\bar{(\cdot)}$ will be used. For example, the reference state trajectories will be denoted by $\bar{x}[0, T]$, with $\bar{x}(t)$ being the corresponding state values at time t. Perturbations denoted by $\Delta(\cdot)$ will be considered, e.g. $\Delta\bar{x}[0, T]$ is a perturbation of $\bar{x}[0, T]$.

2.2 Control strategies

There are two types of control objectives (run-time outputs $y_k(t)$ or $y_k[0, T]$, and run-end outputs z_k), and also different ways of reaching them (on-line with $u_k^{on}(t)$ and run-to-run with $u_k^{rtr}[0, T]$). Each objective can be met either on-line or on a run-to-run basis, this choice being dependent on the type of measurements available. The control strategies are classified in Figure 1 and discussed next.

Implementation aspect	*Control objectives*	
	Run-time outputs $y_k(t)$ or $y_k[0,T]$	Run-end outputs z_k
On-line	1 On-line control $u_k^{on}(t) \rightarrow y_k(t) \rightarrow y_k[0,T]$ PID	2 Predictive control $u_k^{on}(t) \rightarrow z_{pred,k}(t)$ MPC
Run-to-run	3 Iterative learning control $u_k^{rtr}[0,T] \rightarrow y_k[0,T]$ ILC	4 Run-to-run control $\mathcal{U}(\pi_k) = u_k^{rtr}[0,T] \rightarrow z_k$ R2R

Fig. 1. Control strategies resulting from consideration of the control objectives (run-time or run-end outputs) and the implementation aspect (on-line or run-to-run).

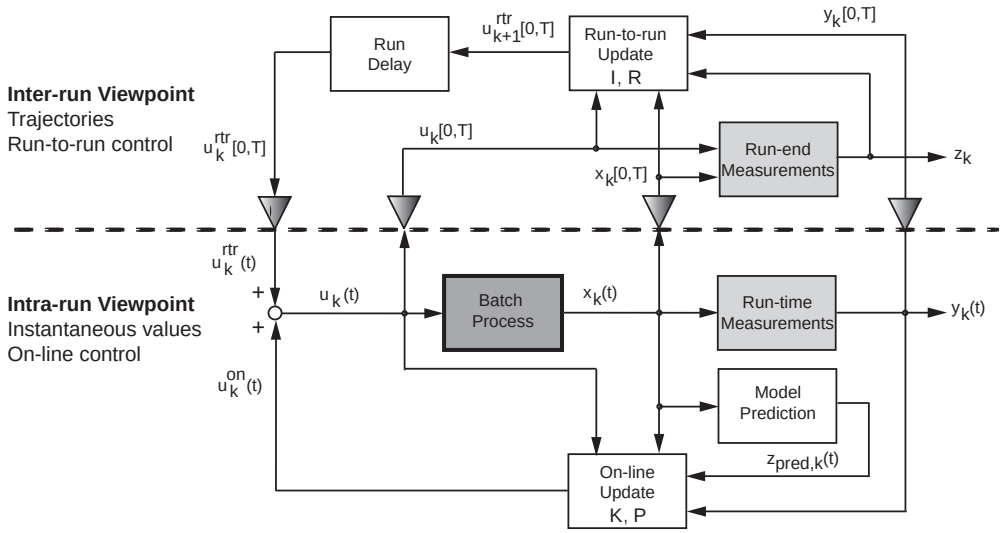


Fig. 2. Batch process with the inputs being updated both on-line (intra-run, use of the run-time measurements $y_k(t)$) and on a run-to-run basis (inter-run, use of the run-end measurements z_k). The symbol ∇ is used to indicate a change in viewing the time argument, e.g. from a trajectory to an instantaneous value when going downward and conversely when going upward.

- *On-line control of run-time outputs.* The approach is similar to that used in the traditional control literature. Control is typically done using PID techniques or more sophisticated alternatives whenever necessary. Formally, this controller can be written as

$$u_k^{on}(t) = \mathcal{K}(y_k(t), y_{sp}(t)) \tag{4}$$

  where $\mathcal{K}$ is the on-line controller for the run-time outputs $y_k(t)$, and $y_{sp}(t)$ the setpoint.
- *On-line control of run-end outputs.* It is necessary here to predict the run-end outputs based on measurement of the run-time outputs. Model predictive control (MPC) is well suited to that task (Nagy and Braatz 2003). The controller can be written as

$$u_k^{on}(t) = \mathcal{P}(z_{pred,k}(t), z_{sp}) \tag{5}$$

  where $\mathcal{P}$ is the on-line controller for the run-end outputs z_k, and $z_{pred,k}(t)$ the prediction of z_k available at time instant t.
- *Run-to-run control of run-time outputs.* In batch processing, key process characteristics such as process gain and time constants can vary considerably. Hence, the need to provide adaptation in a run-to-run manner to compensate the effect of these variations.

  The run-to-run part of the manipulated variable profiles can be generated using Iterative Learning Control (ILC) that exploits information from previous runs (Moore 1993). The controller has the structure

$$u_k^{rtr}[0,T] = \mathcal{I}(y_{k-1}[0,T], y_{sp}[0,T]) \tag{6}$$

  where $\mathcal{I}$ is the iterative learning controller for the run-time outputs $y_k[0,T]$. It processes the entire profile of the previous run to generate the entire manipulated profile for the current run.
- *Run-to-run control of run-end outputs.* The input profiles are parameterized using the input parameters π_k, $u_k^{rtr}[0,T] = \mathcal{U}(\pi_k)$. Control is then implemented using simple discrete integral control laws, that is $\pi_k = \pi_{k-1} + K(z_{sp} - z_{k-1})$ (Francois *et al.* 2005). Formally, the controller can be written as

$$u_k^{rtr}[0,T] = \mathcal{U}(\pi_k),\ \pi_k = \mathcal{R}(z_{k-1}, z_{sp}) \tag{7}$$

  where $\mathcal{R}$ is the run-to-run controller for the run-end outputs z_k, and $\mathcal{U}$ the input parametrization.

Note that, except for predictive control that involves prediction, all the other control schemes use only measurements and thus do not necessitate a process model for implementation, i.e. a very nice feature for batch processes, where detailed accurate models are seldom available (Bonvin 1998).

By combining strategies for the various types of outputs, the control inputs can have contributions from both run-to-run and on-line updates:

$$u_k(t) = u_k^{rtr}(t) + u_k^{on}(t) \tag{8}$$

The term $u_k^{rtr}(t)$ stems from the trajectories $u_k^{rtr}[0,T]$ and represent the 'feedforward' operating policies that are not altered within a run. However, $u_k^{rtr}[0,T]$ may change between runs (via run-to-run update), leading to inter-run dynamics. On the other hand, $u_k^{on}(t)$ represents the 'feedback' correction during the run (via on-line update). This combination of strategies is illustrated in Figure 2.

Applying only run-to-run control exhibits the limitations of being open-loop in run time, in particular for run-time disturbances. In general, a combination of these four strategies is used. However, in such a combined scheme, care should be taken that the on-line and run-to-run corrective actions do not oppose each other. Hence, the stability issue is critical.

In formulating the control strategy, controllability is important since it informs whether or not open-loop inputs exist that can provide the desired performance. Once a controller is designed, stability issues are of uppermost importance. Stabilization (and more appropriately finite-time stabilization), which is the issue of designing a controller that achieves stability and desired performance, will not be addressed in this paper.

3. INTRA- AND INTER-RUN STABILITY

Due to the presence of the two time scales t and k, both intra-run (in run time t) and inter-run (in run index k) stability need to be addressed.

3.1 Intra-run stability

Stability in run time t is important for repeatability and reproducibility reasons. The problem addressed therein is whether the trajectories of various runs with initial conditions sufficiently close will remain close during the rest of the run.

System (1) under on-line closed-loop operation using the feedback law (4) or (5) can be written as:

$$\dot{x}_k(t) = \tilde{F}(x_k(t), t), \qquad x_k(0) = x_k^{ic} \qquad (9)$$

The standard definition of Lyapunov stability is typically used around an equilibrium point (Vidyasagar 1978). To extend this definition to finite-time systems without an equilibrium point, it is first necessary to introduce the concept of a tube around the nominal trajectory in the $(n+1)$-dimensional space of states and time.

Definition 1. The trajectories $x_k[0,T]$ are defined to be inside the (a,b)-tube $\mathcal{B}_{a,b}$ around the reference trajectories $\bar{x}[0,T]$, i.e. $x_k[0,T] \in \mathcal{B}_{a,b}$, if they satisfy $\|x_k(t) - \bar{x}(t)\| < ae^{bt}$, $\forall\, t \in [0,T]$.

The tube consists of a ball of radius a in the n-dimensional state space at time $t = 0$, which shrinks or expands with time at a rate determined by b.

Definition 2. System (9) is locally intra-run β-**tube stable** around the trajectories $\bar{x}[0,T]$ if there exists a $\delta > 0$ such that, for all $x_k^{ic} = \bar{x}(0) + \Delta\bar{x}(0)$ with $\|\Delta\bar{x}(0)\| < \delta$, the state evolution $x_k[0,T] \in \mathcal{B}_{\delta,\beta}$.

A diverging (converging) system has a positive (negative) value of β. Note that a system that initially diverges to eventually converge has a positive β. In addition to its sign, the value of β is quite useful since, with finite-time systems, the dividing line between stability and instability is not whether the trajectories converge or diverge, but by how much they come together or grow apart in the interval of interest. Hence, in the context of batch processes, stability is not a yes-no result, but rather a measure quantified by β.

Definition 3. System (9) is locally intra-run α-**terminal-time stable** around the trajectories $\bar{x}[0,T]$ if there exists a $\delta > 0$ such that, for all $x_k^{ic} = \bar{x}(0) + \Delta\bar{x}(0)$ with $\|\Delta\bar{x}(0)\| < \delta$, the terminal states statisfy $\|x_k(T) - \bar{x}(T)\| < \alpha\delta$.

Terminal-time stability is the counterpart of asymptotic stability for finite-time systems. Again, stability is not simply determined by whether α is greater or less than 1, but instead it is quantified by the value of α.

It is possible to give results similar to the two theorems of Lyapunov (one based on linearization and the other on the existence of a non-increasing Lyapunov function) for tube stability.

Theorem 1. Let $\Delta\dot{x}_k(t) = A(t)\Delta x_k(t)$ with the initial conditions $\Delta x_k(0) = \Delta\bar{x}(0)$ be a bounded linearization of System (9) along $\bar{x}[0,T]$ for run k. Let $\sigma_{max}(t)$ be the maximum of the real parts of the eigenvalues of the time-dependent matrix $\frac{1}{t}\int_0^t A(\tau)d\tau$. Also, let $\bar{\sigma}_{max} = \max_t \sigma_{max}(t)$. Then, System (9) is tube stable around $\bar{x}[0,T]$ with $\beta = \bar{\sigma}_{max}$. Furthermore, the system is locally terminal-time stable around $\bar{x}[0,T]$ with $\alpha = e^{\sigma_{max}(T)T}$.

The proof of the theorem uses Bellman-Gronwall's Lemma (Vidyasagar 1978). Note that the eigenvalues of the integral of A are studied rather than the eigenvalues of A themselves. In most optimally operated finite-time systems (e.g. using a finite-time linear quadratic regulator), though the eigenvalues of the integral are negative, some of the eigenvalues of A might become positive toward the end of the run. This phenomenon caused by on-line control of z_k is referred to as the 'batch kick' in the optimization of batch processes. Intuitively, this means that little can go wrong toward the end since the 'time-to-go' is small.

Turning to the second Lyapunov method, the following result can be stated.

Theorem 2. Let $V(x,t) : \mathbb{R}^n \times \mathbb{R}_+ \rightarrow \mathbb{R}$ be a continuously differentiable function such that $V(\bar{x},t) = 0$ and $V(x,t) > 0$ for all $x(t) \neq \bar{x}(t)$, $\forall t$. If $\dot{V}(x,t) \leq \sigma(t)V(x,t)$ along the system trajectories for all $x(t) = \bar{x}(t) + \Delta\bar{x}(t)$, $\forall t$, $\|\Delta\bar{x}(t)\| < \delta$, then System (9) is tube stable with $\beta = \max_t \frac{1}{t}\int_0^t \sigma(\tau)\, d\tau$.

Note that the definition of stability presented by (Lohmiller and Slotine 1998) using contraction of deviations around pre-specified trajectories is a special case of Definition 2 above and requires contraction at every time instant, i.e. $\sigma(t) < 0$ for all t. This measure is clearly inadequate for batch systems that exhibit a batch kick. Information regarding the overall performance is better related to the integral of σ as given in Theorems 1 and 2 than to its instantaneous value.

3.2 Inter-run stability

The interest in studying stability in run index k arises from the necessity to guarantee convergence of run-to-run adaptation schemes. Here, the standard notion of stability applies as the independent variable k goes to infinity. The main conceptual difference with the stability of continuous processes is that 'equilibrium' refers to entire trajectories. Hence, the norms have to be defined in the space of functions $\mathbb{L}$ such as the integral squared error $\mathbb{L}_2$.

For studying stability with respect to run index k, System (1) is considered under closed-loop operation. At the k^{th} run, the trajectories of the $(k-1)^{st}$ run are known, which fixes $u_k^{rtr}[0,T]$ according to (6) or (7). These input profiles, along with the on-line feedback law (4) or (5), are applied to (1) to obtain $x_k(t)$ for all t and thus $x_k[0,T]$. All these operations can be represented formally as:

$$x_k[0,T] = \tilde{\mathcal{F}}(x_{k-1}[0,T]),\ x_0[0,T] = x_{init}[0,T] \quad (10)$$

where $x_{init}[0,T]$ are the initial state trajectories. Inter-run stability is considered around the equilibrium trajectory computed from (10), i.e. $\bar{x}[0,T] = \tilde{\mathcal{F}}(\bar{x}[0,T])$.

Definition 4. System (10) is locally inter-run Lyapunov stable around the equilibrium trajectories $\bar{x}[0,T]$ if there exist $\delta > 0$ and $\epsilon > 0$ such that, for all $x_0[0,T] = \bar{x}[0,T] + \Delta\bar{x}[0,T]$ with $\|\Delta\bar{x}[0,T]\| < \delta$, $\|x_k[0,T] - \bar{x}[0,T]\| < \epsilon, \forall\, k$. If, in addition, $\lim_{k\rightarrow\infty} \|x_k[0,T] - \bar{x}[0,T]\| = 0$, then the system is locally inter-run asymptotically stable.

This stability definition is fairly standard but in a discrete setting. Thus, in principle, either one of the two Lyapunov methods (via linearization or Lyapunov function) can be used to analyze stability. However, the linearization method has problems since differentiation has to be performed in the space of functions. The Lyapunov-function method can be used once a norm is appropriately defined (Vidyasagar 1978).

Theorem 3. Let $V : \mathbb{L}^n \rightarrow \mathbb{R}$ be a continuously differentiable functional such that $V(\bar{x}[0,T]) = 0$ and $V(x[0,T]) > 0$ for $x[0,T] \neq \bar{x}[0,T]$.

System (10) is locally inter-run Lyapunov stable if, for all $x_0[0,T] = \bar{x}[0,T] + \Delta\bar{x}[0,T]$ with $\|\Delta\bar{x}[0,T]\| < \delta$, $V(x_{k+1}[0,T]) \leq V(x_k[0,T])$, $\forall\ k$.

If, in addition, $\bar{x}[0,T]$ is the largest invariant set satisfying $V(x_{k+1}[0,T]) = V(x_k[0,T])$, then the system is locally inter-run asymptotically stable.

Again, the choice of a Lyapunov function is a major difficulty. The norm of the input error $\|u[0,T] - \bar{u}[0,T]\|_{\mathbb{L}_2}$ has served as a useful Lyapunov function in some of our studies, although the output error has been widely used in the literature.

4. CONTROLLABILITY OF RUN-TIME AND RUN-END OUTPUTS

One of the definitions of controllability for infinite-time dynamic systems requires that there exists an input vector $u[t_0,\tau]$ with which the equilibrium state can be reached from any arbitrary state $x(t_0)$ in the neighborhood of the equilibrium.

There are two difficulties with extending this definition to batch processes. Firstly, the controllability of finite-time systems needs to be defined around trajectories. Therein, the relevant question is whether or not some neighborhood of given trajectories can be reached. Clearly, not all state trajectories can be fixed independently because the state vector $x[0,T]$ contains a lot of redundant information. For example, since a position trajectory enforces the velocity, the trajectories of position and velocity cannot be chosen independently of each other [1]. Hence, only controllability in terms of *independent output* trajectories can be investigated (y-controllability).

Secondly, the above definition of controllability mentions the existence of a time τ, which however might be larger than the terminal time T. This aspect becomes important when considering the

[1] In contrast, when instantaneous values are considered, arbitrary position and velocity values can be specified.

controllability with respect to the run-end outputs (z-controllability).

Here, controllability addresses the problem of the existence of inputs that can implement the desired action and thus is independent of whether the correction is made on-line or on a run-to-run basis.

4.1 Controllability of run-time outputs

Let y_k^i, $i = \{1, \cdots, p\}$, be the i^{th} run-time output of System (1)-(2) and let its relative degree[2] be r^i, i.e. $\frac{\partial}{\partial u_k}\frac{d^j y_k^i}{dt^j} = 0$, $\forall j < r^i$.

Definition 5. System (1)-(2) is locally **y-controllable** around the arbitrary trajectories $\bar{y}[0,T]$ if there exists a $\delta > 0$ such that, for all $\|\Delta\bar{y}[0,T]\| < \delta$, $\Delta\bar{y}^i[0,T] \in \mathbb{C}^{(r^i-1)}$ for $i = \{1, \cdots, p\}$, there exists $u_k[0,T] \in \mathcal{U}$ that leads to $y_k[0,T] = \bar{y}[0,T] + \Delta\bar{y}[0,T]$.

Note that if the first $(r^i - 1)$ derivatives of $\Delta\bar{y}^i$ are discontinuous, Dirac impulses are required at the inputs to meet the outputs. Thus, the perturbations $\Delta\bar{y}^i$ that are considered cannot have discontinuities in their first $(r^i - 1)$ derivatives, i.e. $\Delta\bar{y}^i \in \mathbb{C}^{(r^i-1)}$, where $\mathbb{C}^r$ denotes the space of functions that have continuous derivatives up to order r.

Note also that the trajectories $\bar{y}[0,T]$ are assumed feasible, i.e. they respect the initial conditions and they can be implemented through $\bar{u}[0,T]$ (the condition under which $\bar{u}[0,T]$ exist for a given $\bar{y}[0,T]$ is not addressed here). The question asked in this definition regards only the neighboring trajectories. This is clearly a local inversion problem for which standard conditions for inverting a multi-input multi-output system can be used (Hirschorn 1979).

Theorem 4. Let u_k^j, $j = \{1, \cdots, m\}$, be the j^{th} input of System (1)-(2). Let the relative degrees r^i, $i = \{1, \cdots, p\}$, remain constant around $\bar{y}[0,T]$, and $\mathcal{M}(t)$ be defined as $\mathcal{M}_{i,j}(t) = \frac{\partial}{\partial u_k^j}\frac{d^{r^i} y_k^i}{dt^{r^i}}$. If $\mathcal{M}(t)$ is of rank p, $\forall\ t$, then System (1)-(2) is locally y-controllable around $\bar{y}[0,T]$.

4.2 Controllability of run-end outputs

A similar definition can be provided for system controllability in terms of reaching specified run-end outputs.

[2] The relative degree of an output is the minimal degree of its time derivative for which at least one input appears.

Definition 6. System (1,3) is locally **z-controllable**, from time t_0 on, around an arbitrary operating point $\bar{z}$ if there exists a $\delta > 0$ such that, for all $\|\Delta\bar{z}\| < \delta$, there exists $u_k[t_0,T] \in \mathcal{U}$ that leads to $z_k = \bar{z} + \Delta\bar{z}$.

Here, the notion of controllability is linked to a given time t_0. The question asked is the following: Is it possible to change the outcome of the run if, at time instant t_0 in the run, one wishes so? To answer this question, consider the linearization of System (1,3) around a trajectory, resulting in the linear time-varying system (Friedland 1986):

$$\Delta\dot{x}_k = A(t)\Delta x_k + B(t)\Delta u_k,\ \Delta x(t_0) = 0 \quad (11)$$

$$\Delta z_k = C(t)\Delta x_k \quad (12)$$

Theorem 5. Consider the output controllability Grammian $\mathcal{G}(t)$ for System (11)-(12):

$$P(\tau) = C(\tau)e^{\int_{t_0}^{\tau} A(\kappa)\ d\kappa}B(\tau)$$

$$\mathcal{G}(t_0) = \int_{t_0}^{T} P(\tau)P^T(\tau)\ d\tau \quad (13)$$

If $\mathcal{G}(t_0)$ is of rank q, then System (1,3) is locally z-controllable from time t_0 on.

For on-line control of run-end outputs, Theorem 5 can be used to indicate until what time t_0 in the batch the control of run-end outputs is feasible.

For run-to-run control of run-end outputs, it is important to study the case where the inputs are parameterized. Consider the parameterization $u_k[0,T] = \mathcal{U}(\pi_k)$, where $\pi_k \in \mathbb{R}^{n_\pi}$ are the input parameters. This way, the batch process can be seen as a static map between the input parameters π_k and the run-end outputs z_k. To assess controllability, the transfer matrix between π_k and z_k needs to be computed. The equivalent of Theorem 5 using input parametrization is given next.

Theorem 6. Consider the $q \times n_\pi$ transfer matrix between π and z calculated for System (11)-(12):

$$\mathcal{T}(t_0) = \int_{t_0}^{T} C(\tau)e^{\int_{t_0}^{\tau} A(\kappa)\ d\kappa}B(\tau)\frac{\partial\mathcal{U}}{\partial\pi}\ d\tau \quad (14)$$

If $\mathcal{T}(t_0)$ is of rank q, then System (1,3) with the parametrization $u_k[0,T] = \mathcal{U}(\pi_k)$ is locally z-controllable from time t_0 on.

Note that run-to-run control requires only the evaluation of the matrix $\mathcal{T}(0)$. The rank condition (or invertibility) of $\mathcal{G}$ or $\mathcal{T}$ follows from the

fact that the inputs that can create the necessary change in the run-end outputs are obtained by inversion. However, note that as t_0 approaches T, the Grammian approaches singularity, with $\mathcal{G}(T) = 0$. Similarly, if a piecewise parametrization is used, after a certain time, some of the parameters will have no influence on the outputs, thus making a few columns zero. As t_0 proceeds toward T, more and more columns will become zero. Hence, as $t \to T$, inverting $\mathcal{G}$ or $\mathcal{T}$ requires larger and larger inputs for control. Also rank deficiency may occur, and the system may lose controllability.

5. ILLUSTRATIVE EXAMPLE

Consider the scale-up, from the laboratory to production, of a semi-batch reactor in which several reactions take place. The desired and main side reactions are

$$A + B \to C, \qquad 2B \to D$$

with C the desired product and D an undesired side product. The reactions are fairly exothermic and the reactor is equipped with a jacket for heat removal. The control objective is twofold: (i) Operate isothermally at 50°C by manipulating the jacket temperature, and (ii) match the final concentrations that have been obtained in the laboratory, $c_B(T) = c_{B,max}$ and $c_D(T) = c_{D,max}$, by manipulating the feed rate of reactant B.

The control structure used is illustrated in Figure 3. It implements on-line feedback temperature control. In addition, the feedforward profile for the jacket temperature $T_j^{ff}[0,T]$ is adjusted on a run-to-run basis by means of ILC. In this case, $\mathcal{M} = \frac{d\dot{T}_r}{dT_j}$ is a constant non-zero scalar irrespective of the trajectory chosen (hence, satisfies y-controllability - Theorem 4). The controller reads

$$T_{j,k}(t) = T_{j,k}^{ff}(t) + K_R e_k(t) + \frac{K_R}{\tau_I} \int_0^t e_k(\tau) d\tau,$$

$$T_{j,k+1}^{ff}[0, T-\Delta] = T_{j,k}^{ff}[\Delta, T] + K_{ILC}\, e_k[\Delta, T],$$

with $e_k(t) = T_{r,ref}(t) - T_{r,k}(t)$, K_R the proportional gain and τ_I the integral time constant of the PI master controller. It can be easily verified that the system is tube stable with a negative β. K_{ILC} is the gain of the ILC controller and $\Delta \geq 0$ the value of the input shift. The second equation allows adapting the feedforward term for the jacket temperature setpoint on a run-to-run basis based on ILC with input shift. In Theorem 3, the integral squared output error $\int_0^T e_k^2(\tau)\ d\tau$ is used as the Lyapunov function in run index k. The value of the input shift is tuned for convergence (Welz *et al.* 2004). Due to the presence of the shift, the error does not converge asymptotically to zero.

In addition, the feed rate profile $u[0,T]$ is parameterized using the two feed-rate levels u_1 and u_2, each valid over half the batch time. The final concentrations $c_B(T)$ and $c_D(T)$ are met, on a run-to-run basis, by adjusting the two parameters $\pi = \{u_1, u_2\}$. The transfer matrix $\mathcal{T}$ is evaluated around the current operating point using (14), with $\frac{\partial \mathcal{U}}{\partial \pi} = [1\ 0]^T$ during the first half of the batch and $\frac{\partial \mathcal{U}}{\partial \pi} = [0\ 1]^T$ in the second half. With the matrix $\mathcal{T}$ being full rank (satisfies z-controllability - Theorem 6), the discrete integral control law reads

$$\pi_{k+1} = \pi_k + \mathcal{T}^+ K_{R2R}\,[z_{ref} - z_k], \qquad (15)$$

where $\mathcal{T}^+$ is the pseudo-inverse of $\mathcal{T}$, and K_{R2R} the gain of the run-to-run controller. The run-to-run convergence of this scheme can be shown using Theorem 3 with the squared input error $\|\pi - \pi^*\|^2$ as the Lyapunov function in run index k (Francois *et al.* 2005).

The evolution of the manipulated and controlled variables are illustrated in Figures 4.

6. CONCLUSIONS

The control of batch processes is characterized by run-time and run-end objectives on the one hand, and by actions that can be implemented on-line and on a run-to-run basis on the other. It has been shown that the concepts of stability and controllability, which are well understood for infinite-time systems operating around an equilibrium point, are not directly applicable to finite-time batch processes.

With regard to stability, the concept of tube stability, by which the state trajectories remain within a given tube, has been introduced. The special case of terminal-time stability has also been discussed. Two theorems that help evaluate tube stability have been proposed.

As for controllability with respect to specified trajectories, it was observed that the entire state space cannot be studied due to the fact that there is considerable redundancy in the state trajectories. Hence, only controllability with respect to two types of outputs have been addressed. Controllability was studied from the point-of-view of inversion, and results were adapted from the existing literature.

REFERENCES

Abel, O., A. Helbig, W. Marquardt, H. Zwick and T. Daszkowski (2000). Productivity optimiza-

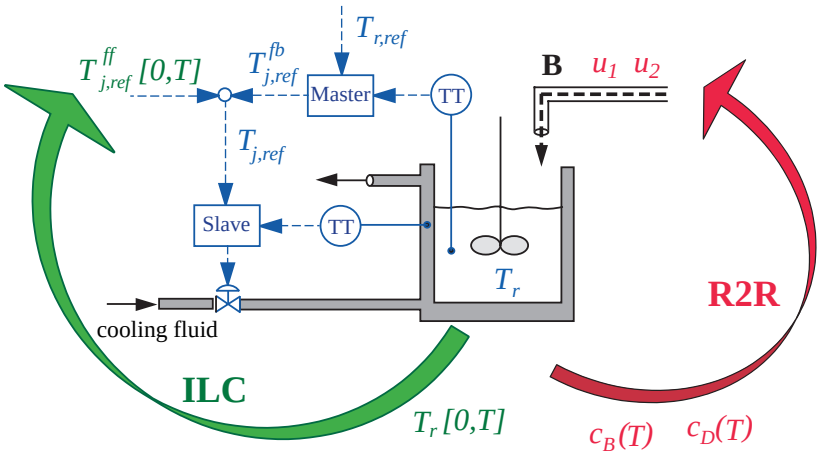


Fig. 3. On-line and run-to-run strategies to control the reactor temperature and the final concentrations.

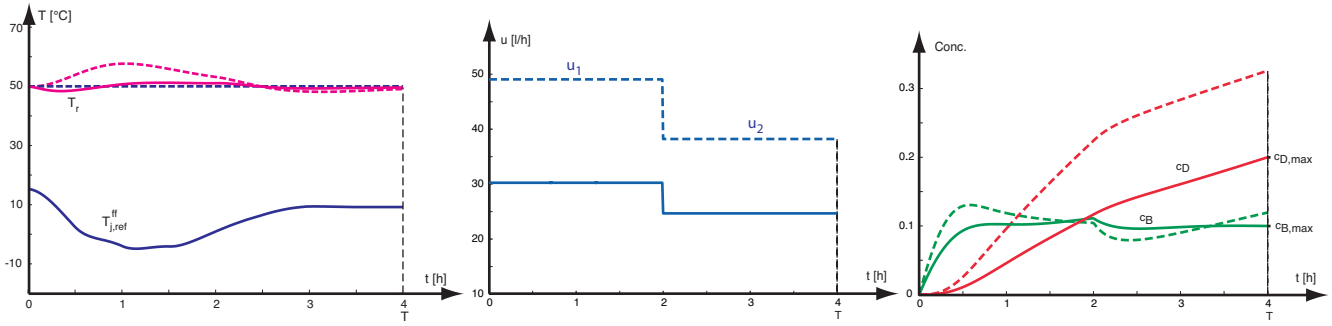


Fig. 4. Evolution of the reactor and jacket temperatures (left), of the feed rate (middle) and of the concentrations c_B and c_D (right), initially (dotted lines) and after 3 iterations (solid lines).

tion of an industrial semi-batch polymerization reactor under safety constraints. *J. Process Contr.* **10**(4), 351–362.

Bonvin, D. (1998). Optimal operation of batch reactors - A personal view. *J. Process Contr.* **8**(5–6), 355–368.

Campbell, W.J., S.K. Firth, A.J. Toprac and T.F. Edgar (2002). A comparison of run-to-run control algorithms. In: *American Control Conference*. Anchorage, Alaska. pp. 2150–2155.

Chin, I.S., S.J. Qin, K.S. Lee and M. Cho (2004). A two-stage iterative learning control technique combined with real-time feedback for independent disturbance rejection. *Automatica* **40**(11), 1913–1922.

Flores-Cerrillo, J. and J.F. MacGregor (2003). Within-batch and batch-to-batch inferential-adaptive control of semibatch reactors: A partial least squares approach. *Ind. Eng. Chem. Res.* **42**, 3334–3335.

Francois, G., B. Srinivasan and D. Bonvin (2005). Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty. *J. Process Contr.* **15**(6), 701–712.

Friedland, B. (1986). *Control System Design – An Introduction to State-Space Methods*. McGraw-Hill, New York.

Hirschorn, R.M. (1979). Invertibility of multivariable nonlinear control systems. *IEEE Trans. automat. Contr.* **24**, 855–865.

Lohmiller, W. and J.J.E. Slotine (1998). On contraction analysis for nonlinear systems. *Automatica* **34**(6), 683–696.

Moore, K.L. (1993). *Iterative Learning Control for Deterministic Systems*. Springer-Verlag, Advances in Industrial Control, London.

Nagy, Z.K. and R.D. Braatz (2003). Robust nonlinear model predictive control of batch processes. *AIChE Journal* **49**(7), 1776–1786.

Srinivasan, B., D. Bonvin, E. Visser and S. Palanki (2003). Dynamic optimization of batch processes: II. Role of measurements in handling uncertainty. *Comp. Chem. Eng.* **44**, 27–44.

Vidyasagar, M. (1978). *Nonlinear Systems Analysis*. Prentice-Hall, Englewood Cliffs.

Welz, C., B. Srinivasan and D. Bonvin (2004). Iterative learning control with input shift. In: *IFAC Symp. DYCOPS-7*. Boston, MA. pp. 187–192.

Session 4.1

Biomedical Systems Modeling, Analysis and Control

Identification of Linear Dynamic Models for Type 1 Diabetes: A Simulation Study

D. A. Finan and D. E. Seborg
University of California, Santa Barbara

Dynamic Modeling of Exercise Effects on Plasma Glucose and Insulin Levels

A. Roy and R. S. Parker
University of Pittsburgh

Pathways for Optimization-Based Drug Delivery Systems and Devices

L. Bleris, P. Vouzis, M. V. Arnold and M. V. Kothare,
Lehigh University

Flexible Run-to-Run Strategy for Insulin Dosing in Type 1 Diabetic Subjects

C. C. Palerm, H. Zisser, L. Jovanovic and F. J. Doyle, III
University of California, Santa Barbara

Nonlinear Model Predictive Control for Optimal Discontinuous Drug Delivery

N. Hudon, M. Guay, M. Perrier and D. Dochain
Queen's University

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

IDENTIFICATION OF LINEAR DYNAMIC MODELS FOR TYPE 1 DIABETES: A SIMULATION STUDY

Daniel A. Finan[†] Howard Zisser[‡] Lois Jovanovic[‡]
Wendy C. Bevier[‡] Dale E. Seborg[†]

[†] *Department of Chemical Engineering*
University of California, Santa Barbara
[‡] *Sansum Diabetes Research Institute*
Santa Barbara, CA

Abstract: Models of type 1 diabetes with accurate prediction capabilities can help to achieve improved glycemic control in diabetic patients when used in a monitoring or model predictive control framework. In this research, empirical models are identified from a simulated physiological model. ARX and Box–Jenkins models of various orders are investigated and evaluated for their description of calibration and validation data that are characteristic of normal operation. In addition, model accuracy is determined for abnormal situations, or "faults." The faults include changes in model parameters (insulin sensitivities), an insulin pump occlusion, underestimates in the carbohydrate content of meals, and mismatches between the actual and patient–reported timing of meals. The models describe normal operating conditions accurately, and can also detect significant faults.

Keywords: Autoregressive models, biomedical systems, computer simulation, dynamic models, identification, stochastic modelling

1. INTRODUCTION

Diabetes mellitus is a disease characterized by insufficient production of insulin by the pancreatic β-cells, leading to prolonged elevated concentrations of blood glucose (Ashcroft and Ashcroft, 1992). Type 1 diabetics in particular rely on exogenous insulin for survival. This exogenous insulin typically enters the body in the subcutaneous tissue. A slow, constant or *basal* infusion helps the body metabolize glucose in times of fasting. Rapid or *bolus* injections complement the basal and are administered coincidentally with a meal to help the body metabolize large loads of carbohydrates (CHO).

Over the past few decades, many dynamic models have been formulated to describe glucose–insulin interactions in type 1 diabetes. The development of such models is relevant to a model predictive control (MPC) approach to diabetes, in which past outputs (*i.e.*, glucose measurements), past inputs (*i.e.*, insulin infusion rates), and model predictions are used to determine the appropriate insulin infusion rate at the current sampling instant (Bequette, 2005).

Physiological diabetes models include the widely used "minimal model" of Bergman *et al.* (1981), which was developed to estimate insulin sensitivity from an intravenous glucose tolerance test. The model describes glucose–insulin dynamics with only three differential equations and

few parameters. This simplicity, however, begets many limitations. For instance, the original model excluded exogenous insulin infusion as an input. Although it has been easily altered to include this input (Parker and Doyle III, 2001), the modified minimal model still does not include the dynamics of subcutaneous insulin infusion. A more recent model by Cobelli *et al.* (1998) is significantly more detailed than the model of Bergman *et al.* (1981), but its details are thus far unpublished. A physiologically rigorous 19–state model was developed by Sorensen (1985) to describe glucose–insulin pharmacokinetics/pharmacodynamics, and includes compartments representative of various bodily organs. Shortcomings of this model include its inability to capture the realistic hyperglycemic extremes characteristic of type 1 diabetes (Lynch and Bequette, 2002). A model developed by Hovorka (Hovorka *et al.*, 2002; Hovorka *et al.*, 2004) presents an attractive tradeoff between simplicity and accuracy. This model is the focus of the current research.

In addition to physiological diabetes models, empirical diabetes models have also been reported, although to a much lesser extent. Autoregressive models have been used to predict the next glucose value (ten minutes ahead) from previous glucose measurements (Desai *et al.*, 2002). The novelty of the current paper is that different types of models are considered (namely ARX and Box–Jenkins) and only "infinite–step ahead" model predictions are evaluated. That is, the empirical models predict future outputs based only upon the process inputs and the previous empirical model outputs; thus, the actual outputs (from the Hovorka model) are not used to update the empirical model predictions.

2. PHYSIOLOGICAL MODEL

The diabetes model considered in this research is the model reported by Hovorka *et al.* (2004) and extended by Wilinska *et al.* (2005). The model inputs are the rate of subcutaneously infused insulin lispro (fast acting insulin), and meal amount and time. The output is the plasma glucose concentration. The model is comprised of three subsystems representing plasma glucose, subcutaneous and plasma insulin, and insulin action. The glucose subsystem is divided into two compartments, a plasma compartment and a "non–accessible" compartment; subcutaneous insulin absorption is also partitioned into two compartments. The insulin action subsystem takes into account the physiological effects of insulin on glucose transport, removal, and endogenous production. These insulin "actions" manifest themselves in the form of time–varying rate constants corresponding to each of these metabolic processes. Model "constants" were taken to be those quantities which were difficult to identify, while model "parameters" were *a priori* identifiable. Nonlinearity arises in the model not only from the insulin actions but also from physiological saturation effects. For example, renal glucose excretion is zero below a certain threshold (160 mg/dL) and insulin–independent peripheral glucose uptake is constant above, and proportional to glucose concentration below, another threshold (80 mg/dL). The model also includes gut absorption dynamics which describe the appearance of glucose in the blood resulting from a meal. The model's subcutaneous insulin absorption subsystem includes parallel fast and slow channels as well as a degree of insulin degradation at the injection site.

Figure 1 shows the steady–state map of plasma glucose concentration (G) and insulin infusion rate (u) predicted by the Hovorka model for three different patient weights. Three operating regions are evident in the model. In Region 1 ($G \geq 160$ mg/dL), renal glucose excretion is present and proportional to G; in Regions 2 and 3 ($G \geq 80$ mg/dL), non–insulin–dependent glucose uptake is constant; and in Region 3 ($G < 80$ mg/dL) non–insulin–dependent glucose uptake is proportional to G. Figure 1 indicates that the model produces negative glucose concentrations for high insulin infusion rates. Although negative glucose concentrations are unrealistic, the model is intended to be operated at physiological glucose levels, *e.g.*, 40 mg/dL $< G <$ 400 mg/dL.

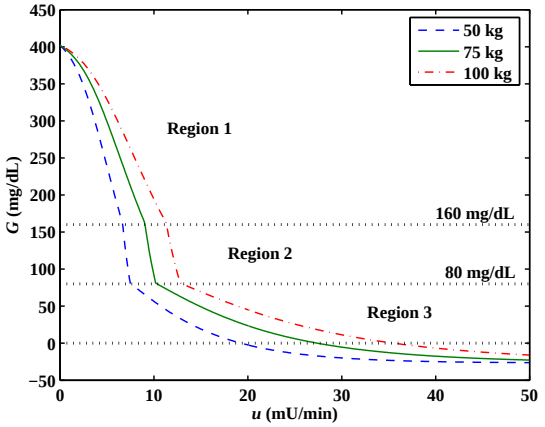


Fig. 1. Steady–state G–u map for three patient weights. Regions 1-3 represent different operating regions for the model.

Transient responses to open–loop changes in the insulin infusion rate u were simulated in order to characterize the insulin–to–glucose dynamics of the Hovorka model. Figure 2 shows the responses of G to step changes in u (*i.e.*, basal changes) and an impulse in u (*i.e.*, a bolus). The step

and impulse magnitudes were chosen so that the process operated entirely within Region 2 (see Figure 1), the most physiologically significant region. Figure 2 indicates that ∼75 h are required for G to reach steady state in response to the step decrease in u, but only about half this time is required to reach steady state after the step increase in u.

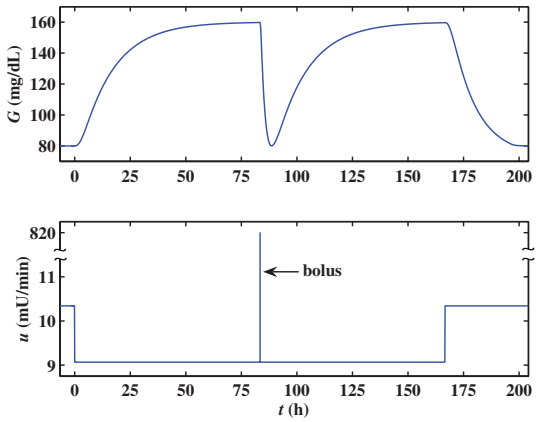


Fig. 2. Transient glucose responses for a 75 kg patient to open–loop step (basal) and impulse (bolus) inputs in u. Note broken scale in the input plot.

Transient responses to a meal disturbance were simulated in order to characterize the *postprandial* (*i.e.*, post–meal) glucose concentration dynamics. Figure 3 shows two responses to a 40 g CHO meal: an open–loop response for which no counteracting bolus is delivered, and the response when an appropriate bolus is delivered coincident with the meal. The rate of appearance of glucose in the blood from the meal U_G is the prediction of the model's gut absorption subsystem. Since using impulse inputs for both boluses and meals could cause identifiability problems, the input for a meal is considered to be U_G. The insulin–to–carbohydrate ratio for the bolus was determined by trial and error such that the bolus significantly reduced the postprandial peak and returned G to its steady–state value quickly, without significant undershoot.

2.1 Normal Operation

In order to simulate days of normal operation, certain assumptions had to be made regarding what is "normal." All runs simulated a 24 h period, starting at 8 AM and ending the following day at 8 AM. The sampling period was 5 min, a realistic interval for the current generation of continuous glucose sensors. Gaussian noise was added to the glucose measurement, with a standard deviation of $\sigma = 3.3$ mg/dL. Breakfast, lunch, and dinner were administered at 8 AM, 12 PM, and 6 PM,

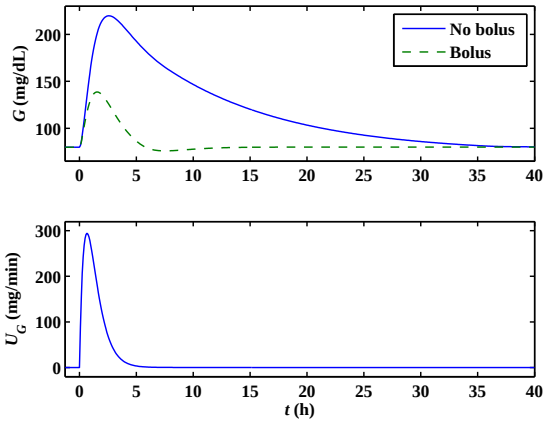


Fig. 3. Transient glucose responses for a 75 kg patient to meal input U_G, with and without a coincident bolus.

respectively. All normal runs used the nominal insulin sensitivities reported by Hovorka *et al.* (2002) and Hovorka *et al.* (2004). The patient weight was 75 kg. Three "normal" datasets were simulated corresponding to an *average meal day* (N_A), a *light meal day* (N_L), and a *heavy meal day* (N_H). In the light meal day, the simulated patient consumes 50% less CHO in each meal compared to the average meal day; in the heavy meal day, the patient consumes 50% more CHO in each meal compared to the average meal day. The insulin–to–carbohydrate ratio was constant for each meal, and was determined as described above.

2.2 Faults

The four realistic "faults" in Table 1 were simulated, representing changes in model parameters (insulin sensitivities), an insulin pump occlusion, patient underestimates of the amount of CHO consumed in meals, and mismatches between the actual timing of meals and that which the simulated patient reports to a hypothetical monitoring system. The same Gaussian noise level, meal times, and patient weight were used in the fault datasets as in the normal datasets.

Table 1. Fault descriptions.

Fault	Description
$F1$	50% reduction in insulin sensitivities
$F2$	100% occlusion of insulin pump for one hour (not during a meal)
$F3$	50% underestimate in CHO content of lunch and dinner
$F4$	15 min mismatch between actual and patient–reported lunch and dinner times

For the $F1$ fault simulation, an appropriate basal rate and insulin–to–carbohydrate ratio were recalculated to compensate for the decreased insulin

sensitivities. For $F2$, the basal insulin infusion rate was completely stopped for one hour, from 12 AM to 1 AM. The u input data used to generate the empirical model prediction, however, do not include this hour–long cessation in the basal. Here the assumption is that an online monitoring system would not be aware of the occlusion, and thus would be challenged to infer it from the available input–output data. Faults $F1$ and $F2$ were generated using the meal magnitudes for a normal, average meal day. For $F3$, lunch and dinner were two times larger than average and were taken with coincident boluses appropriate for the average–sized meals (*i.e.*, the patient underbolused for lunch and dinner). Since the patient underestimated these actual meal amounts by 50%, both the u and U_G input data used to generate the empirical model prediction for $F3$ are the same as for a normal average meal day. For fault $F4$, the lunch and dinner boluses were taken at their nominal times, but the meals were taken 15 min late (*i.e.*, after the boluses). Again, the assumption here is that an online monitoring system would not be aware of these mismatches in meal timing, and thus would be challenged to infer them. Therefore, the input data used to generate the empirical model prediction for $F4$ were the same as for a normal average meal day.

3. EMPIRICAL MODELS

The two types of linear dynamic models investigated in this research are autoregressive models with exogenous input (ARX) and Box–Jenkins (BJ) models. The MATLAB System Identification Toolbox (Ljung, 2005) was used to identify the models. The ARX model is a difference equation in which the current output depends on previous outputs and inputs,

$$\begin{aligned} A(q^{-1})G(t) = B_1(q^{-1})u(t) \\ + B_2(q^{-1})U_G(t) + \varepsilon(t) \end{aligned} \quad (1)$$

where q^{-1} is the backward shift operator (*i.e.*, $q^{-1}G(t) = G(t-1)$). A is a scalar polynomial in ascending powers of q^{-1}, starting with $q^0 = 1$ and B_1 and B_2 are scalar polynomials in ascending powers of q^{-1}, starting with q^{-1}. The ε term represents the Gaussian process noise. Low–order and high–order ARX models were identified from the Hovorka model simulation data. The "low–order" models were second order in the autoregressive and exogenous inputs; the "high–order" models were chosen according to the Akaike information criterion (AIC), which chooses model orders based on a compromise between model simplicity and model accuracy.

The BJ model is a transfer function model that models both deterministic inputs (*i.e.*, u and U_G) and stochastic inputs (*i.e.*, the noise ε) according to

$$\begin{aligned} G(t) = \frac{B_1(q^{-1})}{F_1(q^{-1})}u(t) \\ + \frac{B_2(q^{-1})}{F_2(q^{-1})}U_G(t) + \frac{C(q^{-1})}{D(q^{-1})}\varepsilon(t) \end{aligned} \quad (2)$$

where B_1, B_2, C, D, F_1, and F_2 are scalar polynomials in ascending powers of q^{-1}, starting with q^{-1}. Again, low–order and high–order BJ models were identified from Hovorka model simulation data. The "low–order" BJ models were first order in all inputs while the "high–order" models were fourth order.

For the identification studies, deviation variables were used. Because the physiological model does not account for the diurnal variations in insulin sensitivities, the basal insulin infusion rate was constant for the entire day. Since this steady–state infusion rate was subtracted in forming deviation variables, the input u consisted only of impulses (*i.e.*, boluses). The steady–state value of $G = 80$ mg/dL was subtracted from the output to give the deviation variable ΔG.

4. SIMULATION RESULTS

Model accuracy for both calibration and validation data was quantified by the standard coefficient of determination,

$$R^2 = \left(1 - \frac{\sum_{i=1}^{N}(G_i - \hat{G}_i)^2}{\sum_{i=1}^{N}(G_i - \bar{G})^2}\right) \times 100\% \quad (3)$$

where N is the number of samples, G is the output simulated by the Hovorka model, $\hat{G}$ is the output predicted by the identified model, and $\bar{G}$ is the average of the output simulated by the Hovorka model.

Low–order and high–order ARX and BJ models were identified from each of the three datasets representative of normal operation (*i.e.*, N_A, N_L, and N_H). Each of the twelve identified models was validated on the other two normal datasets. Figure 4 compares the predictions of the high–order models identified from N_A with all three normal datasets. The corresponding R^2 values are listed in Table 2 and range from 46-77%.

The R^2 values of the high–order ARX and BJ models for the three normal datasets are shown in Table 2. Both types of models predict their calibration data accurately ($R^2_{cal} \geq 66\%, \bar{R}^2_{cal} = 74.5\%$) In general, the BJ models consistently

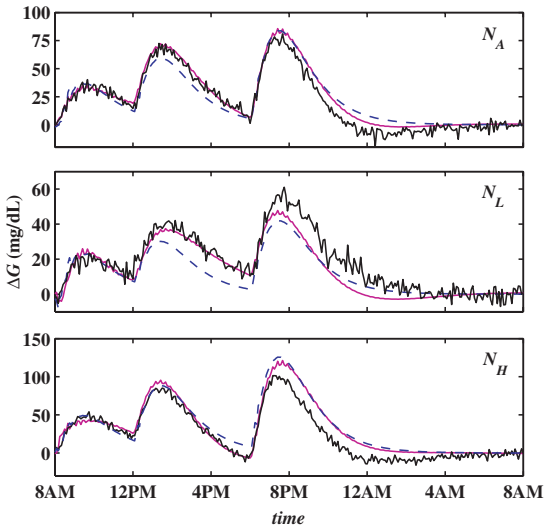


Fig. 4. Comparison of Hovorka model responses and predictions of high–order models identified from N_A. Top: calibration data; middle (N_L) and bottom (N_H): validation data. BJ: thin solid line; ARX: dashed line. Note different scales in the y–axes.

explain more variability in the data than the ARX models. The worst prediction occurs when the ARX model identified from N_L is validated on N_H. Here $R^2 = 0$, implying that the model prediction is no more accurate than the average value of the simulation data.

Table 2. R^2 values of predictions of high–order models identified from N_A for all normal datasets. Boldface values denote results for calibration data.

Day	Model	N_A	N_L	N_H
N_A	ARX	**66**	46	57
	BJ	**77**	57	63
N_L	ARX	32	**74**	0
	BJ	52	**78**	42
N_H	ARX	62	27	**74**
	BJ	71	68	**78**

The R^2 values of the low–order ARX and BJ models for the three normal datasets are shown in Table 3. The low–order model fits are comparable to those of the high–order models ($\bar{R}^2_{high} = 56.9; \bar{R}^2_{low} = 59.7$).

The models identified from the N_A dataset were evaluated on the fault datasets simulated by the Hovorka model. The predictions of the identified high–order models are shown in Figure 5, and the corresponding R^2 values are listed in Table 4. Very low R^2 values imply that the identified model is not accurate and that an abnormal situation (*i.e.*,

Table 3. R^2 values of predictions of low–order models identified from N_A for all normal datasets. Boldface values denote results for calibration data.

Day	Model	N_A	N_L	N_H
N_A	ARX	**66**	43	58
	BJ	**71**	70	79
N_L	ARX	38	**70**	7
	BJ	62	**78**	58
N_H	ARX	61	27	**72**
	BJ	71	66	**80**

fault) has occurred. The high–order ARX model easily detects $F3$, while the high–order BJ model detects $F1$ and $F3$. Although both identified models still explain ∼50% of the variance in $F2$, these R^2 values are significantly lower than the corresponding calibration values (see Table 2). It should also be noted that most of the unexplained variance comes in the last ∼6 h of the $F2$ run, *i.e.*, after the pump occlusion occurs at 12 AM. In an online monitoring situation, recent data would typically be weighted more heavily, and this fault would likely be detected soon after it occurs. Finally, $F4$ goes undetected, illustrating a degree of insensitivity of the identified models. A reasonable amount of insensitivity is advantageous because the detection of an insignificant fault situation is undesirable.

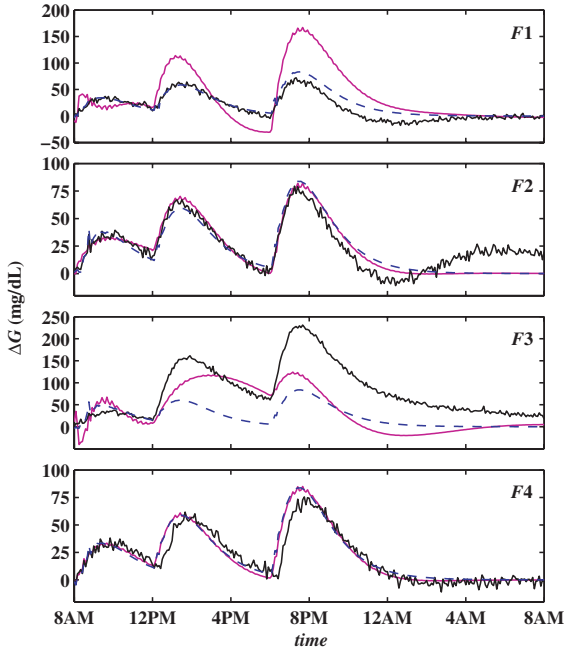


Fig. 5. Validation results for high–order models identified from N_A for $F1$ (top), $F2$ (middle top), $F3$ (middle bottom), and $F4$ (bottom). BJ: thin solid line; ARX: dashed line. Note different scales in the y–axes.

Table 4. R^2 values of predictions of all models identified from N_A, evaluated for the fault datasets.

Order	Model	$F1$	$F2$	$F3$	$F4$
High	ARX	53	48	-20	61
Order	BJ	-63	51	-2	64
Low	ARX	54	47	-21	60
Order	BJ	68	52	61	63

The R^2 values of the low–order ARX and BJ models identified from N_A evaluated on the four fault datasets are also shown in Table 4. The R^2 values are significantly higher than those of the high–order models ($\bar{R}^2_{high} = 11.2; \bar{R}^2_{low} = 43.5$).

5. CONCLUSIONS

Accurate linear dynamic models have been identified from a simulated physiological diabetes model. The simulations represented realistic conditions by incorporating measurement noise, reasonable meal times and magnitudes, insulin–to–carbohydrate ratios, and faults. The high–order ARX and BJ models identified from the normal data provide accurate predictions of the normal datasets. The low–order model predictions of the normal datasets are comparable to those of the high–order models.

The models identified from N_A were applied to the four fault datasets to determine whether a distinction could be made between normal operation and faults. Two of the four faults ($F1$ and $F3$) were detected readily by the high–order models. $F2$ was more difficult to detect by the high–order models, in part because the fault occurred near the end of the run. Finally, $F4$ went undetected due to the insignificance of this fault. The low–order models developed from N_A largely failed to distinguish between normal operation and faults.

ACKNOWLEDGEMENTS

This research was supported by the National Institutes of Health (grant R21–DK069833–02). Their financial support is gratefully acknowledged. The research has been performed in collaboration with Francis J. Doyle III and Cesar C. Palerm of UC Santa Barbara. Their advice and expertise is greatly appreciated.

REFERENCES

Ashcroft, F.M. and S.J.H. Ashcroft (1992). *Insulin: Molecular Biology to Pathology*. Oxford University Press. New York, NY.

Bequette, B.W. (2005). A critical assessment of algorithms and challenges in the development of a closed–loop artificial pancreas. *Diabetes Technol Ther* **7**(1), 28–47.

Bergman, R.N., L.S. Philips and C. Cobelli (1981). Physiological evaluation of factors controlling glucose tolerance in man. *J Clin Invest* **68**(6), 1456–1467.

Cobelli, C., G. Nucci and S. Del Prato (1998). A physiological simulation model of the glucose–insulin system in type I diabetes. *Diabetes Nutr Metab* **11**(1), 78.

Desai, S., J. Tamada, R. Kurnik and R. Potts (2002). Predicting glucose values from previous measurements. *Diabetes Technol Ther* **4**, 215.

Hovorka, R., F. Shojaee-Moradie, P.V. Carroll, L.J. Chassin, I.J. Gowrie, N.C. Jackson, R.S. Tudor, A.M. Umpleby and R.H. Jones (2002). Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT. *Am J Physiol Endocrinol Metab* **282**(5), E992–1007.

Hovorka, R., V. Canonico, L.J. Chassin, U. Haueter, M. Massi-Benedetti, M.O. Federici, T.R. Pieber, H.C. Schaller, L. Schaupp, T. Vering and M.E. Wilinska (2004). Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiol Meas* **25**(4), 905–20.

Ljung, L. (2005). *System Identification Toolbox*. 6 ed. The MathWorks, Inc. 3 Apple Hill Drive, Natick, MA 01760-2098.

Lynch, S.M. and B.W. Bequette (2002). Model predictive control of blood glucose in type I diabetics using subcutaneous glucose measurements. In: *Proceedings of the American Control Conference*. pp. 4039–4043.

Parker, R.S. and F.J. Doyle III (2001). Control–relevant modeling in drug delivery. *Adv Drug Deliv Rev* **48**, 211–228.

Sorensen, J.T. (1985). A physiologic model of glucose metabolism in man and its use to design and assess improved insulin therapies for diabetes. PhD thesis. Massachusetts Institute of Technology.

Wilinska, M.E., L.J. Chassin, H.C. Schaller, L. Schaupp, T.R. Pieber and R. Hovorka (2005). Insulin kinetics in type–1 diabetes: Continuous and bolus delivery of rapid acting insulin. *IEEE Trans Biomed Eng* **52**(1), 3–12.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DYNAMIC MODELING OF EXERCISE EFFECTS ON PLASMA GLUCOSE AND INSULIN LEVELS

Anirban Roy * **Robert S. Parker** *,[1]

* *Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, PA*

Abstract: A mathematical model of the changes in plasma glucose and insulin concentrations during mild-to-moderate physiological exercise was developed for insulin dependent diabetic patients. From a metabolic prospective, the significant exercise induced effects are: increased glucose uptake rate by the working tissues; increased hepatic glucose production to maintain overall glucose homeostasis; and decreased plasma insulin concentration. The minimal mathematical model developed by Bergman *et al.* (1981) was extended to include the major exercise effects on plasma glucose and insulin levels. Model predictions of glucose and insulin dynamics were consistent with the existing literature data. This extended model provides a new disturbance test platform for the development of closed-loop glucose control algorithms.

Keywords: diabetes, glucose, insulin, exercise, minimal model.

1. INTRODUCTION

Diabetes mellitus is a metabolic disease caused by either the loss of pancreatic insulin secretion (Type-I) or resistance developed by the body towards the glucoregulatory action of insulin (Type-II). In order to prevent major health complications, it is important to maintain plasma glucose concentration within the normoglycemic range (70 - 120 $\frac{mg}{dl}$) (DCCT - The Diabetes Control and Complications Trial Research Group, 1993; DCCT - The Diabetes Control and Complications Trial Research Group, 1996). The major long term effects of diabetes are caused due to hyperglycemia, where the plasma glucose concentration exceeds 120 $\frac{mg}{dl}$ due to insufficient endogenous insulin secretion (DCCT - The Diabetes Control and Complications Trial Research Group, 1993; DCCT - The Diabetes Control and Complications Trial Research Group, 1996). Prolonged hyperglycemia causes kidney disease, blindness, loss of limbs, etc (DCCT - The Diabetes Control and Complications Trial Research Group, 1993; DCCT - The Diabetes Control and Complications Trial Research Group, 1996). Of more immediate concern is hypoglycemia when the plasma glucose concentration falls below 70 $\frac{mg}{dl}$. Such conditions can lead to dizziness, coma or even death (DCCT - The Diabetes Control and Complications Trial Research Group, 1993; DCCT - The Diabetes Control and Complications Trial Research Group, 1996).

Since the 1960s, mathematical models have been used to describe glucose-insulin dynamics (Bolie, 1961). Bergman *et al.* (1981) proposed a three compartment minimal model to analyze the glucose disappearance and insulin sensitivity during an intra-venous glucose tolerance test (IVGTT). Modifications have been made to the original minimal model to incorporate various physiological effects of glucose and insulin. Cobelli *et al.* (1986) developed a revised minimal model in order to separate the effects of glucose production from utilization. The overestimation of glucose effectiveness and underestimation of insulin sensitivity by the minimal model was addressed in yet another publication by Cobelli *et al.* (1999), where a second non-accessible glucose compartment was added on to the original model. Hovorka *et al.* (2002) extended the original minimal model by adding three glucose and insulin sub-compartments in order to capture absorption, distribution and disposal dynamics, respectively. However, none of

[1] To whom correspondence should be addressed: rparker@pitt.edu; +1-412-624-7364; 1249 Benedum Hall, Pittsburgh, PA 15261 USA

these models captures the changes in glucose and insulin dynamics due to exercise.

Physiological exercise induces several fundamental metabolic changes in the body (Wasserman and Cherrington, 1991). An increase in exercise intensity amplifies glucose uptake by the working tissues (Wasserman *et al.*, 1991). In order to maintain plasma glucose homeostasis, hepatic glucose production also increases with increasing work intensity (Wahren *et al.*, 1971). During prolonged exercise, hepatic glycogen stores begin to deplete, which leads to a reduction in hepatic glucose production as the glucose production mechanism shifts from glycogenolysis to gluconeogenesis (Ahlborg *et al.*, 1974). Since the energy requirement at a given exercise intensity is approximately constant, the overall plasma glucose concentration tends to fall well below the normoglycemic range for prolonged exercise durations (Wasserman and Cherrington, 1991). Elevated physical activity also promotes a drop in plasma insulin concentration from its basal level (Wolfe *et al.*, 1986; Wasserman *et al.*, 1989). The goal of the present work is to incorporate the fundamental effects of physiological exercise into the Bergman minimal model (Bergman. *et al.*, 1981) in order to capture the plasma glucose and insulin dynamics during mild-to-moderate exercise.

2. BERGMAN MINIMAL MODEL

Bergman *et al.* (1981) successfully quantified the pancreatic responsiveness and insulin sensitivity of a diabetic patient using a three compartmental mathematical model, as shown in Figure 1.

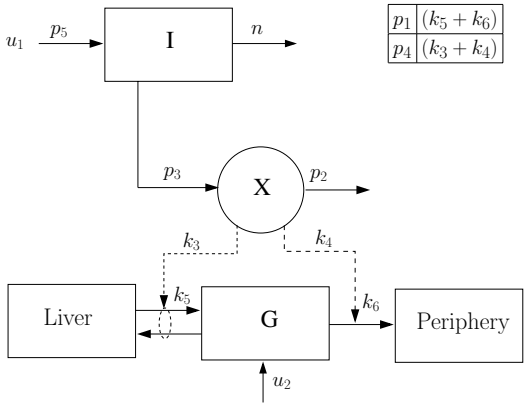


Figure 1: Bergman minimal model of insulin and glucose dynamics, adapted from (Bergman. *et al.*, 1981)

Compartments I, X, and G, represent the plasma insulin ($\frac{\mu\text{U}}{\text{ml}}$), remote insulin ($\frac{\mu\text{U}}{\text{ml}}$), and plasma glucose ($\frac{\text{mg}}{\text{dl}}$) concentrations, respectively. The model assumes that all the necessary insulin is infused exogenously (u_1), thereby modeling the insulin-dependent diabetic patients. A portion of the infused insulin enters into the remote compartment, X, from the circulatory system. The remote insulin (X) actively takes part in promoting uptake of plasma glucose (G) into the hepatic and extra-hepatic tissues.

The deviation form of Bergman minimal model is mathematically given by (Bergman. *et al.*, 1981):

$$\frac{dI}{dt} = -n(I(t) - I_b) + p_5 u_1(t) \quad (1)$$

$$\frac{dX}{dt} = -p_2(X(t) - X_b) + p_3(I(t) - I_b) \quad (2)$$

$$\frac{dG}{dt} = -p_1 G(t) - p_4 X(t) G(t) + p_1 G_b + \frac{u_2(t)}{Vol_G} \quad (3)$$

Here, I_b, X_b and G_b are the basal plasma insulin, basal remote insulin, and basal plasma glucose concentrations, respectively. The rate constant n represents disappearance of plasma insulin above its basal level. The rates of appearance of insulin in, and disappearance of remote insulin from, the remote insulin compartment are governed by the parameters p_3 and p_2, respectively. Dietary absorption or external infusion of glucose is indicated by $u_2(t)$, and the glucose distribution space is indicated by Vol_G. Parameter p_1 represents the rate at which plasma glucose above its basal level is removed from the plasma space independent of the influence of insulin. Glucose uptake under the influence of insulin is governed by the parameter p_4. Parameter values for the minimal model are provided in Table 1.

Table 1: Parameters of the Bergman minimal model, from (Bergman. et al.*, 1981)*

Parameter	Value	Unit
p_1	0.04	1/min
p_2	0.037	1/min
p_3	0.000012	1/min
p_4	1.0	ml/min·μU
p_5	0.000568	1/ml
n	0.142	1/min
G_b	80.0	mg/dl
Vol_G	117.0	dl

3. QUANTITATING EXERCISE INTENSITY

The maximum rate of oxygen consumption for an individual is VO_2^{max} ($\frac{\text{ml}}{\text{kg-min}}$). Oxygen consumption is approximately linearly proportional to the energy expenditure (Åstrand, 1960). Hence, it is possible to indirectly measure an individual's maximum capacity to do aerobic work by measuring oxygen consumption. When physical activity is expressed as a percentage of VO_2^{max}, (PVO_2^{max}), exercise effects may be compared between individuals of the same sex and similar body weight at the same PVO_2^{max}. The average PVO_2^{max} for a person in the basal state is 8% (Felig and Wahren, 1975). Ahlborg *et al.* (1974) demonstrated that PVO_2^{max} increases rapidly at the onset of exercise,

reaches its ultimate value within 5-6 minutes and remains constant for the duration of exercise. The exercise model developed in this study uses PVO_2^{max} to quantify exercise level. The ordinary differential equation (ODE) capturing the exercise intensity is given by:

$$\frac{dPVO_2^{max}}{dt} = -0.8PVO_2^{max}(t) + 0.8u_3(t) \quad (4)$$

Here, $PVO_2^{max}(t)$ is the exercise level as experienced by the individual, and $u_3(t)$ is the ultimate exercise intensity – an input to the model. The parameter value of 0.8 ($\frac{1}{\text{min}}$) was selected to achieve a $PVO_2^{max}(t)$ settling time of approximately 5 minutes.

4. MINIMAL EXERCISE MODEL

The aim is to capture the effects of exercise on plasma glucose and insulin concentrations in response to mild-to-moderate aerobic exercise. A rise in glucose uptake by the working tissues occurs in response to exercise, and this is followed by an increase in hepatic glucose production. However, the rate of liver glucose production decreases with prolonged exercise due to the depletion of liver glycogen stores. As the muscle energy demand remains approximately constant for a given exercise level, the overall plasma glucose concentration eventually declines after an initial rise (Ahlborg *et al.*, 1974; Ahlborg and Felig, 1982). The initial rise in plasma glucose concentration with the onset of exercise is due to a several fold of increase in hepatic glucose production exceeding the glucose demand by working tissues (Marliss and Vranic, 2002). There is also an exercise induced decline in plasma insulin level. Wolfe *et al.* (1986) demonstrated that the application of a pancreatic clamp to maintain basal insulin levels during exercise (at 40 PVO_2^{max}) significantly increases the plasma glucose uptake, thereby disrupting glucose homeostasis. This is consistent with the idea that an increase in exercise level elevates the stimulating effect of insulin on glucose uptake (Wolfe *et al.*, 1986; Wasserman *et al.*, 1989). The ODEs for the exercise minimal model are as follows:

$$\frac{dI}{dt} = -n(I(t) - I_b) + p_5u_1(t) - I_e(t) \quad (5)$$

$$\frac{dX}{dt} = -p_2(X(t) - X_b) + p_3(I(t) - I_b) \quad (6)$$

$$\frac{dG}{dt} = -p_1G(t) - p_4X(t)G(t) + p_1G_b + G_{prod}(t) - G_{up}(t) + \frac{u_2(t)}{Vol_G} \quad (7)$$

$$\frac{dG_{prod}}{dt} = a_1PVO_2^{max}(t) - a_2G_{prod}(t) \quad (8)$$

$$\frac{dG_{up}}{dt} = (a_3PVO_2^{max}(t) - a_4)PVO_2^{max}(t) - a_5G_{up}(t) \quad (9)$$

$$\frac{dI_e}{dt} = a_6PVO_2^{max}(t) - a_7I_e(t) \quad (10)$$

The insulin dynamics, (5), have been modified from the Bergman minimal model, (1), by the addition of the final term. Here $I_e(t)$ is the rate of insulin removal from the circulatory system due to exercise. The plasma glucose dynamics, (7), differ from (3) of the Bergman minimal model by the terms, $(G_{prod}(t) - G_{up}(t))$. Variables $G_{up}(t)$ and $G_{prod}(t)$ represent rates of glucose uptake and hepatic glucose production induced by exercise, respectively. The dynamics of hepatic glucose production, glucose uptake, and removal of plasma insulin, induced by exercise are given by (8), (9), and (10), respectively.

Parameters for the minimal exercise model were estimated using the nonlinear 'Least Square' technique as described in (Carson and Cobelli, 2001). The normalized residual is obtained as:

$$\chi^2 \equiv \sum_{i=1}^{N} \left[\frac{y_i - y(t_i, a_1 \ldots a_M)}{\sigma_i} \right]^2 \quad (11)$$

Here y_i is the measured data at time t_i which has standard deviation of σ_i. The model prediction is given by $y(t_i, a_1 \ldots a_M)$, where a_i represent model parameters. Equation (11) can be considered a weighted minimization using $(\frac{1}{\sigma})^2$ as the weights. N is the number of data points and M is the total number of model parameters. χ^2 is often denoted as 'weighted sum squared error'. The aim is to estimate a_i in order to minimize χ^2.

Table 2: Parameters of the minimal exercise model in addition to those in Table 1

Parameter	Value	Unit
a_1	0.011	mg/dl·min
a_2	0.9	–
a_3	0.000001	mg/dl·min
a_4	0.00013	mg/dl·min
a_5	0.00002	–
a_6	0.00025	μU/ml·min
a_7	0.009	–

Data from (Ahlborg *et al.*, 1974) and (Ahlborg and Felig, 1982) were used to estimate the parameters a_6 and a_7, thereby quantifying the exercise induced removal of plasma insulin from the circulatory system. After fixing a_6 and a_7, the insulin model was validated by comparing with the data from (Wolfe *et al.*, 1986). In fitting plasma glucose parameters, initially a_3 was set to zero. Parameters a_1, a_2, a_4, and a_5 were estimated from (Ahlborg *et al.*, 1974) and (Marliss and Vranic, 2002). The model was validated by comparing the predictions of plasma glucose concentration with the data from (Ahlborg and Felig, 1982). In order to improve the model fit, the bilinearity was introduced in the glucose uptake ODE (9) by estimating a_3 (where, $a_3 \neq 0$) and re-estimating a_4, using (Ahlborg

et al., 1974) and (Ahlborg and Felig, 1982). The parameter values for the minimal exercise model are given in Table 2.

5. RESULTS

5.1 Plasma Insulin Dynamics During Exercise

To study the plasma insulin dynamics during prolonged exercise periods, Ahlborg *et al.* (1974) conducted an experiment where healthy subjects were studied in a continuous bicycle exercise for 4 hours (PVO_2^{max} = 30). Blood samples were taken at regular intervals to measure the plasma insulin level. With the onset of exercise, plasma insulin level declined from its basal level ($14 \pm 1.9 \frac{\mu U}{ml}$), and continued to do so until the end of the experiment, as shown in Figure 2. It can be observed that the model under predicts insulin concentration in short terms, followed by slight over prediction at long times. Quantitatively, however, the model was consistently within one standard deviation of the mean. Parameters a_6 and a_7 had values as given in Table 2.

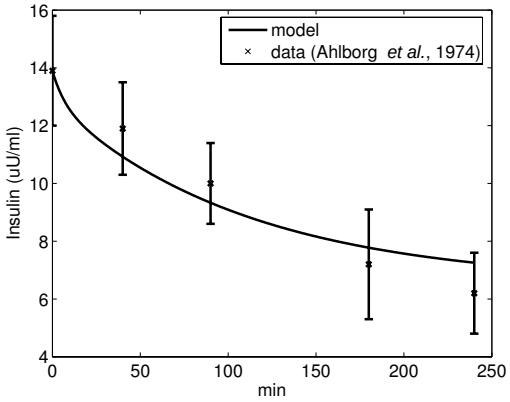


Figure 2: Published (mean ± std. dev.) and model fit to data of plasma insulin concentration in response to mild exercise (PVO_2^{max} = 30)

In another study, Ahlborg *et al.* (1982) conducted a similar leg exercise experiment at a higher exercise level (PVO_2^{max} = 60). Plasma insulin concentration declined consistently with the onset of exercise, as shown in Figure 3. It can be observed that the model prediction for plasma insulin concentration was generally better than that of Figure 2, although under prediction occurred at longer times. Parameters a_6 and a_7 had values as given in Table 2. The model predictions were within one standard deviation of the mean, thus validating the insulin model.

Data from a separate study by Wolfe *et al.* (1986) was used to validate the insulin model. Light exercise (PVO_2^{max} = 40) was performed to observe the removal of plasma insulin from the circulatory system. With the onset of exercise the insulin level declined well below the basal level ($13.2 \pm 3 \frac{\mu U}{ml}$), and this hypoinsulinemic state persisted until end of the experiment, as shown in Figure 4.

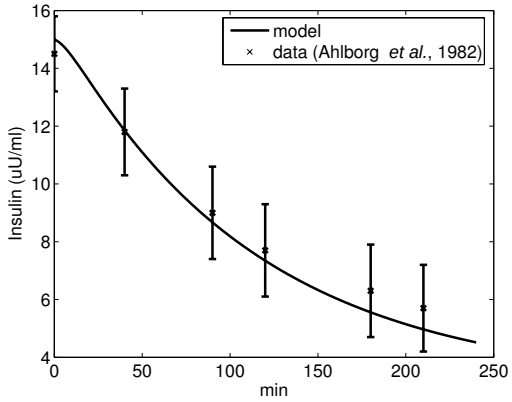


Figure 3: Published (mean ± std. dev.) and model fit to data of plasma insulin concentration in response to moderate exercise (PVO_2^{max} = 60)

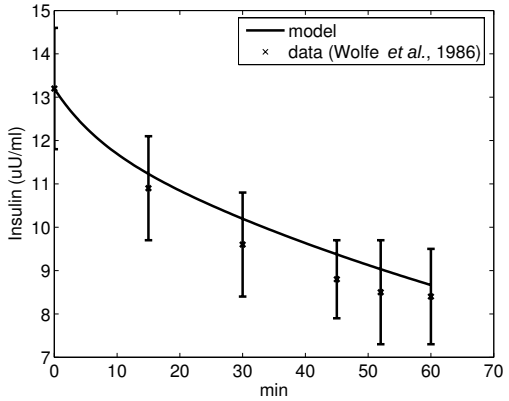


Figure 4: Model simulation validation versus published data (mean ± std. dev.) of plasma insulin concentration in response to mild exercise (PVO_2^{max} = 40)

5.2 Plasma Glucose Dynamics During Exercise

The plasma glucose dynamics during a short duration (40 minute) of moderate level exercise (PVO_2^{max} = 50), are shown in Figure 5 (Marliss and Vranic, 2002). Healthy subjects performed a full-body exercise, where blood samples were collected at regular intervals to measure the plasma glucose concentration. With the onset of exercise, the plasma glucose level increased slightly from its basal state ($89 \pm 4.5 \frac{mg}{dl}$) to $95 \pm 5 \frac{mg}{dl}$ and then started to decrease towards the end of the experiment. The resulting model predictions of plasma glucose were quantitatively consistent with the published data. Parameters a_1, a_2, a_4 and a_5 had values as given in Table2.

The experiment conducted by Ahlborg *et al.* (1974) was considered to observe the plasma glucose dynamics during prolonged exercise periods. Blood samples were taken at regular intervals to measure the plasma glucose level. With the onset of exercise, the plasma glucose level increased slightly from $81.8 \pm 2.5 \frac{mg}{dl}$ to $83.5 \pm 3 \frac{mg}{dl}$ at the 40 minute mark. This

degree of change is consistent with the data of Marliss *et al.* (2002) shown in Figure 5.

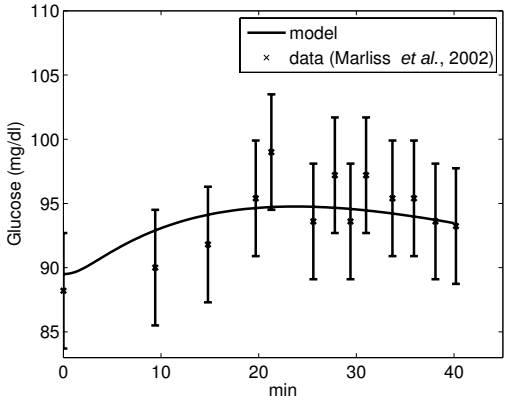


Figure 5: Published (mean ± std. dev.) and model fit to data of plasma glucose concentration in response to moderate exercise (PVO_2^{max} = 50)

Beyond that, plasma glucose level decreased consistently until the end of the experiment, as shown in Figure 6. It can be observed that, during the short term response the model under predicted the mean data, and with eventual over prediction for t $\geq$ 180 min. However, quantitatively the model was consistently within one standard deviation of the mean. Again, parameters a_1, a_2, a_4 and a_5 had values as given in Table 2. Note that a_3 was set to zero, as discussed in Section 4.

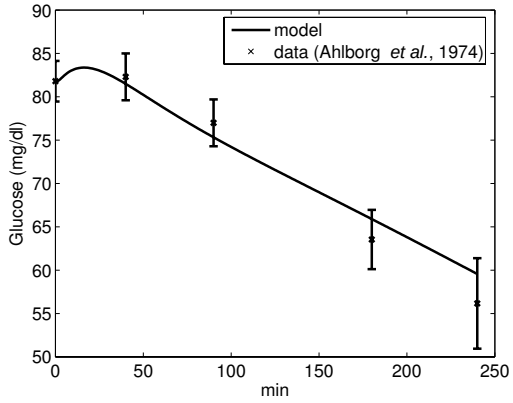


Figure 6: Published (mean ± std. dev.) and model fit to data of plasma glucose concentration in response to mild exercise (PVO_2^{max} = 30)

To test the glucose model for moderate exercise intensity (PVO_2^{max} = 60), data from another similar leg exercise study conducted by Ahlborg *et al.* (1982) was considered. Throughout the duration of the experiment, glucose uptake was higher than splanchnic glucose production. Hence, from the onset of exercise, overall plasma glucose concentration decreased consistently, as shown in Figure 7. With a_3 = 0, the model over predicted plasma glucose dynamics significantly at all times (Figure 7).

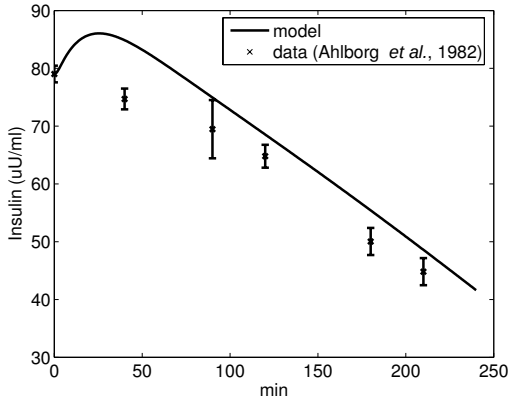


Figure 7: Published (mean ± std. dev.) and model fit to data of plasma glucose concentration in response to moderate exercise (PVO_2^{max} = 60), where a_3 = 0 and a_4 = 0.00017

In order to capture the plasma glucose dynamics during moderate intensity exercise (PVO_2^{max} = 60), the parameter a_4 was modified to incorporate an effect linear with exercise intensity, $a_3 PVO_2^{max}(t) - a_4$. Parameters a_3 and a_4 in the glucose uptake ODE (9) were re-estimated with $a_3 \neq 0$ using the data from (Ahlborg and Felig, 1982). Comparison of plasma glucose concentrations between the revised model and literature data (Ahlborg and Felig, 1982) are shown in Figure 8. It can be observed that the short term response of the model does not capture the data adequately.

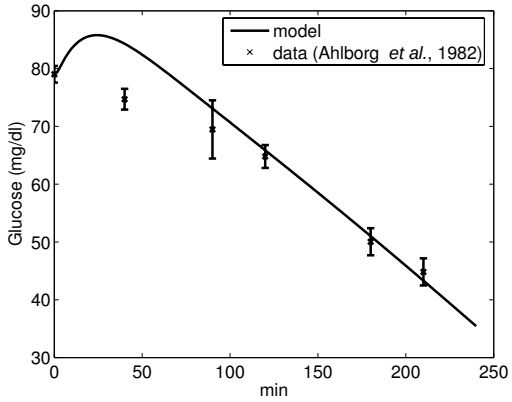


Figure 8: Published (mean ± std. dev.) and model fit to data of plasma glucose concentration in response to moderate exercise (PVO_2^{max} = 60), where a_3 = 0.000001 $\frac{mg}{dl \cdot min}$ and a_4 = 0.00013 $\frac{mg}{dl \cdot min}$

It is interesting to note that the short term data from (Ahlborg and Felig, 1982), in Figure 8, is inconsistent with that of (Marliss and Vranic, 2002) and (Ahlborg *et al.*, 1974), in Figures 5 and 6, respectively. Hence the model behaviors are inherently limited by the data quality, and additional experimental and simulation studies are required to refine and validate the model.

6. SUMMARY AND DISCUSSION

A minimal model of exercise effects on plasma glucose-insulin dynamics was developed. The model successfully captured the effects of mild-to-moderate aerobic exercise on plasma glucose and insulin concentrations. Inclusion of separate dynamics in the model for glucose uptake (9) and hepatic glucose production (8) induced by exercise made it possible to capture the initial rise (due to higher hepatic glucose production than tissue glucose uptake) and eventual decline of plasma glucose level with prolonged exercise (due to shortage of glycogen storage and a shift in glucose production mechanism). It was necessary to introduce a bilinear term in the glucose uptake ODE (9) in order to capture the full set of observed glucose dynamics induced by mild-to-moderate exercise. Given the small values for a_3 and a_4, changes in model structure may be warranted. The model also successfully captured the removal of plasma insulin from the circulatory system during physical exercise. This model provides the control community with an alternative benchmark problem in glucose control for diabetic patients by allowing the analysis of meal and exercise disturbances alone or in combination.

Some of the parameters (a_6 and a_7) were estimated using data from (Ahlborg *et al.*, 1974) and (Ahlborg and Felig, 1982), where healthy subjects were used in the experiments. Ideally, this data would be from a diabetic population; however the majority of exercise studies are on healthy subjects. Additional experimental studies, ideally in a diabetic population, would improve model fidelity.

7. ACKNOWLEDGMENT

Support for this work was provided by ACS PRF # 38068-G9.

REFERENCES

Ahlborg, G. and P. Felig (1982). Lactate and glucose exchange across the forearm, legs, and splanchnic bed during and after prolonged exercise. *J. Clin. Invest.* **69**, 45–54.

Ahlborg, G., P. Felig, L. Hagenfeldt, R. Hendler and J. Wahren (1974). Substrate turnover during prolonged exercise in man. Splanchnic and leg metabolism of glucose, free fatty acids, and amino acids. *J. Clin. Invest.* **53**, 1080–1090.

Åstrand, I. (1960). Aerobic work capacity in men and women with special reference to age. *Acta. Physiol. Scan.* **49**, 1–92.

Bergman., R. N., Phillips L. S. and Cobelli C. (1981). Physiologic evaluation of factors controlling glucose tolerance in man. *J. Clin. Invest.* **68**, 1456–1467.

Bolie, V. W. (1961). Coefficients of normal blood glucose regulation. *J. Appl. Physiol.* **16**, 783–788.

Carson, E. and C. Cobelli (2001). Modelling methodology for physiology and medicine. *Academic Press, San Diego, CA*.

Cobelli, C., Caumo A. and Omenetto M. (1999). Minimal model sg overestimation and si underestimation: improved accuracy by a bayesian two-compartment model. *J. Physiol - Endo.* **277**, 481 – 488.

Cobelli, C., Pacini G., Toffolo G. and Sacca L. (1986). Estimation of insulin sensitivity and glucose clearance from minimal model: New insights from labeled ivgtt. *Am. J. Physiol.* **250**, E591 – E598.

DCCT - The Diabetes Control and Complications Trial Research Group (1993). The effect of intensive treatment of diabetes on the development and progression of long–term complications in insulin–dependent diabetes mellitus. *N. Engl. J. Med.* **329**, 977–986.

DCCT - The Diabetes Control and Complications Trial Research Group (1996). The absence of a glycemic threshold for the development of long–term complications: The perspective of the diabetes control and complications trial. *Diabetes* **45**, 1289–1298.

Felig, P. and J. Wahren (1975). Fuel homeostasis in exercise. *N. Engl. J. Med.* **293**, 1078–1084.

Hovorka, R., Shojaee-Moradie F., Carroll P. V., Chassin L. J., Gowrie I. J., Jackson N. C., Tudor R. S., Umpleby A. M. and Jones R. H. (2002). Partitioning glucose distribution/transport, disposal, and endogenous production during ivgtt.. *Am. J. Physiol.* **282**, E992 – E1007.

Marliss, E. B. and M. Vranic (2002). Intense exercise had unique effects on both insulin release and its roles in glucoregulation. *Diabetes* **51**, S271 – S283.

Wahren, J., Felig P., Ahlborg G. and Jorfeldt L. (1971). Glucose metabolism during leg exercise in man. *J. Clin. Invest.* **50**, 2715–2725.

Wasserman, D. H. and A. D. Cherrington (1991). Hepatic fuel metabolism during muscular work: Role and regulation. *Am. J. Physiol.* **260 (Endocrinol. Metab. 23)**, E811–E824.

Wasserman, D. H., Geer R. J., Rice D. E., Bracy D., Flakoll P. J., Brown L. L., Hill J. O. and Abumrad N. (1991). Interaction of exercise and insulin action in humans. *Am. Physiol. Society* **34**, E37 – E45.

Wasserman, D. H., Williams P. E., Lacy D. B., Goldstein R. E. and Cherrington A. D. (1989). Exercise-induced fall in insulin and hepatic carbohydrate metabolism during muscular work. *Am. Physiol. Society* **32**, E500 – E509.

Wolfe, R. R., Nadel E. R., Shaw J. F., Stephenson L. A. and Wolfe M. H. (1986). Role of changes in insulin and glucagon in glucose homeostasis in exercise. *Am. Soc. Clin. Invest.* **77**, 900 – 907.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PATHWAYS FOR OPTIMIZATION-BASED DRUG DELIVERY SYSTEMS AND DEVICES

Leonidas Bleris [*] **Panagiotis Vouzis** [***]
Mark V. Arnold [***] **Mayuresh V. Kothare** [**,1]

[*] *Department of Electrical and Computer Engineering*
Lehigh University, Bethlehem, PA, 18015
[**] *Department of Chemical Engineering*
Lehigh University, Bethlehem, PA, 18015
[***] *Department of Computer Engineering*
Lehigh University, Bethlehem, PA, 18015

Abstract: Drug synthesis and discovery represents today one of the most rapidly evolving scientific areas. This is primarily due to the interdisciplinary collaboration between chemists, pharmacologists, molecular biologists, and biochemists. A direct implication of the developments in drug discovery is the need for novel drug delivery systems and devices. Considering the advances in engineering disciplines and micro/nano technology the potential for producing new drug delivery devices is substantial. Notably, so is the necessity of these devices in creating solid commercial value propositions for the medical markets. In this work we present research results related to embedding optimization-based control on-a-chip for drug delivery applications.

Keywords: Embedded Control Systems, Real-Time Model Predictive Control, Drug Delivery Devices

1. INTRODUCTION

Today, drugs can be delivered in many different ways including: orally (pills or suspensions), through the vein (intravenously), through the artery (arterially), topically through the skin (transdermally), through the rectum (suppository), through the eye (ocular), through the lungs (inhaled), by injection into the skin (subcutaneously), by injection into the muscle (intramuscularly), and under the tongue (sublingually). By means of improving the drug delivery devices, companies and researchers aim to meet particular goals (Brunner, 2004), most of them related to the state of the patients. These goals include: improved efficacy, reduced side effects, continuous dosing, reduced pain from administration, increased ease of use, increased use compliance, improved mobility, and decreased involvement of healthcare workers.

Controlled drug delivery (Langer, 2004) currently involves control of the time course or the location of drug delivery. In this work we examine alternative pathways for the implementation of optimal drug delivery control (of the time course) on an embedded target, for glucose regulation by means of insulin delivery. We provide research results of this implementation on an Application Specific Integrated Circuit (ASIC), on a general purpose processor, and finally on a Field Programmable Gate Array (FPGA).

[1] Partial financial support for this research from the US National Science Foundation grant CTS-0134102 and the Pittsburgh Digital Greenhouse is gratefully acknowledged.

Optimization-based control schemes can be used in order to effectively control nonlinear and multivariable models, and to impose constraints on both the control action and the states. The constraints are usually set on the bounds and the rate of change, and can be incorporated in the form of equalities or inequalities. These control algorithms typically involve the solution of an optimization problem that explicitly incorporates knowledge of a dynamic model of a process, with the addition of design and process objectives. This concept has seen widespread use, particularly through applications of predictive control. The advantages of Model Predictive Control (MPC), such as the ability to handle constraints, the applicability to nonlinear processes and to multivariable problems, constitute this control method an ideal choice for satisfying most of the design objectives of a drug delivery device.

Currently, companies that want to sell insulin delivery devices must illustrate to the Food and Drug Administration (FDA) that their devices are "substantially equivalent" to drug delivery devices already for sale. MPC controllers (device) and the drug (insulin) are approved and are on the market individually. The use of insulin is the standard method for regulating glucose. Thus the toxicology of the pharmaceutical is known. Nevertheless, implementations of MPC have been on the market mainly for chemical plants, and on workstation implementations. In 2004, the FDA announced that the worlds first implantable Radio Frequency Identification (RFID) microchip for human has been cleared for medical uses in the United States. About the size of a grain of rice, "VeriChip" is a subdermal radio frequency microchip. The device has no power supply, and it is activated when a scanning device runs across the skin above it. A tiny transmitter on the chip then releases patient-specific information. Although not capable of carrying out arithmetic operations such technology is available, and one of the biggest regulatory thresholds was surpassed. Thus, it is a matter of time before new generations of similar devices seek FDA approval. There are however other issues to be examined: the performance robustness of the chip, the biocompatibility of the device, and that the stability and bioavailability of the drug have not been compromised by the drug-device combination.

Although the solution to glucose regulation problems may appear in the form of an implantable capsule (Leoni and Desai, 2001) that continuously produces insulin and releases it to the bloodstream, there are numerous potential applications for medical controllers on-a-chip including: control of physiological processes, muscle control, respiration control, drug infusion control (for instance during anesthesia), cardiac pacemakers and defibrillators, heart rate control, blood flow and pressure control, HIV control, and neurological implants. For example, in the Human Immunodeficiency Virus (HIV) control case, where the drug cocktail provided is currently expensive, the regulation (with the solution of an optimization problem) of the dosage to the absolute minimum is highly desirable. Another example is a device that detects irregular heart pulses and uses a control algorithm to regulate doses of a blood thinning drug (for example aspirin). This device maintains the risks of thrombosis at low levels while minimizing the chances of internal bleeding.

2. BLOOD GLUCOSE CONTROL IN DIABETIC PATIENTS

Diabetes mellitus is a chronic health condition where the human body is unable to produce insulin and properly breakdown sugar (glucose) in the blood. The insufficient insulin production or lack of responsiveness to insulin, results to hyperglycemia (blood glucose levels over 120mg/dL). There are two primary types of diabetes mellitus, type I (insulin-dependent or juvenile-onset), which may be caused by an autoimmune response, and type II (non-insulin-dependent or adult-onset). Symptoms include hunger, thirst, excessive urination, dehydration and weight loss. Complications can include heart disease, stroke, neuropathy, poor circulation leading to loss of limbs, hearing impairment, vision problems and death.

The treatment of diabetes requires regular insulin injections, proper nutrition and exercise in order to maintain normoglycemia, defined as blood glucose $70-100$mg/dL. Insulin and glucagon are the hormones responsible for glucose regulation. Both insulin and glucagon are secreted from the pancreas, and thus are referred to as pancreatic endocrine hormones. Most of the long-term complications associated with diabetes result from sustained hyperglycemia, but hypoglycemia can result in very acute symptoms such as coma and death.

A significant effort has been devoted to the development of closed-loop controllers for blood glucose control. For approaches to applying control to diabetic subjects the reader is referred to (Parker *et al.*, 1999; Doyle-III *et al.*, 2000; Rubb and Parker, 2003; Parker and Doyle-III, 2001).

2.1 Minimal Glucose Model

Minimal models of glucose and insulin plasma levels have been developed (Bergman *et al.*, 1979; Pacini and Bergman, 1986) for humans using

Frequently-Sampled Intravenous Glucose Tolerance (FSIGT) tests. During a FSIGT test, a single intravenous injection of glucose is given to a fasting subject and blood samples are collected at regular timed intervals. The blood samples are then analyzed for glucose and insulin concentration. Figure 1 shows a typical response from a normal subject.

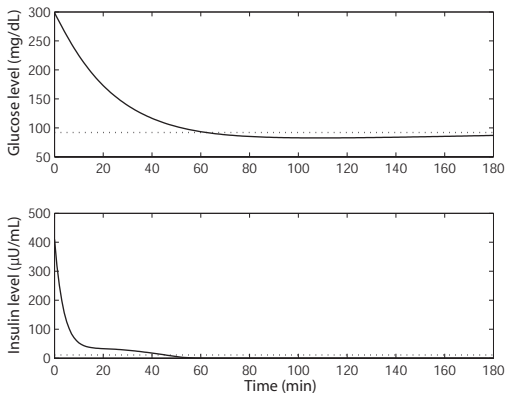


Fig. 1. Typical glucose and insulin response from a normal subject.

As illustrated in Figure 1, the glucose level in plasma is at a peak at the time of the injection, drops to a minimum which is below the basal glucose level, and then gradually returns to the basal level (dashed line). The insulin level in plasma rapidly rises to a peak immediately after the injection, drops to a lower level which is still above the basal insulin level, rises again to a lesser peak, and then gradually drops to its basal level. Depending on the state of the subject, there can be wide variations from this response that can determine the condition of the patient (G. M. Steil and Bergman, 1993).

The glucose minimal model involves two physiologic compartments: an interstitial tissue compartment and a plasma compartment. The differential equations corresponding to the two compartments are:

$$\frac{dG(t)}{dt} = k_1(G_b - G(t)) - X(t)G(t) \qquad (1)$$

$$\frac{dX(t)}{dt} = k_2(I(t) - I_b) - k_3X(t) \qquad (2)$$

where t is time, G(t) is the plasma glucose concentration at time t, I(t) is the plasma insulin concentration at time t, and X(t) is the interstitial insulin at time t, with G(t_0)=G_0 and X(t_0)=0. G_b is the basal plasma glucose concentration and I_b is the basal plasma insulin concentration. The insulin sensitivity is defined as $S_I = k_2/k_3$ and the glucose effectiveness is defined as $S_G = k_1$. Basal plasma concentrations of glucose and insulin are typically measured either before or 180 minutes after the administration of glucose. There are four unknown parameters in this model: k_1, k_2, k_3, and G_0 that depend on the particular subject and can be estimated experimentally. We use the parameters adopted from (Pacini and Bergman, 1986; Riel, 2004).

3. MODEL PREDICTIVE CONTROL

Model Predictive Control originated in the chemical process industries. The main advantages of MPC are the ability to handle constraints and its applicability to multivariable nonlinear processes. Because of the computational requirements of the optimizations associated with MPC, it has primarily been applied to plants in the process industry, with slow dynamics. Furthermore, existing implementations of MPC typically perform numerical calculations using workstations in 64-bit Floating Point (FP) arithmetic, which is too expensive, power demanding and large in size. Therefore the implementation of real-time embedded model predictive control, for systems with fast dynamics, where the size and the application precludes the use of a dedicated workstation, presents new technological challenges.

Controllers belonging to the MPC family are generally characterized by the following steps: initially the future outputs are calculated at each sample interval over a predetermined horizon N, the prediction horizon, using a process model. These outputs $y(t+k|t)$ for $k = 1...N$ depend up to the time t on the past inputs and on the future signals $u(t+k|t)$, $k = 0...N-1$ which are those to be sent to the system. The next step is to calculate the set of future control moves by optimizing a determined criterion, in order to keep the process as close as possible to a predefined reference trajectory. This criterion is usually a quadratic function of the difference between the predicted output signal and the reference trajectory. In some cases, in order to minimize the control effort the control moves $u(t+k|t)$ are included in the objective function:

$$J_P(k) = \sum_{k=0}^{P}\{[y(t+k|t)-y_{ref}]^2+Ru(t+k|t)^2\} \quad (3)$$

$$|u(t+k|t)| \leq b \quad , \quad k \geq 0 \qquad (4)$$

where $y(t+k|t)$ are the predicted outputs, y_{ref} is the desired set reference output, $u(t+k|t)$ the control sequence and R is a the weighting on the control moves, a design parameter. This system is subject to input constraints given by the vector b. Finally, the first control move $u(t|t)$ is sent to the system while the rest are rejected. At the next sampling instant the output $y(t+1)$ of the system is used in the optimization using feedback and the procedure is repeated so that we get an updated control sequence.

4. EMBEDDED MODEL PREDICTIVE CONTROL

Some initial results have been reported to the direction of embedding Model Predictive Control. Further analysis and references can be found at (Bleris *et al.*, 2006*a*). With the following subsections we report our recent research results, and we focus on the FPGA implementation.

4.1 General purpose processor

We have examined a general purpose processor implementation (Bleris and Kothare, 2005*b*; Bleris and Kothare, 2005*a*). We used a single board computer phyCORE-MPC555 that packs the power of Motorola's embedded 32-bit MPC555 microcontroller within a miniature footprint. The MPC555 is a high-speed 32-bit Central Processing Unit (CPU) that contains a 64-bit floating point unit designed to accelerate advanced algorithms, running at 40MHz. To implement the optimization algorithm of MPC, we used a combination of software tools: CodeWarrior Integrated Development Environment (IDE), MATLAB, Real-Time Workshop, and SIMULINK.

In order to test the performance of a multi-model MPC formulation in real time we used Processor in the Loop (PIL) co-simulations; the MPC chip in closed loop with the monitored patient (the minimal model on the host workstation). To examine the influence of the control horizon on the computational costs of MPC running on the Motorola processor, we set the number of optimizations at a fixed number. In Figure 2 we provide profiling results for different cases of prediction horizon and optimization steps. Additionally, in Figure 3 we provide the performance of MPC for different number of optimizations, keeping fixed the prediction and control horizons. As expected the computational time for this case grows linearly with the optimizations. Both these profiling results illustrate that we remain in all examined cases at under one second for the computation of the optimal insulin dosage using MPC. From interpolation of the results of Figure 3 we obtain the required time for one optimization loop, (a) $53msec$ for prediction horizon of 21 and control horizon of 7, (b) $95msec$ for prediction horizon of 31 and control horizon of 7.

4.2 Application-Specific Instruction Processor

For the Application Specific Integrated Circuit (ASIC) implementation (Bleris *et al.*, 2006*a*; Bleris *et al.*, 2005) we proposed the following design framework. By emulating the microcontroller arithmetic operations, we reduce the precision of

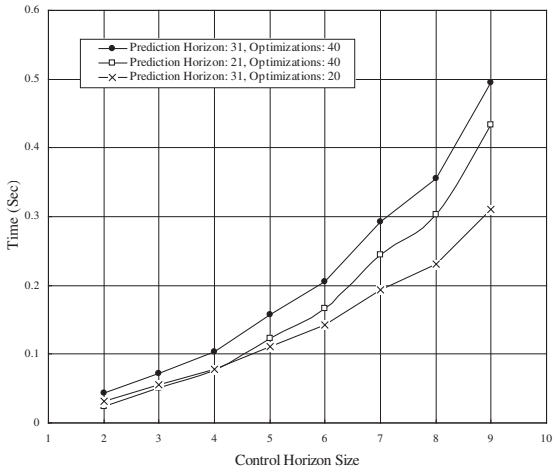


Fig. 2. Profiling results using as a variable the control horizon for different number of optimizations and prediction horizons.

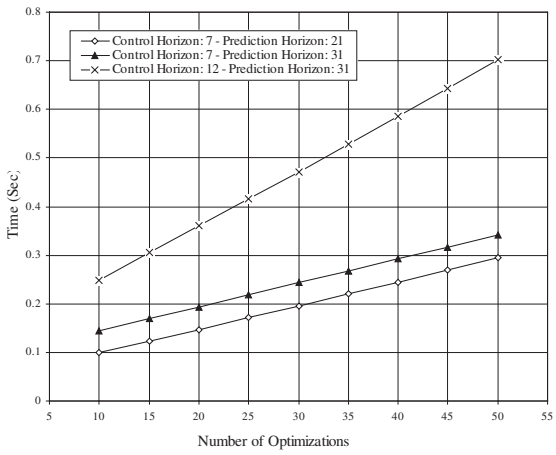


Fig. 3. Profiling results for fixed prediction and control horizon sizes and variable number of optimizations.

the microprocessor to the minimum, while maintaining stable control performance for a particular control application. This reduction is accomplished by series of parametric tests using different word sizes and utilizing computational tools to simulate the controlled model. Taking advantage of the low precision, a Logarithmic Number System (LNS) based micro-processor architecture was proposed (Garcia *et al.*, 2004) (Figure 4) that provides energy, computational cost, and price savings (Figure 5). This reduced-precision ASIC can achieve sampling speeds as low as 32msec for relatively large problems. Additionally, to quantify the advantage of reducing the precision, estimations for both 64-bit FP and 16-bit LNS circuits showed that for an arithmetic unit that computes addition, subtraction, multiplication and division, the size required is about 17 times larger for 64-bit FP.

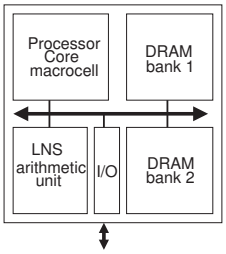


Fig. 4. Architecture block.

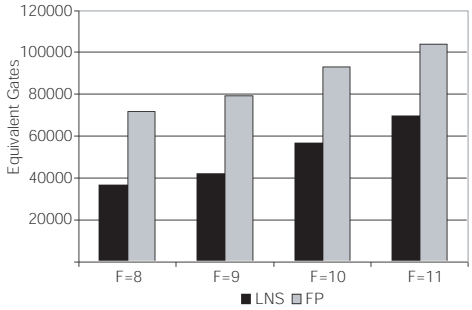


Fig. 5. Arithmetic Logic Unit (ALU) areas for LNS and FP at different precisions.

4.3 Mixed software-hardware embedded controller

For the mixed software-hardware embedded controller we examine a codesign (Vouzis *et al.*, 2005; Bleris *et al.*, 2006*b*) step before the actual implementation that decomposes the algorithm into two parts. One that fits into the host processor and one that fits into the custom made unit that performs all the (repetitive and computationally demanding) arithmetic operations. The selected microprocessor acting as a host for our design is the 16-bit Extensible Instruction Set Computer (EISC) from ADCUS, Inc. For prototyping we use the Field Programmable Gate Array (FPGA) Virtex-4 XC4VLS25 device of Xilinx interfacing with Mablab, running on a PC workstation, in order to implement Processor-In-the-Loop (PIL) techniques that help to test and debug the embedded system. Both the ADCUS microprocessor and the matrix co-processor are described in Verilog and the whole design is synthesized with the ISE 7.1 design environment of Xilinx.

A field-programmable gate array is a large-scale integrated circuit that can be programmed after it is manufactured rather than being limited to a predetermined, unchangeable hardware function. FPGAs come in a wide variety of sizes and with many different combinations of internal and external features. What they have in common is that they are composed of small blocks of programmable logic. These basic blocks may be replicated many thousands of times to create a large programmable hardware fabric. In more complex FPGAs these general-purpose logic blocks are combined with higher level arithmetic and control structures, such as multipliers and counters, in support of common types of applications such as signal processing.

Defining the behavior of an FPGA (the hardware that it contains) has traditionally been done either using a Hardware Description Language (HDL) such as VHDL or Verilog or by arranging blocks of pre-existing functions, whether gate-level logic elements or higher-level macros, using a schematic– or block diagram–oriented design tool. Hardware applications implemented in FPGAs are generally slower and consume more power than the same applications implemented in custom ASICs. Nonetheless, the dramatically lowered risk and cost of development for FPGAs have made them excellent alternatives to custom Integrated Circuits (ICs). The reduced development times associated with FPGAs often makes them compelling platforms for ASIC prototyping as well.

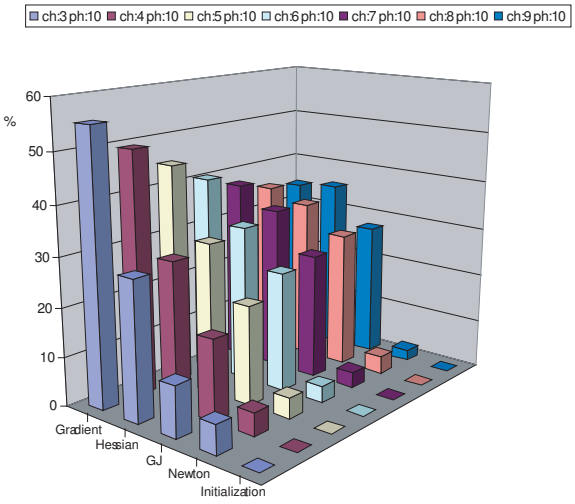


Fig. 6. Profiling results for prediction horizon 10 and variable control horizon.

We use Newtons algorithm to solve the optimization problem by combining the constraints into the cost function using barrier functions. This optimization algorithm consists of five functions which can be considered as the five basic operational blocks. These functions are: the initializations prior to each iteration loop of the optimization, the calculation of the Gradient vector and the Hessian matrix, the Gauss-Jordan matrix inversion, and finally the optimal move calculation using Newtons iteration.

In order to partition the algorithm to hardware and software we examine the behavior of these operational blocks using a profiler. In Figure 6 we present the profiling results of the five operational blocks, for a prediction horizon of 10 and variable control horizons. A direct observation is that the computation of the Gradient and the Hessian requires approximately $70 - 80\%$ of the

total optimization time. The next most expensive function is the matrix inversion which can take up to 30% of the total time, for large control horizons. Furthermore, we observe that for small control horizons the Gradient function uses almost half of the total optimization time and double the time of the Hessian function. Finally, we observe that by increasing the control horizon size, the matrix inversion becomes more expensive, and the computational time required by the Gradient and Hessian functions converge. This higher level analysis of the MPC optimization code reveals that the repetitive matrix operations of the Gradient and Hessian, comprise the major part of the processing. Therefore these specific matrix operations are the main part that has to be implemented efficiently, while the rest of the operations can be performed by a general purpose microprocessor.

5. CONCLUDING REMARKS

A selection of research results on embedding MPC were presented in this paper. These include profiling results of a general purpose processor, synthesis estimates of an ASIC chip, and codesign considerations for a co-processor FPGA implementation. During the presentation of the paper we intend to provide the results of the MPC running on the FPGA.

REFERENCES

Bergman, R. N., Y. Z. Ider, C. R. Bowden and C. Cobelli (1979). Quantitative estimation of insulin sensitivity. *American Journal of Physiology* **236**, E667–E677.

Bleris, L. G. and M. V. Kothare (2005*a*). Implementation of Model Predictive Control for Glucose Regulation using a General Purpose Microprocessor. In: *44th IEEE Conference on Decision and Control and European Control Conference*. Seville, Spain.

Bleris, L. G. and M. V. Kothare (2005*b*). Real-time implementation of model predictive control. In: *2005 American Control Conference*. Portland, OR. pp. 4166–4171.

Bleris, L. G., J. G. Garcia and M. V. Kothare (2005). Model predictive hydrodynamic regulation of microflows. In: *2005 American Control Conference*. Portland, OR. pp. 1752–1757.

Bleris, L. G., M. V. Kothare, J. G. Garcia and M. G. Arnold (2006*a*). Towards embedded model predictive control for system-on-a-chip applications. *Journal of Process Control* **16**, 255–264.

Bleris, L. G., P. Vouzis, M. G. Arnold and M. V. Kothare (2006*b*). Submitted: A Co-Processor FPGA Platform for the Implementation of Real-Time Model Predictive Control. In: *2006 American Control Conference*. Minneapolis, MI.

Brunner, C. S. (2004). Challenges and opportunities in emerging drug delivery technologies. Product Genesis Inc.

Doyle-III, F. J., R. S. Parker and E. P. Gatzke (2000). Advanced model predictive control for type I diabetic glucose control. In: *Proceedings of the 2000 American Control Conference*. Chicago, IL.

G. M. Steil, A. Volund, S. E. Kahn and R. N. Bergman (1993). Reduced sample number for calculation of insulin sensitivity and glucose effectiveness from the minimal model. *Diabetes* **42**, 250–256.

Garcia, J. G., M. G. Arnold, L. G. Bleris and M. V. Kothare (2004). LNS architectures for embedded model predictive control processors. In: *2004 International Conference on Compilers, Architectures and Synthesis for Embedded Systems*. Washington, D.C.. pp. 79 – 84.

Langer, R. (2004). Transdermal drug delivery: past progress, current status, and future prospects. *Advanced Drug Delivery Reviews* **56**, 557–558.

Leoni, L. and T. A. Desai (2001). Nanoporous biocapsules for the encapsulation of insulinoma cells: Biotransport and biocompatibility considerations. *IEEE Transactions on Biomedical Engineering* **48**, 1335–1341.

Pacini, G. and R. N. Bergman (1986). A computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intravenous glucose tolerance test. *Computer Methods and Programs in Biomedicine* **23**, 113–122.

Parker, R. S. and F.J. Doyle-III (2001). Control-relevant modeling in drug delivery. *Advances in Drug Delivery Reviews* **48(2)**, 211–248.

Parker, R. S., F. J. Doyle and N. A. Peppas (1999). A model-based algorithm for blood glucose control in type I diabetic patients. *IEEE Transactions on Biomedical Engineering* **46(2)**, 148157.

Riel, N. V. (2004). Minimal models for glucose and insulin kinetics. In: *Technique Report*. Eindhoven University of Technology.

Rubb, J. D. and R. S. Parker (2003). Glucose control in type I diabetic patients: A volterra model-based approach. In: *Proceedings of the 2003 ADCHEM*. Hong Kong, China.

Vouzis, P., L. G. Bleris, M. V. Kothare and M. G. Arnold (2005). Towards a Co-design Implementation of a System for Model Predictive Control. In: *2005 AIChE Annual Meeting*. Cincinnati, OH.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

FLEXIBLE RUN–TO–RUN STRATEGY FOR INSULIN DOSING IN TYPE 1 DIABETIC SUBJECTS

Cesar C. Palerm [*] **Howard Zisser** [**]
Lois Jovanovic [**,***] **Francis J. Doyle, III** [*,***,1]

[*] *Dept. of Chemical Engineering, University of California Santa Barabara*
[**] *Sansum Diabetes Research Institute, Santa Barbara, CA*
[***] *Biomolecular Science and Engineering Program, University of California Santa Barabara*

Abstract: People with type 1 diabetes require frequent adjustment of their insulin dose to maintain as near normal glycemia as possible. This process is not only burdensome, but for many difficult to achieve. As a result, control algorithms to facilitate the insulin dosage have been proposed, but have not been completely successful in normalizing glycemia. Here we present a novel run–to–run control algorithm to adjust the meal related insulin dose using only postprandial blood glucose measurements.

Keywords: biomedical control systems, batch control, insulin sensitivity, medical systems, diabetes, run–to–run control

1. INTRODUCTION

The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (2003) defines diabetes mellitus as a group of metabolic diseases which are characterized by hyperglycemia. This hyperglycemia results from defects in insulin secretion, insulin action, or both. Type 1 diabetes is caused by an absolute deficiency of insulin secretion. It includes cases primarily due to β cell destruction, and who are prone to ketoacidosis. These cases are those attributable to an autoimmune process, as well as those with β cell destruction for which no pathogenesis is known (i.e. idiopathic). People with type 1 diabetes fully depend on exogenous insulin. It is estimated that 17.1 million people world wide had type 1 diabetes in 2000 (Wild *et al.*, 2004; Eiselein *et al.*, 2004).

The chronic hyperglycemia in diabetes is associated with long-term complications due to damage, dysfunction and failure of various organs, especially the eyes, kidneys, nerves, heart and blood vessels. The three main complications being retinopathy, nephropathy and neuropathy. These can eventually lead to renal failure, blindness, amputation and other types of morbidity. Subjects with diabetes are at higher risk of cardiovascular disease, and face increased morbidity and mortality when critically ill.

The efficacy of intensive treatment in preventing diabetic complications has been established by the Diabetes Control and Complications Trial (DCCT) (Diabetes Control and Complications Trials Research Group, 1993) and the United Kingdom Prospective Diabetes Study (UKPDS) (UK Prospective Diabetes Study Group, 1998). In both trials the treatment regimens that reduced average glycosylated hemoglobin (a clinical

[1] Corresponding author (frank.doyle@icb.uscb.edu)

measure of glycemic control, which reflects average blood glucose levels over the preceding 2-3 months) A_{1C} to approximately 7% (normal range is 4-6%) were associated with fewer long term microvascular complications. Recent evidence even suggests that these target levels might not be low enough (Khaw *et al.*, 2001).

Intensive treatment requires multiple (3 or more) daily injections of insulin, or treatment with an insulin infusion pump. In any case, this tight control (*i.e.* as close to normal as possible) should be maintained for life in order to accrue the full benefits. Many factors influence the insulin dose requirements over time, including weight, physical condition and stress levels. Due to this, frequent blood glucose monitoring is required. Based on these measurements the insulin dosage must be modified, dietary changes implemented (such as alteration in the timing, frequency and content of the meals), as well as changes in activity and exercise patterns.

With the advent of home blood glucose monitoring technologies becoming available, physicians started to seek ways to use this information to fine–tune the therapeutic regimen. Among the first heuristic algorithms in the literature, we can highlight those of Skyler *et al.* (1981) and Jovanovic and Peterson (1982). Both set heuristic rules based on practical experience; the main difference between these two is that Skyler *et al.* (1981) relies on pre–prandial blood glucose measurements exclusively, while Jovanovic and Peterson (1982) uses prandial measurements as well to adjust the insulin dosing.

The algorithm proposed by Jovanovic and Peterson (1982) is taken as the basis to program a pocket computer, which was tested in 5 type 1 diabetic subjects. They demonstrate that computer–assisted insulin–delivery decision making is feasible (Chanoch *et al.*, 1985). This computer program was then compared to the standard approach for new continuous subcutaneous insulin infusion pump users. Peterson *et al.* (1986) found the approach to be feasible, although it did not fully normalize blood glucose levels. Still, computer users achieved lower average blood glucose and A_{1C}values over the course of the study.

Schiffrin *et al.* (1985) programmed a portable computer to adjust dosing of short and intermediate acting insulin in a 2–injection per day strategy, using pre–prandial blood glucose measurements. Even within the limitations of the therapy regimen used, they saw marked improvements in glycemic control when using the computer. Chiarelli *et al.* (1990) compared this computer method with a manual method; while they found no differences in glycemic control, they did notice fewer instances of hypoglycemia in the computer users. Peters *et al.* (1991) adapts this algorithm and compared its effectiveness against manual adjustments, finding that metabolic control and safety were comparable in both.

Taking the heuristic algorithm of Skyler *et al.* (1981) as their starting point, Beyer *et al.* (1990) create their own algorithms; as the original, they use pre–prandial blood glucose measurements. In a clinical trial of 50 subjects they clearly show that the computer group did much better than the regular intensive treatment group (Schrezenmeir *et al.*, 2002).

So far, none of these computer algorithms make use of the newer monomeric insulins. Owens *et al.* (2005) propose a run–to–run control algorithm to adjust the timing and dose of meal related insulin boluses, taking advantage of these fast acting insulin formulations. The basic assumption is that there is a sensor available from which frequent blood glucose measurements can be taken, and thus the maximum and minimum blood glucose excursions in the prandial period can be determined. The feasibility of the algorithm was studied in a clinical setting, making some changes to allow for fingerstick blood glucose determinations at 60 and 90 minutes after the start of the meal, in lieu of the maximum and minimum. Two-thirds of the subjects maintained acceptable glycemic control, but the rest diverged in their responses due to various factors (Zisser *et al.*, 2005).

In this work we modify the algorithm to overcome the difficulties encountered in clinical practice. The run–to–run formulation described here gives more flexibility to the subject, as blood glucose measurements are not required to be taken at specific times. In section 2 we present the basis of the run–to–run algorithm, followed by the specific implementation for insulin dosing. We present simulation results using this method in section 3.

2. RUN–TO–RUN ALGORITHM

The original formulation for the run–to–run control applied to insulin bolus dosing and timing is described in (Owens *et al.*, 2005). It is based on the application of a constraint control scheme in the run–to–run framework to optimize the operation of batch processes in the chemical industry (Srinivasan *et al.*, 2003*a*; Srinivasan *et al.*, 2003*b*).

The general run–to–run control algorithm is:

(1) Parameterize the input profile for run k, $u_k(t)$, as $\mathcal{U}(t, \nu_k)$. Also consider a sampled version, ψ_k, of the output $y_k(t)$, such that it has the same dimension as the controlled variable vector ν_k. Thus we have

$$\psi_k = F(\nu_k) \tag{1}$$

(2) Choose an initial guess for ν_k (when $k = 1$).
(3) Complete the run using the input $u_k(t)$ corresponding to ν_k. Determine ψ_k from the measurements $y_k(t)$.
(4) Update the input parameters as

$$\nu_{k+1} = \nu_k + K\left(\psi^r - \psi_k\right) \tag{2}$$

where K is an appropriate gain matrix and ψ^r represents the reference values to be attained. Increment k for the next run, and repeat steps 3-4 until convergence.

In the context of diabetes management, we use the natural day–to–day cycle as a run; within this run, there are three separate meals (namely breakfast, lunch and dinner), for which an appropriate insulin bolus has to be determined. The objective is to minimize the prandial glycemic excursion, without overdosing insulin. Thus, our manipulated variable, $u_k(t)$, corresponds to the insulin profile, and the measurement profile, $y_k(t)$, corresponds to glucose measurements. Time, t, is within a given day, k, which is also a run. Owens *et al.* (2005) show, using an RGA analysis, that there is effectively no coupling between the meals; we also use this assumption in the new algorithm.

There were two drawbacks to the original implementation when evaluated in a clinical setting. The first was the changing of the timing of the insulin bolus with respect to the start of the meal. Many times this resulted in a bolus being administered in the middle of a meal; at other times, the administration before the start of the meal was inconvenient to the subject, and was not adhered to. Besides, when using monomeric insulin, the timing of the bolus makes a negligible difference in the postprandial profile when compared with the effect of the dose. For these reasons it was decided to fix the timing to always coincide with the beginning of the meal. The second drawback was the need for blood glucose determinations at 60 and 90 minutes after the start of the meal; if the subject for some reason forgot to take either of them, then the algorithm was not able to correct for the following day (Zisser *et al.*, 2005).

The main change is in the selection of the performance measure used. To have the flexibility of taking blood glucose measurements at different times, we can no longer use a fixed glucose level. Instead, we use an approximation of the slope of the glycemic response. The only restrictions we place on the patient is that the first glucose measurement must be taken at least 60 minutes after the start of the meal, and the second one be at least 30 minutes after the first, but not more than 180 minutes after the start of the meal. We denote these times, for each meal, as: T_{B_1}, T_{B_2}, T_{L_1}, T_{L_2}, T_{D_1}, T_{D_2}. Then, our sampled output vector is

$$\psi_k = \begin{bmatrix} G(T_{B_1}) - G(T_{B_2}) \\ G(T_{L_1}) - G(T_{L_2}) \\ G(T_{D_1}) - G(T_{D_2}) \end{bmatrix} \tag{3}$$

As the times can change from one meal to the next, and from run to run, we need a reference value that is normalized with respect to time. We define this reference in terms of units of glucose per minute for each meal, ψ_0^r, and then scale by the actual time between the two measurements. We can write this as

$$\psi^r = \psi_0^r \circ \begin{bmatrix} T_{B_2} - T_{B_1} \\ T_{L_2} - T_{L_1} \\ T_{D_2} - T_{D_1} \end{bmatrix} \tag{4}$$

where $\circ$ denotes the Hadamard (element–wise) product.

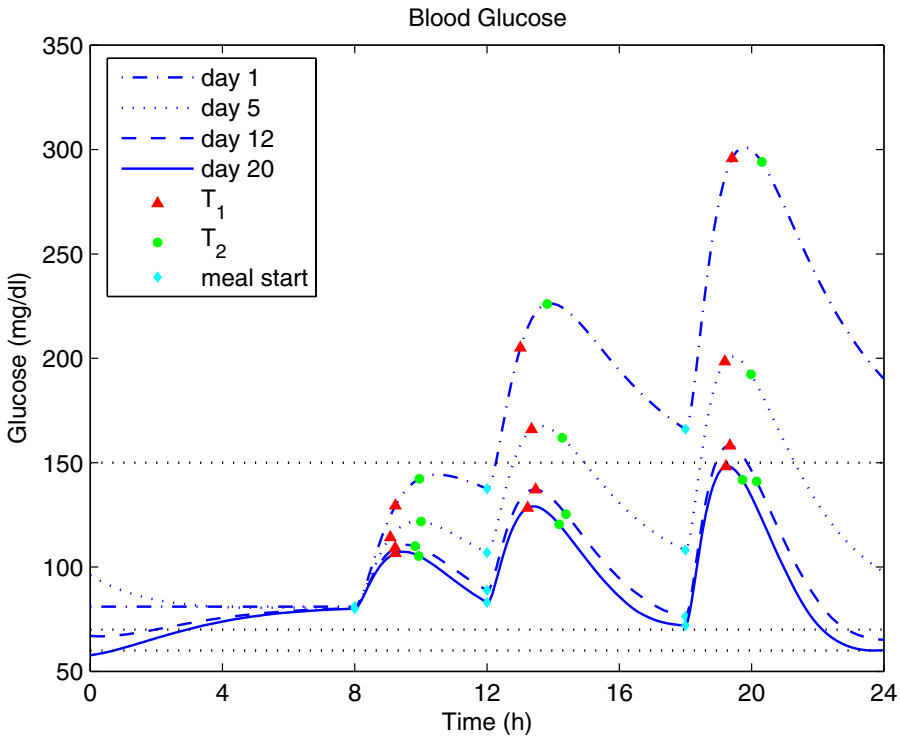
The manipulated variable ν_k is simply the dose of insulin corresponding to each meal of day k, $\nu_k = \begin{bmatrix} Q_B & Q_L & Q_D \end{bmatrix}^T$. The controller gain, K is set depending on the insulin sensitivity of the patient.

The reasoning for this performance measure is based on the blood glucose response seen for different doses. For a bolus that is correctly dosed, we expect the peak glucose excursion to be around 60 minutes, and to drop from that point on until it reaches the basal level. If the bolus is under–dosed, this moves the peak into the future. Thus, if we have under–bolused, the difference in blood glucose levels between the first and second measurements will be negative, or positive but very small. As the dose approaches the ideal level, this difference will increase. This is all illustrated in figure 1(a).

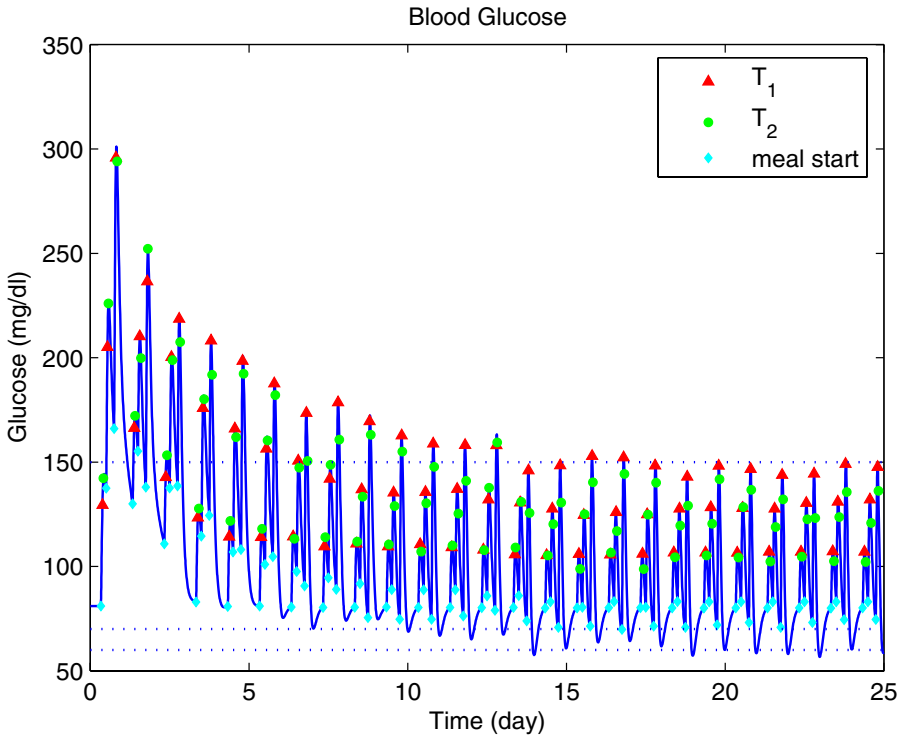
3. SIMULATION RESULTS

There are several published models of glucose and insulin dynamics in the literature. For this particular study we have selected the one published by Hovorka *et al.* (2004), replacing the subcutaneous insulin infusion model with the one described in (Wilinska *et al.*, 2005). The model captures not only the dynamics of glucose and insulin, but also the absorption of insulin from a subcutaneous delivery (as is the case with insulin infusion pumps), and the appearance of glucose in plasma from a mixed meal.

For each day, the simulation has the meals at 8:00, 12:00 and 18:00 hours, with a carbohydrate content of 20, 40 and 70 grams, respectively. For each day and meal, the timepoints at which blood glucose measurements are taken are selected randomly (using a uniform distribution); the first one can take place from 60 to 90 minutes after the start of the meal, the second one follows 30 to 60 minutes later.



(a) Glucose profile for selected days



(b) Glucose profile over a period of 25 days

Fig. 1. In (a) it can clearly be seen that the time between sampling times changes for the different meals, and shows how the run–to–run algorithm is able to bring the dosing within the desired bounds. (b) shows the full profile over 25 consecutive days.

The reference drop in blood glucose (per minute), was selected for each meal separately, considering the typical amount of carbohydrate consumed in each meal as the main guideline. We have selected $\psi_0^r = \begin{bmatrix}0.058 & 0.104 & 0.30\end{bmatrix}^T$. The controller gain is set at $K = 0.0005$, and is scaled by 2, 3 or 4 for subjects with lower insulin sensitivities. The

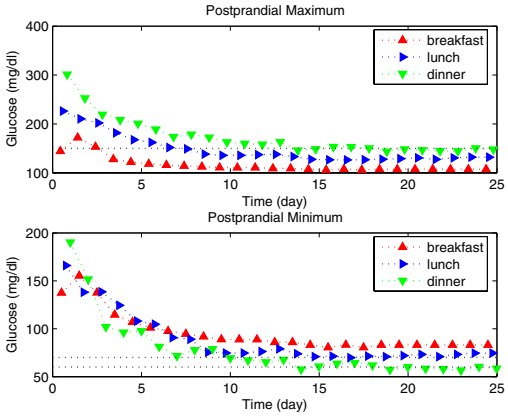


Fig. 2. Maximum and minimum glucose excursions after a meal converge to clinically acceptable bounds.

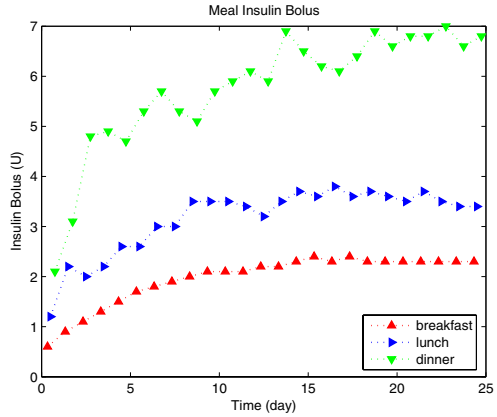


Fig. 3. Meal insulin bolus converges to the optimal amount for the given meal.

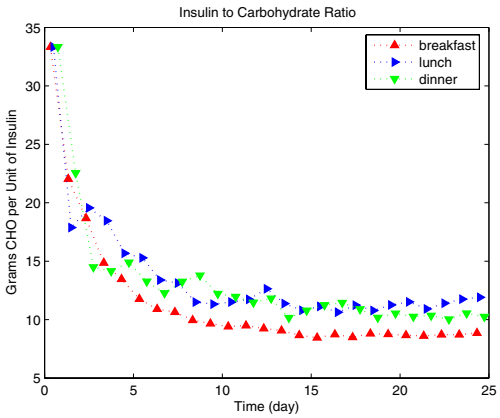


Fig. 4. The algorithm converges to the same insulin to carbohydrate ratio, regardless of the carbohydrate content of the meal.

amount of the insulin bolus is rounded to the nearest 0.1 U of insulin, which is the resolution of most infusion pumps.

The initial guess for the insulin requirement for each meal is set at an insulin to carbohydrate ratio of 1:33 (a more typical value is around 1:10). Thus we start giving much less insulin than is actually required for the first run ($k = 0$). Figure 1(b) shows the simulation for 25 days, with figure 1(a) highlighting a couple of days only. The dotted lines show the desired bounds for the blood glucose excursions; note that we are more aggressive in keeping blood glucose below 150 mg/dl than preventing it from going below 70 mg/dl.

Even though the algorithm does not directly consider the minimum and maximum excursions after a meal, these are still relevant clinical markers. Figure 2 shows the maximum and minimum values after each meal, where once again the dotted lines represent the desirable bounds. The amount of the insulin bolus and the corresponding insulin to carbohydrate ratios are shown in figures 3 and 4, respectively. The insulin to carbohydrate ratio is what the patients and physicians use to calculate their insulin requirements for a given meal; this shows clearly that the algorithm converges to the ideal ratio. It is important to note that although in this case they converge to approximately the same value, it is not necessarily the case in real life, as insulin sensitivity has a circadian variation which is not captured by the simulation model used.

4. CONCLUSIONS

The feasibility of using run–to–run control to determine the optimal insulin bolus dose and timing was shown by Zisser *et al.* (2005), but some hurdles were identified. Changing the timing of the insulin bolus was one of them, which coupled with the small difference it makes when using monomeric insulin, it was decided to keep it fixed to coincide with the beginning of the meal. The second was the requirement that blood glucose measurements be taken at 60 and 90 minutes; besides imposing additional burden on the patient to keep close track of time after a meal, it also meant that when the patient missed these time points the algorithm could no longer make a correction for the dosing the following day.

We have proposed a new performance measure, which gives the patient the freedom of taking post–prandial glucose measurements at times that are more flexible and do not require them to become slaves to the clock. We have shown that even with this variation in the timing, the controller is able to converge within a couple of days, significantly improving the degree of glycemic control.

Further simulation studies must be done to incorporate other sources of variability that are expected, including measurement noise, mismatch between the estimated carbohydrate content of

the meal and the actual value, and variation in the timing and carbohydrate content of the meals. Initial results (not shown) are quite encouraging. We are currently undertaking a robustness analysis that takes into account all of these sources of uncertainty.

5. ACKNOWLEDGEMENTS

We acknowledge the support from the National Institutes of Health (grants R01–DK068706–02 and R01–DK068663–02) that have made this work possible.

REFERENCES

Beyer, J., J. Schrezenmeir, G. Schulz, T. Strack, E. Kstner and G. Schulz (1990). The influence of different generations of computer algorithms on diabetes control. *Comput Methods Programs Biomed* **32**(3-4), 225–232.

Chanoch, L. H., L. Jovanovic and C. M. Peterson (1985). The evaluation of a pocket computer as an aid to insulin dose determination by patients.. *Diabetes Care* **8**(2), 172–176.

Chiarelli, F., S. Tumini, G. Morgese and A. M. Albisser (1990). Controlled study in diabetic children comparing insulin-dosage adjustment by manual and computer algorithms. *Diabetes Care* **13**(10), 1080–1084.

Diabetes Control and Complications Trials Research Group (1993). The effect of intensive treatment of diabetes on the development and progression of long–term complications in insulin–dependent diabetes mellitus. *N Engl J Med* **329**, 977–986.

Eiselein, L., H. J. Schwartz and J. C. Rutledge (2004). The challenge of type 1 diabetes mellitus. *ILAR J* **45**(3), 231–236.

Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (2003). Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* **26**(s1), s5–s20.

Hovorka, R., V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. O. Federici, T. R. Pieber, H. C. Schaller, L. Schaupp, T. Vering and M. E. Wilinska (2004). Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiol Meas* **25**(4), 905–20.

Jovanovic, L. and C. M. Peterson (1982). Home blood glucose monitoring. *Compr Ther* **8**(1), 10–20.

Khaw, K.T., N. Wareham, R. Luben, S. Bingham, S. Oakes, A. Welch and N. Day (2001). Glycated haemoglobin, diabetes, and mortality in men in Norfolk cohort of European Prospective Investigation of Cancer and Nutrition (EPIC-Norfolk). *British Medical Journal* **322**(7277), 15–18.

Owens, C. L., H. Zisser, L. Jovanovic, B. Srinivasan, D. Bonvin and F. J. Doyle, III (2005). Run–to–run control of blood glucose concentrations for people with type 1 diabetes mellitus. *IEEE Trans Biomed Eng*, submitted.

Peters, A., M. Rübsamen, U. Jacob, D. Look and P. C. Scriba (1991). Clinical evaluation of decision support system for insulin-dose adjustment in IDDM. *Diabetes Care* **14**(10), 875–880.

Peterson, C. M., L. Jovanovic and L. H. Chanoch (1986). Randomized trial of computer-assisted insulin delivery in patients with type I diabetes beginning pump therapy. *Am J Med* **81**(1), 69–72.

Schiffrin, A., M. Mihic, B. S. Leibel and A. M. Albisser (1985). Computer-assisted insulin dosage adjustment. *Diabetes Care* **8**(6), 545–552.

Schrezenmeir, J., K. Dirting and P. Papazov (2002). Controlled multicenter study on the effect of computer assistance in intensive insulin therapy of type 1 diabetics. *Comput Methods Programs Biomed* **69**(2), 97–114.

Skyler, J. S., D. L. Skyler, D. E. Seigler and M. J. O'Sullivan (1981). Algorithms for adjustment of insulin dosage by patients who monitor blood glucose. *Diabetes Care* **4**(2), 311–318.

Srinivasan, B., D. Bonvin, E. Visser and S. Palanki (2003*a*). Dynamic optimization of batch processes: II. role of measurements in handling uncertainty. *Comput Chem Eng* **27**(1), 27–44.

Srinivasan, B., S. Palanki and D. Bonvin (2003*b*). Dynamic optimization of batch processes: I. characterization of the nominal solution. *Comput Chem Eng* **27**(1), 1–26.

UK Prospective Diabetes Study Group (1998). Intensive blood–glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* **352**, 837–853.

Wild, S., G. Roglic, A. Green, R. Sicree and H. King (2004). Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* **27**(5), 1047–1053.

Wilinska, M. E., L. J. Chassin, H. C. Schaller, L. Schaupp, T. R. Pieber and R. Hovorka (2005). Insulin kinetics in type–1 diabetes: Continuous and bolus delivery of rapid acting insulin. *IEEE Trans Biomed Eng* **52**(1), 3–12.

Zisser, H., L. Jovanovic, F. Doyle, III, Paulina Ospina and Camelia Owens (2005). Run–to–run control of meal–related insulin dosing. *Diabetes Technol Ther* **7**(1), 48–57.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

NONLINEAR MODEL PREDICTIVE CONTROL FOR OPTIMAL DISCONTINUOUS DRUG DELIVERY

Nicolas Hudon * **Martin Guay** *,[1] **Michel Perrier** **
Denis Dochain ***

* *Queen's University, Kingston, ON, Canada*
** *École Polytechnique Montréal, QC, Canada*
*** *Université Catholique de Louvain,Louvain-la-Neuve, Belgium*

Abstract: This paper exploits a gradient-based model predictive control technique to solve an optimal switching time problem over periodic orbits. Drug delivery scheduling applications, where it is desired to maximize the averaged effect of a drug over time, motivate the study for this type of online optimization problem. The objective is to find the optimal time-switching policy between full treatment and no treatment periods. It is shown, by a numerical application to a simple drug delivery problem, that the resulting predictive algorithm drives the system to the optimal periodic orbit in the state space.

Keywords: Model Predictive Control, Time-switching Control, Optimal Drug Delivery, Periodic Orbits.

1. INTRODUCTION

The usual task of nonlinear model predictive control is to find and track the steady-state optimum of a cost functional, subject to the system dynamics and state constraints. In some applications, however, a steady-state optimization may not be feasible nor optimal with respect to a given measure. For example, as outlined in (Varigonda *et al.*, 2004*a*; Varigonda *et al.*, 2004*b*), the steady-state optimization of some drug delivery problems yield optimum conditions that do not lead to therapeutic drug treatments. From a practical point of view, optimal drug delivery problems can be seen as optimal time-switching problems, alternating full treatment periods and no treatment periods. For HIV control, optimal scheduling policies were proposed in (Zurakowski and Teel, 2003; Zurakowski *et al.*, 2004). In (Guay *et al.*, 2005), differential flatness was used to parameterize the trajectories of the system and to compute, in real-time, optimal periodic trajectories using an extremum-seeking method. Since the problem is periodic by nature, we will study it as an optimal control problem over periodic orbits.

In this paper, the problem of optimal drug delivery by periodic injections is solved as an optimal time-switching control problem. Each control cycle includes periods of treatment, or short input impulses, and periods with no treatment. The control is parameterized by a time-switching parameter that should converge to the optimal time interval length between full treatment periods. This problem was treated in the linear case by (Yastreboff, 1969) and more recently using heuristics in (Grognard and Sepulchre, 2001). Asymptotic stabilization for linear output feedback systems was studied in (Allwright *et al.*, 2005). Gradient-based algorithm to solve an anal-

[1] Corresponding Author guaym@chee.queensu.ca

ogous problem was also presented in (Egerstedt *et al.*, 2003).

In this paper, we use a nonlinear model predictive approach, recently developed in (DeHaan and Guay, 2005) that generalize the approach given by (Magni and Scattolini, 2004). The key idea is to see the optimal control moves (here parameterized by switching-time between two control values) as unknown parameters that can be identified on-line via the model predictive control algorithm. The method relies on discrete transitions of the control action based on real-time evolution of those parameters. We allow maximization calculations throughout the entire sampling interval and update the control parameters at a fixed sampling time. The assumptions needed for this implementation relax the flatness assumption used in (Guay *et al.*, 2005) and (Varigonda *et al.*, 2004*b*).

The paper is divided as follows. In Section 2, we formulate the optimization problem and parameterize the single input to the system using time at which control switches occur as the control parameter. In Section 3, we present the nonlinear model predictive algorithm based on (DeHaan and Guay, 2005) and discuss stability of the closed loop scheme. Numerical application of the resulting law to the drug delivery problem from (Varigonda *et al.*, 2004*b*) is presented in Section 4. Conclusions and future investigations are outlined in Section 5

2. CONTROL PROBLEM FORMULATION

We consider a single input nonlinear dynamical system of the form:

$$\dot{x} = f(x, u) \tag{1}$$

where $x \in \mathbb{X} \subset \mathbb{R}^n$ are the states of the system, $u \in \mathbb{U} = \{u_{\min},\ u_{\max}\}$ is the control input and $f : \mathbb{X} \times \mathbb{U} \to \mathbb{R}^n$ is a sufficiently smooth function. We assume that $\mathbb{X}$ is a compact subset of $\mathbb{R}^n$. We also assume that the user-defined objective functional is a convex differentiable function on $\mathbb{X}$. The control design objective is to optimize a cost functional given by

$$J = \frac{1}{T}\int_t^{t+T} L(x(\tau))d\tau \tag{2}$$

with respect to $u(\tau)$ for $\tau \in [t,\ t+T]$, where T is the fixed period of the system, approximated here as the length of the horizon considered later for the optimization problem. We seek to maximize J subject to the system dynamics (1) and inequality constraints

$$x_{\min} \le x(\tau) \le x_{\max}, \quad \tau \in [t,\ t+T] \tag{3}$$

We consider the problem of finding an optimal switching time between two known values of the control inputs, i.e. $u_{\min}$ and $u_{\max}$. To represent this type of behavior, we parameterize the control $u(\tau)$ as a finite sum of Heaviside functions $u(\theta)$:

$$u(\theta) = \sum_{i=1}^{m} [H(\tau - i\theta) - H(\tau - i\theta - \varepsilon)] \tag{4}$$

where θ is the switching time parameter. This parameter will be determined on-line by the optimization algorithm. The parameter ε is the known duration of each full-treatment period. In practice, m would be chosen such that $m \cdot \theta < T$ with a meaningful prediction horizon, T. Other parameterizations could have been considered, however Heaviside functions clearly represent the physical application of discontinuous drug infusion. Another advantage here is that the calculations will be greatly simplified using some elementary proprieties of Heaviside functions and Dirac delta operator. We now state the following assumption:

Assumption 2.1. The unforced dynamics (1) with $u(t) \equiv u_{\min} \equiv 0$ is stable.

This assumption is needed to ensure stability of the close loop dynamics as discuss in the next Section.

3. NONLINEAR MODEL PREDICTIVE CONTROL

3.1 Interior-Point Method

In order to find the optimal control policy that steers the system (1) to the periodic orbit maximizing the cost functional (2), we have to encode the state constraints (3). We propose log-barrier functions (Nash and Sofer, 1996). The cost functional (2) becomes the following:

$$J_c = \frac{1}{T}\int_t^{t+T} \left(L\left(x(\tau)\right)\right) + R_1(x(\tau)) + R_2(x(\tau))\right) d\tau \tag{5}$$

where

$$R_1(x(\tau)) = \sum_{j=1}^{n} \mu_{1,j} \log(x_j(\tau) - x_{j,\max} - \epsilon_{1,j})$$

$$R_2(x(\tau)) = \sum_{j=1}^{n} \mu_{2,j} \log(x_{j,\min} - x_j(\tau) + \epsilon_{2,j})$$

and $\mu_{\cdot,j} > 0$, $\epsilon_{\cdot,j} > 0$, $j = 1, \ldots, n$, are tuning constants for the barrier functions. Given that the functional is convex with respect to the unknown control parameter θ, we can rely on the first order conditions for optimality, given by

$$\nabla_\theta J_c(\theta^*) = 0 \tag{6}$$

where $\nabla_\theta J_c(\theta^*)$ is the gradient of the functional J_c with respect to θ evaluated at the minimizer θ^*. From the definition of the cost functional (5), this gradient is expressed as

$$\nabla_\theta J_c(\theta) = \frac{1}{T}\int_t^{t+T} \Gamma_1 \frac{\partial x}{\partial u}\frac{\partial u}{\partial \theta}\, d\tau \tag{7}$$

where Γ_1 is the n-row vector defined by

$$\Gamma_1 = \left(\frac{\partial L}{\partial x} + \frac{\partial R_1}{\partial x} + \frac{\partial R_2}{\partial x}\right)^T \tag{8}$$

with each j^{th}-element, $j = 1, \ldots, n$, is given by

$$\Gamma_{1_j} = \frac{\partial L}{\partial x_j} + \frac{\mu_{1,j}}{x_j(\tau) - x_{j,\max} - \epsilon_{1,j}} - \frac{\mu_{2,j}}{x_{j,\min} - x(\tau) + \epsilon_{2,j}} \tag{9}$$

The first derivative of x with respect to u can be evaluate along the trajectories of the following tractable dynamics

$$\frac{d}{dt}\left(\frac{\partial x}{\partial u}\right) = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial u} \tag{10}$$

By the parametrization (4) of $u(\theta)$, we have :

$$\frac{\partial u}{\partial \theta} = -\sum_{i=1}^{N} i\Big(\delta(t - i\theta) - \delta(t - i\theta - \varepsilon)\Big) \tag{11}$$

where $\delta(\cdot)$ is the Dirac delta function. By definition,

$$\int_\Lambda f(x)\delta(x-a)dx = f(a), \quad \text{if } a \in \Lambda \tag{12}$$

Therefore, we can rewrite (7) as

$$\nabla_\theta J_c(\theta) = \frac{1}{T}\sum_{i=1}^{N} i\left[\Gamma_1 \frac{\partial x}{\partial u}\right]_{i\theta}^{i\theta+\varepsilon} \tag{13}$$

where $[F(\cdot)]_a^b = F(b) - F(a)$. Equations (13) shows the dependence of the cost functional J_c on the control parameter θ. We will use this information in the next section to derive a stable updating law for θ. The main advantage of the proposed parametrization is that the cost function gradient to evaluate is now given by a finite sum of trackable terms.

3.2 Parameter Update Law

In this section, we apply the nonlinear model predictive control procedure proposed in (DeHaan and Guay, 2005). The idea is to assume that we can compute the model prediction instantaneously solving the closed-loop dynamics:

$$\begin{aligned} \dot{x} &= f(x, u(\theta)) \\ \dot{\theta} &= \Psi(t, x) \end{aligned} \tag{14}$$

The continuous update law $\Psi(t,x)$ must be chosen such that $\langle \nabla_\theta J_c, \Psi(t,x)\rangle \leq 0$. In (DeHaan and Guay, 2005), one way to achieve this criterion is to use a general descent continuous update law:

$$\Psi(t,x) = \text{Proj}\{\vartheta(t,x)\} \tag{15}$$

$$\vartheta(t,x) = k_\theta \Upsilon \nabla_\theta J_c \tag{16}$$

In the present paper, Υ is set to I (gradient-based method). The projection algorithm Proj$(\cdot)$ is designed to ensure that the value of the parameter remains in the convex set

$$\Omega_w = \{\theta \in \mathbb{R} : |\theta| \leq w_m\} \tag{17}$$

This algorithm is given by

$$\dot{\theta} = \text{Proj}\{\theta, \vartheta\} = \begin{cases} \vartheta & \text{if } |\theta| < w_m \\ & \text{or } |\theta| = w_m \\ & \text{and } \nabla\mathcal{P}(\theta)\vartheta \leq 0 \\ \vartheta - \vartheta\dfrac{\lambda \nabla\mathcal{P}(\theta)\nabla\mathcal{P}(\theta)^T}{\|\nabla\mathcal{P}(\theta)\|_\lambda^2} & \text{otherwise} \end{cases} \tag{18}$$

where $\mathcal{P}(\theta) = \theta^2 - w_m \leq 0$, λ is a positive constant gain for the projection algorithm. General properties of this projection algorithm are presented in (Krstic *et al.*, 1995) and (Pomet and Praly, 1992).

3.3 Convergence to the Optimal Cycle

Following the extremum seeking procedure proposed in (Guay and Zhang, 2003), we use the following Lyapunov function:

$$V = \frac{1}{2}|J_c(\theta)|^2 \geq 0 \tag{19}$$

The derivative of V with respect to time is

$$\dot{V} = |J_c(\theta)|\left(\nabla_\theta J_c(\theta)\dot{\theta} + \Gamma_1(T+t) - \Gamma_1(t)\right) \tag{20}$$

It is possible to show by a simple Lyapunov argument that the convergence of the algorithm to the optimal cycle is ensured if $\dot{J}_c \to 0$ as $t \to 0$. To achieve stability, the open-loop dynamics must be such that:

$$\frac{\partial \Gamma_1}{\partial x} f(x, 0) < 0 \tag{21}$$

Since by assumption the open-loop dynamics are stable for $u \equiv 0$, the last condition is met for $\frac{\partial \Gamma_1}{\partial x} > 0$. Therefore, the optimum is reached when the cost stabilizes to a constant value, that is

when we reach the optimal closed orbit. Since the dynamic for $\dot{\theta}$ is stable, the algorithm ensures convergence to the optimal periodic orbit whenever θ reaches is optimal value.

3.4 Receding Horizon Implementation

To implement the algorithm derived above in real-time, we need to evaluate on-line sensitivity information of the state trajectories with respect to the control. Moreover, to evaluate the gradient, we need the sensitivity of those equations with respect to time, as expressed in the equations (10). Following ideas presented in (DeHaan and Guay, 2005), the proposed method uses the simulation of the system (1) with u generated by a fix switching time θ over the receding horizon $\tau \in [t,\ t + T]$, corresponding to one assumed period of the system. This enables us to generate the gradient information and to update θ according to equation (14). The new value of the parameter θ is changed at the end of the cycle to generate another free dynamic of the system.

To summarize the algorithm:

(1) Assume fixed switching time θ and prediction horizon T.
(2) Simulate the system with discontinuous inputs for t to $t + T$.
(3) Compute the gradient $\nabla_\theta J_c(\theta)$ as the finite sum 13.
(4) At time $t = \theta$, update θ and the horizon T.

Results from (DeHaan and Guay, 2005) show a computational advantage of this method over existing receding horizon techniques. We now turn our attention to a simple numerical example to show the potential of the method.

4. APPLICATION TO DRUG DELIVERY

To illustrate, we apply the algorithm developed in Section 3 to a drug delivery problem studied in (Guay *et al.*, 2005; Varigonda *et al.*, 2004*b*). The objective is to maximize the time average of some indicator function of the drug concentration, c and the drug antagonist concentration, a:

$$J = \frac{1}{T}\int_0^T I(E(c,a))d\tau \tag{22}$$

where the indicator function is defined as

$$I(E) = \frac{(E/E_1)^\gamma}{[1+(E/E_1)^\gamma][1+(E/E_2)^{2\gamma}]} \tag{23}$$

and the drug effect E is

$$E(c,a) = \frac{c}{(1+c)(1+a/a^*)} \tag{24}$$

where a^* is the relative potency of the antagonist. The prescribed range of the effect of the drug during the therapy is enforce by the parameters E_1 and E_2 in the indicator function. To be effective, the therapy must lie within the interval $[E_1,\ E_2]$ during the cycle. The parameter γ is used to increase the sensitivity of the indicator to changes in the drug effect. The non-dimensional linear dynamics of the systems are given by

$$\dot{c} = -c + u \tag{25}$$
$$\dot{a} = K_a(c - a) \tag{26}$$

with K_a, the rate constant of the antagonist elimination. We constrain the states with $0 \leq a \leq 1$ and $0 \leq c \leq 1$. In this region, the unforced dynamics converges to the origin. The first derivative of x with respect to u is given by:

$$\frac{d}{dt}\begin{pmatrix}\frac{dx}{du}\end{pmatrix} = \begin{pmatrix}-1 & 0\\ K_a & -K_a\end{pmatrix}\begin{pmatrix}\frac{dx}{du}\end{pmatrix} + \begin{pmatrix}1\\0\end{pmatrix} \tag{27}$$

Simulations and control parameters are given in Tables 1 and 2 respectively.

Table 1. Drug Delivery Problem Parameters

K_a	0.1
a^*	1
E_1	0.3
E_2	0.6
γ	10
$u_{\min}$	0
$u_{\max}$	1
ε	1

Table 2. Control Parameters

κ	0.1
w_m	10
λ	1
T	10
$\mu_{1,2}$	1
$\epsilon_{1,2}$	0.01

Simulation results of the state variable trajectories are given in Figure 1. A phase plane diagram is shown in Figure 2. From that figure, we see how the system is driven to a stable periodic orbit and how this periodic orbit is moved to the optimal one.

The trajectory of the switching-time parameter in Figure 3 shows that the control procedure with the optimal parameter θ.

The cost function value over time is represented in Figure 4. The effect of the drug over time and the value of the indicator function are presented in Figure 5. From this figure, we see that the therapeutic range $[E_1,\ E_2]$ is reached at each cycle.

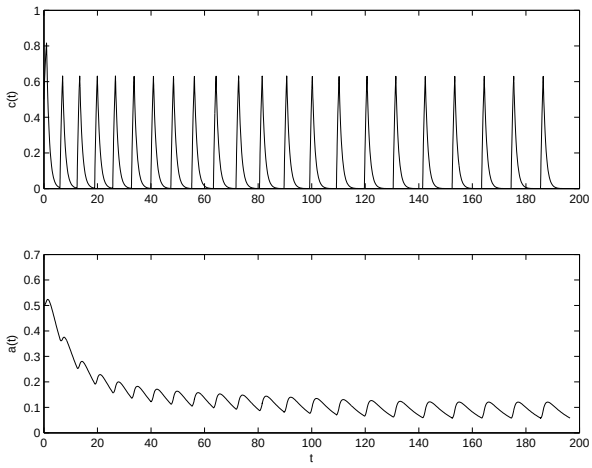


Fig. 1. States trajectories over time

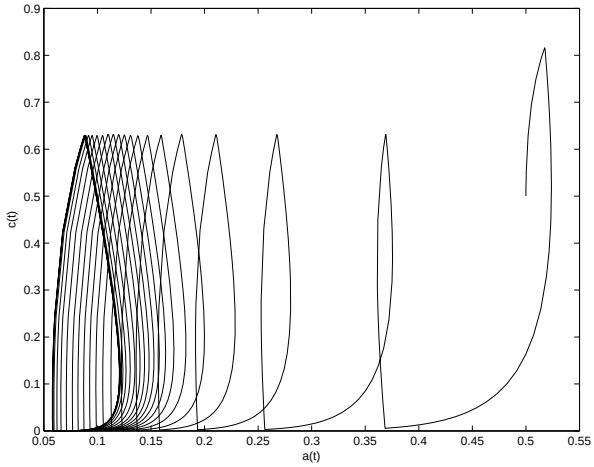


Fig. 2. States in the phase space

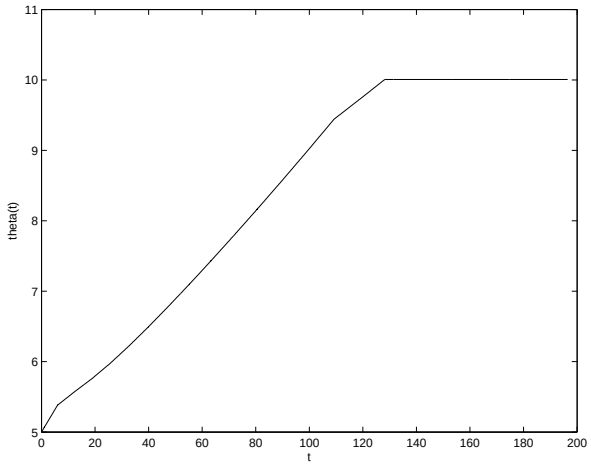


Fig. 3. Evolution of the switching time θ

5. CONCLUSION

In this paper, we posed and solved a single input optimal control problem over periodic orbits using extremum seeking and receding horizon techniques. The method proposed used parametrization of the control as a sequence of time switching.

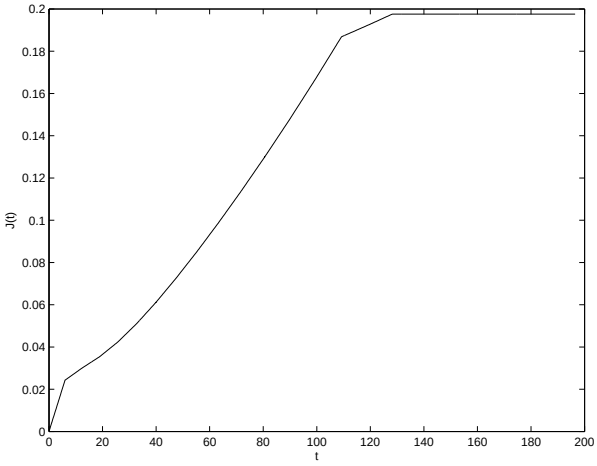


Fig. 4. Cost function over time

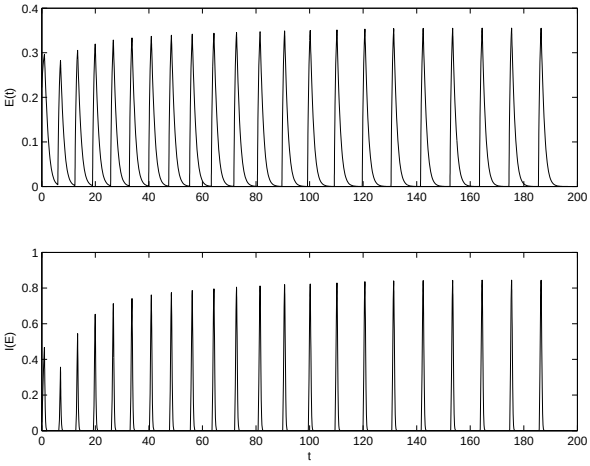


Fig. 5. Drug effect and indicator function over time

Future investigations will focus on adaptive extensions of the method with applications to drug delivery optimal problems with unknown parameters. Other applications with impulse controls can also be considered (see for example the bipedal robot application (Morris and Grizzle, 2005) and the wedge billiard (Sepulchre and Gerard, 2004)).

REFERENCES

Allwright, J.C., A. Astolfi and H.P. Wong (2005). A Note on Asymptotic Stabilization of Linear Systems by Periodic, Piecewise Constant, Output Feedback. *Automatica* **41**, 339–344.

DeHaan, D. and M. Guay (2005). A New Real-Time Framework for Nonlinear Model Predictive Control of Continuous-Time Nonlinear Systems. In: *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference*. Seville, Spain. pp. 957–962.

Egerstedt, M., Y. Wardi and F. Delmotte (2003). Optimal Control of Switching Times in Switched Dynamical Systems. In: *Proceedings*

of the 42nd IEEE Conference on Decision and Control. Maui, Hawaii. pp. 2138 – 2143.

Grognard, F. and R. Sepulchre (2001). Global Stability of a Continuous-time Flow which Computes Time-optimal Switchings. In: *Proceedings of the 40th IEEE Conference on Decision and Control*. Orlando, FA. pp. 3826–3831.

Guay, M. and T. Zhang (2003). Adaptive Extemum Seeking Control of Nonlinear Dynamic Systems with Parametric Uncertainties. *Automatica* **39**, 1283–1293.

Guay, M., D. Dochain, M. Perrier and N. Hudon (2005). Extremum Control Over Periodic Orbits. In: *Proceedings of the 16th IFAC World Congress*. Prague, Czech Republic.

Krstic, M., I. Kanellakopoulos and P. Kokotovic (1995). *Nonlinear and Adaptive Control Design*. Wiley and Sons. NY.

Magni, L. and R. Scattolini (2004). Model Predictive Control of Continuous-time Nonlinear Systems with Piecewise Constant Control. *IEEE Transactions on Automatic Control* **49**(6), 900–906.

Morris, B. and J.M. Grizzle (2005). A Restricted Poincaré Map for Determining Exponentially Stable Periodic Orbits in Systems with Impulse Effects: Application to Bipedal Robots. In: *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference*. Seville, Spain. pp. 4199–4206.

Nash, S.G. and A. Sofer (1996). *Linear and Nonlinear Programming*. McGraw-Hill.

Pomet, J.B. and L. Praly (1992). Adaptive Nonlinear Regulation: Estimation from the Lyapunov Equation. *IEEE Transactions on Automatic Control* **37**(6), 729–740.

Sepulchre, R. and M. Gerard (2004). Stabilization Through Weak and Occasional Interactions : A Billiard Benchmark. In: *Proceedings of the 6th IFAC Symposium on Nonlinear Control Systems*. Stuttgart, Germany. pp. 75–80.

Varigonda, S., T.T. Georgiou and P. Daoutidis (2004*a*). Numerical Solution of the Optimal Periodic Control Problem using Flatness. *IEEE Transactions on Automatic Control* **49**(2), 271–275.

Varigonda, S., T.T. Georgiou, R.A. Siegel and P. Daoutidis (2004*b*). Optimal Periodic Control of a Drug Delivery System. In: *Proceedings of IFAC DYCOPS*. Boston, MA.

Yastreboff, M. (1969). Synthesis of time-optimal control by time intreval adjustment. *IEEE Transactions on Automatic Control* **14**(6), 707–710.

Zurakowski, R. and A.R. Teel (2003). Enhancing Immune Response to HIV Infection Using MPC-Based Treatment Scheduling. In: *Proceedings of the American Control Conference*. Denver, CO.

Zurakowski, R., A.R. Teel and D. Wodarz (2004). Utilizing Alternate Target Cells in Treating HIV Infection Through Scheduled Treatment Interruptions. In: *Proceedings of the American Control Conference*. Boston, MA.

Session 4.2

Bioprocess Modeling and Identification

Optimal Experiment Design in Bioprocess Modelling:From Theory to Practice

A. M. Cappuyns, K. Bernaerts, I. Y. Smets, O. Ona, E. Prinsen, J. Vanderleyden and J. F. Van Impe
Katholieke Universiteit Leuven

Dynamic Modelling of a Biofilter Used for Nitrification of Drinking Water at Low Influent Ammonia Concentrations

Queinnec, J. C. Ochoa, E. Paul and A. VandeWouwer
Le Centre National de la Recherche Scientifique - Facult´e Polytechnique de Mons

Dynamic PCA for Phase Identification of Rifamycin B Fermentation in Multi-Substrate Complex Media

X. T. Doan, R. Srinivasan, P. M. Bapat, and P. P. Wangikar
Institute of Chemical and Engineering Sciences

A New Model of Phenol Biodegradation and Activated Sludge Growth in Fedbatch Cultures

C. Ben-Youssef, J. Waissman and G. Vázquez
Univerdidad Politécnica de Pachuca

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

OPTIMAL EXPERIMENT DESIGN IN BIOPROCESS MODELING: FROM THEORY TO PRACTICE

Astrid M. Cappuyns * **Kristel Bernaerts** *
Ilse Y. Smets * **Ositadinma Ona** ** **Els Prinsen** ***
Jos Vanderleyden ** **Jan F. Van Impe** *

* *BioTeC - Katholieke Universiteit Leuven,*
W. de Croylaan 46, B-3001 Leuven (Belgium)
Fax: +32-16-32.29.91
email: jan.vanimpe@cit.kuleuven.be
** *CMPG - Katholieke Universiteit Leuven,*
Kasteelpark Arenberg 20, B-3001 Leuven (Belgium)
*** *Department of biology, University of Antwerp,*
Universiteitsplein 1, B-2610 Antwerpen (Belgium)

Abstract: In this paper the problem of parameter identification for the Monod model is considered. As known for a long time, noisy batch measurements do not allow unique and accurate estimation of the kinetic parameters of the Monod model. Techniques of optimal experiment design are, therefore, addressed to design informative experiments and improve the parameter estimation accuracy. During the design process, practical feasibility has to be kept in mind. In this paper it is demonstrated how a theoretical optimal design can successfully be translated to a feasible optimal design. Both design and validation of informative fed batch experiments are illustrated with a case study that models the growth of the nitrogen fixing bacteria *Azospirillum brasilense*.

Keywords: optimal experiment design, parameter identification, Monod kinetics, *Azospirillum brasilense*, bioreactor

1. INTRODUCTION

When modeling (bio)chemical processes, some limitations which will hamper the model identification process, have to be kept in mind (Bernaerts and Van Impe, 2004). To overcome these problems, an accurate design of the experiments is needed. Experimental data should contain sufficient information in order to enable correct model structure characterization, and accurate and unique parameter estimation. It has been demonstrated that the use of optimal experiment design for parameter estimation can contribute to an improvement of the parameter estimation accuracy (e.g., Walter and Pronzato, 1997; Versyck and Van Impe, 1999).

Azospirillum brasilense belongs to a group of bacteria that exert beneficial effects on plant growth. One of the factors responsible for the plant growth promotion is the production of phytohormones, e.g., the auxin indole-3-acetic acid (Baldani *et al.*, 1983). This characteristic opens perspectives to exploit the nitrogen fixing bacteria of the genus *Azospirillum* as alternative for, or supplement to chemical fertilization. Therefore, a quantitative analysis of growth and phytohormone production by *Azospirillum brasilense* is very interesting

(Smets *et al.*, 2004). In this paper the modeling of growth and more specifically the estimation of the Monod kinetic parameters, is addressed.

In this case study the feed rate profile of a fed batch is optimized to enable accurate estimation of the growth model parameters. The designed optimal experiment requires advanced control to be realized in practice. The case study illustrates how, in a second stage, a trade-off can be found between maximum information content and practical feasibility of the experiment.

The structure of this paper is as follows. First, the material and methods and the theoretical background of optimal experiment design are introduced in Section 2. In Section 3 the implementation of optimal experiments is discussed. Finally, Section 4 summarizes the major conclusions.

2. MATERIALS AND METHODS

2.1 Bioreactor experiments

Experiments were performed in a computer controlled BioFlo 3000 benchtop fermentor (New Brunswick Scientific, USA) with an autoclavable vessel of 1.25 to 5L working volume. 100 mL of a preculture containing *Azospirillum brasilense* was transferred to the vessel containing a minimal malate medium (MMAB) (Vanstockem *et al.*, 1987). L-malate is provided as sole carbon source. PID cascade controllers ensure that the fermentation temperature is kept constant at 30°C, pH at 6.3 and the dissolved oxygen concentration at 3% (micro-aerobic range).

Culture media samples were removed at regular intervals. Cell density was obtained through measurement of optical density (OD) at 600 nm (Genesis 10S, Thermo Spectronic). L-malate was measured using test kits from Roche (R-biopharm, Germany).

2.2 Growth model

The evolution of biomass C_X [OD] and substrate concentration C_S [g/L] in a fed batch reactor can be described by following mass balance type equations:

$$\begin{aligned} \frac{dC_X}{dt} &= \mu \cdot C_X - \frac{U}{V} \cdot C_X \\ \frac{dC_S}{dt} &= -\sigma \cdot C_X + \frac{U}{V} \cdot (C_{S,in} - C_S) \\ \frac{dV}{dt} &= U \end{aligned} \qquad (1)$$

with

$$\mu = \mu_{max} \cdot \frac{C_S}{C_S + K_M} \qquad (2)$$

$$\sigma = \frac{\mu}{Y_{X/S}} + m \qquad (3)$$

with V [L] the volume of the liquid phase and $C_{S,in}$ [g/L] the substrate concentration in the volumetric feed rate U [L/h]. μ [h^{-1}] is the specific growth rate and is specified by the Monod equation (2) with μ_{max} [h^{-1}] the maximum specific growth rate and K_M [g/L] the half-saturation constant. The relation between specific growth rate μ [h^{-1}] and specific consumption rate σ [(g/(OD·L))·h^{-1}] is given by the linear law (3). $Y_{X/S}$ [(OD·L)/g] is a yield coefficient of biomass over substrate and m [(g/(OD·L))·h^{-1}] represents a maintenance factor. In this case study, maintenance is, in a first stage, neglected (m=0).

2.3 Parameter estimation

The identification cost imposed for parameter estimation is the sum of squared errors SSE:

$$SSE = \sum_{i=1}^{n} (y_{exp}(t_i) - y_{model}(t_i))^2 \qquad (4)$$

with $y_{model}(t_i)$ the model predictions, $y_{exp}(t_i)$ the experimental observations and n the number of samples.

As m is set equal to zero, the yield coefficient $Y_{X/S}$ can be estimated separately by eliminating the specific growth rate from the growth model:

$$\frac{dZ}{dt} = \frac{U}{V} \cdot (Y_{X/S} \cdot C_{S,in} - Z) \qquad (5)$$

with

$$Z = Y_{X/S} \cdot C_S + C_X$$

This leaves two growth parameters to be estimated, i.e., K_M and μ_{max}, together with the initial conditions $C_X(0)$ and $C_S(0)$.

The implemented identification routines for model parameter identification are the `E04UCF` routine from the NAG library (Numerical Algorithms Group) in Fortran and the `lsqnonlin` routine in Matlab (The Mathworks Inc., Natick). Numerical integration is performed with the NAG-routine `D02EJF` in Fortran.

2.4 Optimal experiment design

The information content of an experiment, with respect to parameter identification, can be evaluated through the Fisher information matrix **F** (e.g., Walter and Pronzato, 1997):

$$\mathbf{F} \triangleq \int_0^{t_f} \left(\frac{\partial \mathbf{y}}{\partial \theta}\right)^T \mathbf{Q} \left(\frac{\partial \mathbf{y}}{\partial \theta}\right) \cdot dt \qquad (6)$$

The main components of the Fisher information matrix $\mathbf{F}$ are the model output sensitivities $\frac{\partial \mathbf{y}}{\partial \theta}$, and the uncertainty of the measurements. The latter is represented by the weighting matrix $\mathbf{Q}$, which is set equal to the inverse of the measurement error covariance matrix. The model output sensitivities reflect the sensitivity of the model output $\mathbf{y}$ to small variations of the parameters θ.

Depending on the requirements imposed by the application, a specific scalar function of $\mathbf{F}$ is used as performance index for optimal experiment design. Different design criteria are available and the choice of the criterion will influence the resulting design (Walter and Pronzato, 1997; Vanrolleghem and Dochain, 1998). In this case study, the modified E-criterion is selected. This criterion aims at the minimization of the condition number of $\mathbf{F}$, i.e., the ratio of the largest to the smallest eigenvalue of $\mathbf{F}$.

$$\Lambda(\mathbf{F}) = \frac{\lambda_{max}(\mathbf{F})}{\lambda_{min}(\mathbf{F})} \qquad (7)$$

The objective is to have eigenvalues as close as possible to each other. When the condition number reaches its minimal value, i.e., $\Lambda(\mathbf{F}) = 1$, the contour lines of the cost surface for a two parameter problem will be circular. Such highly informative experiment allows unique parameter estimation. Values of $\Lambda(\mathbf{F})$ greater than 1 induce ellipsoid contour lines.

Given the nonlinear model structure, the design also depends on the nominal parameter values, i.e., the initial guess for the unknown parameters, used in the optimization. The closer the nominal values approach the true parameter values, the better the obtained design. Optimal experiment design is, therefore, an iterative procedure. After evaluation, the design is implemented. The resulting experimental data and identified parameters are used as a basis for a next round of optimal experiment design.

3. RESULTS AND DISCUSSION

3.1 Parameter identification from batch data

Nihtilä and Virkkunen (1977) showed that the parameters of the Monod kinetics cannot be uniquely identified from noisy batch measurements. This is illustrated in Figure 1. The upper plot depicts the experimental data of a preliminary batch experiment together with one of the many possible solutions of the parameter estimation problem. The identification problem is illustrated in the lower plot. Joint and individual uncertainty are very large for the kinetic parameters K_M and μ_{max}. The contour plot reveals a valley with different parameter combinations which result in an equally low cost. This means that a change in one of the parameters can be compensated by a change in the other parameter.

To overcome this problem, new experiments need to be designed which are more informative in the sense of accurate parameter estimation. Techniques of optimal experiment design for parameter estimation are addressed to tackle this problem.

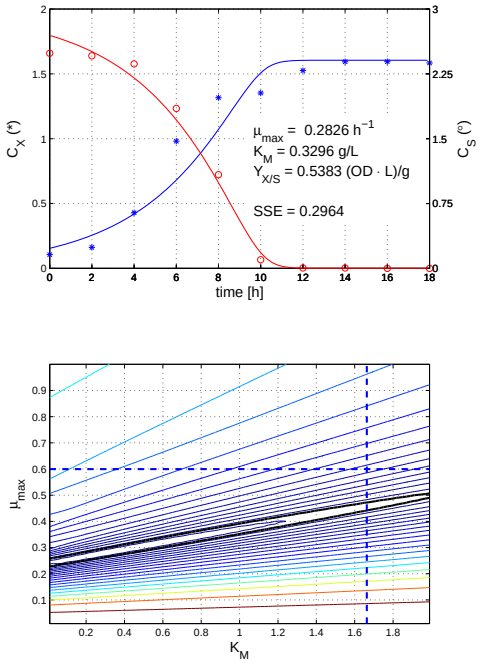


Fig. 1. Model parameter identification from a batch experiment. Upper plot: experimental data (∗,∘) and model predictions (−) for biomass and substrate concentrations. Lower plot: contours of equal identification cost (SSE) as function of the model parameters K_M and μ_{max}. The bold line is the 95% joint confidence region, the dashed lines depict the 95% individual confidence intervals on K_M and μ_{max}.

3.2 Feeding profile

The optimal control problem is to find the best possible admissible feed rate profile $U(t)$ with respect to the quality of the estimates for the Monod parameters K_M and μ_{max}. Van Impe et al. (1995) formulated following conjecture:

> *A feed rate strategy which is optimal in the sense of process performance, is an excellent starting point for feed rate optimization with respect to estimation of those parameters with large influence upon process performance.*

A feeding profile optimal for process performance is one in which substrate concentration is kept constant from the beginning. For unique parameter estimation an extra perturbation is required, which can be achieved by preceding the singular feeding phase by an initial batch phase (Versyck and Impe, 1999). The structure of this feeding profile is depicted in Figure 2 (dashed line). In the feeding phase the feed rate $U(t)$ is given by a feed forward control law of the form:

$$U_{sing}(t) = \frac{\sigma C_X V}{C_{S,in} - C_S^*} \tag{8}$$

with $C_{S,in}$ the limiting substrate concentration in the feeding solution and C_S^* the constant substrate concentration aimed at during the feeding phase. There are two degrees of freedom in this feed rate optimization problem, i.e., the initial substrate concentration $C_S(0)$ and the substrate concentration during the feeding phase C_S^*.

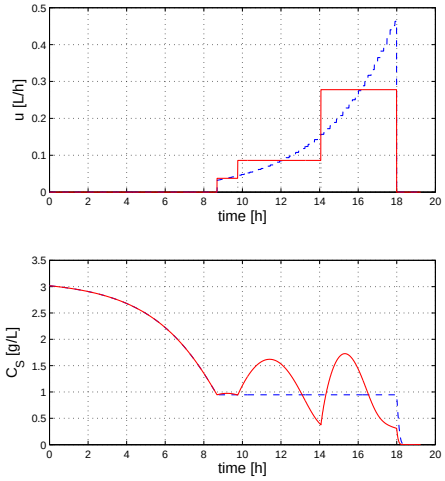


Fig. 2. Singular (- -) and 3-step (−) feed rate profile for optimal growth parameter identification.

3.3 optimization of the feeding profile

The parameters obtained from an initial batch experiment are used as nominal values for the design of a new and more informative experiment. Different combinations of $C_S(0)$ and C_S^* were found which give a condition number equal to 1. These profiles entail, however, some important practical problems.

A first problem are the values found for $C_S(0)$ and C_S^*. The obtained concentrations are very low and hard to realize in practice. For fixed values of $C_S(0)$, the optimal C_S^* and corresponding condition number were computed. A higher concentration for $C_S(0)$ yields a higher C_S^*. This way a range of designs with suboptimal ($C_S(0)$,C_S^*) combinations was defined, which are still informative enough with regard to parameter identification.

As a second step to increase practical feasibility, the time-varying feeding profile was simplified by replacing the singular feeding phase by steps of constant feed rate. This step approximation is done in such a way that the volume added per step of feeding is the same as for the time-varying feeding profile in that period:

$$U_{cte}(t,\theta) = \frac{\int_{t_i}^{t_{i+1}} u_{sing}(t,\theta)dt}{t_{i+1} - t_i} \tag{9}$$

Profiles with three as well as with one step were computed. An example of a feed rate profile with three steps is illustrated in Figure 2 (solid line). For this example there are two degrees of freedom to be optimized, i.e., the time points for switching from one feed rate to the next (t_1 and t_2). A profile with one step of constant feeding has only one degree of freedom, i.e., the time instant t_f at which the feeding stops. The resulting condition numbers for the different optimal and suboptimal feeding profiles are listed in Table 1.

Table 1. Overview of different designs.

Feeding profile	Condition number $\Lambda(\mathbf{F})$
Unconstrained singular profile	
$C_S(0)$= 0.4340 g/L	1.00
C_S^*= 0.1702 g/L	
Constrained singular profile	
$C_S(0)$= 3.0167 g/L	50.29
C_S^*= 0.9463 g/L	
Simplifications of the constrained singular profile	
3 steps:	
t_1= 9.76 h	62.13
t_2= 14.07 h	
1 step:	
t_f= 16.35 h	321

with $C_{S,in}$ = 50 g/L and nominal parameter values μ_{max}=0.421 h^{-1}, K_M=0.439 g/L, Y_{XS}=0.777 (OD·L)/g

To evaluate the loss of information content through simplification of the feed rate profile, the different optimal and simplified designs were extensively tested through identification of the growth parameters on simulated noisy data. All four designs listed in Table 1 delivered satisfying results concerning accurate parameter identification. Therefore, the most simple design with regard to practical feasibility, i.e., the design with only one step of constant feeding, was selected for implementation.

3.4 Implementation and validation

The results presented in this section were obtained in a second round of optimal experiment design. The results of the performed fed batch experiment are illustrated in the upper plot of Figure 3 and the identified growth parameters are summarized in Table 2. A malate solution of 10 g/L was added to the reactor with a feed rate of 0.07

L/h starting three hours after the start of the experiment. The period of feeding is represented by the vertical dashed lines on the plot. The lower plot depicts the contour plot together with the 95% individual confidence intervals and 95% joint confidence region for μ_{max} and K_M. The 95% joint confidence region now forms a closed ellips. Comparing the confidence intervals with the ones calculated for the batch experiments (see Figure 1, lower plot), confirms that the estimation accuracy of the parameters K_M and μ_{max} is significantly improved.

The predictive quality of the obtained model parameters was subsequently evaluated by comparing simulations with the identified model parameters and experimental data [1]. Hereto, a new fed batch experiment with a different feeding profile was performed. The feed rate for this experiment was also 0.07 L/h, but the period of feeding was shifted. Additionally, the data of the initial batch experiment were used for evaluation. These validation results are depicted in Figure 4.

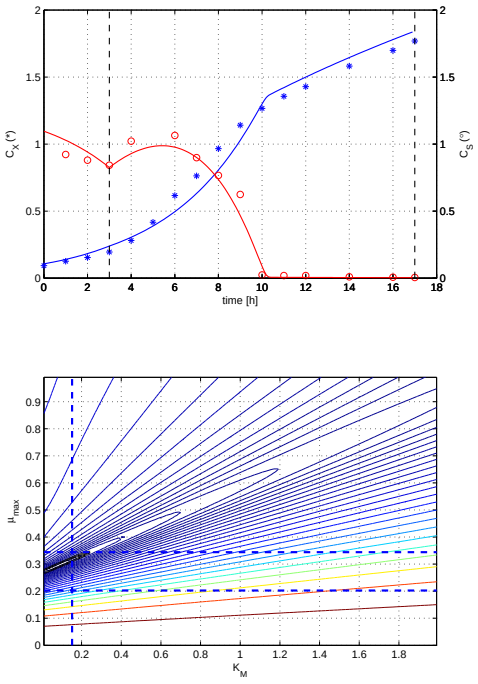


Fig. 3. Model parameter identification from an informative fed batch experiment. Upper plot: experimental data (∗,○) and model predictions (−) for biomass and substrate concentrations. Lower plot: contours of equal identification cost (SSE) as function of the model parameters K_M and μ_{max}. The bold line is the 95% joint confidence region, the dashed lines depict the 95% individual confidence intervals on K_M and μ_{max}.

[1] The initial conditions ($C_X(0)$ and $C_S(0)$) have been reestimated for each simulation.

Table 2. Parameter values for the model (1,2,3) with and without maintenance.

	no maintenance	with maintenance
$Y_{X/S}$	0.4905	0.5468
μ_{max}	0.2733	0.2961
K_M	$1.441{\cdot}10^{-2}$	$4.163{\cdot}10^{-2}$
m_S	-	$1.445{\cdot}10^{-2}$
SSE	0.1626	0.1362

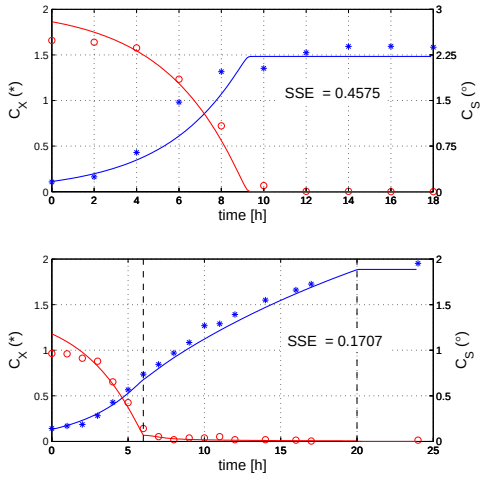


Fig. 4. Experimental data (∗,○) and simulation of the growth model (−) on batch (upper plot) and fed batch experiments (lower plot). Parameters are taken from Table 2.

3.5 Remarks concerning the growth model

The growth model for *Azospirillum brasilense* presented in this paper, started from the assumption that maintenance can be neglected. The available experimental data, however, do not allow to determine whether maintenance can be omitted or not.

This shortcoming is illustrated by identification of the parameters for the model (1,2,3) taking maintenance into account. Expression (5) cannot be used, in this case, to estimate the yield coefficient $Y_{X/S}$. Here, the four parameters (μ_{max}, K_M, $Y_{X/S}$ and m) have to be identified simultaneously. The results are presented in Table 2 and in Figure 5. The model with maintenance seems to provide a better description of the last hours (10h till 17h) of the fed batch experiment, while the validation results (see Figure 6) are less good for that same period. The model with maintenance predicts a decrease in biomass concentration after depletion of the substrate. This phenomenon is, however, not observed in the data.

Another problem of the model is the overestimation of the initial substrate concentration $C_S(0)$. The estimated values for $C_S(0)$ are consistently higher than the experimental values. The consumption of malate in the first hours of the experiments seems to exhibit a delay or lag which cannot be described by the model.

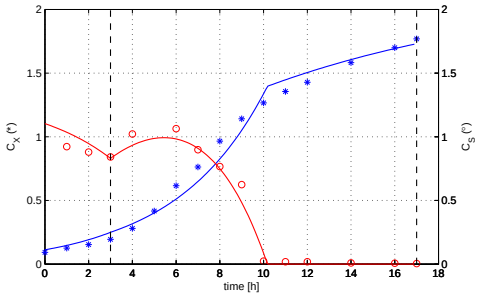


Fig. 5. Identification of parameters for a model including maintenance: experimental data (∗,○) and model prediction (−) for biomass and substrate concentrations.

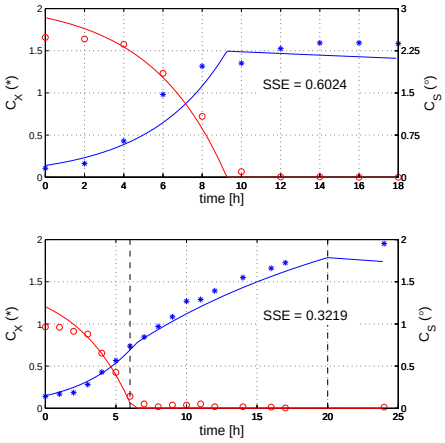


Fig. 6. Experimental data (∗,○) and simulation of the growth model including maintenance (–) on batch (upper plot) and fed batch experiments (lower plot).

Although the model does have some shortcomings as mentioned above, it provides an accurate description of the transitions of one growth phase to another. The current model is in its most simple form, and shall be extended in a further stage to overcome these problems.

4. CONCLUSIONS

This paper presents a successful validation of optimal experiment design for parameter identification for the Monod kinetics. Due to some practical limitations, a trade-off has to be made between maximum information content and practical feasibility. Theoretical optimal designs have successfully been translated to feasible (sub)optimal designs by imposing constraints on substrate concentrations and simplifying the feeding phase. With only a few additional experiments the accuracy of the kinetic parameters was significantly increased as illustrated by the individual confidence intervals and joint confidence regions.

ACKNOWLEDGEMENTS

Work supported in part by the IWT-GBOU-20160 project, the FWO-G.0085.03 project, OT/03/30 of the Research Council of the Katholieke Universiteit Leuven, the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian Federal Science Policy Office and by the K.U.Leuven-BOF EF/05/006 Center-of-Excellence Optimization in Engineering. K. Bernaerts and I.Y. Smets are Postdoctoral Fellows with the Fund for Scientific Research Flanders (FWO).

REFERENCES

Baldani, V.L.D., J.I. Baldani and J. Döbereiner (1983). Effects of *Azospirillum* inoculation on root infection and nitrogen incorporation in wheat. *Canadian Journal of Microbiology* **29**, 924–929.

Bernaerts, K. and J.F. Van Impe (2004). Data-driven approaches to the modelling of bioprocesses. *Transactions of the Institute of Measurement and Control* **26**(5), 349–372.

Nihtilä, M. and J. Virkkunen (1977). Practical identifiability of growth and substrate consumption models. *Biotechnology and Bioengineering* **19**, 1831–1850.

Smets, I., K. Bernaerts, A. Cappuyns, O. Ona, J. Vanderleyden, E. Prinsen and J.F. Van Impe (2004). A prototype model for indole-3-acetic acid (IAA) production by *Azospirillum brasilense* sp 245. In: *Proceedings of the 7th International Conference on DYnamics and COntrol of Process Systems, Dycops 7, CDROM* (S.L. Shah and J.F. MacGregor, Eds.). 6p.

Van Impe, J.F., J.E. Claes and G. Bastin (1995). Optimal feed rate profile for combined bioprocess modeling and optimization. In: *Proceedings of the 1995 European Control Conference* (A. Isidori, S. Bittani, E. Mosca, A. De Luca, M.D. Di Benedetto and G. Oriolo, Eds.). pp. 3510–3515.

Vanrolleghem, P.A. and D. Dochain (1998). Bioprocess model identification. In: *Advanced instrumentation, data interpretation and control of biotechnological processes* (J.F.M. Van Impe, P.A. Vanrolleghem and D.M. Iserentant, Eds.). pp. 251–318. Kluwer Academic Publisher, Dordrecht.

Vanstockem, M., K. Michiels, J. Vanderleyden and A. Van Gool (1987). Transposon mutagenesis of *Azospirillum brasilense* and *Azospirillum lipoferum*: physical analysis of Tn5 and Tn5-mob insertion mutants. *Applied and Environnemetal Microbiology* **53**, 410–415.

Versyck, K.J. and J.F. Van Impe (1999). Feed rate optimization for fed-batch bioreactors: from optimal process performance to optimal parameter estimation. *Chemical Engineering Communications* **172**, 107–124.

Walter, E. and L. Pronzato (1997). *Identification of parametric models from experimental data.* Springer, Masson.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DYNAMIC MODELLING OF A BIOFILTER USED FOR NITRIFICATION OF DRINKING WATER AT LOW INFLUENT AMMONIA CONCENTRATIONS

Isabelle Queinnec, [***,1] **Juan-Carlos Ochoa,** [**]
Etienne Paul [**] **and Alain Vande Wouwer** [*]

[*] *Service d'Automatique, Faculté Polytechnique de Mons, Boulevard Dolez 31, B-7000 Mons, Belgium*
[**] *LIPE, Institut National des Sciences appliquées, 125 avenue de Rangueil, F-31077 Toulouse cedex 4, France*
[***] *LAAS-CNRS, 7 avenue du Colonel Roche, F-31077 Toulouse cedex 4, France*

Abstract: This paper reports on the development of a mathematical model of a packed bed biofilter operating at low influent ammonia concentrations. It is initially filled with biomass-free media, the adhesion by filtration of the bacteria present in the groundwater allowing colonization of the filter. The mathematical model is intended for simulation/optimization purposes, and should describe sufficiently well the start-up phase, as well as nominal operation. Unknown model parameters are estimated using experimental data collected on pilot plants. Validation and cross-validation results are discussed.

Keywords: Mathematical modelling; Distributed parameter systems; Parameter estimation; Biotechnology

1. INTRODUCTION

Generally, groundwater contains ammonia and is thus unsuitable for direct use as drinking water. Packed-bed biofilters enable a combination of biodegradation and physical retention, which ensures the capture of nitrifying bacteria carried by groundwater. Cell attachment and growth at the carrier surface create the biofilm. However, biofilters used for drinking water nitrification operate at lower ammonia concentrations than those usually observed in industrial wastewater treatment plants, and in most cases, the ammonia concentration is so low that it becomes the rate-limiting factor of biological nitrification. Moreover, a one- or two- month period is usually necessary to capture a sufficient amount of nitrifying bacteria, so as to reach the expected removal efficiency. For safe process operation, biofilter disinfection is also regularly performed, involving long stand-by phases where it is again necessary to wait for the biofilter colonization. Therefore, improving start-up of biofilters operating at low substrate concentrations is a major challenge related to the drinking water industry. Nitrites in the outflow of the biofilter must be avoided in all operating conditions.

In order to design a biofilter and optimize its operation, appropriate mathematical models would undoubtedly be very useful. The model should be complex enough to give a reliable representation of the physical and biological processes but simple enough to allow parameter identification from experimental data (practical parameter identifiability problem). A review of the published literature shows that only limited information is available on modelling of drinking water biofilters. A few papers report works on ammonia removal through biological filtration in aqua-

[1] Author to whom correspondence should be addressed: e-mail: queinnec@laas.fr

culture industry (Grommen *et al.*, 2002), (Zhu and Chen, 1999). Numerical models were also developed to simulate the transient behaviour of biofilters used for biodegradable organic matter removal (Hozalski and Bouwer, 2001).

In the present work, experiments have been carried out under different conditions to explore the behavior of a packed-bed biofilter in the start-up and steady-state (nominal operation) phases. Taking into account the main biological processes, filtration and adsorption, a dynamic model based on a set of mass-balance partial differential equations (PDEs) is derived. Unknown model parameters are inferred from experimental data by minimizing an output-error criterion. Validation and cross-validation results are discussed.

2. MATERIALS AND METHODS

Two different filters were used in order to cover a larger range of operating conditions. The first filter structure, related in the following to experiments C1, is 21.5 *cm* in diameter and 180 *cm* in length, with a bed depth of 140 *cm*. The second filter (experiments C2) is 15 *cm* in diameter and 180 *cm* in length with a bed depth of 140 *cm*. Both beds are composed of (1.0-2.0 *mm*) manganese dioxide (ρ , 1.75-1.85 $g.cm^{-3}$; diameter, 0.36-1 *mm*). The filters are provided in different sites for water sampling (distributed vertically).

Groundwater is used in all experiments, whose average composition is ammonium: $0.2 \pm 0.02 mgN - NH_4^+.l^{-1}$; nitrite: $0.01 \pm 0.005 mgN - NO_2^-.l^{-1}$; nitrate $0.01 \pm 0.005 mgN - NO_3^-.l^{-1}$. In experiments C2, additional ammonium was added to increase the influent water composition to 1.1 or 5 $mgN - NH_4^+.l^{-1}$. Flow rates were set to 254 $l.h^{-1}$ and 140 $l.h^{-1}$ for filters 1 and 2, respectively, so as to impose the same liquid superficial velocity of about 8-10 $m.h^{-1}$. The water temperature inside the biofilter was 16^oC and 24^oC for experiments C1 and C2, respectively. The monitored variables were the dissolved oxygen concentration, conductivity and pH. In all cases, the biofilter was initially filled with biomass-free media, before eventually being uniformly inoculated with nitrifying bacteria previously grown in an aerated batch reactor.

Ammonia, nitrite and nitrate concentrations in the bulk phase were measured according to French standards (Afnor, 1994).

3. MODELLING

3.1 Bacterial growth and inactivation

Nitrification is a reaction chain oxidizing ammonia into nitrate, which consists of two main biological reactions (Henze *et al.*, 2002) associated to bacterial growth. It is commonly assumed that the production of nitrite and nitrate is associated to the growth of *Nitrosomonas* and *Nitrobacter*, respectively, which are formed with yields Y_{NS} and Y_{NB}.

The reaction rates are known to be limited by their nitrogeneous substrates at low concentrations as well as by oxygen. The temperature is also known to have a strong influence on these rates. The specific growth rates are then formulated according to classical Monod laws, where the dependency on the temperature is given by:

$$\mu_{\mathrm{i,max}}(T) = \mu_{\mathrm{i,max}}(20\,^{\circ}\mathrm{C}) 1.103^{T-20} \quad , \ \mathrm{i} = \mathrm{NS \ or \ NB}$$

The biomass forms a biofilm around the particles. The growth of bacteria is counterbalanced by an inactivation process, i.e., part of the biomass can be considered as inactive despite its presence in the biofilm. This leads to a maximum active biomass concentration $X^{\mathrm{F}}_{\mathrm{max}}$ of both bacteria types, which compete for a place at the interface of the biofilm (Haag *et al.*, 2004). In agreement with the growth and inactivation kinetics introduced by Jacob (Jacob, 1994), the balance of growth/inactivation for each type of bacteria present in the biofilm is expressed as follows:

$$\dot{X}^{\mathrm{F}}_{\mathrm{NS}} = \mu_{\mathrm{NS}} X^{\mathrm{F}}_{\mathrm{NS}} - \frac{X^{\mathrm{F}}_{\mathrm{NS}}}{X^{\mathrm{F}}_{\mathrm{max}}} \left(\mu_{\mathrm{NS}} X^{\mathrm{F}}_{\mathrm{NS}} + \mu_{\mathrm{NB}} X^{\mathrm{F}}_{\mathrm{NB}} \right) \quad (1)$$

$$\dot{X}^{\mathrm{F}}_{\mathrm{NB}} = \mu_{\mathrm{NB}} X^{\mathrm{F}}_{\mathrm{NB}} - \frac{X^{\mathrm{F}}_{\mathrm{NB}}}{X^{\mathrm{F}}_{\mathrm{max}}} \left(\mu_{\mathrm{NS}} X^{\mathrm{F}}_{\mathrm{NS}} + \mu_{\mathrm{NB}} X^{\mathrm{F}}_{\mathrm{NB}} \right) \quad (2)$$

3.2 Filtration

The adhesion of the bacteria present in the groundwater to the solid bed is mainly due to filtration. The basic equation used in filtration theory to represent the removal of particles (suspended particle concentration $X^{\mathrm{B}}_{\mathrm{tot}}$) with distance z in a packed filter was first empirically derived by (Iwasaki, 1937). Various attempts were made to find a simple correlation between the filter coefficient k_f and key variables such as particle size, filtration velocity, porous media. In the simplest case, the filter coefficient is assumed to be constant. Assuming that transport by dispersion and detachment process can be neglected, the filtration process is described by two equations:

$$\frac{\partial X^{\mathrm{F}}_{\mathrm{tot}}}{\partial t} = k_{\mathrm{f}} \frac{Q}{A} X^{\mathrm{B}}_{\mathrm{tot}} \quad (3)$$

$$\frac{\partial (\varepsilon X^{\mathrm{B}}_{\mathrm{tot}})}{\partial t} = -\frac{Q}{A} \frac{\partial X^{\mathrm{B}}_{\mathrm{tot}}}{\partial z} - k_{\mathrm{f}} \frac{Q}{A} X^{\mathrm{B}}_{\mathrm{tot}} \quad (4)$$

where superscript B and F refer to the bulk and solid phases, respectively, and k_f is the filtration constant.

3.3 Decay

The decay cycle involves a loss of bacteria, part of which is transformed into ammonia by hydrolysis. The whole cycle of decay is fully described in (Henze *et al.*, 2002). It involves the decay with specific rate *b*,

to produce particulate biodegradable organic nitrogen X_{ND}, with yield $\nu = i_{XB} - f_p i_{XP}$, then its transformation into soluble biodegradable organic nitrogen and finally into ammonia nitrogen. The less favorable case where the hydrolysis and ammonification are assumed to be instantaneous is considered here:

$$X_{NS} + X_{NB} \xrightarrow{b} \nu X_{ND} \xrightarrow{\infty} \nu NH_4^+ \qquad (5)$$

Of course, this assumption does not reflect the reality, but allows the number of equations to be to reduced, while considering an approximate decay cycle.

3.4 Adsorption

A common way to describe sorption processes is based on the boundary layer theory which assumes an adsorption equilibrium at the interface between the mobile and stationary phases. A widely used isotherm for the sorption equilibrium was proposed in (Freundlich, 1906), involving the Freundlich constant K_{Fr}, the exponent $0 < n_{Fr} \leq 1$ and adsorption specific rate k_{ads}, which all depend on the media used.

3.5 The PDE model

The biological and physical transformations described in the previous subsections are schematically represented in Figure 1.

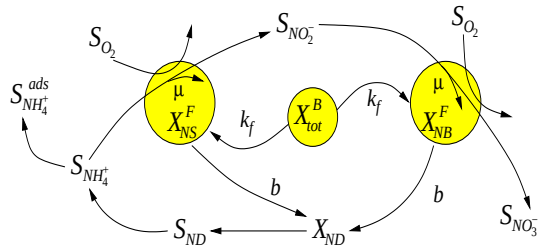


Fig. 1. Scheme of biological and physical phenomena involved in the biofilter nitrification process

The dynamic model equations are derived from mass balances. Since the biofilter is a spatially distributed system, these balances have to consider the state variables as functions of time and space. Height different states are considered in the proposed modeling approach leading to the state vector:

$$x^T = [S_{NH_4^+} \;\; S_{NO_2^-} \;\; S_{NO_3^-} \;\; S_{O_2} \;\; X_{tot}^B \;\; S_{NH_4^+}^{ads} \;\; X_{NS}^F \;\; X_{NB}^F \;]$$

The model PDEs are derived by expressing the dynamic mass balances around a infinitesimal slice along the column axis (Haag *et al.*, 2004). Under the assumption that the biofilm density is large enough so that the variation of porosity ε related to biomass growth can be neglected, i.e. ε is constant, the system of partial differential equations describing the biofilter is given by:

$$\dot{S}_{NH_4^+} = -\frac{Q}{\varepsilon A}\frac{\partial S_{NH_4^+}}{\partial z} + \nu b \frac{(X_{NS}^F + X_{NB}^F)}{\varepsilon} - \frac{\mu_{NS}}{Y_{NS}}\frac{X_{NS}^F}{\varepsilon} - \frac{k_{ads}}{\varepsilon}\left(S_{NH_4^+} - \left(\frac{S_{NH_4^+}^{ads}}{K_{Fr}}\right)^{1/n_{Fr}}\right) \qquad (6)$$

$$\dot{S}_{NO_2^-} = -\frac{Q}{\varepsilon A}\frac{\partial S_{NO_2^-}}{\partial z} + \frac{Y_{NO_2^-}}{Y_{NS}}\mu_{NS}\frac{X_{NS}^F}{\varepsilon} - \frac{1}{Y_{NB}}\mu_{NB}\frac{X_{NB}^F}{\varepsilon} \qquad (7)$$

$$\dot{S}_{NO_3^-} = -\frac{Q}{\varepsilon A}\frac{\partial S_{NO_3^-}}{\partial z} + \frac{Y_{NO_3^-}}{Y_{NB}}\mu_{NB}\frac{X_{NB}^F}{\varepsilon} \qquad (8)$$

$$\dot{S}_{O_2} = -\frac{Q}{\varepsilon A}\frac{\partial S_{O_2}}{\partial z} - \left(\frac{Y_{O_2,NS}}{Y_{NS}}\mu_{NS}\frac{X_{NS}^F}{\varepsilon} + \frac{Y_{O_2,NB}}{Y_{NB}}\mu_{NB}\frac{X_{NB}^F}{\varepsilon}\right) \qquad (9)$$

$$\dot{X}_{tot}^B = -\frac{Q}{\varepsilon A}\frac{\partial X_{tot}^B}{\partial z} - k_f\frac{Q}{\varepsilon A}X_{tot}^B \qquad (10)$$

$$\dot{S}_{NH_4^+}^{ads} = \frac{\varepsilon}{1-\varepsilon}k_{ads}\left(S_{NH_4^+} - \left(\frac{S_{NH_4^+}^{ads}}{K_{Fr}}\right)^{1/n_{Fr}}\right) \qquad (11)$$

$$\dot{X}_{NS}^F = \mu_{NS}X_{NS}^F + f_{NS,in}k_f\frac{Q}{A}X_{tot}^B - bX_{NS}^F - \frac{X_{NS}^F}{X_{max}^F}\left(\mu_{NS}X_{NS}^F + \mu_{NB}X_{NB}^F - b(X_{NS}^F + X_{NB}^F) + k_f\frac{Q}{A}X_{tot}^B\right) \qquad (12)$$

$$\dot{X}_{NB}^F = \mu_{NB}X_{NB}^F + (1 - f_{NS,in})k_f\frac{Q}{A}X_{tot}^B - bX_{NB}^F - \frac{X_{NB}^F}{X_{max}^F}\left(\mu_{NS}X_{NS}^F + \mu_{NB}X_{NB}^F - b(X_{NS}^F + X_{NB}^F) + k_f\frac{Q}{A}X_{tot}^B\right) \qquad (13)$$

The derived PDE system has to be supplemented with appropriate initial and boundary conditions:

- initial spatial profile: $x(t_0, z) = x_0(z)$,
- inflow ($z = z_0$) boundary conditions: $x(t, z_0) = x_{in}(t)$, which have to be consistent at $(t_0, z = 0)$.

3.6 Model simulation

The non-linear PDE system described above is solved numerically using a *Method of Lines* strategy, which proceeds in two steps: (a) the spatial domain is discretized and the spatial derivatives are approximated by finite differences, (b) the resulting system of semi-discrete ODEs is integrated in time.

4. MODEL IDENTIFICATION

Several model parameters are involved in the PDE model, whose values have been published in the literature. The parameters relative to nitrification are generally considered as well known in the case of fully stirred bioreactors (Henze *et al.*, 2002). However, it is considered in the present study that the affinity of the micro-organisms for their substrates is influenced by the porous support.

Another key parameter of the model is the maximum active biomass concentration X_{max}^F, the value of which is unknown and has to be identified.

The parameters relative to adsorption can be either identified, or evaluated through specific experiments. The second way has been used in this study. The set of model parameters is summarized in Table 1.

Table 1. Model parameters

symbol (unit)	value
Y_{NS} ($^{\mathrm{gDCO}}/_{\mathrm{gN-NH_4^+}}$)	0.142
Y_{NB} ($^{\mathrm{gDCO}}/_{\mathrm{gN-NO_2^-}}$)	0.084
$Y_{\mathrm{NO_2^-}}$ ($^{\mathrm{gN-NO_2^-}}/_{\mathrm{gN-NH_4^+}}$)	0.988
$Y_{\mathrm{NO_3^-}}$ ($^{\mathrm{gN-NO_3^-}}/_{\mathrm{gN-NO_2^-}}$)	0.993
$Y_{\mathrm{O_2,NS}}$ ($^{\mathrm{gDCO}}/_{\mathrm{gN-NH_4^+}}$)	$3.42 - Y_{\mathrm{NS}}$
$Y_{\mathrm{O_2,NB}}$ ($^{\mathrm{gDCO}}/_{\mathrm{gN-NO_2^-}}$)	$1.14 - Y_{\mathrm{NB}}$
$\mu_{\mathrm{NS,max}}$ (d^{-1})	0.7
$\mu_{\mathrm{NB,max}}$ (d^{-1})	0.8
K_{NS} ($mg\mathrm{N}-\mathrm{NH_4^+}.\mathrm{l}^{-1}$)	to be estimated
K_{NB} ($mg\mathrm{N}-\mathrm{NO_2^-}.\mathrm{l}^{-1}$)	to be estimated
$K_{\mathrm{O_2}}$ ($mg\mathrm{O_2}.\mathrm{l}^{-1}$)	0.8
b (d^{-1})	0.05
$\nu = i_{\mathrm{XB}} - f_{\mathrm{p}} i_{\mathrm{XP}}$ ($^{\mathrm{gN}}/_{\mathrm{gDCO}}$)	0.0812
X_{max} ($g\mathrm{DCO}.\mathrm{m}^{-3}$)	to be estimated
k_{f} (m^{-1})	0.2
ε (-)	0.12
dp (dm)	0.0094
$K_{\mathrm{F_r}}$ (-)	0.26
k_{ads} (d^{-1})	162
n_{Fr} (-)	1

Besides the model parameters, the initial and input conditions have to be specified for each experiment. At the initial time, the concentration of bacteria in the bulk phase, the free and adsorbed ammonia, nitrite and nitrate concentrations are set equal to 0. The oxygen is saturated (10 $mg.l^{-1}$). In the cases where additional nitrifying bacteria are inoculated, the initial concentration of *Nitrosomonas* and *Nitrobacter* have to be estimated. The initial conditions are summarized in Table 2.

Table 2. Initial concentrations

Initial concentration	value
$S_{NH_4^+}(t=0,z)$	0 or known
$S_{NO_2^-}(t=0,z)$	0 or known
$S_{NO_3^-}(t=0,z)$	0 or known
$S_{0_2}(t=0,z)$	9
$X^{\mathrm{B}}_{\mathrm{tot}}(t=0,z)$	0
$S^{ads}_{NH_4^+}(t=0,z)$	0 or known
$X^{\mathrm{F}}_{\mathrm{NS}}(t=0,z)$	to be estimated
$X^{\mathrm{F}}_{\mathrm{NB}}(t=0,z)$	to be estimated

The boundary conditions for the state variables in the bulk phase are fixed by the column inflow. Influent concentrations of ammonia, nitrite, nitrate and oxygen are potentially time-varying, but measured. The influent concentration of bacteria in the bulk phase $X^{\mathrm{B}}_{\mathrm{tot,in}}$ is unknown, but constant. The influent concentrations are summarized in Table 3.

For parameter estimation, a classical least-squares criterion is used, which is minimized using the Nelder-Mead simplex method, implemented in MATLAB routines.

Table 3. Influent concentrations

Initial concentration	value
$S_{NH_4^+}(t,z=0)$	measured - time-varying
$S_{NO_2^-}(t,z=0)$	measured - time-varying
$S_{NO_3^-}(t,z=0)$	measured - time-varying
$S_{0_2}(t,z=0)$	10
$X^{\mathrm{B}}_{\mathrm{tot,in}} = X^{\mathrm{B}}_{\mathrm{tot}}(t,z=0)$	to be estimated
$f_{NS,in}$	to be estimated

5. RESULTS AND DISCUSSION

5.1 Few considerations about the estimation procedure

Parameter estimation problem is particularly delicate in biological water treatment processes, due to the complexity of the models (and associated number of parameters) and the difficulty of collecting experimental data in well-defined and reproducible conditions. Particularly, real-life experiments involve unmodeled phenomena, random perturbations, sampling errors, and limited accuracy of the analysis procedures. For all these reasons, it is illusory to estimate accurate parameter values, and in the present study, our objective is mostly to validate the proposed model structure and to determine representative parameter estimates. Of course, in order to alleviate the above mentioned difficulties, independent experiments corresponding to various operating conditions have been carefully conducted. More precisely, two different datasets, corresponding to two different packed bed biofilters, are available. The first dataset involves two experiments:

- C1Exp1: $X^{\mathrm{F}}_{\mathrm{NS}}(t=0,z)=0$, $X^{\mathrm{F}}_{\mathrm{NB}}(t=0,z)=0$, $X^{\mathrm{B}}_{\mathrm{tot,in}} \neq 0$ unknown, known low level of ammonia concentration at input (around 0.2 $mg\mathrm{N}-\mathrm{NH_4^+}.\mathrm{l}^{-1}$);
- C1Exp2: $X^{\mathrm{F}}_{\mathrm{NS}}(t=0,z) \neq 0$ unknown, $X^{\mathrm{F}}_{\mathrm{NB}}(t=0,z) \neq 0$ unknown, $X^{\mathrm{B}}_{\mathrm{tot,in}} \neq 0$ unknown but the same as in the previous experiment, known low level of ammonia concentration at input (around 0.2 $mg\mathrm{N}-\mathrm{NH_4^+}.\mathrm{l}^{-1}$).

For such experiments of about one month, initiated with very low concentrations of bacteria and rather slow growth rates, the maximum active biomass concentration has a minor effect on the model transients, i.e., sensitivity with respect to X_{max} is very low in a wide range of values above 100 $mg\mathrm{DCO}.\mathrm{l}^{-1}$. On the other hand, the limiting conditions of these experiments allow the half-saturation constants to be estimated. This can be seen on Figure 2 where the cost function measuring the deviation between simulated and measured outputs is plotted for different values of the half-saturation constants. The minimum of the function is achieved for $K_{\mathrm{NS}} = 0.4 mg\mathrm{N}-\mathrm{NH_4^+}.\mathrm{l}^{-1}$ and $K_{\mathrm{NB}} = 0.175 mg\mathrm{N}-\mathrm{NO_3^-}.\mathrm{l}^{-1}$. These values are not affected for any X_{max} belonging to the interval $[100...500]$.

Moreover, experiment C1Exp1 can be used to estimate the concentration of particles in the influent water, and the fraction $f_{\mathrm{NS,in}}$ of *Nitromosonas*. Ac-

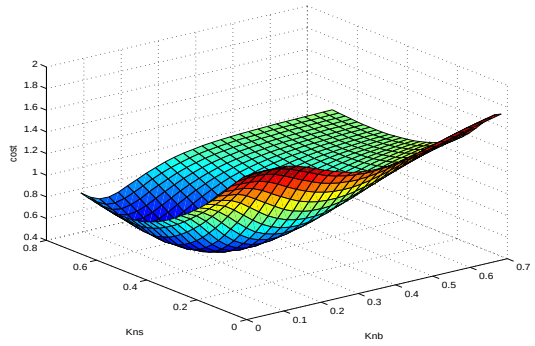


Fig. 2. C1Exp1 - Cost function evolution with the half-saturation constants K_{NS} and K_{NB} for given values $X_{\max} = 200 mg\mathrm{DCO}.\mathrm{l}^{-1}$, $X^{\mathrm{B}}_{\mathrm{tot,in}} = 0.0001 mg\mathrm{DCO}.\mathrm{l}^{-1}$ and $f_{\mathrm{NS,in}} = 0.9$

cording to the growth yields Y_{NS} and Y_{NB}, an initial guess would be $f_{\mathrm{NS,in}} = 0.63$. However, this fraction is strongly influenced by the conditions of conservation of the groundwater. Figure 3 illustrates this fact, i.e. the output least-square cost function is plotted for different values of the influent particle concentration and repartition between *Nitrosomonas* and *Nitrobacter*. The minimum of the cost function corresponds to $X^{\mathrm{B}}_{\mathrm{tot,in}} = 0.0012 mg\mathrm{DCO}.\mathrm{l}^{-1}$ and $f_{\mathrm{NS,in}} = 0.9$.

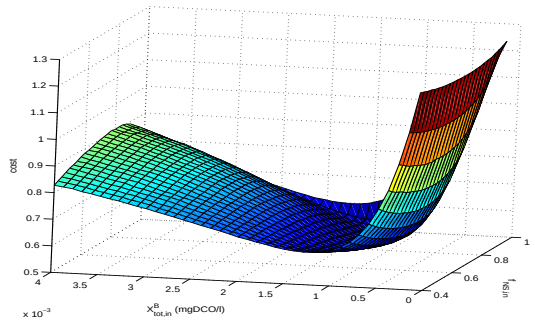


Fig. 3. C1Exp1 - Cost function evolution with the influent concentration of bacteria $X^{\mathrm{B}}_{\mathrm{tot,in}}$ and fraction of *Nitrosomonas* $f_{\mathrm{NS,in}}$ for given $X_{\max} = 200 mg\mathrm{DCO}.\mathrm{l}^{-1}$, $K_{\mathrm{NS}} = 0.4 mg\mathrm{N} - \mathrm{NH}^{+}_{4}.\mathrm{l}^{-1}$ and $K_{\mathrm{NB}} = 0.2 mg\mathrm{N} - \mathrm{NO}^{-}_{3}.\mathrm{l}^{-1}$

The second dataset contains one experiment:

- C2Exp1: High but unknown initial concentration $X^{\mathrm{F}}_{\mathrm{NS}}(t=0,z)$ and $X^{\mathrm{F}}_{\mathrm{NB}}(t=0,z)$, $X^{\mathrm{B}}_{\mathrm{tot,in}} = 0$, known high level of ammonia concentration at input (over 1 $mg\mathrm{N} - \mathrm{NH}^{+}_{4}.\mathrm{l}^{-1}$).

When studying this second dataset, the half-saturation constants K_{NS} and K_{NB} are assumed to be known (values determined from the first series of experiments at low influent ammonia concentration) and attention is focused on the estimation of $X_{\max}$, $X^{\mathrm{F}}_{\mathrm{NS}}(t=0,z)$ and $X^{\mathrm{F}}_{\mathrm{NB}}(t=0,z)$. Moreover, since the initial concentration of bacteria is provided by a nitrifying sludge previously acclimated from an activated sludge reactor for 40 days, the fraction of *Nitrosomonas* is set to its standard value $f_{\mathrm{NS}} = 0.63$. The biofilter is in fact inoculated with a so high concentration of *Nitrosomonas* and *Nitrobacter* that it can be verified in Figure 4 that this initial concentration corresponds more or less to the maximum active biomass concentration, i.e. the optimal cost function is given for $X^{\mathrm{F}}_{\mathrm{tot}}(t=0,z) = X_{\max} = 200 mg\mathrm{DCO}.\mathrm{l}^{-1}$.

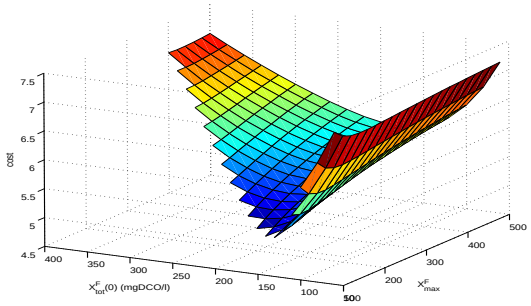


Fig. 4. C2Exp1 - Cost function evolution with the initial concentration of bacteria $X^{\mathrm{F}}_{\mathrm{tot}}(t=0,z)$ and maximum active biomass concentration $X_{\max}$ (only cases where $X^{\mathrm{F}}_{\mathrm{tot}}(t=0,z) \leq X_{\max}$ are considered), for given $f_{\mathrm{NS}} = 0.63$, $K_{\mathrm{NS}} = 0.4 mg\mathrm{N} - \mathrm{NH}^{+}_{4}.\mathrm{l}^{-1}$ and $K_{\mathrm{NB}} = 0.2 mg\mathrm{N} - \mathrm{NO}^{-}_{3}.\mathrm{l}^{-1}$

5.2 *Model fitting*

The numerical values of the estimated model parameters and particle concentrations are given in Table 4 and 5, respectively. Bounds on the standard deviations and correlations between parameters can be computed using the inverse of the Fisher information matrix. The main correlations are between $X^{\mathrm{F}}_{\mathrm{NS}}(t=0,z)$ and K_{NS} on the one hand, and between $X^{\mathrm{F}}_{\mathrm{NB}}(t=0,z)$, K_{NB}, $X^{\mathrm{B}}_{\mathrm{tot,in}}$ and $f_{NS,in}$ on the other hand. $X^{\mathrm{F}}_{\max}$ is also partly correlated with K_{NS} and K_{NB}. This shows that it is more suitable to estimate the half-saturation constants based on experiments without inoculation of biomass if the inoculum concentrations are not precisely known.

Table 4. Estimated model parameters

Model parameter	value
K_{NS} ($mg\mathrm{N} - \mathrm{NH}^{+}_{4}.\mathrm{l}^{-1}$)	0.4
K_{NB} ($mg\mathrm{N} - \mathrm{NO}^{-}_{3}.\mathrm{l}^{-1}$)	0.2
$X^{\mathrm{F}}_{\max}$ ($mg\mathrm{DCO}.\mathrm{l}^{-1}$)	200

Table 5. Estimated influent and initial particles concentrations

Particle ($mg\mathrm{DCO}.\mathrm{l}^{-1}$)	C1Exp1	C1Exp2	C2Exp1
$X^{\mathrm{F}}_{\mathrm{NS}}(0,z)$	0	0.25	126
$X^{\mathrm{F}}_{\mathrm{NB}}(0,z)$	0	0.25	74
$X^{\mathrm{B}}_{\mathrm{tot,in}}$		0.001	0
$f_{NS,in}$		0.9	0

Figures 5 to 7 show the spatial profiles of the nitrogenous components in the liquid phase and biomass fixed on the porous bed as snapshots of Experiment C1Exp1, at time t = 6 days, 9 days and 22 days, respectively. Figure 8 and 9 show the spatial profiles at time t = 3 days and 18 days relative to Experiment C2Exp1. The graphical results confirm the good agreement between the model and the experimental data.

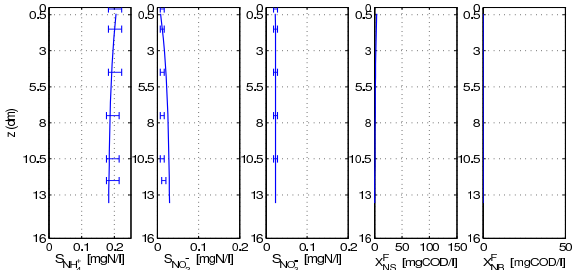


Fig. 5. C1Exp1 - Spatial profiles at t = 6 days. model prediction (solid line) and measurements (dash symbol)

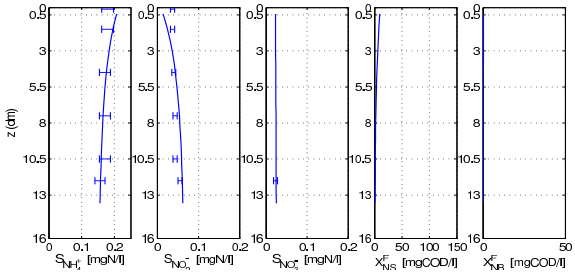


Fig. 6. C1Exp1 - Spatial profiles at t = 9 days. model prediction (solid line) and measurements (dash symbol)

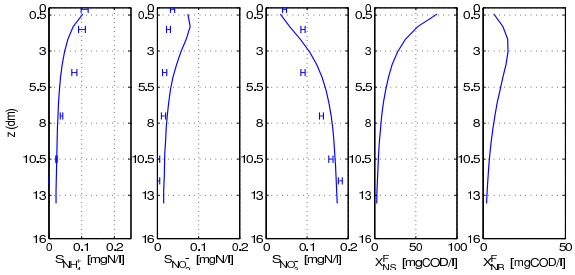


Fig. 7. C1Exp1 - Spatial profiles at t = 22 days. model prediction (solid line) and measurements (dash symbol)

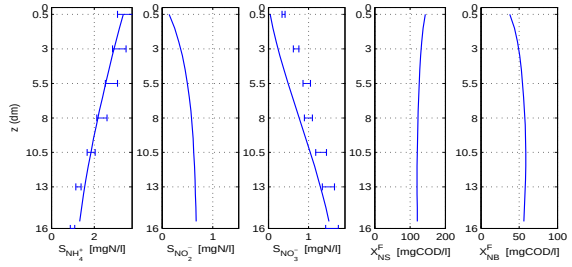


Fig. 8. C2Exp1 - Spatial profiles at t = 3 days. model prediction (solid line) and measurements (dash symbol)

6. CONCLUSION

A mass-balance PDE model has been set up, based on main biological reactions, filtration and adsorption phenomena, and calibrated with experiments carried out with two packed bed biofilters operating under different conditions (e.g., influent water composition, temperature, inoculum, operational events). Parameter estimation was discussed, taking into account the

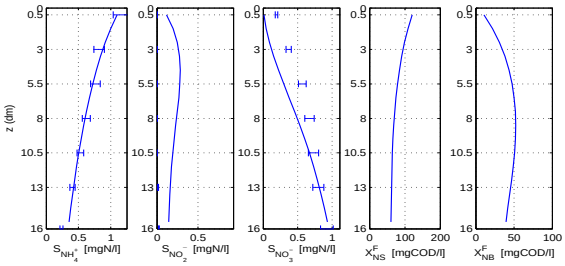


Fig. 9. C2Exp1 - Spatial profiles at t = 18 days. model prediction (solid line) and measurements (dash symbol)

published results, parameter sensitivity analysis, and identification of selected parameters. Validation results show that the model is in good agreement with experimental data (accepting the idea that experiments with biological water treatment systems are delicate to achieve, and are unavoidably corrupted by random perturbations and measurement errors).

Acknowledgements: This research was partially granted by the CNRS/CGRI-FNRS exchange program between Isabelle Queinnec and Alain Vande Wouwer.

REFERENCES

Afnor (1994). Recueil de normes françaises: qualité del'eau.. Technical report.

Freundlich, H. (1906). Über die adsorption in lösungen. *Z. Phys. Chem. A* **57**, 385–470.

Grommen, R., I. Van Hauteghem, M. Van Wambeke and W. Verstraete (2002). An improved nitrifying enrichment to remove ammonium and nitrite from freshwater aquaria systems. *Aquaculture* **211**(1-4), 115–124.

Haag, J., A. Vande Wouwer, E. Paul and I. Queinnec (2004). Experimental modeling of a biofilter for combined adsorption and nitrification. In: *Proc. of 9th IFAC Symposium on Computer Applications in Biotechnology (CAB'9)*. Nancy (France).

Henze, M., P. Harremoës, J. la Cour Jansen and E. Arvin (2002). *Waste Water Treatment*. Biological and Chemical Processes. 3rd ed.. Springer. Berlin.

Hozalski, R.M. and E.J. Bouwer (2001). Non steady-state simulation of bom removal in drinking water biofilters: model development. *Wat. Res.* **35**(1), 198–210.

Iwasaki, T. (1937). Some notes on sand filtration. *J. Am. Wat. Works Assoc.* **29**(10), 1591–1602.

Jacob, J. (1994). Modélisation et Optimisation Dynamique de Procédés de Traitement des Eaux de Type Biofiltre: Traitement dy Systèmes d'Équations Differentielles Partielles et Algebriques (EDPA). PhD thesis. Institut National Polytechnique de Toulouse. Toulouse.

Zhu, S. and S. Chen (1999). An experimental study on nitrification biofilm performances using a series reactor system. *Aquacultural engineering* **20**, 245–259.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DYNAMIC PCA FOR PHASE IDENTIFICATION OF RIFAMYCIN B FERMENTATION IN MULTI–SUBSTRATE COMPLEX MEDIA

Xuan–Tien Doan * and **R. Srinivasan** *,**

* *Institute of Chemical and Engineering Sciences, 1 Pesek Road, Jurong Island, Singapore 627833, e-mail: doan_xuan_tien@ices.a-star.edu.sg*

** *Department of Chemical and Biomolecular Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, e-mail: chergs@nus.edu.sg*

Prashant M Bapat *** and **Pramod P Wangikar** ***

*** *Department of Chemical Engineering, Indian Institute of Technology, Bombay, Powai Mumbai 400076 INDIA, e-mail: {prashan, pramodw}@iitb.ac.in*

Abstract: Information regarding when and how a fermentation process changes from one phase to the next is very useful to its modelling and hence control and optimization. In this study, we demonstrated that such information could be obtained by applying DPCA to online measurements of the fermentation process. The process under study is fermentation of Rifamycin B in a multi–substrate complex medium. We compare our observation to the results obtained from the simulation developed for the same system (Bapat *et al.*, in press). The analysis showed that for the first 100 hours or so, the progress of the fermentation experiment in the DPCA score space matched very well to the developed simulation, which had been validated with actual off–line data (Bapat *et al.*, in press). After that (ie. 100 hours onward), there is a significant difference between DPCA analysis result and the simulation result. The reason seemed to be that the simulation did not capture the effects of the secondary metabolism which becomes dominant at later stage of the fermentation.

Keywords: multivariate statistics, dynamic PCA, fermentation, cybernetic model, substitutable substrates, Rifamycin B

1. INTRODUCTION

There are a number of reasons which necessitate phase identification of fermentation. The first reason lies in the improved understanding of the process. The knowledge of when and how the process change from one phase to the next could give insights into which metabolic pathways the fermentation is undertaking. This is especially relevant to fermentation with multi–substrate complex media where there are many metabolic pathways (corresponding to multiple substrates) for the microorganism to proceed. In addition, phase recognition of fermentation process might also be useful in its optimization and control. A model with high accuracy and high robustness for a fermentation process is always desired but more often than not unavailable. The difficulty in modelling such a process is blamed on the complex dynamics of microorganisms, the variable/ill–defined fermentation media (Lopes and Menezes, 2004), and the multi–phase characteristic of the fermentation itself (Hanai and Honda, 2004). Accurate state identification could help to enable phase–wise process modelling for improved performance.

Multivariate statistical techniques and particularly Principal Component Analysis (PCA) have been used in many areas such as monitoring and supervision of continuous processes (MacGregor *et al.*, 1991) as well as batch processes (Nomikos and MacGregor, 1995); improving process understanding (Kosanovich *et al.*, 1996). In addition, PCA applications have been reported in (Gregersen and Jorensen, 1999; Albert and Kinley, 2001; Lopes and Menezes, 2004) for monitoring and supervision of fermentation process. In this paper, we will use dynamic PCA (DPCA) approach to analyze online data from the fermentation of Rifamycin B in a multi–substrate complex medium. DPCA, a variant of PCA technique, was proposed by (Ku *et al.*, 1995) to account for process dynamic behaviors more effectively. Results from DPCA analysis are compared to the corresponding ones from a simulation developed for the same system and described in (Bapat *et al.*, in press).

2. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is a linear dimensionality reduction technique, optimal in terms of capturing the variability of the data. It determines a set of orthogonal vectors, called loading vectors, ordered by the amount of variance explained in the loading vector directions. The new variables, often referred to as *principal components* are uncorrelated (with each other) and are weighted, linear combinations of the original ones. The total variance of the variables remains unchanged from before to after the transformation. Rather, it is redistributed so that the most variance is explained in the first principal component (PC), the next largest amount goes to the second PC and so on. In such a redistribution of total variance, the least number of PCs is required to account for the most variability of the data sets. The development of PCA model, which can be found in numerous published literature including (Ralston *et al.*, 2001; Russell *et al.*, 2000) is summarized as follows. For a given data matrix $\mathbf{X}^o$ (raw data), which has n samples and m process variables as in (1), each row $\mathbf{x_i^T}$ is a sample of m variables associated with a given time.

$$\mathbf{X}^o = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \tag{1}$$

where: x_{ij} is the data value for the j^{th} variable at the i^{th} sample.

Initially, some scaling is required. The most common approach is to scale the data using its mean and standard deviation

$$\mathbf{X} = (\mathbf{X}^o - \mathbf{1}_n \mu^T)\mathbf{D}^{-1} \tag{2}$$

where: $\mathbf{X}^o$ is a $n \times m$ data set of m process variables and n samples.

μ is the $m \times 1$ mean vector of the dataset.

$\mathbf{1}_n = [1, 1, \dots, 1]^T \in \mathbf{R}^n$.

$\mathbf{D} = diag(sd_1, sd_2, \dots, sd_m)$ whose i^{th} element is standard deviation of the i^{th} variable.

After appropriate scaling, the loading vectors can be determined by singular value decomposition (SVD) of the data matrix

$$\frac{1}{\sqrt{n-1}}\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{3}$$

where: $\mathbf{U} \in R^{n \times n}$ and $\mathbf{V} \in R^{m \times m}$ are unitary matrices.

$\mathbf{\Sigma} \in R^{n \times m}$ is diagonal matrix.

Solving Equation 3 is equivalent to solving an eigenvalue decomposition of the sample covariance matrix $\mathbf{S}$

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^T \tag{4}$$

The matrix $\mathbf{\Sigma}$ contains the nonnegative real singular values of decreasing magnitude along its main diagonal ($\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{min(m,n)}$), and zero off–diagonal elements. Column vectors

in the matrix $\mathbf{V}$ are the loading vectors. Upon retaining the first a singular values, the loading matrix $\mathbf{P} \in R^{m \times a}$ is obtained by selecting the corresponding loading vectors.

The projections of the observations in $\mathbf{X}$ into the lower dimensional space are contained in the score matrix

$$\mathbf{T} = \mathbf{XP} \tag{5}$$

and the projection $\hat{\mathbf{X}}$ of $\mathbf{T}$ back into the m–dimensional observation space

$$\hat{\mathbf{X}} = \mathbf{TP}^T \tag{6}$$

The residual matrix $\mathbf{E}$ is the difference between $\mathbf{X}$ and $\hat{\mathbf{X}}$

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} \tag{7}$$

The residual matrix $\mathbf{E}$ contains that part of the data not explained by the PCA model with a principal components and usually associated with "*noise*", the uncontrolled process and/or instrument variation arising from random influences. The removal of this data from $\mathbf{X}$ can produce a more accurate representation of the process, $\hat{\mathbf{X}}$ (Russell *et al.*, 2000).

Dynamic Principal Component Analysis (DPCA)

Ordinary PCA presented above is essentially a linear technique, and hence its best applications are limited to steady state data with linear relationships between variables (Misra *et al.*, 2002). To analyze a dynamic system, *Dynamic Principal Component Analysis* (DPCA) is required. The concept of DPCA was based on applying PCA to time lagged input data (Ku *et al.*, 1995).

Mathematically, DPCA starts with forming a time–lagged version of the input data $\mathbf{X}$

$$\mathbf{X}_d^o = \begin{pmatrix} \mathbf{x}(d+1)^T & \mathbf{x}(d)^T & \dots & \mathbf{x}(1)^T \\ \mathbf{x}(d+2)^T & \mathbf{x}(d+1)^T & \dots & \mathbf{x}(2)^T \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \mathbf{x}(n)^T & \mathbf{x}(n-1)^T & \dots & \mathbf{x}(n-d)^T \end{pmatrix} \tag{8}$$

where: $\mathbf{x}(k) = [x_{k,1} x_{k,2} \dots x_{k,m}]^T$ is the m–dimensional observation vector at time k. n is the number of data samples. d is the time lag.

The corresponding covariance matrix $\mathbf{S}$ for the time–lagged data is

$$\mathbf{S} = \frac{(\mathbf{X}_d)^T (\mathbf{X}_d)}{n - d - 1} \tag{9}$$

Solving the eigen–decomposition of the covariance matrix $\mathbf{S}$ (Equation 4) and retaining a principal components gives the DPCA model for $\mathbf{X}$.

3. RIFAMYCIN B FERMENTATION MODEL

The fermentation model that we used in this study was developed by P. Wangikar and his colleagues at Indian Institute of Technology (Chemical Engineering Department) and reported in (Bapat *et al.*, in press). It is a dynamic model for the fermentation of Rifamycin B, an antibiotic which is produced on industrial scale, in a multi–substrate complex medium. The model considers the organism to be an optimal strategist (maximizing growth and product formation) with a built–in mechanism that regulates the sequential and simultaneous uptake of multiple substrate combinations. The uptake of individual substrate is assumed to be dependent on the level of a key enzyme or a set of enzymes. In addition, the fraction of flux through a given metabolic branch is estimated by solving the constraint multivariate optimization problem.

4. EXPERIMENT DATA

A detailed description of the Rifamycin B fermentation experiment can be found in (Bapat *et al.*, in press). In the experiment, a combination of different substrates were employed. In this study, we analyze *GLU_AMS_SFCSL_FEDBATCH* experiment which had *GLU*cose, *AM*monium *S*ulphate, *S*oya *F*lour and *C*orn *S*teep *L*iquor. Initial conditions for the experiment are outlined in Table 1.

Table 1. Initial conditions of Rifamycin B fermentation experiment

Variables (g/L)	GLU_AMS_SFCSL_FEDBATCH
Biomass	0.65
Amino acid	4
Glucose	70.43
$(NH_4)_2SO_4$	3.4
Insoluble	20

From the online data collected from the experiments, we selected the measurements for a number of variables which correspond to the experimental conditions (cf. Table 2), to form a data matrix input to DPCA analysis.

5. DPCA ANALYSIS

5.1 Methodology

Procedure to carry out DPCA analysis is summarized below

Table 2. Variables in DPCA analysis

No.	Variables
1	Age (hour)
2	exhaust CO_2 concentration (%)
3	exhaust O_2 concentration (%)
4	pH
5	dissolved O_2 concentration (%)
6	stirring rate (rpm)

(1) The data set is initially augmented (ie. transform into the lagged form $\mathbf{X}_d$) as shown in Equation 8. Several time lags were studied and based on the findings in (Bapat *et al.*, in press), the time lag is set at $t = 8$ hour.
(2) Auto–scaling is applied to $\mathbf{X}_d$ (Equation 2).
(3) The covariance matrix $\mathbf{S}$ of the augmented data is evaluated (Equation 9)
(4) Eigen–decomposition of $\mathbf{S}$ is performed and $a = 2$ principal component vectors are retained.

5.2 Results and Discussion

Fig. 1 shows the fermentation progress in DPCA score space. When there is a change in the progress's direction, the point is marked as a red dot and corresponding time is shown. The simulation developed by P. Wangikar and his colleagues was run for the same initial condition as that for the experiment. Its results are presented in Fig. 3. For better visualization, the result for amino acid predicted by the simulation is shown in a separate plot (ie. Fig. 2).

Observing Figs. 1, 2 and 3, we can conclude that the results from DPCA analysis agree very well with the simulation results for the first 100 hours. As the simulation results shown in Fig. 3 indicates, among the three substrates, amino acid has the largest consumption rate at the beginning of the fermentation. When its consumption rate slows down, corresponding rates of other substrates start to increase. This is reflected in Fig. 1 as a turning point at $t = 20$ hr. The next significant change in the fermentation progress occurs at $t = 27$ hr when the amino acid actually starts being reproduced. Around $t = 60$ hr, Fig. 3 shows that the fermentation media runs out of ammonium sulfate and this results in the turning point at $t = 60$ hr in the score plot Fig. 1. During 60 to 92 hr, both amino acid and glucose are consumed but from 92 hr, the prior substrate is reproduced while the latter continues being consumed. Again, DPCA detects the change and reflects in a turning in the fermentation progress (cf. Fig. 1). At $t = 135$ hour it seems that DPCA results could be implying the depletion of amino acid in the media, which is also predicted by the developed simulation.

However, from $t = 97$ hour, DPCA results start to deviate from what is predicted by the simulation. For example, DPCA score plot clearly indicates that phase changes occur at $t = 105$ hr, $t = 125$ hr, $t = 146$ hr but no such changes could be observed from the simulation results shown in Fig 3.

The reason for this discrepancy needs further investigation and especially verification directly with actual experimental data (which will be available in the near future), instead of simulation results. Nevertheless, it should be noted that as the nitrogen source starts to deplete (i.e., both ammonia and amino acids) around $t = 90$ hr, the fermentation goes into a mode of endogenous metabolism, where some cell lysis occurs and cells grow on the nitrogen available from protein released by lysis. Glucose uptake continues for growth and for maintenance. The secondary metabolic product formation, which is more significant in this phase, was not accounted for by the developed model. This explains the observation that the simulation model fit experimental data very well until the depletion of nitrogen source from the medium (Bapat *et al.*, in press). Toward the end of fermentation run, as the secondary metabolism becomes dominant, the simulation results appear significantly deviate from the actual data (Bapat *et al.*, in press). Consequently, comparison between DPCA results and the simulation results for close–to–end fermentation experiment might not give any valid conclusion.

6. CONCLUSION

We applied DPCA to online measurements of Rifamycin B fermentation data to study the fermentation progress. We compared our observation to the results obtained from the simulation developed for the same system (Bapat *et al.*, in press). The analysis showed that for the first 100 hours or so, the progress of the fermentation experiment in the DPCA score space matched very well to the developed simulation, which had been validated with actual off–line data (Bapat *et al.*, in press). After that (ie. 100 hours onward), there is a significant difference between DPCA analysis result and the simulation result. The reason seemed to be that the simulation did not capture the effects of the secondary metabolism which becomes dominant at later stage of the fermentation.

The study demonstrated the capability of DPCA in identifying phase changes, which could be useful in fermentation process optimization and control. For further work, we are going to validate the DPCA results with the actual off–line data, which as believed will further support the capability of DPCA. In addition, data from fermentation of Rifamycin B in other complex media are

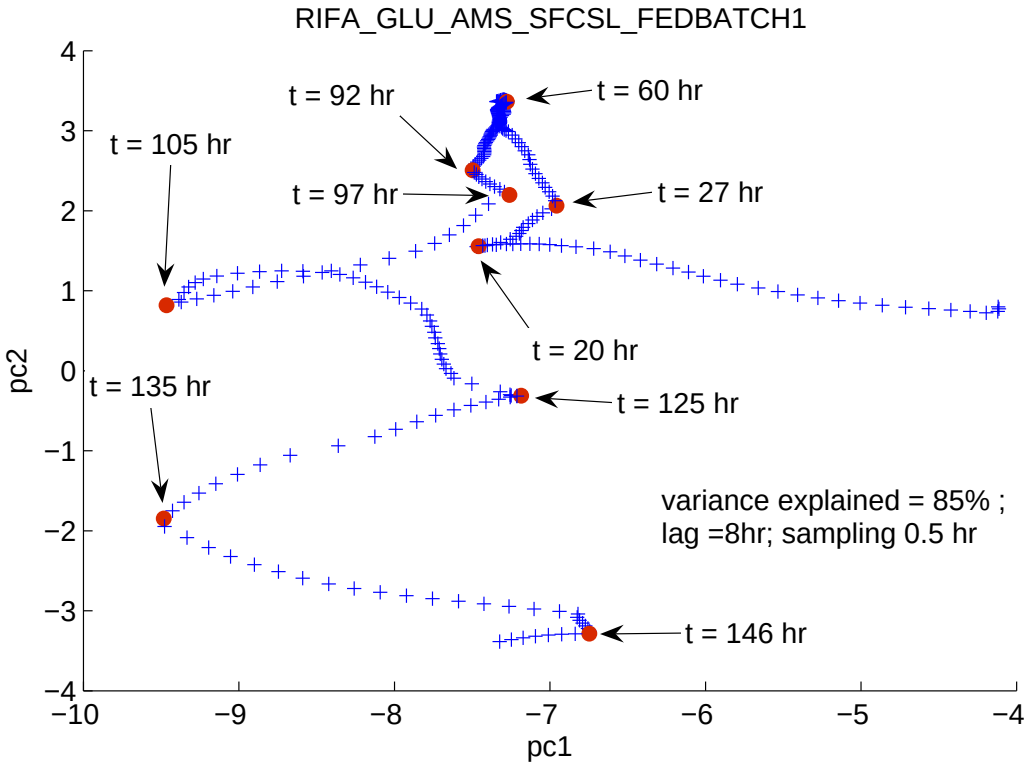


Fig. 1. Score plot from DPCA analysis: all red dots correspond to the time where phase changes in fermentation are likely to occur

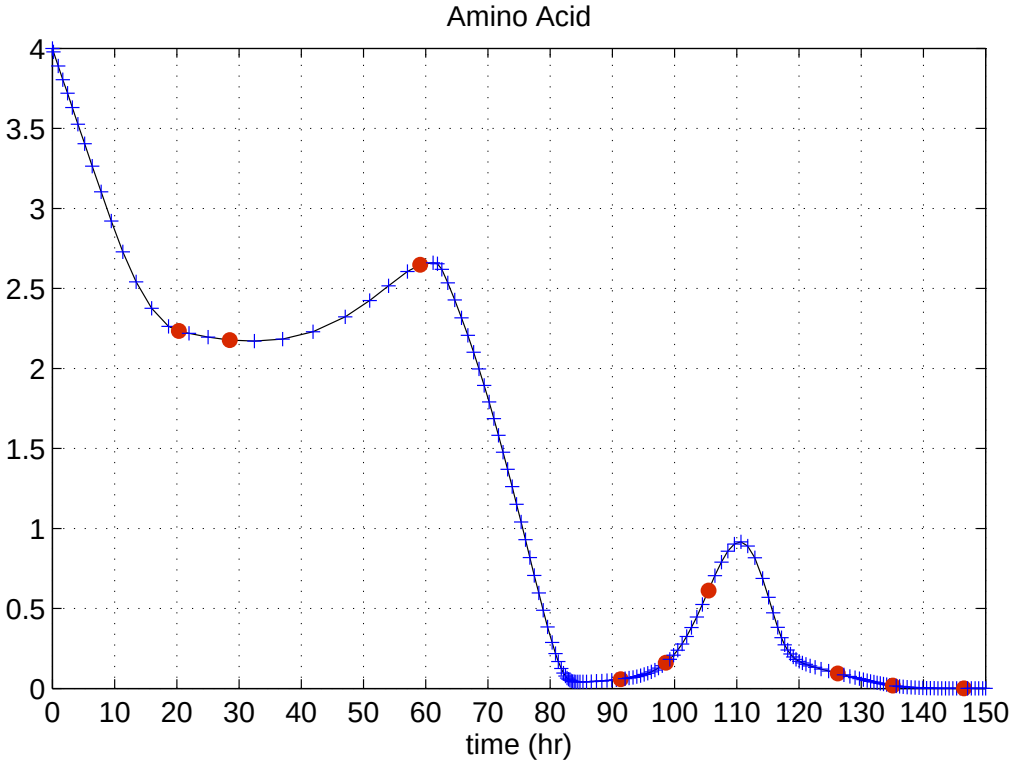


Fig. 2. Simulation results for amino acids in the same experiment with corresponding red dots as in Figure 1

also available and will be analyzed in the same way. These works would establish the ground for further studies such as building inferential PLS model and integrate it with the developed simulation for better optimization and control.

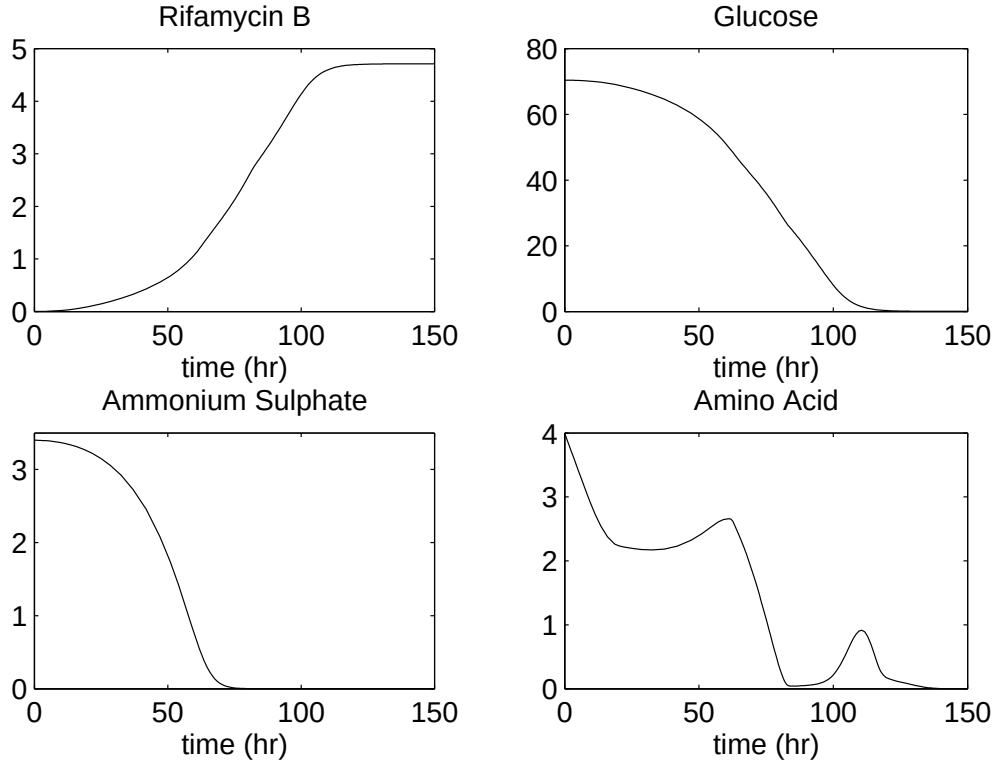


Fig. 3. Simulation results for the same system

REFERENCES

Albert, S. and R. D. Kinley (2001). Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. TRENDS *in Biotechnology* **19**, 53–62.

Bapat, Prashant M., Sharad Bhartiya, K. V. Venkatesh and Pramod P. Wangikar (in press). A structured kinetic model to represent the utilization of multiple substrates in complex media during rifamycin b fermentation. *Biotechnology & Bioengineering*.

Gregersen, Lars and Sten Bay Jorensen (1999). Supervision of fed–batch fermentations. *Chemical Engineering Journal* **75**, 69–76.

Hanai, Taizo and Hiroyuki Honda (2004). Application of knowledge information processing methods to biochemical engineering biomedical and bioinformatics fields. *Adv Biochem Engin/Biotechnol* **91**, 51–73.

Kosanovich, Karlene A., Kenneth S. Dahl and Michael J. Piovoso (1996). Improved process understanding using multiway principal component analysis. *Ind. Eng. Chem. Res.* **35**, 138–146.

Ku, W., R. H. Storer and C. Georgakis (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **30**, 179–196.

Lopes, J. A. and J. C. Menezes (2004). Multivariate monitoring of fermentation processes with non–linear modelling methods. *Analytica Chimica Acta* **515**, 101–108.

MacGregor, J. F., T. E. Marlin, J. Kresta and B. Skagerberg (1991). Multivariate statistical methods in process analysis and control. In: *Cheical process control* – CPCIV. pp. 79–100.

Misra, M., H. H. Yue, S. J. Qin and C. Ling (2002). Multivariate process monitoring and fault diagnosis by multi-scale PCA. *Computers and Chemical Engineering* (26), 1281–1293.

Nomikos, P. and J. F. MacGregor (1995). Multivariate SPC charts for monitoring batch processes. *Technometrics* **37**, 41–59.

Ralston, P., G. DePuy and J. H. Graham (2001). Computer–based monitoring and fault diagnosis: a chemical process case study. *ISA Transactions* (40), 85–98.

Russell, E. L., L. H. Chiang and R. D. Braatz (2000). *Data–driven Techniques for Fault Detection and Diagnosis in Chemical Process*. Springer–Verlag London.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

A NEW MODEL OF PHENOL BIODEGRADATION AND ACTIVATED SLUDGE GROWTH IN FEDBATCH CULTURES

Cherif Ben-Youssef * **Julio Waissman** **
Gabriela Vázquez **

* *Universidad Politécnica de Pachuca, Carr. Pachuca-Cd. Sahagún, 43830, Hidalgo Mexico*
** *CITIS/CIQ, Universidad Autónoma del Estado de Hidalgo, Pachuca, 42084, Hidalgo, Mexico*

Abstract: Phenol biodegradation using acclimated activated sludge was investigated in batch and fedbatch cultures with variable initial concentrations of phenol and biomass. The batch experimental data show a cell growth after phenol exhaustion, making the use of a conventional Haldane model, for which specific growth rate $\mu = 0$ when substrate $S = 0$, inadequate to describe biomass growth profiles. On the other hand, the fedbatch experiments have shown enhanced inhibitory effects on phenol degradation when compared to batch cultures. Both phenomena were attributed to the metabolic intermediates accumulation and to their later consumption during phenol degradation. Consequently, a new Haldane-based model that explicitly integrates the kinetic evolution of a main intermediate was designed. The model allowed accurate predictions of both phenol and biomass concentration courses in both operating modes over a wide range of initial conditions.

Keywords: Batch, Fedbatch, Phenol biodegradation, Substrate inhibition, Activated sludge, Mathematical modelling

1. INTRODUCTION

Phenol is a pollutant commonly found in industrial wastewaters and the aquatic environment due mainly to its numerous applications as intermediate in chemical processes. This molecule is toxic to several biochemical functions and to fish life and causes severe odor and taste problems in drinking water even at low concentrations. Since regulatory agencies impose strict limits on the phenol content in industrial discharges, the development and application of efficient treatment technologies are necessary.

Phenol has usually been removed by costly physicochemical methods, as adsorption, ion exchange or chemical oxidation. However, biological treatments, mostly based on activated sludge systems, are well-suited decontamination techniques because they have shown to be economical and lead to complete mineralization of phenol to innocuous products (CO_2 and H_2O). The major drawback of these processes is their sensitivity to variations in the pollutant load, as at certain concentrations phenol is an inhibitory substrate even for the bacterial species able of using it as an energy and carbon source. In this paper, the attention is focused on *fedbatch* reactors, in which substrates are fed either intermittently or continuously during the process. This operating mode offers many advantages –at least from an industrial viewpoint– over batch and continuous cultures. The main advantage is concretely economical, since improved pro-

ductivity may be obtained by providing controlled conditions in the supply of inhibitory substrates (Shioya, 1992).

Mathematical modelling is a powerful tool for understanding the behavior of biological processes and optimizing their efficiency through the design of model-based control strategies. Although a simple Monod equation has been successfully applied to describe phenol degradation (Reardon *et al.*, 2000), the Haldane substrate-inhibition model has been the most frequently used in the various operating modes (batch, fedbatch and continuous) for both mixed and pure cultures (Leonard *et al.*, 1999; Spigno *et al.*, 2004).

However, as it has been often described in the literature, the Haldane model predictions may present large discrepancies between data generated by batch runs and their application to fedbatch or continuous cultures (Garcia-Sanchez *et al.*, 1998). Even in batch cultures, evident disagreements between measured and estimated phenol concentration profiles were observed when using higher initial concentrations of phenol. Explanations of these discrepancies have often been attributed to the non-stationary behavior of the microorganisms, resulting in changing values of the model's kinetic parameters, including the biomass-to-substrate yield coefficient.

Other authors have pointed out the fact that inhibition parameters can only be estimated at higher phenol concentrations and that the Haldane model does not take into account the effects due to the production and consumption of several metabolic intermediates during phenol degradation. Among these, the major intermediate, the 2-hydroxymuconic acid semialdehyde, has been reported to be associated with the quantity of phenol consumed during the culture, likewise its accumulation has been correlated to the appearance of a yellow color in the culture medium (Leonard *et al.*, 1999; Mörsen and Rehm, 1990). However, the quantification of each of the intermediates is obviously a highly difficult task to carry on, especially in the case of mixed populations.

In terms of mathematical modelling, some authors have proposed modified Haldane models without explicit integration of the inhibitory intermediate effects on the model structure (Nuhoglu and Yalcin, 2005). A very few studies reported a dynamic model that explicitly integrates the intermediate production kinetics, e.g. the work of (Garcia-Sanchez *et al.*, 1998) for chemostat cultures with a pure culture of *P. putida* and none have been applied in the case of activated sludge in fedbatch cultures.

The aim of this work is to develop a new model of phenol biodegradation and biomass growth suited to design fedbatch flow rates strategies able to optimize the phenol consumption and minimize the cell growth. In this study, various batch and fedbatch experiments regarding the growth of an acclimatized activated sludge culture on phenol were carried out, and a new model able to efficiently fit the corresponding experimental data and that explicitly takes into account the inhibitory effects of the metabolic intermediates on phenol biodegradation was proposed.

2. PROCESS DESCRIPTION

2.1 Reactor system

A laboratory scale reactor system that could operate in batch and fedbatch modes was used for the study. A diagram of the bioreactor system is shown in Fig. 1. The plant was linked to a monitoring PC through a NI-PCI-6024E data acquisition board from the National Instrument family. The Labview 7 Express was used as programming language and development tool for data acquisition and storage, graphic display, digital implementation of the feed flow rate profiles in fedbatch cultures, control of the steered speed, pH, aeration and peristaltic pumps.

2.2 Materials and methods

Phenol (99.5% purity) was purchased from Sigma-Aldrich (Germany); 4-amino antipyrine and other chemicals were purchased from J. T. Baker (U.S.A.). Samples of activated sludge were obtained from a plant treating both municipal and industrial wastewater (San Juan Ixhuatepec, Mexico). Raw activated sludge was acclimated to increasing phenol concentrations in daily semi-continuous cycles. Each day, a liter of activated sludge was mixed with the same volume of settled domestic wastewater sampled from the University sewer ($340 - 400$ mg DBO_5/l) and a variable volume of concentrate phenol solution (20 g/l). After 23 hours of aeration and 0.75 hours of sedimentation, the supernatant was drained off and new sewage added. The phenol concentrations were $5 - 700$ mg/l starting and finishing the acclimation period, respectively. The solids content in the reactor was maintained at 4–5 gTSS/l. The medium used had the following composition (mg/l): K_2HPO_4 (404), KH_2PO_4 (220), $(NH_4)_2SO_4$ (50), $MgSO_47H_2O$ (10), $CaCl_22H_2O$ (1.85), $MnCl_24H_2O$ (1.5), $FeCl_36H_2O$ (0.3). Batch experiments were conducted at ambient temperature (21^0C) in 1 l glass bottles containing 600 ml of medium and variable volumes of a phenol solution (20 g/l) and acclimatized activated sludge, in order to provide different initial contents of

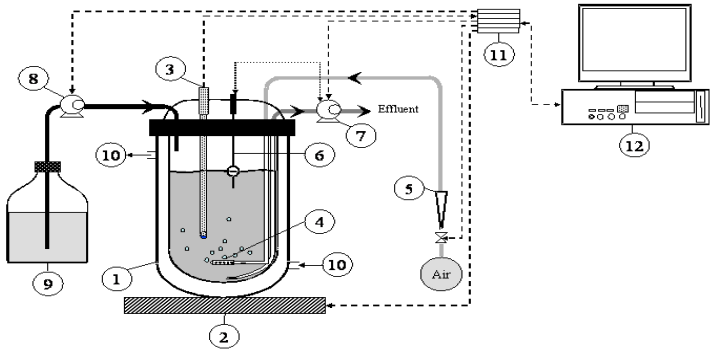


Fig. 1. Diagram of the laboratory bioreactor system; 1 = bioreactor; 2 = magnetic stirrer; 3 = pH meter; 4 = air sprinkler; 5 = air flux meter; 6 = level controller; 7 = controlled peristaltic pump; 8 = controlled peristaltic pump; 9 = phenol feed tank (1.036 g/l); 10 = thermostated water jacket; 11 = data acquisition card (stirred speed, pH, aeration, pumps); 12 = monitoring computer

substrate and biomass. Air was introduced to the bottles by means of aquarium sprinklers, which assured also a good mixture of the cultures. During the experiments, samples were obtained at different times to assess the phenol and biomass concentrations. For phenol analysis, samples were first centrifugated at 3500 rpm for 10 minutes. Phenol was determined colorimetrically by using the 4-aminoantipyrine method (Woolard and Irvine, 1995). In short, 0.2 ml of a 0.1 M glycine solution containing 5% (w/v) $K_3Fe(CN)_6$ was added to 2 ml of centrifuged sample in a 10 ml vial. After mixing, the content was allowed to react for 5 minutes. Then, 2 ml of glycine buffer containing 0.25% (w/v) 4-aminoantipyrine was added. The glycine buffer was prepared by mixing 5.58 g of glycine hydrochloride, 3.75 g of glycine and 900 ml of distilled water, and by adjusting the pH to 9.7 with 6 N NaOH and finally diluting to 1 l. The content of the vial was mixed and allowed to react for 20 minutes. The absorbance of the mixture at 506 nm was measured in a Perkin Elmer Lambda 40 UV-vis spectrophotometer within the next 30 minutes, in order to avoid a decrease in the assay response. The calibration curves were made and found linear up to a concentration of 25 mg/l ($r^2 = 0.999$). This method had a detection threshold limit of 0.07 mg/l and a variation coefficient of 1.06% of the measured values ($n = 9$). For measuring biomass, the total suspended solids (TSS) concentration was determined gravimetrically by filtering 10 ml-samples through a 0.2 μm-pore-size membrane and drying for 24 hours at 105 ^{0}C. The detection threshold limit of the method was 0.284 gTSS/l, with a variation coefficient of 2.68% of the measured values ($n = 9$).

3. KINETIC MODELLING CONSIDERING THE ROLE OF A METABOLIC INTERMEDIATE

3.1 Model structure

In this work, a mathematical model that explicitly takes into account the main intermediate effect on both phenol degradation and cell growth is proposed. The process is assumed to be described by the following simplified reaction scheme:

$$S_1 \overset{\curvearrowleft}{\rightarrow} X + S_2 \tag{1}$$

$$S_2 \overset{\curvearrowleft}{\rightarrow} X \tag{2}$$

where S_1 is the phenol concentration, S_2 is the main metabolic intermediate concentration, X is the total microbial population concentration growing on either S_1 and S_2. Since the main intermediate S_2 was not identified, its true concentrations are unknown, so S_2 is actually expressed in pseudo-concentration. The mass-balance equations for the various constituents of phenol biodegradation gives the following first-order set of differential equations:

$$\frac{dX}{dt} = \mu X - \frac{Q_{in}}{V} X \tag{3}$$

$$\frac{dS_1}{dt} = -q_{s_1} X + \frac{Q_{in}}{V}\left(S_1^{in} - S_1\right) \tag{4}$$

$$\frac{dS_2}{dt} = \nu_{s_2} X - q_{s_2} X - \frac{Q_{in}}{V} S_2 \tag{5}$$

$$\frac{dV}{dt} = Q_{in} - Q_{out} \tag{6}$$

where μ is the specific biomass growth rate, q_{s_1} and q_{s_1} are, respectively, the specific consumption rate of phenol and the intermediate; ν_{s_2} is the specific intermediate production rate. The different

modes of culture can be directly coupled to this general set of mathematical equations by setting ($Q_{in} = Q_{out}$) in continuous cultures, ($Q_{in} = Q_{out} = 0$) in batch cultures and ($Q_{out} = 0$) in fedbatch cultures.

3.2 Modelling of the specific reaction rates

In this work, since it was assumed that the activated sludge is growing on both phenol and the metabolic intermediate, the global specific growth rate μ may be expressed as:

$$\mu = \mu_1 + \mu_2 \tag{7}$$

where μ_1 is modelled according to a modified (in order to integrate the inhibitory effect of intermediate accumulation on phenol consumption) Haldane-type equation whilst a modified Monod-type one is used for μ_2:

$$\mu_1 = \frac{\mu_{\max_1} S_1}{K_{s_1} + S_1 + S_1^2/K_{i_1}} \cdot \frac{K_2}{K_2 + S_2} \tag{8}$$

$$\mu_2 = \frac{\mu_{\max_2} S_2}{K_{s_2} + S_2} \cdot \frac{K_1}{K_1 + S_1} \tag{9}$$

One may note from (8) and (9) the dual-substrate type of structure of the phenol and intermediate kinetics. Constant values of biomass-to-phenol Y_1 and biomass-to-intermediate Y_2 yield coefficients were obtained, so the specific growth and consumption rates where correlated by the following linear relationships:

$$q_{s_1} = \frac{\mu_1}{Y_1} \tag{10}$$

$$q_{s_2} = \frac{\mu_2}{Y_2} \tag{11}$$

The specific production rate of the metabolic intermediate was linearly correlated to the specific growth rate of biomass on phenol as follows:

$$\nu_{s_2} = \alpha \mu_1 \tag{12}$$

Table 1. Parameter values used for model simulation

Symbol	Batch	Fedbatch	SI Units
$\mu_{\max_1}$	0.39	0.4	1/h
Y_1	0.57	0.67	mg/mg
K_{s_1}	30	2	mg/l
K_{i_1}	170	17	mg/l
K_2	160	91	mg/l
$\mu_{\max_2}$	0.028	0.3	1/h
Y_2	0.67	0.75	mg/mg
K_{s_2}	350	75	mg/l
K_1	-	66	mg/l
α	1.6	6.7	mg/l

There are several methods of parameter estimation for model tuning. A model may be fitted either via a numerical method, such as least squares or a quadratic estimation method, or via a heuristic method. In this work, the kinetic parameters were estimated using a direct search method (Hooke and Jeeves, 1961). This common method is often used for models with large numbers of variables and parameters, e.g. the proposed phenol degradation model has 10 parameters.

Two sets of experiments, experiments (A_1, A_2) and (B_1, B_2, B_3), corresponding to two generation of acclimatized activated sludge were achieved (the A-generation of activated sludge was accidently lost). The difference between the acclimation period, i.e. 6 months for the A-generation and 2 months for the B-generation, explains the difference between the kinetic parameters values used for model simulation (see Table 1). The fact that the kinetic parameters vary according to the *history* of the microorganisms has already been pointed out (Sokol, 1988).

Simulations were obtained by using the general dynamical model given by (3-6) together with the kinetic expressions given by (7-12) and a fourth-order Runge-Kutta algorithm for numerical integration of the ordinary differential equations.

4. MODEL VALIDATION

4.1 Batch cultures

The experimental and model predicted data of phenol (S_1), the metabolic intermediate (S_2) and biomass (X) concentrations in batch cultures (A_1 and A_2 experiments) are illustrated in Figs. 2 to 3. From these results, it can be noticed that the fitting between the model-simulated and the experimental data is satisfactory. In particular, the phenomenon of biomass growth after phenol exhaustion is clearly observed, i.e. after $t = 6$ h for the A_1 experiment (Fig. 2) and after $t = 9$ h for the A_2 experiment (Fig. 3). When phenol is totally removed, the main contribution to the total biomass growth is due to the intermediate degradation as shown by the simulated profiles of this variable. This confirms the choice of a model based on the reaction scheme given by Eqs. (1-2). Notice that the effect of the K_1 inhibitory constant in (9) was not included in the model for the A-batch cultures (see Table 1).

4.2 Fedbatch cultures

An exponential feed profile was chosen for the fedbatch experiments (Ben-Youssef *et al.*, 2004).

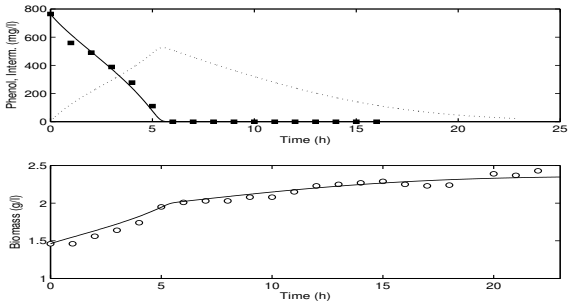


Fig. 2. Batch A_1 experimental and model-simulated phenol, biomass and metabolic intermediate (dotted line) concentrations versus time.

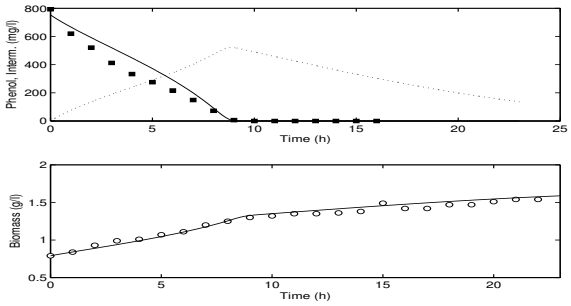


Fig. 3. Batch A_2 experimental and model-simulated phenol, biomass and metabolic intermediate (dotted line) concentrations versus time.

The general expression of the feed flow rate profile is given by:

$$Q_{in}(t) = \begin{cases} 0 & 0 \leq t < t_1 \\ Q_0 e^{a(t-t_1)} & t_1 \leq t \leq t_2 \\ 0 & t_2 < t \leq t_f \end{cases} \qquad (13)$$

Two fedbatch experiments were carried out with two different feeding profiles. The corresponding input parameters of (13) are summarized in Table 2.

Table 2. Parameter values of the fedbatch feed flow rate

Experiment	a	Q_0	t_1	t_2	t_f
Fedbatch F_1	0	350	0	3	6.5
Fedbatch F_2	0.0617	335	0.38	3	6.5

A preliminary batch experiment was necessary to adjust the new set of kinetic parameters and to tune the model (see Fig. 4) for the B-generation of acclimated activated sludge.

The experimental and model-predicted data of phenol, the intermediate and biomass concentrations in fedbatch cultures are illustrated in Fig. 5 and Fig. 6.

The first fedbatch experiment (B_2) was arbitrarily designed with a constant feed flow rate $Q_{in} = 350$ ml/h during 3 hours at a feeding phenol concentration $S_1^{in} = 1036$ mg/h in order to test the ability of the activated sludge to consume approximatively 1 g of phenol. As illustrated by the experimental data plotted in Fig. 5, total phenol degradation was observed whatever the time culture, which is indicative of the substrate-limiting mode of the experiment. The proposed model was able to predict adequately both biomass and phenol concentration evolutions.

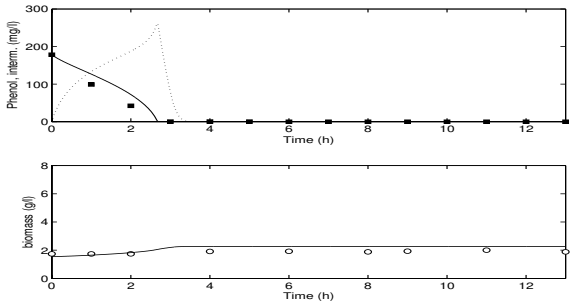


Fig. 4. Batch B_1 experimental and model-simulated phenol, biomass and metabolic intermediate (dotted line) concentrations versus time.

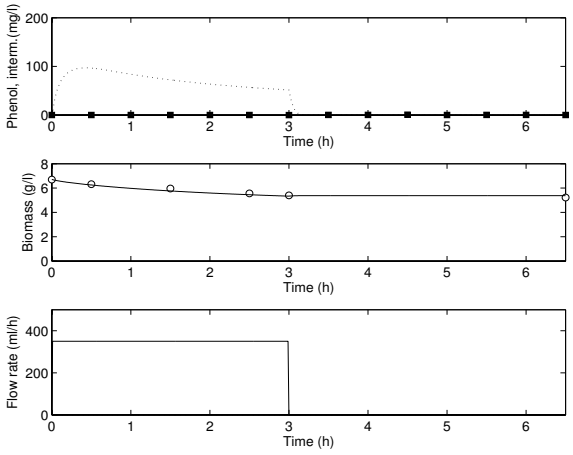


Fig. 5. Fedbatch B_2 experimental and model-simulated evolution: phenol, metabolic intermediate (dotted line) and biomass concentrations; feed flow rate.

The second fedbatch experiment (B_3) was then designed in order to try to switch from phenol degradation under substrate-limitation, as realized in the B_2 experiment, to phenol degradation under substrate-inhibition. However, the principal objective was not only to create inhibitory conditions, which may be easily obtained under high phenol initial and/or high feed flow rate operating conditions, but at the same time to try to perform an inhibitory experiment close to the unstable operating point corresponding to the transition between the limitation and inhibition branches. The *a priori* estimation of the B_3 experiment operating conditions that may satisfy our objective was made by a predictive simulation study using the available proposed phenol degradation model.

The chosen strategy consisted in increasing the initial phenol concentration up to 125 mg/l, then starting a short culture in batch mode during 23 min. (e.g. until the intermediate accumulates enough) and finally activating the fedbatch mode with a slightly different feed flow profile than that used in the B_2 experiment.

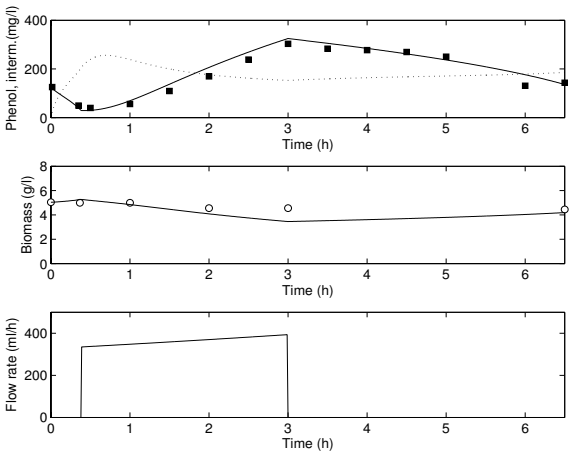


Fig. 6. Fedbatch B_3 experimental and model-simulated evolution: phenol, metabolic intermediate (dotted line) and biomass concentrations; feed flow rate.

The model predictions were accurate as shown in Fig. 6 for both phenol and biomass concentrations. From these figures, it is interesting to note how, despite the slight difference between the two fedbatch policy, higher levels of intermediate concentrations were reached for the B_3 fedbatch experiment. This is in complete accordance with the proposed model structure, which predicts a higher inhibition effect on phenol biodegradation due to the intermediate accumulation.

5. CONCLUSIONS

The main scope of this paper was to derive a reliable new kinetic model for a phenol biodegradation process using acclimatized activated sludge in batch and fedbatch cultures. While the conventional Haldane model appeared to be inaccurate, the explicit integration of the main metabolic intermediate production kinetics in the new model allowed the obtention of adequate fitting results between the experimental and simulated data. The complete model has been experimentally validated for the entire set of cultures under different initial concentrations of biomass and phenol. The model predictions allowed the design of substrate-limitation and substrate-inhibition fedbatch experiments that may be useful to design fedbatch strategies with improved productivity for phenol biodegradation systems.

6. ACKNOWLEDGMENTS

Financial support for this work from CONACyT (research funding SEP-2003-C02-45394) is highly appreciated.

REFERENCES

Ben-Youssef, C., J. Waissman and G. Vázquez (2004). Iterative learning control with input parameterization of a fedbatch lactic acid fermentation process. *WSEAS Trans. on Mathematics* **3**(1), 49–54.

Garcia-Sanchez, J. L., B. Kamp and K. A. Onysko (1998). Double inhibition model for degradation of phenol by *Pseudomonas putida* Q5. *Biotechnol. Bioeng.* **60**(5), 560–567.

Hooke, R. and T. A. Jeeves (1961). Direct search solution of numerical and statistical problems. *J. Assoc. Comp. Machinery* **8**, 212–229.

Leonard, D., C. Ben-Youssef, C. Destruhaut, N. D. Lingley and I. Queinnec (1999). Phenol degradation by *Ralstonia eutropha*: colorimetric determination of 2-hydroxymuconate semialdehyde accumulation to control feed strategy in fed-batch fermentations. *Biotechnol. Bioeng.* **65**(4), 407–415.

Mörsen, A. and H.-J. Rehm (1990). Degradation of phenol by a defined mixed culture immobilized by adsorption on activated carbon and sintered glass. *Appl. Microbiol. Biotechnol.* **33**, 206–212.

Nuhoglu, A. and B. Yalcin (2005). Modelling of phenol removal in a batch reactor. *Process. Biochem.* **40**, 1233–1239.

Reardon, K. F., D. C. Mosteller and J. D. B. Rogers (2000). Biodegradation kinetics of benzene, toluene and phenol as single and mixed substrates for *Pseudomonas putida* F1. *Biotechnol. Bioeng.* **69**, 385–400.

Shioya, S. (1992). Optimization and control in fed-batch bioreactors. *Adv. Biochem. Eng. Biotechnol.* **46**, 111–142.

Sokol, W. (1988). Uptake rate of phenol by *Pseudomonas putida* grown in unsteady state. *Biotechnol. Bioeng.* **32**, 1097–1103.

Spigno, G., M. Zilli and C. Nicolella (2004). Mathematical modelling and simulation of phenol degradation in biofilters. *Biochem. Eng. J.* **19**, 267–275.

Woolard, C. R. and R. L. Irvine (1995). Treatment of hypersaline wastewater in the sequencing batch reactor. *Water Res.* **29**(4), 1159–1168.

Session 4.3

Estimation and Adaptive Control

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

Tuning an adaptive controller using a robust control approach

Jesse Huebsch and Hector Budman

Department of Chemical Engineering, Waterloo, ON, Canada

Abstract: This paper proposes a technique for tuning of a discrete adaptive controller that is designed based on Lyapunov stability concepts. The tuning is based on the minimization of a performance index that can be calculated from a generalized eigenvalue problem (GEVP). The resulting controller, tuned with the proposed methodology, provides better performance than an adaptive controller based on a Recursive Least Squares Estimator (RLS) during sudden changes in model parameters. *Copyright © 2005 IFAC*

Keywords: Adaptive control, robustness, tuning characteristics

1. INTRODUCTION

Adaptive controllers have been proposed for systems that cannot be accurately modelled using available off-line data. This lack of model accuracy often arises for time-varying systems and for systems for which the model structure is not known a priori. For example, the growth rate term in the mass balance equations of bioreactors is often not accurately known (Zhang and Guay, 2002). Therefore, an adaptive estimator maybe used to estimate this term and then a controller can be designed based on this estimated term.

When the model structure of a process is unknown a priori, a commonly used empirical model suitable for adaptive control design, is given as follows (Sanner and Slotine, 1992):

$$y_{k+1} = \sum_{i=0}^{n} a_{i,k} g(y_{k-i}) + \sum_{j=0}^{n} b_{j,k} h(u_{k-j}) \qquad (1)$$

where y is the state, u is the input or manipulated variable, g and h are pre-specified basis functions that can be linear or nonlinear with respect to y and u and *a's* and *b's* are *a priori* unknown parameters to be estimated on-line from input-output data. When linear basis functions are chosen, the model in equation (1) results in the common ARMA model given as follows:

$$y_{k+1} = \sum_{i=0}^{n} a_{i,k} y_{k-i} + \sum_{j=0}^{n} b_{j,k} u_{k-j} \qquad (2)$$

The fact that the models in (1) and (2) are linear with respect to estimated parameters *a's* and *b's* facilitates the design of estimators with proper convergence and performance properties as explained later in the manuscript. Different types of estimators have been used to estimate parameters using the models given by equations (1) and (2). The most common type is the recursive least square estimator (RLS) that minimizes the sum of square errors between the measured and estimated values of y. The recursive least squares estimator is defined by the following two recursive equations:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \frac{\mathbf{P}_{k-2}\mathbf{X}_{k-1}}{c + \mathbf{X}_{k-1}{}^T \mathbf{P}_{k-2}\mathbf{X}_{k-1}} \left[y_k - \mathbf{X}_{k-1}{}^T \hat{\boldsymbol{\theta}}_{k-1} \right] \qquad (3)$$

$$\mathbf{P}_{k-1} = \mathbf{P}_{k-2} - \frac{\mathbf{P}_{k-2}\mathbf{X}_{k-1}{}^T \mathbf{X}_{k-1}\mathbf{P}_{k-2}}{c + \mathbf{X}_{k-1}{}^T \mathbf{P}_{k-2}\mathbf{X}_{k-1}} \qquad (4)$$

Where,

$\boldsymbol{\theta} = [a_1,...,a_n,b_1,...,b_n]$ is the actual vector of parameters assuming that equation (1) is an accurate model of the actual process.

$\hat{\boldsymbol{\theta}}_k = [\hat{a}_{1,k},...,\hat{a}_{n,k},\hat{b}_{1,k},...,\hat{b}_{n,k}]$ is the vector of estimated parameters at time k

$\mathbf{X} = [y_k, y_{k-1}...y_{k-n}, u_k, u_{k-1},...,u_{k-m}]$ is referred as to the regression vector and is a function of past input-output data.

c is the forgetting factor and it is used to assign a larger weight to new data versus older data.

P is the estimation covariance matrix and it is an indicator of the uncertainty in the parameter estimates.

The estimates obtained with the RLS estimator can be used for control using for example a one-step-ahead controller. When the parameter estimates are assumed to be equal to the actual parameters, i.e. the certainty equivalence principle is applied, the one step-ahead controller is:

$$u_k = \frac{ysp_{k+1} - [0,u_{k-2},...,u_{k-n},-y_{k-1},...,-y_{k-n}]\hat{\boldsymbol{\theta}}_{k+1}}{\hat{b}_{1,k+1}} \quad (5)$$

On the other hand, if a more robust controller is desired that accounts for uncertainty in the parameters, a controller referred to as *cautious* (Wittenmark, 1995) can be used that takes into account the uncertainty given by the elements of $\boldsymbol{P_k}$:

$$u_k = g_{cautious}(ysp_{k+1}, \mathbf{X}, \hat{\boldsymbol{\theta}}_{k+1}, \mathbf{P_k}) \quad (6)$$

Under persistent excitation conditions and for *c=0*, the parameter estimates converge to their actual values whereas the matrix **P** converges to zero. This is a desirable outcome for time invariant systems, i.e. systems for which **θ**=constant. However, this is highly undesirable for time varying systems since, according to equation (3), the parameter's adaptation stops when **P**=0. This undesirable scenario can be partially addressed by selecting a nonzero forgetting factor *c*. However, when *c* is nonzero, may cause to very large or infinite values of **P** in the absence of excitation which may ultimately lead to poor control if the controller given by equation (5) is used or alternatively to controller turn-off if the cautious controller in equation (6) is used. To address this issue, resetting of **P**, based on specific algorithms (Huzmezan et al, 2003)) or resetting at ad hoc selected time intervals, has been proposed. In summary, the tuning of an RLS estimator is challenging for chemical systems where parameters are expected to change in a gradual or step-like fashion, due to for instance occasional changes in operating conditions**.**

An alternative estimator that avoids some of the difficulties related to the RLS estimator is the gradient estimator (Sanner and Slotine, 1992). This type of estimator has been proposed for some chemical engineering applications including adaptive control of bioreactors (Perrier and Dochain, 1993). For this estimator the parameter update equation is as follows:

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_{k-1} + \mathbf{K}_{k-1}\mathbf{X}_{k-1}^T S_k \quad (7)$$

$$S_k = f(S_{k-1}, X_{k-1}, K_D) \quad (8)$$

Where, **K** is diagonal and is the adaptation gain matrix S_k is the tracking error and is calculated, as shown later in the manuscript, based on Lyapunov stability concepts and, K_D is a tuning constant that determines the rate of convergence of *S*.

The adaptation gain matrix may be constant or time varying. In this study, **K** will be allowed to change with time and it will be referred to as $\mathbf{K}_k$ to indicate its value at time interval *k.*

The obvious advantage of the gradient estimator is that it does not depend on an adapting covariance matrix **P** that presents inherent difficulties as discussed above**.** On the other hand, the algorithm requires proper tuning of an adaptation gain matrix that has great impact on the estimator performance. Often, in the literature, researchers have selected the gain matrix **K** ad-hoc or based on numerical simulations provided that a suitable model is available. However, there are no available techniques to systematically select the elements of **K.** Additionally, the tracking error equation also introduces an additional tuning parameter K_D as shown above. Therefore, a methodology is needed to tune adaptive controllers that used the gradient estimator given by equation (7).

The objective of the current study is to propose a tuning methodology for this type of adaptive controller with a gradient estimator.

A logical choice to select the tuning parameters ***K*** and K_D is to solve, using the information up to interval *k*-1, the following optimization problem:

$$\min_{\mathbf{K},K_D} E\left\{\frac{1}{N}\sum_{k=1}^{N}(y(k) - y_{setpoint}(k))^2\right\} \quad (9)$$

The expectation of the sum of squares instead of the actual sum is used to account for model uncertainty in the parameters during adaptation and unmeasured disturbances. This problem is closely related to the optimal dual adaptive control problem that search for the optimal trade-off between sufficient excitation for fast model parameter identification versus good tracking properties. In the classical dual adaptive control formulation the minimization is done with respect to the future inputs whereas in equation (9) the cost is minimized with respect to the tuning parameters. However, the two problems are closely related in the sense that the control actions are directly dependent on the tuning parameters.

The problem given by (9) is difficult due to the mathematical expectation that has to be computed for all possible disturbances and in the presence of model uncertainty. Thus, only numerical solutions

have been reported for relatively simple problems and under certain assumptions (Wittenmark, 1995).
In this paper an approximate solution to the problem given by equation (9), based on robust control ideas, is proposed. The idea is to represent the closed loop system by a nominal model and model uncertainty. Using this representation, it will be shown that the problem in (9) can be formulated as an optimization of a set of linear matrix inequalities (LMI's). The paper is organized as follows. Section 2 describes the adaptive control algorithm and the stability and convergence proofs. Section 3 presents the formulation of the tuning problem as an optimization using a set of LMI's. Results and comparisons between the proposed method and an adaptive controller based on RLS estimation are presented in Section 4. Section 5 provides a brief summary and conclusions.

2. ADAPTIVE CONTROLLER ALGORITHM

The controller algorithm presented in this section is a discrete version of an algorithm proposed by Sanner (1992) for continuous systems.

2.1 Definitions

Given a DARMA (Discrete Autoregressive Moving Average) model of a system that is n^{th} order with respect to the state and m^{th} order with respect to the input:

$$y_{k+1} = \sum_{i=0}^{n-1} a_i y_{k-i} + \sum_{j=0}^{m-1} b_j u_{k-j} \qquad (10)$$

The vectors of the parameters a_i and b_j are defined as follows:

$$\mathbf{A} = \begin{bmatrix} a_0 & \cdots & a_{n-1} \end{bmatrix}^T \qquad (11a)$$

$$\mathbf{B} = \begin{bmatrix} b_0 & \cdots & b_{m-1} \end{bmatrix}^T \qquad (11b)$$

The parameter estimate vectors are defined as follows:

$$\hat{\mathbf{A}}_k = \begin{bmatrix} \hat{a}_{0,k} & \cdots & \hat{a}_{n-1,k} \end{bmatrix}^T \qquad (12a)$$

$$\hat{\mathbf{B}}_k = \begin{bmatrix} \hat{b}_{0,k} & \cdots & \hat{b}_{m-1,k} \end{bmatrix}^T \qquad (12b)$$

Let the values of past input and output data be given by the following vectors:

$$\mathbf{Y}_k = \begin{bmatrix} y_k & \cdots & y_{k-n+1} \end{bmatrix}^T \qquad (13a)$$

$$\mathbf{U}_k = \begin{bmatrix} u_k & \cdots & u_{k-m+1} \end{bmatrix}^T \qquad (13b)$$

Then, using equations (11)-(13), the DARMA model given by equation (10), can be reformulated in terms of the input and output vectors as follows:

$$y_{k+1} = \mathbf{A}^T \mathbf{Y}_k + \mathbf{B}^T \mathbf{U}_k \qquad (14)$$

Also for the purpose of designing an implementable controller, let

$$\hat{\mathbf{B}}_k^{old} = \begin{bmatrix} \hat{b}_{1,k} & \cdots & \hat{b}_{m-1,k} \end{bmatrix}^T \qquad (15a)$$

$$\mathbf{U}_k^{old} = \begin{bmatrix} u_{k-1} & \cdots & u_{k-m+1} \end{bmatrix}^T \qquad (15b)$$

A filtered feedback error, to be justified by the stability proof given in the following section, is given as follows:

$$s_k = \frac{-\hat{\mathbf{A}}_k{}^T \mathbf{Y}_{k-1} - \hat{\mathbf{B}}_k{}^T \mathbf{U}_{k-1} + 2y_k + (1-K_D)s_{k-1} - \hat{\mathbf{A}}_{k-1}{}^T \mathbf{Y}_{k-1} - \hat{\mathbf{B}}_{k-1}{}^T \mathbf{U}_{k-1}}{1+K_D} \qquad (16)$$

For simplicity, an adaptive algorithm based on a one-step-ahead controller will be used. A term proportional to the filtered error, s_k, is added to tune the closed loop response, as follows:

$$u_k = \left(y_{sp} - \hat{\mathbf{A}}_k{}^T \mathbf{Y}_k - \hat{\mathbf{B}}_k^{old\,T} \mathbf{U}_k^{old} + (1-K_D)s_k \right) \cdot \hat{b}_{0,k}{}^{-1} \qquad (17)$$

In the particular case that $\hat{b}_{0,k}$ is zero during adaptation, the parameters are reset to the values in the previous time interval. In the presence of persistent excitation, it can be shown that the system will eventually converge to the correct values.

The errors in the estimated parameters are defined in the form of deviation variables as follows:

$$\hat{\mathbf{A}}_k = \tilde{\mathbf{A}}_k + \mathbf{A} \qquad (18a)$$

$$\hat{\mathbf{B}}_k = \tilde{\mathbf{B}}_k + \mathbf{B} \qquad (18b)$$

The gradient descent method is used to formulate the parameter update equations where the error used for updating is the sum of the current and past value of the filtered errors, $(s_k + s_{k-1})$:

$$\hat{\mathbf{A}}_k = \hat{\mathbf{A}}_{k-1} + \mathbf{K}_\mathbf{A} \mathbf{Y}_{k-1} (s_k + s_{k-1}) \qquad (19a)$$

$$\hat{\mathbf{B}}_k = \hat{\mathbf{B}}_{k-1} + \mathbf{K}_\mathbf{B} \mathbf{U}_{k-1} (s_k + s_{k-1}) \qquad (19b)$$

$\mathbf{K}_\mathbf{A}, \mathbf{K}_\mathbf{B}$ are matrices of adaptation gains, with diagonal structure. Thus, the tuning parameters of the controller are K_D and $\mathbf{K} = [\mathbf{K}_\mathbf{A}\ \mathbf{K}_\mathbf{B}]$.

2.2 Stability

Assumptions: For simplicity, it is assumed for the following proof that the system is time invariant or its parameters change in a step-like fashion where the time between changes is long enough such as the parameters converge to a steady state value. Also, for the first proof, the tuning parameters K_D and $\mathbf{K}$ are assumed to be constant with time. Later in this section, the proof is expanded to account for tuning parameters that change in time within a finite set of values. For brevity, only a brief description of the proof is presented.

2.3 Stability with constant tuning parameters:

A Lyapunov function made by the combination of the squares of the estimation errors and the filtered error is defined as follows:

$$V_k = \tilde{\mathbf{A}}_k{}^T \mathbf{K_A}^{-1} \tilde{\mathbf{A}}_k + \tilde{\mathbf{B}}_k{}^T \mathbf{K}_B{}^{-1} \tilde{\mathbf{B}}_k + s_k^2 \quad (20)$$

Substituting equation (17) into equation (14) results in the following:

$$y_{k+1} = \mathbf{A}^T \mathbf{Y}_k + \mathbf{B}_k^{old\,T} \mathbf{U}_k^{old} + b_0 \cdot \hat{b}_{0,k}{}^{-1} \left(y_{sp} - \hat{\mathbf{A}}_k{}^T \mathbf{Y}_k - \hat{\mathbf{B}}_k^{old\,T} \mathbf{U}_k^{old} + (1 - K_D) s_k \right) \quad (21)$$

When the Lyapunov energy converges to zero, $\hat{\mathbf{A}}_k = \mathbf{A}, \hat{\mathbf{B}}_k = \mathbf{B}$ and $s_k = 0$. and then it can be easily shown from equation (21) that: $y_{k+1} = y_{sp}$. For Lyapunov stability it is required:

$$V_{k+1} - V_k \le 0 \quad (22)$$

Combining equations (18), (20) and (22) and after collecting like terms and completing squares:

$$V_{k+1} - V_k = \left(\hat{\mathbf{A}}_{k+1} - \hat{\mathbf{A}}_k\right)^T \left(\mathbf{K_A}^{-1} \hat{\mathbf{A}}_{k+1} + \mathbf{K_A}^{-1} \hat{\mathbf{A}}_k - 2\mathbf{K_A}^{-1} \mathbf{A}\right) + \left(\hat{\mathbf{B}}_{\mathbf{k}+1} - \hat{\mathbf{B}}_{\mathbf{k}}\right)^T \left(\mathbf{K_B}^{-1} \hat{\mathbf{B}}_{\mathbf{k}+1} + \mathbf{K_B}^{-1} \hat{\mathbf{B}}_{\mathbf{k}} - 2\mathbf{K_B}^{-1} \mathbf{B}\right) + (s_{k+1} + s_k)(s_{k+1} - s_k) \quad (22)$$

After evaluating expressions (19a) and (19b) at interval k+1 and substituting the result into equation (22), the following results:

$$V_{k+1} - V_k = -K_D (s_{k+1} + s_k)^2 \quad (23)$$

If $K_D > 0$, Equation (23) guarantees that the Lyapunov function is decreasing with time.

2.4 Stability with time varying tuning parameters

For the optimization problem given by equation (9) it is advantageous to add additional degrees of freedom to the problem by allowing the decision variables, i.e. the tuning parameters K_D and $\mathbf{K}$ ($\mathbf{K_A}$ and $\mathbf{K_B}$) to change with time whereas the stability proof in the previous section assumed that this parameters are constant in time. In this section the stability proof will be extended to account for the situation that the tuning parameters change with time. To the knowledge of the authors it is not possible to prove stability and to solve the optimization in equation (9) for infinite possible values of the tuning parameters. Therefore, it will be assumed that these parameters can only acquire a finite number of values and then a suboptimal solution of (9) will be sought using a combination of these values. Accordingly, a set of tuning parameter values will be defined as follows:

$$\Theta = \{\{K_{D1}\ \mathbf{K_{A1}}\ \mathbf{K_{B1}}\}, \{K_{D2}\ \mathbf{K_{A2}}\ \mathbf{K_{B2}}\} \ldots \{K_{Dn}\ \mathbf{K_{An}}\ \mathbf{K_{Bn}}\}\} \quad (24)$$

For each element of the set Θ, the system is guaranteed to be stable and the parameters converge as per the proof given in the previous subsection. The key idea to ensure stability when different elements of Θ are considered along time, is to calculate simultaneously on-line the evolution the parameter estimates $\hat{B}_k$ and $\hat{A}_k$ and the error s_k for all elements of Θ. However, only one specific control action based on one of the elements is actually implemented at any given time. In figure 1 the curves labelled '1' through '5' refer to the Lyapunov function given by equation (20) corresponding to parameters $[K_{A,1}, K_{B,1}, K_{D,1}]$ through $[K_{A,5}, K_{B,5}, K_{D,5}]$ respectively when, for example, n=5 in definition (24). Then when a new element of Θ is considered, i.e. for a specific $\{K_{Di}\ \mathbf{K_{Ai}}\ \mathbf{K_{Bi}}\}$, the parameter estimates $\hat{B}_k$ and $\hat{A}_k$ and the filtered error s_k are reset to the values corresponding to this element of the set. This switch occurring at each time interval may cause a local increase in Lyapunov energy. As an example, refer to the jump in Lyapunov energy between letters 'A' and 'B' in Fig 1. Each curve in Figure 1 corresponds to the progression of the Lyapunov function for a different element of Θ. Clearly, the Lyapunov function ultimately converges to zero despite that temporary increases in this function may occur.

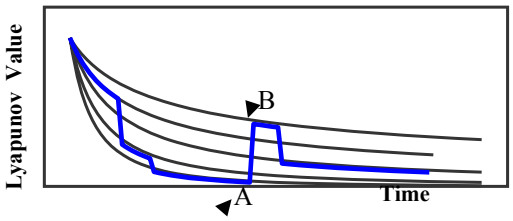


Fig 1: Lyapunov function as a function of time (each line correspond to a different element of the set Θ defined by equation (24))

3. AN APPROXIMATE SOLUTION FOR THE PROBLEM STATED IN EQUATION (9).

As mentioned above, the optimization problem given by equation (9) is very difficult since it has to be solved for all possible disturbances and model uncertainty. Therefore, in this section, an approximated solution to this problem based on a robust control approach is proposed. In order to apply this approach, a nonlinear state space model is formulated, based on definitions (10)-(19) presented above, as follows:

$$
\begin{aligned}
y_{i,k+1} &= l_i\left(y_k, d_{i,k}, \widetilde{\mathbf{A}}_k, \hat{\mathbf{A}}_k, \widetilde{\mathbf{B}}_k, \hat{\mathbf{B}}_k, s_k, k\right) \\
s_{i,k+1} &= f_i\left(y_k, y_{sp,k}, d_{i,k}, \widetilde{\mathbf{A}}_k, \hat{\mathbf{A}}_k, \widetilde{\mathbf{B}}_k, \hat{\mathbf{B}}_k, s_k, \mathbf{K}_A, \mathbf{K}_B, K_D, k\right) \\
\hat{\mathbf{A}}_{i,k+1} &= g_i\left(y_k, y_{sp,k}, d_{i,k}, \widetilde{\mathbf{A}}_k, \hat{\mathbf{A}}_k, \widetilde{\mathbf{B}}_k, \hat{\mathbf{B}}_k, s_k, \mathbf{K}_A, \mathbf{K}_B, K_D, k\right) \\
\hat{\mathbf{B}}_{i,k+1} &= h_i\left(y_k, y_{sp,k}, d_{i,k}, \widetilde{\mathbf{A}}_k, \hat{\mathbf{A}}_k, \widetilde{\mathbf{B}}_k, \hat{\mathbf{B}}_k, s_k, \mathbf{K}_A, \mathbf{K}_B, K_D, k\right) \\
y_{sp,k+1} &= \tau_d^{-1} \cdot y_{sp,k} + \left(1 - \tau_d^{-1}\right) y_{ref,k}
\end{aligned}
\tag{25}
$$

Where the disturbance $d_{i,k}$ is an output bounded disturbance and τ_D in the last equation in (25) is the desired closed loop time constant. For example, for a first order system, the first equation in (25) is:

$$y_{k+1} = a \cdot y_k + b \cdot u_k + d_k \tag{26}$$

$$d_k \in [-\delta_D, \delta_D] \tag{27}$$

The other state equations in (25) are derived from equations (10)-(19) and are omitted here for brevity.

The deviations in the model parameters $\widetilde{\mathbf{A}}_{i,k}$ and $\widetilde{\mathbf{B}}_{i,k}$, defined by equation (19), are not known because the actual values of these parameters are not known a priori. On the other hand, bound on these parameters can be obtained on-line by calculating confidence intervals of these parameters based on regression of current and past input output data. Then, based on definition (18) the deviations in parameters with respect to their nominal values $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{B}}_k$ are assumed to be bounded by the identified confidence intervals as follows:

$$\widetilde{\mathbf{A}}_k \in [-\boldsymbol{\delta}\mathbf{A}_\mathbf{k}, \boldsymbol{\delta}\mathbf{A}_\mathbf{k}], \ \widetilde{\mathbf{B}}_k \in [-\boldsymbol{\delta}\mathbf{B}_\mathbf{k}, \boldsymbol{\delta}\mathbf{B}_\mathbf{k}], \tag{28}$$

Equations (27) and (28) define an uncertainty set :

$$\Delta = \left\{[-\boldsymbol{\delta}\mathbf{A}_\mathbf{k}, \boldsymbol{\delta}\mathbf{A}_\mathbf{k}], [-\boldsymbol{\delta}\mathbf{B}_\mathbf{k}, \boldsymbol{\delta}\mathbf{B}_\mathbf{k}], [-\delta_D, \delta_D]\right\} \tag{29}$$

Liu (1968) has shown that bounds on the stability and performance properties of a nonlinear system can be found from the properties of an equivalent linear time varying system that is given as follows:

$$
\begin{aligned}
&\begin{bmatrix} \boldsymbol{\eta}_{k+1} \\ e_k \end{bmatrix} = \begin{bmatrix} \mathbf{E}_k(\boldsymbol{\delta}_i) & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_k \\ v_k \end{bmatrix} \\
&\boldsymbol{\eta}_0 \text{ is known} \\
&\boldsymbol{\eta}_k = \begin{bmatrix} y_k & s_k & \hat{\mathbf{A}}_k & \hat{\mathbf{B}}_k & y_{sp,k} \end{bmatrix} \\
&v_k = \begin{bmatrix} y_{ref,k} \end{bmatrix} \quad e_k = y_{sp,k} - y_k \\
&\mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 - \tau_d^{-1} \end{bmatrix}^T \\
&\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \end{bmatrix} \\
&\mathbf{H} = [0]
\end{aligned}
\tag{32}
$$

Where the matrix $\mathbf{E}_k$ is given by the linear combinations of matrices $\mathbf{E}_{i,k}$ obtained from the Jacobian of the nonlinear system given by equations (25) calculated at each one of the possible combinations of the uncertainty set vertices defined by equation (29) as follows:

$$\left[\frac{\partial \boldsymbol{\eta}_{k+1}(\delta_i)}{\partial \boldsymbol{\eta}_k}\right] = \left[\mathbf{E}_{i,k}(\delta_i)\right] \tag{33}$$

Since the derivatives in (33) can be shown to be linear with respect to the uncertainty elements described in (29), then:

$$
\begin{aligned}
[\mathbf{E}_k] &= \sum_{i=1}^{L} \alpha_{i,k} [\mathbf{E}_i(\delta_i)], \\
\sum_{i=1}^{L} \alpha_{i,k} &= 1, \ \alpha > 0
\end{aligned}
\tag{34}
$$

For example, for a first order system, there are 8 possible combinations of the uncertainty bounds according to (34a) and (35) and correspondingly 8 possible matrices $\mathbf{E_i}$ are calculated at each time interval *k*.

Defining the ratio between the feedback errors to the input setpoint changes $y_{ref,k}$:

$$\gamma > \frac{\|e\|_2}{\|v\|_2} \tag{35}$$

Then, a bound on γ for all the models defined by equation (32) can be found from a General Eigenvalue Problem (GEVP) defined as follows:

$$\Phi = \min_{\mathbf{P}} \gamma$$

s.t.

$$
\begin{bmatrix}
\mathbf{E}_k(\boldsymbol{\delta}_i)^T \mathbf{P} E_k(\boldsymbol{\delta}_i) - \mathbf{P} & \mathbf{E}_k(\boldsymbol{\delta}_i)^T \mathbf{P} F & \mathbf{G}^T \\
\mathbf{F}^T \mathbf{P} \mathbf{E}_k(\boldsymbol{\delta}_i) & \mathbf{F}^T \mathbf{P} \mathbf{F} - \gamma^2 \mathbf{I} & \mathbf{H}^T \\
\mathbf{G} & \mathbf{H} & -\mathbf{I}
\end{bmatrix} < 0 \tag{36}
$$

This GEVP can be solved by using the LMI toolbox of Matlab. Then, an approximated solution to the problem given by equation (9) can be obtained from:

$$\min_{\Theta} \Phi \tag{37}$$

Where, Φ is calculated from (36) and Θ is a finite combination of tuning parameters defined by (24) and Δ is the uncertainty set defined by equation (29). The minimization in equation (37) is done using the function *fmin* in Matlab.

After initializing the values $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{B}}_k$ and control action vector $\mathbf{U}_k^{old}$, the tuning procedure includes the following steps at each time interval *k*:

1. Calculate the uncertainty bounds in the uncertainty set Δ using available data up to the current time..
2. Update the parameter values according to equations (25) for each one of the tuning parameters combinations in the set Θ defined by equation (24).
3. Find the best tuning parameter combination in the set Θ by solving the optimization problem stated in equation (37).
4. Implement into the process the control action that corresponds to the best set of tuning parameters found in step 3 (It should be noticed that simultaneous calculations for all the combinations in the set Θ are carried on at each interval *k* but only

one control action corresponding to the best combination is actually implemented.
5. Go to step 1.

4. EXAMPLE

To illustrate the tuning method, a first order system was investigated as described by equations (38). First order filtered white noise disturbance d and square wave input $y_{ref,k}$ are considered and step changes in the parameters are assumed to occur at k=100 and 150 respectively as described in (38). The tuning parameters in the set Θ defined by (32) are limited to all the combinations of the values [0.4, 1.6, 3]. For example:

$$\Theta = \{K_A, K_B, K_D\} = \{\{1.6,0.4,3\},\{1.6,1.6,3\},\{3,3,0.4\}....etc\}$$
$$x_{k+1} = a_k x_k + b_k u_k$$
$$y_k = x_k + \eta_k$$
$$where$$
$$k \in \{1,\cdots,200\}$$
$$a_k = \begin{cases} 1.05 & k < 100 \\ 0.95 & 100 \le k \le 200 \end{cases}$$
$$b_k = \begin{cases} 0.5 & k < 150 \\ 0.6 & 150 \le k \le 200 \end{cases}$$
$$\eta_k = (1-\beta)d_k + \beta\eta_{k-1},\ \beta = 0.75, d \in N(0,0.005)$$
$$y_{sp,k+1} = (1-\alpha)y_{sp,k} + \alpha y_{ref,k}, \alpha = 0.65,$$
$$y_{ref,k} : squarewave\ (amplitude = 0.5, period = 20) \quad (38)$$

Figure 2 shows the evolution of the tuning parameters as a function of time following the solution of the optimization given by (38) at each time interval in the neighbourhood of the parameter step change. This figure shows that the tuning parameters have to change frequently in time both due to the change in parameters and the oscillating setpoint. Table 1 shows the normalized sum of square errors along the simulation. For comparison, the sum of squared errors obtained from a simulation for an arbitrarily tuned adaptive controller with fixed in time tuning parameters (Ka=1,Kb=1 and Kd=1) is shown in Table 1. The sum of errors is significantly larger for the arbitrary tuning as compared to the LMI method illustrating the need for proper tuning. Finally an adaptive controller based on an RLS estimator is simulated. The resulting normalized error, significantly larger than the error obtained with the LMI based method, is tabulated in Table 1. As expected the RLS based controller does not perform as well as the LMI controller especially after the step change in the parameters. The reason is that before this step change occurs, at k=100, the covariance matrix P converges almost to zero and consequently the RLS estimator responds very slowly to the sudden change in the model parameters as compared to the proposed estimator tuned according to equation (38). This is clearly shown in Figure 3 where the adaptation of parameter a is shown for both the proposed tuning method and for the RLS based method.

CONCLUSIONS

A method is proposed to tune an adaptive controller based on a gradient estimator. The method uses a robust control approach to minimize a cost function in the presence of model uncertainty and disturbances. The controller based on the proposed tuning method is shown to be superior to a controller based on an RLS estimator during step changes in parameters.

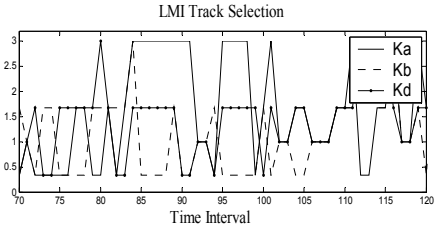


Fig. 2: tuning parameters as a function of time resulting from the proposed optimization (eq. (37))

Tuning	Ka=1,Kb=1 KD=1	Proposed method	RLS
$\|y_{sp} - y\| / \|y_{sp}\|$	0.0084	0.0074	0.0109

Table 1: Comparison of the normalized sum of squared errors.

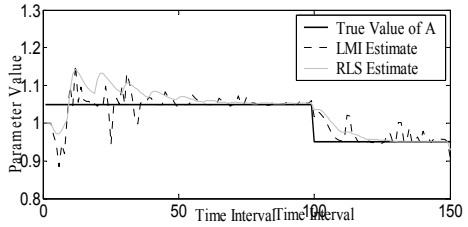


Fig. 3: adaptation of parameter a for the proposed estimator (eq.(37)) and for RLS.

REFERENCES

Liu, R. W. (1968) Convergent Systems, IEEE Trans. On Aut. Control, 13, 4, 384-391.

Huzmezan, M.,G.A.Dumont,W.A.Gough and S. Kovak (2003) Adaptive Control of Integrating Time Delay Systems: A PVC Batch Reactor, IEEE Trans. On Contr. Syst. Tech., 11, 3, 390-398.

Perrier, M. and D. Dochain (1993) Evaluation of Control Strategies for Anaerobic Digestion Processes. International J. of Adaptive and Signal Processing, 7, 309-321.

Sanner, R.M. and J.J. Slotine, (1992), Gaussian Networks for Direct Adaptive Control, IEEE Trans. On Neural Networks, 3, 837-862.

Wittenmark, B. (1995) Adaptive dual control methods: An overview. In 5th IFAC Symposium on Adaptive Systems in Control and Signal Processing, 67-72, Budapest, Hungary.

Zhang, T. and and M. Guay, (2002) Adaptive nonlinear observers of microbial growth processes, Journal of Process Control, 12, 633-643.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PARAMETER CONVERGENCE IN ADAPTIVE EXTREMUM SEEKING CONTROL

V. Adetola and M. Guay [1]

Queen's University, Kingston ON, K7L 3N6 Canada

Abstract: This paper addresses the problem of parameter convergence in adaptive extremum seeking control design. An alternate version of the popular persistence of excitation condition is proposed for a class of nonlinear systems with parametric uncertainties. The condition is translated to an asymptotic sufficient richness condition on the reference set-point. Since the desired optimal set-point is not known *a priori* in this type of problem, the proposed method includes a technique for generating perturbation signal that satisfies this condition in closed loop. This demonstrates its superiority in terms of parameter convergence. The method guarantees parameter convergence with minimal but sufficient level of perturbation. The effectiveness of the proposed method is illustrated with a simulation example.

Keywords: Extremum seeking; Persistence of excitation; Sufficient richness.

1. INTRODUCTION

Extremum seeking control (ESC) is a class of adaptive control that deals with regulation to unknown set points. This type of control has been proposed by a number of authors to handle optimization problems in nonlinear control systems and a number of applications of this method have been reported in the literature ((Krstic and Wang, 2000; Wang *et al.*, 1998; Guay and Zhang, 2003; Guay *et al.*, 2004) for example). The controller finds the operating set-points that optimize a performance or cost function. The uncertainty associated with the function makes it necessary to use some sort of adaptation and perturbation to search for the optimal operating condition.

One of the main challenges with model based or adaptive extremum-seeking control and most deterministic adaptive control approach is the ability to recover the true unknown values of the parameters. In most approaches, parameter convergence to their true values can only be ensured if the closed-loop trajectories provide sufficient excitation for the parameter estimation routine. In standard linear adaptive control approaches, this problem is tractable (Ioannou and Sun, 1996) and can be solved satisfactorily. A dither signal can be introduced momentarily in the control system to achieve the necessary excitation. For nonlinear systems, the problem of determining appropriate excitation conditions remains open. Although some limited persistence of excitation (PE) conditions have been derived, they remain difficult to apply. Such conditions appear naturally in (Guay and Zhang, 2003) for the solutions of an adaptive extremum-seeking control problem. In fact, the fulfillment of such conditions dictates the performance of the optimization routine.

This study is focused on model based extremum seeking techniques. In particular, we consider the class of adaptive ESC problems introduced in (Guay and Zhang, 2003) where the structure of the objective function is employed in the design. In contrast to non-model based approaches (see

[1] corresponding author. Email: guaym@chee.queensu.ca

(Krstic and Wang, 2000) for example), no direct measurement of the objective function is available but must be inferred through the measurements of the state variables and the estimation of model parameters. Examples of this type of problem arise when the economic function involves quantities such as costs of raw materials, operating costs and values of products aside from system's states and unknown parameters.

In the previous works in this area (for example (Guay and Zhang, 2003; Adetola and Guay, 2005; DeHaan and Guay, 2005)), convergence to the optimum is guaranteed only by assuming the satisfaction of a PE condition. Apart from the fact that it is difficult to choose a signal that satisfies such assumptions, it is necessary to select one that achieves a good compromise between the conflicting objectives of identification and control. This paper complements the previous works by translating the PE condition, which depends on the nonlinear closed loop signals, into a sufficient richness condition on the desired set-point signals. However, since the desired optimal set-point is uncertain in this type of problem, the design of a perturbation signal that satisfies this condition cannot be carried out off-line. The proposed method includes a technique for generating such signal in closed loop. The design guarantees parameter convergence with a minimum loss of regulation performance.

2. PROBLEM DESCRIPTION

Consider the following optimization problem

$$\min_{x_p} \; p(x_p, \theta) \tag{1}$$

subject to the system's dynamics

$$\begin{aligned} \dot{x}_p &= f_p(x) + \phi(x_p)\theta + G_p(x)u \\ \dot{x}_q &= f_q(x) \end{aligned} \tag{2}$$

where $x = [x_p^T \;\; x_q^T]^T \in \mathbb{R}^n$ are the systems states, $u \in \mathbb{R}^m$ is the control input. The vector $x_p \in \mathbb{R}^m$ represents the system states involved in the objective function, θ represents unknown parameter vector assumed to be uniquely identifiable and to lie in a known, convex set $\theta \in \Omega_\theta \subseteq \mathbb{R}^{n_\theta}$. The mappings $f(x) : \mathbb{R}^n \to \mathbb{R}^n$ and $G_p(x) : \mathbb{R}^n \to \mathbb{R}^{m\times m}$ are smooth. The following assumptions are made about (1) and (2).

Assumptions

A1. The function p is C^2 in its arguments and $\partial^2 p/\partial x_p^2 \geq c_0 I > 0, \forall (x_p, \theta) \in (\mathbb{R}^m \times \Omega_\theta)$.

A2. $\exists G_p(x)^{-1} \; \forall x \in \mathbb{R}^n$.

A3. The state $x_q \in \mathbb{R}^{m-n}$ belongs to a positively invariant set for any bounded x_p.

A4. The mapping $\phi(x_p) : \mathbb{R}^n \to \mathbb{R}^{m\times n_\theta}$ is a sufficiently smooth - $C^{\beta-1}$ matrix valued function; $\beta \geq \max\{2,\ \text{ceil}(\frac{n_\theta}{m})\}$, where ceil(.) rounds its argument to the nearest integer towards infinity. Moreover, $\phi(x_p)$ is assumed bounded for bounded x_p.

Assumption $A1$ state that the cost surface is strictly convex in x_p and the simplifying assumption $A2$ is only made in order to allow for a direct design of the adaptive controller.

3. EXTREMUM SEEKING SET-POINT AND CONTROLLER DESIGN

3.1 Set-point update law

Considering the fact that the cost function contains unknown parameter θ, the desired set-point measurement cannot be obtained off-line. However, if the function $p(x_p, \theta)$ is not complex, the optimal value can be determined as a function of θ by solving for x_p in $\partial p/\partial x_p = 0$. When the analytical expression of x_p is not available, the desired set-point may be obtained online using Lyapunov method.

Let $x_p^r \in \mathbb{R}^m$ denote a reference set-point for x_p and $\hat{\theta}$ denote an estimate of the unknown parameter θ. An online update law is designed such that $x_p^r(t)$ approaches the optimum value $x_p^*(\hat{\theta})$ exponentially. Let us consider an optimization Lyapunov function candidate

$$V_{sp} := \frac{1}{2}\left\| \frac{\partial p(x_p^r, \hat{\theta})}{\partial x_p^r} \right\|^2 \triangleq \frac{1}{2}\left\| z_r \right\|^2 \tag{3}$$

Taking the time derivative of V_{sp}, we have

$$\dot{V}_{sp} = \frac{\partial p}{\partial x_p^r}\left[\frac{\partial^2 p}{\partial x_p^r \partial x_p^r}\dot{x}_p^r + \frac{\partial^2 p}{\partial x_p^r \partial \hat{\theta}}\dot{\hat{\theta}} \right]. \tag{4}$$

Choosing the update law as

$$\dot{x}_p^r = -\left(\frac{\partial^2 p}{(\partial x_p^r)^2} \right)^{-1} \left[k_r \frac{\partial p}{\partial x_p^r}^T + \frac{\partial^2 p}{\partial x_p^r \, \partial \hat{\theta}}\dot{\hat{\theta}} \right] \tag{5}$$

with $k_r > 0$, (4) becomes

$$\dot{V}_{sp} \leq -k_r \left\| z_r \right\|^2 \tag{6}$$

Proposition 1. The optimal set-point $x_p^r(t)$ generated by (5) is feasible and converges to $x_p^*(\hat{\theta})$ exponentially.

Proof. Assuming (for now, it will be shown later) that $(\hat{\theta}, \dot{\hat{\theta}})$ is bounded. This assumption coupled with assumption $A1$ ensure that (5) exist and it is finite. It follows from (6) that the origin $z_r = 0$ is exponentially stable Applying the inverse function theorem, it can be seen that the mapping z_r is a diffeomorphism. Hence it concluded that $x_p^r(t)$ converges to $\hat{\theta}$-dependent optimal set-point $x_p^*(\hat{\theta})$ exponentially fast. □

3.1.1. Sufficiently rich optimal set-point Since parameter convergence is a vital issue in ESC design, we have to provide some richness condition on the set-point x_p^r to ensure that $\hat{\theta} \to \theta$ as $t \to \infty$. To achieve this, the set-point is appended with a bounded perturbation signal $d(t)$. The rich set-point is given by

$$r(t) := x_p^r(t) + d(t) \tag{7}$$

where $d(t)$ is a sufficiently smooth and uniformly bounded signal. In particular, the signal is parameterized as

$$d(t) := \sum_{k=1}^{\hbar} a_k(t) \sin(\omega_k t) = a(t)\rho(t) \tag{8}$$

where $a(t) = [a_1(t)\ \ a_2(t)\ \ \dots\ \ a_\hbar(t)]$ is the signal amplitude vector and $\rho(t) = [\sin\omega_1 t\ \ \sin\omega_2 t\ \ \dots\ \ \sin\omega_\hbar t]$, (with $\omega_i \neq \omega_j$ for $i \neq j$), is the corresponding sinusoidal function vector. A method for generating the coefficients $a(t)$ is provided in subsection 4.1. The design ensures that $a(t) \to a^*$, the optimal value that satisfies a PE condition.

3.2 Adaptive tracking controller

Let us define the tracking and parameter estimation error vectors

$$z_c = x_p - r \quad \text{and} \quad \tilde{\theta} = \theta - \hat{\theta}. \tag{9}$$

and consider the Lyapunov function candidate

$$V_c := \frac{1}{2}\|z_c\|^2 + \frac{1}{2}\tilde{\theta}^T\Gamma^{-1}\tilde{\theta} \tag{10}$$

with $\Gamma = \Gamma^T > 0$. Taking the time derivative of V_c along the trajectory of (2), we have

$$\begin{aligned}\dot{V}_c =& z_c^T\Big(f_p(x) + \phi(x_p)\hat{\theta} + G_p(x)u - \dot{r}\Big) \\ &- \dot{\hat{\theta}}^T\Gamma^{-1}\tilde{\theta} + z_c^T\phi(x_p)\tilde{\theta}\end{aligned}$$

Considering the control law

$$u = -G_p(x)^{-1}\Big(f_p(x) + \phi(x_p)\hat{\theta} - \dot{r} + k_c z_c\Big), \tag{11}$$

with $k_c > 0$ and the parameter update law

$$\dot{\hat{\theta}} = \Gamma\phi(x_p)^T z_c, \tag{12}$$

it follows from (2) and (11) that

$$\dot{z}_c = \phi(x_p)\tilde{\theta} - k_c z_c, \tag{13}$$

and the time derivative of the Lyapunov function results in

$$\dot{V}_c \leq -k_c\|z_c\|^2. \tag{14}$$

Proposition 2. Consider the closed loop system (13), adaptive control (11) and parameter update law (12), the design is such that

$$\lim_{t\to\infty}\Big(z_c,\ z_c^{(k)},\ \tilde{\theta}^{(k)}\Big) = 0 \tag{15}$$

with $1 \leq k \leq \beta$ and $(.)^{(k)}$ denotes $\frac{d^k}{dt^k}(.)$.

To prove this result, we need the following lemma.

Lemma 3. Barbalat's lemma (Krstic *et al.*, 1995): A signal $\zeta^{(k)} \to 0$ as $t \to \infty$ if (a) $\int_0^\infty \zeta^{(k)}dt$ exist and its finite and (b) the signal $\zeta^{(k)}$ is uniformly continuous.

Condition (a) is evident when $\zeta \in \mathcal{L}_2$ or $\zeta^{(k-1)} \to 0$ asymptotically and condition (b) can be inferred from the boundedness of $\zeta^{(k)}$ and $\zeta^{(k+1)}$.

Proof of Proposition 2. It is known from (10) that V_c is a positive definite function (bounded from below by zero). Since V_c is non-increasing (14), it is concluded that $z_c(t)$ and $\hat{\theta}(t)$ are uniformly bounded. Moreover, there exist a bounded ς such that $-\infty < -\varsigma \leq V_c(\infty) - V_c(0) < 0$. This implies that $-\varsigma \leq \int_0^\infty \dot{V}_c(\tau)d\tau \Rightarrow \varsigma \geq k_c\int_0^\infty \|z_c(\tau)\|^2 d\tau \Rightarrow \|z_c\|_{\mathcal{L}_2}^2 \leq \varsigma/k_c < \infty$. Since $z_c \in \mathcal{L}_2$ and $\phi(x_p)$ is bounded by assumption, it follows from (13) that $\dot{z}_c(t) \in \mathcal{L}_\infty$. Applying the above lemma, we conclude that $z_c \to 0$ as $t \to \infty$.

Also, we know that $\int_0^\infty \dot{z}_c(\sigma)d\sigma = -z_c(0)$ exists and is finite. From the fact that $\dot{z}_c$ is a function of bounded signals we deduce that $\ddot{z}_c$ is bounded, which implies that $\dot{z}_c$, is uniformly continuous and hence $\dot{z}_c \to 0$ as $t \to \infty$. Also, it follows from the adaptive law (12) that $\lim_{t\to\infty}\dot{\tilde{\theta}}(t) = 0$.

Subsequently, it will be shown that (15) holds for $1 < k \leq \beta$ by induction. Suppose $({z_c}^{(k-1)},\ \tilde{\theta}^{(k-1)}) \to 0$, then $({z_c}^{(k)},\ \tilde{\theta}^{(k)})$ satisfies condition (a). Also, condition (b) is satisfied because $({z_c}^{(k)},\ \tilde{\theta}^{(k)})$ and $({z_c}^{(k+1)},\ \tilde{\theta}^{(k+1)})$ are functions of bounded signals. Hence, $({z_c}^{(k)}, \tilde{\theta}^{(k)}) \to 0$. Since, $({z_c}^{(1)},\ \tilde{\theta}^{(1)}) \to 0$ is guaranteed, we conclude that (15) holds. □

4. PARAMETER CONVERGENCE

Consider the state error dynamic (13) and the parameter error dynamic $\dot{\tilde{\theta}} = -\Gamma\phi(x_p)^T z_c$ obtained from (12). By an argument similar to the one used in traditional adaptive control theory, a sufficient condition for parameter convergence is that the regressor $\phi(x_p)$ be persistently exciting. That is, there exists positive constants μ_0 and T such that

$$\int_t^{t+T} \phi(\tau)^T\phi(\tau)d\tau \geq \mu_0 I, \quad \forall t \geq 0$$

Though the matrix $\phi(\tau)\phi(\tau)^T$ is singular for all τ when $(n_\theta > m)$, the PE condition requires that ϕ rotates sufficiently in space that the integral of the matrix $\phi(\tau)^T\phi(\tau)$ is uniformly positive definite over any interval of some length T. However, it is difficult to check that ϕ satisfies the PE condition since the solution of the closed loop trajectories are not known *a priori*.

In the following, an alternative sufficient condition that addresses the above limitations and guarantees parameter convergence is presented. The condition requires an augmented regressor matrix to be sufficiently rich.

By differentiating (13) with respect to time, z_c^k can be written explicitly as

$$z_c^k = -k_c z_c^{(k-1)} + \sum_{j=0}^{k-1} \frac{(k-1)!}{j!(k-j-1)!} \phi(x_p)^{(j)} \tilde{\theta}^{(k-j-1)}, \quad 1 \le k \le \beta \tag{16}$$

Using proposition 2 and the fact that $z_c \to 0$ in the limit as $t \to \infty$, (which implies that $x_p \to r^* = x_p^*(\bar{\theta}) + a^* \rho(t)$), equation (16) results in

$$\lim_{t\to\infty} z_c^k(t) = \lim_{t\to\infty} \phi(r^*)^{(k-1)}(t)\tilde{\theta}(t) = 0, \quad 1 \le k \le \beta \tag{17}$$

Defining

$$Z_c := \left[z_c^{(1)} \; z_c^{(2)} \; \dots \; z_c^{(\Pi)} \right]^T \quad \text{and} \tag{18}$$

$$\Phi := \left[\phi^T \; \phi^{T(1)} \; \dots \; \phi^{T(\Pi-1)} \right]_{n_\theta \times (m*\Pi)} \tag{19}$$

where $\max\{2, \; \text{ceil}(\frac{n_\theta}{m})\} \le \Pi \le \beta$. Equation (17) can then be re-written in a compact form as

$$\lim_{t\to\infty} Z_c = \lim_{t\to\infty} \Phi(r^*)^T(t)\tilde{\theta}(t) = 0. \tag{20}$$

The next step in the analysis is to decompose the time varying signal Φ into a constant matrix and a periodic part. This procedure is similar to the one presented in (Lin and Kanellakopoulos, 1999). Firstly, (19) is expressed as

$$\Phi = \left[\bar{\phi}_1 \; \dots \; \bar{\phi}_m \quad \dots \dots \dots \quad \bar{\phi}_1^{(\Pi-1)} \; \dots \; \bar{\phi}_m^{(\Pi-1)} \right] \triangleq \left[\psi_1 \; \psi_2 \; \dots \; \psi_{m\Pi} \right], \quad m\Pi = m * \Pi \tag{21}$$

where $\bar{\phi}_l^{(.)}$ is the l^{th} column of matrix $\phi^{T(.)}$. The trigonometric (or Fourier) series expansion for each nonlinearity vector ψ_i is computed as follows: Let $\omega_{i1}, \omega_{i2}, \cdots, \omega_{iCi}$ $(0 \le \omega_{i1} < \omega_{i2} < \cdots < \omega_{iCi})$ and $\nu_{i1}, \nu_{i2}, \cdots, \nu_{iSi}$ $(0 < \nu_{i1} < \nu_{i2} < \cdots < \nu_{iSi})$ denote the distinct frequencies appearing in the cosine terms and the sine terms of the Fourier series expansion respectively. If we let

$$\xi_i(t) = \left[\cos\omega_{i1}t \; \cdots \; \cos\omega_{iC_i}t \; \sin\nu_{i1}t \cdots \; \sin\nu_{iSi}t \right]^T \triangleq \left[\xi_{i1}(t) \; \cdots \; \xi_{iC_i}(t) \; \xi_{i(Ci+1)}(t) \cdots \; \xi_{Ci+Si}(t) \right]^T \quad i = 1, \dots, m\Pi \tag{22}$$

Then, each nonlinearity vector ψ_i defined in (21) can be expressed in the form

$$\psi_i = \Upsilon_i \, \xi_i(t) = \sum_{j=1}^{C_i+S_i} \Upsilon_{ij} \, \xi_{ij}(t), \quad i = 1, \dots, m\Pi \tag{23}$$

where Υ_i are $n_\theta \times (C_i + S_i)$ constant matrices whose elements are the real Fourier coefficients of the corresponding signals, and $\Upsilon_{ij}, j = 1, \dots, (C_i + S_i)$ is the j^{th} column of Υ_i. This decomposition method allows one to judge the richness of the vector based on a constant matrix only. However, as pointed out in (Lin and Kanellakopoulos, 1999), the Fourier series expansion employed in the decomposition may contain an infinite number of terms, when the elements of (21) are not polynomial nonlinearities. In this case, the series expansion may be truncated. Combining (20) with equations (21) and (23), we obtain

$$\lim_{t\to\infty} \xi_{ij}(t) \, \Upsilon_{ij}^T \, \tilde{\theta}(t) = 0 \quad i = 1, \dots, m\Pi, \quad j = 1, \dots, C_i + S_i \tag{24}$$

and since the scalar functions ξ_{ij} are all of the form $\cos\omega t$ or $\sin\nu t$, equation (24) is equivalent to

$$\lim_{t\to\infty} \Upsilon_{ij}^T \, \tilde{\theta}(t) = 0 \quad i = 1, \dots, m\Pi, \quad j = 1, \dots, C_i + S_i. \tag{25}$$

Moreover, defining $\Upsilon_1 = \Upsilon_{11} \cdots \Upsilon_{1(C1+S1)}$, $\Upsilon_2 = \Upsilon_{21} \cdots \Upsilon_{2(C2+S2)}$ etc, and $\Upsilon^T = [\Upsilon_1 \; \cdots \; \Upsilon_m]^T$, equation (25) can be written in a more compact form

$$\lim_{t\to\infty} \Upsilon^T \tilde{\theta}(t) = 0 \quad \text{or} \quad \lim_{t\to\infty} \tilde{\theta}^T(t) \, \mathcal{W} \, \tilde{\theta}(t) = 0 \tag{26}$$

Since Υ is a constant matrix containing the set-point x_p^* and a^* in its entries (in the limit as $t \to \infty$), if the n_θ rows of Υ are linearly independent or if $\mathcal{W} = \Upsilon\Upsilon^T$ is positive definite, then $\tilde{\theta} = 0$ is guaranteed. However, it is not possible to verify this conditions *a priori* for a given dither signal because the matrix depends on unknown reference set-point (the θ-dependent solution of (1)). In the next section, we show how to generate optimal size of some pre-selected sinusoids online.

4.1 Dither signal design

It has been shown that the presence of nonlinearities in a regressor vector increase the degree of PE of a given reference signal for nonlinear systems with special structure (Lin and Kanellakopoulos, 1998; Lin and Kanellakopoulos, 1999). However, for a general nonlinear system, this may not be the case, the nonlinearities may detract or add to the excitation (Dasgupta and Shrivastava, 1991). In this work, we propose that the dither signal be chosen as a linear combination of sinusoids with at least n_θ distinct frequencies. However, since such a choice with constant arbitrary amplitude may not be optimal for nonlinear systems, a method for generating optimal coefficients of the different basis functions (sinusoids) is provided. A quadratic objective function is minimized subject to a constraint that optimizes the size of the selected frequency contents in order to ensure positive definiteness of matrix $\mathcal{W} = \Upsilon\Upsilon^T$.

The condition requires all the eigenvalues of $\mathcal{W}$ to be positive. This is true if and only if the determinant of $\mathcal{W}$ (the product of the eigenvalues) is positive since $\mathcal{W}$ is a symmetric positive semidefinite matrix.

The optimum amplitude of the dither signal is proposed as the solution of the following constrained optimization problem.

$$\begin{aligned} \min_{a\in\mathbb{R}^h} \quad & a^T Q a \\ \text{such that} \quad & \mathcal{W}_d = det(\mathcal{W}) > 0 \end{aligned} \tag{27}$$

with $Q \succ 0$. The optimization problem is tackled using an infeasible interior point technique (Vanderbei and Shanno, 1999). Firstly, a slack variable ε is added so that (27) becomes

$$\begin{aligned} \min_{a\in\mathbb{R}^h} \quad & a^T Q a \\ \text{such that} \quad & \mathcal{W}_d - \varepsilon = 0, \quad \varepsilon > 0. \end{aligned} \tag{28}$$

The constraints are then eliminated by augmenting the objective function with high costs for violating them as follows.

$$\begin{aligned} \min_{a,\varepsilon} \; P_a = & a^T Q a - \frac{1}{M_1}\log(\sigma-\varepsilon) \\ & + M_2(\mathcal{W}_d - \varepsilon)^2, \quad \sigma > 0 \end{aligned} \tag{29}$$

with $M_1, M_2 > 0$. By the logarithmic barrier term, the slack variable is required to be greater than a design variable σ at all times. However, the equality constraint ($\mathcal{W}_d - \varepsilon = 0$) can be violated at any instant, its satisfaction is only achieved as the optimum solution is approached. The solution of (29) can be shown to converge to that of (27) in the limit as the positive constants $M_1, M_2 \to \infty$.

Since we assume that system (2) is fundamentally identifiable at the defining parameter values, feasibility of (27) (and hence (29)) is guaranteed by including sufficiently large number of regressor derivatives in (19). The unconstrained optimization problem (29) can be solved with gradient techniques. Let $\bar{a}^* = [a^*, \varepsilon^*]$ be the optimizer of (29), an update law that ensures $\bar{a} \to \bar{a}^*$ as $t \to \infty$ is chosen as

$$\dot{\bar{a}} = \text{Proj}\left\{-k_{\bar{a}}\mathcal{D}z_{\bar{a}},\, \bar{a}\right\}, \quad \bar{a}(0) = [a_0,\; \varepsilon_0] \tag{30}$$

where Proj{.} is a standard projection algorithm (Krstic *et al.*, 1995) used to ensure that the vector $\bar{a}$ is bounded or remains in some given set. The vector $z_{\bar{a}} = \partial P_{\bar{a}}/\partial\bar{a}$ is the gradient function, $k_{\bar{a}} > 0$ is a design parameter and $\mathcal{D}$ is a positive definite matrix function. Matrix $\mathcal{D}$ can be chosen as in steepest descent method where $\mathcal{D} = I$ (identity matrix) or as in trust region where $\mathcal{D} = \left(\partial^2 P_{\bar{a}}/\partial\bar{a}^2 + (\digamma+\kappa)I\right)^{-1}$ with $\digamma$=Frobenius matrix norm of $\partial^2 P_{\bar{a}}/\partial\bar{a}^2$ and $\kappa > 0$ is a small design constant parameter. The initial conditions are to be selected such that $\varepsilon_0 > \sigma$ and some elements of a_0 equals zero to avoid excessive initial perturbation of the system.

Theorem 4. Consider the optimization problem (1) for system (2) satisfying assumptions $A1 - A4$. The controller (11), the update laws (5), (12) and (30) for sufficiently large Π ensure that the system's state $x_p(t)$ converge to a neighborhood of $x_p^*(\theta)$ - the unique minimizer of (1).

Proof. It can be deduced from proposition (2) that $\lim_{t\to\infty}\left\|x_p(t) - x_p^r(t)\right\| \le \lim_{t\to\infty}\|d(t)\|$ and it is known from proposition (1) that $\lim_{t\to\infty}\left\|x_p^r(t) - x_p^*(\hat{\theta})\right\| = 0$. Moreover, (30) ensures $\lim_{t\to\infty} a(t) = a^*$ for large enough Π. Therefore, the only solution of (26) is $\lim_{t\to\infty}\tilde{\theta}(t) = 0$, which implies $\lim_{t\to\infty}\left\|x_p^*(\hat{\theta}) - x_p^*(\theta)\right\| = 0$. Using triangle inequality, we conclude that $\lim_{t\to\infty}\left\|x_p(t) - x_p^*(\theta)\right\| \le \|a^*\|$. □

5. SIMULATION EXAMPLE

Consider two parallel isothermal stirred-tank reactors (DeHaan and Guay, 2005) in which reagent A forms product B and waste-product C

$$A \xrightarrow{1} B$$
$$2A \xrightarrow{2} C$$

The economic steady state cost function to be optimized is given by

$$\begin{aligned} p(x_p,\, \theta) = & \sum_{i=1}^{2}[(p_{i1} + P_A - P_B)k_{i1}A_i V_i^0 \\ & + (p_{i2} + 2P_A)k_{i2}A_i^2 V_i^0], \end{aligned}$$

where P_A, P_B denote component prices, p_{ij} is the net operating cost of reaction j in reactor i. The reaction kinetic constants k_{ij} are only nominally known. A_i is the concentration of reagent A in reactor i with dynamics

$$\frac{dA_i}{dt} = A^{in}\frac{F_i^{in}}{V_i} - A_i\frac{F_i^{out}}{V_i} - k_{i1}A_i - k_{i2}A_i^2$$

The inlet flows are the control inputs, while the outlet flows are governed by PI controllers which regulate reactor volume to V_i^0. Therefore,

$$\dot{x}_p = \underbrace{-\begin{bmatrix} \dfrac{x_{p1}k_{V1}(x_{q1} - V_1^0 + x_{q3})}{x_{q1}} \\ \dfrac{x_{p2}k_{V2}(x_{q2} - V_2^0 + x_{q4})}{x_{q2}} \end{bmatrix}}_{f_p}$$
$$\underbrace{-\begin{bmatrix} x_{p1} & 2x_{p1}^2 & 0 & 0 \\ 0 & 0 & x_{p2} & 2x_{p2}^2 \end{bmatrix}}_{\phi}\theta + \underbrace{\begin{bmatrix} \dfrac{A_i n}{x_{q1}} & 0 \\ 0 & \dfrac{A_i n}{x_{q2}} \end{bmatrix}}_{G_p} u,$$

where $x_p = [A_1,\ A_2]^T$, x_{q1}, x_{q2} are the two tank volumes, x_{q3}, x_{q4} are the PI integrators, and $\theta = [k_{11}, k_{12}, k_{21}, k_{22}]^T$.

Following the design procedure, the optimizing controller, parameter estimates and the set-point signal x_p^r are generated via equations (11), (12) and (5) respectively. For the simulation, the dither signal is selected as $d_1(t) = d_2(t) = a_1(t)\sin(0.3t) + a_2(t)\sin(0.18t)$ and $\Pi = 3$ so that the augmented regressor matrix $\Phi(r^*)^T = [\phi^T\ \dot{\phi}^T\ \ddot{\phi}^T]$. The matrix Υ is obtained via the decomposition method presented in section 4. For simulation purpose, x_p^* is replaced with its estimate x_p^r at each time t and the optimal value of the dither amplitude that ensures the positive definiteness of $\mathcal{W} = \Upsilon\Upsilon^T$ is obtained via (30). Fig. 1(a) shows that the cost function converges

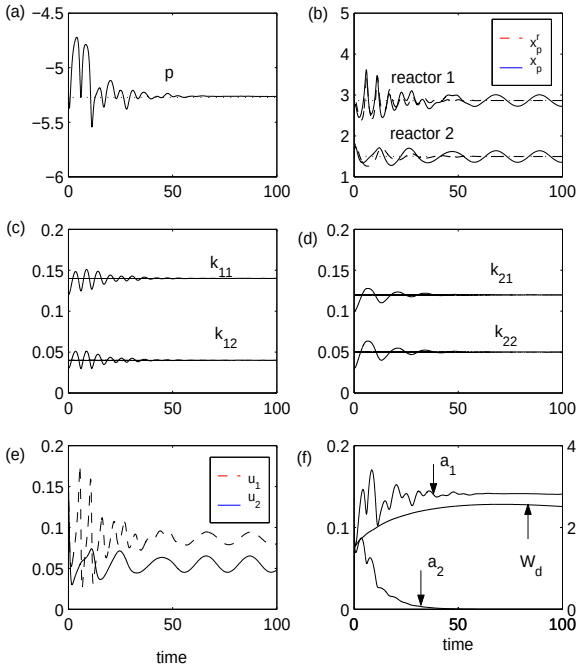


Fig. 1. Simulated system trajectories: (a) cost function, (b) reference set-point and state, (c, d) unknown parameters and estimates for reactor 1 and 2 respectively, (e) control inputs, (f) dither signal amplitude and determinant of matrix $\mathcal{W}$.

to the unknown optimal $p^*(x_p^*, \theta)$. Fig. 1(b) shows that the set-point signal converges to the optimum value $x_p^*(\theta)$ while the state x_p oscillates about the optimum. The parameter estimates converge to the true values as shown in fig. 1(c-d) and the control input, fig. 1(e), is implementable. The trajectories of the dither amplitude and the determinant are shown in fig. 1(f) for completeness. The figure showed that $a(t)$ converges to the required optimum (vertical-axis labelling on the left) and the determinant $\mathcal{W}_d$ remains positive (vertical-axis labelling on the right).

6. CONCLUSION

A persistence of excitation condition is proposed for the ESC of a class of nonlinear systems. An optimization based method is then developed for generating sufficiently rich optimum set-points that satisfies this condition online. The proposed design method guarantees parameter convergence and at the same time ensure small steady-state error in the cost function.

REFERENCES

Adetola, V.A. and M. Guay (2005). Adaptive output feedback extremum seeking control of linear systems. In: *Proceedings of the 16th IFAC world Congress*. Prague.

Dasgupta, Soura and Yash Shrivastava (1991). Persistent exciation in bilinear systems. *IEEE Transactions on Automatic Control* **36**(3), 305–313.

DeHaan, D. and M. Guay (2005). Extremum seeking control of state constrained nonlinear systems. *Automatica* **41**(9), 1567–1574.

Guay, M and T. Zhang (2003). Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties. *Automatica* **39**(7), 1283–1293.

Guay, M., D. Dochain and M. Perrier (2004). Adaptive extremum seeking control of continuous stirred tank bioreactors with unknown growth kinetics. *Automatica* **40**(5), 881–888.

Ioannou, P.A and Jing Sun (1996). *Robust Adaptive Control*. Pentice Hall, Upper Saddle River, New Jersey.

Krstic, M and H.H. Wang (2000). Stability of extremum seeking feedback for general dynamic systems. *Automatica* **36**(4), 595–601.

Krstic, M., I. Kanellakopoulos and P. Kokotovic (1995). *Nonlinear and Adaptive Control Design*. John Wiley and Sons Inc. Toronto.

Lin, Jung-Shan and Ioannis Kanellakopoulos (1998). Nonlinearities enhance parameter convergence in output-feedback systems. *IEEE Transactions on Automatic Control* **43**, 204–222.

Lin, Jung-Shan and Ioannis Kanellakopoulos (1999). Nonlinearities enhance parameter convergence in strict feedback systems. *IEEE Transactions on Automatic Control* **44**, 89–94.

Vanderbei, Robert J. and David F. Shanno (1999). An interior point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications* **13**, 231–252.

Wang, H.H., S. Yeung and M. Krstic (1998). Experimental application of extremum seeking on an axial flow compressor. In: *Proceedings of the American Control Conference*. Philadelphia. pp. 1989–1993.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

GEOMETRIC ESTIMATION OF TERNARY DISTILLATION COLUMNS

[(1)] Anna Pulis, [(2)] Carlos Fernandez, [(1)] Roberto Baratti, [(2)] Jesus Alvarez

[(1)] Dipartimento di Ingegneria Chimica e dei Materiali, Università di Cagliari, Piazza D'Armi, 09123 Cagliari, Italy
[(2)] Departamento de Ingeniería de Procesos e Hidráulica, Universidad Autónoma Metropolitana-Iztapalapa Apdo. 55534, 09340 México, D.F., México

Abstract: The problem of estimating effluent compositions from temperature measurements in ternary distillation columns is addressed within a geometric estimation framework where the estimation structure and the algorithm are jointly designed. The employment of passive estimation structures and error propagation measures yields criteria to choose the sensor number and locations as well as the set of innovated states. The proposed approach is tested with experimental data from a 32-stage pilot column (tert-butanol-ethanol-water system). With 64 on-line dynamical equations and a straightforward tuning scheme, the proposed estimator yields the same behaviour than the one of an Extended Kalman Filter with 2144 equations and an optimization-based tuning procedure

Keywords: distillation column, ternary distillation, passive estimation, process estimation

1. INTRODUCTION

Distillation is an important separation process, and the development of effluent composition estimation schemes is motivated by the need to design or redesign processes to meet more stringent safety, efficiency and quality specifications. In particular, a temperature measurement-driven estimation scheme: (i) is motivated by the cost, reliability and delay drawbacks of composition measurement instruments, and (ii) can be applied to improve the supervisory, advisory or feedback control tasks. The distillation and control problems have been extensively tested with a diversity of techniques, the related state of the art can be seen elsewhere (Baratti *et al*., 1995; Baratti *et al*., 1998; Oisiovici and Cruz, 2000; Venkateswarlu and Avantika, 2001), and here it suffices to mention that the nonlinear extended Kalman filter (EKF): (i) is the most widely used estimation technique, (ii) has been successfully tested in continuous and batch system over a wide range of separations and operating conditions, (iii) has an implementation that requires the tuning of covariance via trial-and-error (Oisiovici and Cruz, 2000;Venkateswarlu and Avantika, 2001) or optimization-based (Baratti *et al*. 1995, Baratti *et al*.1998) searches, and the on-line integration of a set of auxiliary ordinary differential equations (ODEs) whose number grows rapidly with stage number of stages and components. Moreover, it is not clear how the EKF nonlinearity and complexity features can be reconciled with the linearity and simplicity of the majority of the industrial linear (MIMO, decentralized, or one way-decoupled) PI and (MIMO) model predictive control schemes. In principle, the on-line integration of the EKF Riccati equations can be circumvented by employing the nonlinear geometric (Luenberger-like) estimation approach (Alvarez and Lopez, 1999, Alvarez, 2000), and the same approach enables the consideration of the sensor locations and the innovated states as design degrees of freedom, as it can be seen in previous polymerization reactors (Lopez and Alvarez, 2004) and (fixed-sensor) binary distillation column (Tronci *et al*., 2005) estimation studies. In a distillation column, estimation structure means the sensor locations and the set of innovated states, or equivalently, of states whose dynamical model has (direct) measurement injection.

The above considerations motive the scope of the present work: the development of a joint structure (i.e., the sensor location and innovated states-algorithm (i.e., the dynamic data processor) estimation design for ternary distillation columns, with a favorable comparison with the EKF technique in the light of complexity, reliability and implementation-tuning effort considerations. Following the geometric estimation approach (Alvarez, 2000, Lopez and Alvarez 2004), a single-sensor passive structure is chosen on the basis of suitable error propagation measures in conjunction with estimator testing, and the adjustable-algorithm is designed according to a geometric technique, without auxiliary Riccati equations, and with the estimation

structure as design degree of freedom. The proposed approach is tested with experimental data drawn from a 32-stage pilot column with the ethanol- tertbutanol-water system, finding that a passivated estimator with one sensor, 65 ODEs, and a conventional-like tuning scheme yields the same behavior than an EKF with two-sensors, 2144 ODEs, and a tuning drawn from the adjustment of six parameters using off-line optimization.

2. ESTIMATION PROBLEM

Consider an N-stage two-measurement (one per section) *continuous ternary distillation column*, with: (i) molar feed flow F at light- intermediate component composition pair (c^1_F, c^2_F), (ii) bottoms B (or distillate D) effluent rate at composition c_0 (or c_D) of light component, (iii) reboiler heat injection at rate Q (proportional to the vapor flow rate V), and (iii) two measurements, one in the stripping section and one in the enriching section, at locations to be determined. From standard (liquid-vapor equilibrium at each stage, quasi steady-state hydraulics and enthalpy balance with equimolar flow) assumptions, the column behavior is described by the following nonlinear dynamical system (Baratti *et al*. 1995; Skogestad, 1997):

Stripping section ($1 \le i \le n_F$, k = 1, 2)

$$\dot{c}^k_i=[(R+F)\,\Delta^+c^k_i - V\Delta^-\nu_k(c^1_i,c^2_i)]/\eta^i(R+F) \qquad (1a)$$

Feed tray (i = n_F, k = 1, 2)

$$\dot{c}^k_{n_F} =[R\Delta^+c^k_{n_F} -V\Delta^-\nu_k(c^1_{n_F},c^2_{n_F}) + F(c^k_F-c^k_{n_F})]/\eta^i(R+F) \qquad (1b)$$

Enriching section ($n_F + 1 \le i \le N-1$, k = 1, 2)

$$\dot{c}^k_i = [R\Delta^+c^k_i - V\Delta^-\nu_k(c^1_i, c^2_i)]/\eta^i(R) \qquad (1c)$$

Top Tray (i = N, k = 1, 2)

$$\dot{c}^k_N = [R\Delta^+c^k_N - V\Delta^-\nu_k(c^1_N, c^2_N)]/\eta^i(R) \qquad (1d)$$

Measurements

$$y_s = T_s = \beta(c^1_s, c^2_s),\ s \in [1, n_F -1] \qquad (1e)$$

$$y_e = T_e = \beta(c^1_e, c^2_e),\ e \in [n_F +1, N] \qquad (1f)$$

where (k = 1, 2)

$c^1_i + c^2_i + c^3_i = 1$, $\quad \Delta^+c^k_i = c^k_{i+1} - c^k_i$

$\Delta^-\nu^k_i(c^1_i, c^2_i) := \nu^k_i(c^1_i, c^2_i) - \nu_k(c^1_{i-1}, c^2_{i-1})$

$\nu_1(c^1_{-1}, c^2_{-1}) = c^1_0$, $\nu_2(c^1_{-1}, c^2_{-1}) = c^2_0$

$c^1_{N+1} = c^1_D = \nu_1(c^1_N, c^2_N)$, $\quad c^2_{N+1} = c^2_D = \nu_2(c^1_N, c^2_N)$

where c^1_i and c^2_i are the component (molar fraction) compositions in the i-th stage (the third component composition is given by $c^3_i = 1 - c^1_i - c^2_i$), y_s (or y_e) is the measured value of the temperature T_s (or T_e) in the s(or e)-th stage (to be determined) of the stripping (enriching) section, ν_1 (or ν_2) is the nonlinear (liquid-vapor equilibrium) function that determines the i-th component composition in the vapour phase, β is the nonlinear bubble point function that yields the temperature, and η is the tray hydraulics function that sets the exit molar flow rate from the i-th stage.

Knowning that, with at least two sensors, a ternary column is completely locally observable about a steady-state, (Yu and Luyben, 1987; Quintero-Marmol et al., 1991) and that this assesement should be revised in the light of a nonlinear instantaneous observability framework (Alvarez et al., 1999, 2000, 2004), a comment on the consideration of a single-sensor is in order: in a way that is analogous to the design of robust controllers via backstepping (Krstic et al., 1995, Alvarez *et al*., 2004), in the geometric estimation approach one gives up the (possibly illconditioned) complete (nominal) observability structure, in order to favor robustness, diminish observability requirements, at the cost of more sluggish reconstruction rate. From this viewpoint, it makes sense to consider a (possibly with illconditioned complete observability property) ternary distillation column with a robustly detectable single-sensor structure. In this robust partial observability structure case, the estimator has two components, one with measurement innovation and one noninnovated.

Our *estimation problem* consists in jointly designing the estimation structure (i.e., the sensor locations and innovated state set) and the estimation algorithm (the dynamic data processor) for ternary distillation columns. In particular, we are interested in: (i) the developing an estimator design that compares favourably with the EKF, in terms of complexity, reliability and implementation-tuning effort considerations, and (ii) in testing the approach with experimental data.

Experimental run. The experimental data were generated by a pilot distillation column fed by a water-ethanol-tertbutanol mixture (located at University of Padova, Italy). The column has 30 sieve trays, a vertical thermosiphone reboiler, and a total shell-tube condenser (the overhead vapour is totally condensed and the reflux drum is open to the atmosphere). The feed enters the column in the 8-th stage (i.e., n_F = 8), and there are temperature measurements in nine stages (0, 4, 8, 12, 16, 18, 22, 26 and 30). The top and bottom pressures were 814 and 760 mmHg, respectively, a linear pressure drop was assumed, and the liquid feed temperature T_F = 299 K was lower than the one (T-n_F = 366 K) the feed tray. The reflux, feed and vapor rate were (R, F) = (3.486, 3.489)10^{-5}m^3/s and vapor boilup rate V = 1.437 gmol/s, the feed compositions were $c^E_{n_F}$ = 0.0979 and $c^T_{n_F}$ = 0.0630. The initial condition corresponded to a steady-state with higher vapor flowrate V = 1.437 gmol/s. In other words, the column transient was induced by a step decrease in V, to a value of 0.963 gmol/s. For three different times, the resulting temperature profiles, drawn form simulations, are shown in Figure 1, and the corresponding effluent

concentrations over time can be seen in subsequent figures. Henceforth, super index E (or T) in c_i^k denotes the species ethanol (or ter-butanol) in the i-th tray. Due to the presence of close-to-azeotropic compositions in the enriching section, the related temperature profile is rather flat, and some components are in small amount, and consequently, the estimation task in the enriching section should be considerably more difficult than in the stripping section.

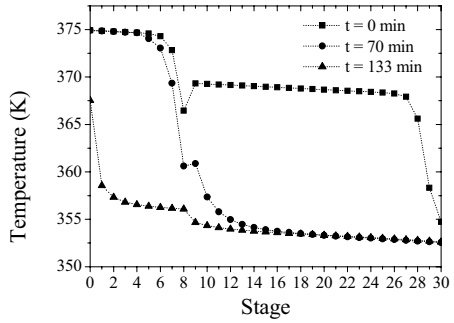


Figure 1: Simulated temperature profiles, at three different times.

3. PASSIVE ESTIMATION

From the adjustable-structure geometric estimation design (Alvarez, 2000; Lopez and Alvarez, 2004) in conjunction with the staged nature of the column, and the recent results of the method application to the binary case (Tronci *et al*. 2005), let us begin the structural assessment by considering single-sensor robustness-oriented *passive structures with one innovated state*. Further motivations for the employment of a passive estimation structure in a combined estimation-control passive design can bee seen elsewhere (Krstic *et al.,* 1995; Alvarez *et al*., 2004; Gonzalez and Alvarez, 2005; Alvarez *et al*., 2005).

3.1 Ethanol as single innovated state

Let us assume that a single sensor is located at the i-th column stage, and that the ethanol composition is the innovated state. The corresponding PI estimator is given by (Alvarez and Lopez, 1999):

$$\dot{\hat{c}}_i^E = f_i^E(\hat{c}_i^E, \hat{c}_i^T, \hat{c}_{i-1}^E, \hat{c}_{i-1}^T, \hat{c}_{i+1}^E, \hat{c}_{i+1}^T) + [1/\beta_{c_E}(\hat{c}_i^E, \hat{c}_i^T)]\{\dot{w} + 2\zeta\omega[y_j - \beta(\hat{c}_i^E, \hat{c}_i^T)]\} \quad (2a)$$

$$\dot{w} = \omega^2[y_i - \beta(\hat{c}_i^E, \hat{c}_i^T)] \quad (2b)$$

$$\dot{\hat{c}}_i^T = f_i^T(\hat{c}_i^E, \hat{c}_i^T, \hat{c}_{i-1}^E, \hat{c}_{i-1}^T, \hat{c}_{i+1}^E, \hat{c}_{i+1}^T), \quad k = E, T \quad (2c)$$

$$\dot{\hat{c}}_j^k = f_i^k(\hat{c}_j^E, \hat{c}_j^T, \hat{c}_{j-1}^E, \hat{c}_{j-1}^T, \hat{c}_{j+1}^E, \hat{c}_{j+1}^T), j \neq i, j \in [1,N] \quad (2d)$$

where

$$\beta_{c_E}(c_E, c_T) = \partial_{c_E}\beta(c_E, c_T), \quad S_i^E = 1/|\beta_{c_E}(c_i^E, c_i^T)| \quad (3a,b)$$

ω (or ζ) is the adjustable characteristic frequency (or damping factor) associated with the underlying nearly linear second-order output error dynamics, w is a dynamical state that estimates and compensates the effect of modelling errors in the predicted output, or equivalently, eliminates the output error mismatch. Due to the almost linear output error dynamics that underlies the preceding estimator construction, the tuning of the pair (ω, ζ) can be performed according to conventional-like techniques and tuning guidelines for second-order linear filters (Alvarez and Lopez, 1999). Typically, ω is from 3-to-15 times larger (faster) than the natural frequency of the measurement response. S_i^E is the asymptotic error propagation measure (Lopez and Alvarez 2004), and will be occasionally referred to as sensitivity measure.

Figure 2 shows the sensitivity measure dependency on the spatial (stage) location, for three different times. As it can be seen in Figure 2: the sensitivity measure profile worsens with time, especially in the enriching section. As expected due to the presence of close-to-azeotropic compositions in the enriching section, that section exhibits more error propagation measure than the stripping section. By far, the column bottom stage is the one with the least error propagation, and the high values of S_i^E in the enriching section question the employment of a sensor in that section, especially towards the top. If a sensor in this section was to be tried, it should be placed in the middle of the section, as a compromise between robustness and closeness to the effluent compositions.

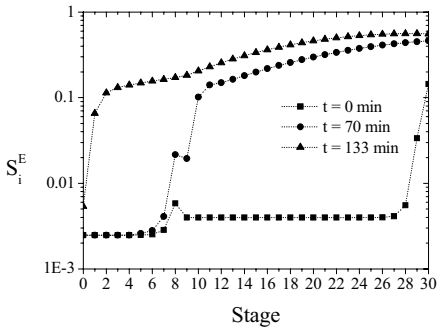


Figure 2: Singularity measure dependency on single-sensor location, when the ethanol composition is the innovated state, at three different times.

The straightforward application of conventional-like tuning guidelines (Alvarez and Lopez, 1999; Alvarez *et al*., 2004; Gonzalez and Alvarez, 2005) for second order linear filters yields the estimator parameters: ζ = 1.5 (to avoid oscillatory error response), ω = 0.03 min^{-1}. The resulting single-sensor (column bottom) estimator behavior is presented in Figure 3, showing that, as predicted by the sensitivity plot of Figure 2, the column bottom sensor exhibits a good data assimilation capability in the light of the uncertainty due to the composition off-line determinations. When the sensor is located in the middle of the enriching section, the bottom compositions estimates worsen, and there is not significant improvement in the distillate estimates. This corroborates the sensitivity measure-based prediction: a measurement in the enriching section hardly provides useful information.

Thus, according to Figure 3, the top effluent composition estimate task is basically being executed by the information content injected in the column bottom in conjunction with the column model.

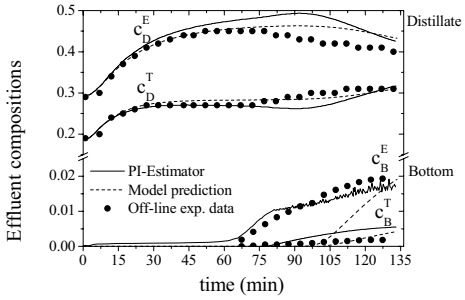


Figure 3. Single-sensor (column bottom) passive estimation (ethanol composition as innovated state).

3.2 Terbutanol as innovated state

When the terbutanol is the innovated state and the sensor in located at the (variable) i-th stage, the estimator (2) becomes:

$$\dot{\hat{c}}_i^T = f_i^T(\hat{c}_i^E, \hat{c}_i^T, \hat{c}_{i-1}^E, \hat{c}_{i-1}^T, \hat{c}_{i+1}^E, \hat{c}_{i+1}^T) + \left[1/\beta_{c_T}(\hat{c}_i^E, \hat{c}_i^T)\right]\{\dot{w} + 2\zeta\omega[y_j - \beta(\hat{c}_i^E, \hat{c}_i^T)]\} \quad (4a)$$

$$\dot{w} = \omega^2[y_i - \beta(\hat{c}_i^E, \hat{c}_i^T)] \quad (4b)$$

$$\dot{\hat{c}}_i^E = f_i^E(\hat{c}_i^E, \hat{c}_i^T, \hat{c}_{i-1}^E, \hat{c}_{i-1}^T, \hat{c}_{i+1}^E, \hat{c}_{i+1}^T), \quad k = E, T \quad (4c)$$

$$\dot{\hat{c}}_j^k = f_i^k(\hat{c}_j^E, \hat{c}_j^T, \hat{c}_{j-1}^E, \hat{c}_{j-1}^T, \hat{c}_{j+1}^E, \hat{c}_{j+1}^T), \; j \neq i, \; j \in [1,N] \quad (4d)$$

where

$\zeta = 1.5, \quad \omega = 0.03 \text{ min}^{-1}$

$\beta_{c_T}(c_E, c_T) = \partial_{c_T}\beta(c_E, c_T), \quad S_i^T = 1/|\beta_{c_T}(c_i^E, c_i^T)|$ (5a,b)

The corresponding asymptotic error propagation measure is presented in Figure 4.

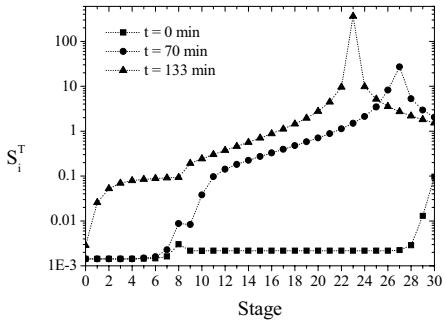


Figure 4. Singularity measure dependency on single-sensor location, when the terbutanol composition is the innovated state, at three different times.

According to Figure 4, the terbutanol should definitely not be chosen as the innovated state when the measurement is located about tray 23, and the same structure should not chosen for the stage interval 20-30. Comparing with Figure 2, when the measurement is located in the stage interval 0-16 either the ethanol or the terbutanol can be chosen as innovated state, and in both cases the location of sensors in the interval 20-30 should be avoided, especially for the case of terbutanol as innovated state. Physically speaking this means that: (i) in the stage interval 0-12 there is a sufficiently large temperature decrease for estimation purposes, with ethanol or terbutanol as innovated state, and (ii) in the stage interval 20-30 the estimation task via masurement injection is more difficult, could be pursued with the ethanol as innovated state but not with the terbutanol as innovated state, because the presence of ethanol (or terbutanol) is mildly (or imperceptible) reflected in the temperature measurement.

The resulting single-sensor (column bottom) estimator behavior is presented in Figure 5, with results that are similar to the ones of the case (Figure 3) with the ethanol as innovated state. Again, when the sensor is located in the middle of the enriching section, the bottom compositions estimates worsen, and there is not significant improvement in the distillate estimates.

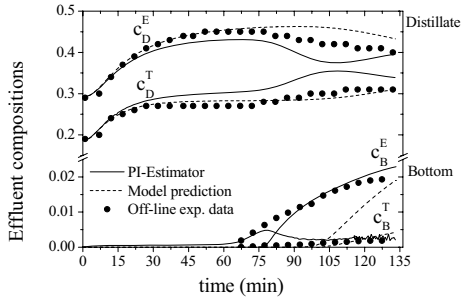


Figure 5. Single-sensor (column bottom) passive estimation (terbutanol composition as innovated state).

3.3 Concluding remarks

Being passivity originally an input-output control concept (Krstic *et al*., 1995), a comment on its interpretation in the estimation case is to the point. In out ternary distillation case, a passive estimation structure (2) [or (4)] signifies: (i) a single-state innovated dynamics, (i) a measured output-estimated input pair (y, w) with relative degree equal to one, and (iii) a stable (restricted) noninnovated dynamics (6) [or (7)]:

Ethanol as innovated state

$$\hat{c}_i^E = \gamma_E(\hat{c}_i^T, y_i) \quad (6a)$$

$$\dot{\hat{c}}_i^T = f_i^T[\gamma(\hat{c}_i^T, y_i), \hat{c}_i^T, \hat{c}_{i-1}^E, \hat{c}_{i-1}^T, \hat{c}_{i+1}^E, \hat{c}_{i+1}^T], \; k = E, T \quad (6b)$$

$$\dot{\hat{c}}_j^k = f_i^k(\hat{c}_j^E, \hat{c}_j^T, \hat{c}_{j-1}^E, \hat{c}_{j-1}^T, \hat{c}_{j+1}^E, \hat{c}_{j+1}^T), \; j \neq i, \; j \in [1,N] \quad (6c)$$

Terbutanol as innovated state

$$\hat{c}_i^T = \gamma_T(\hat{c}_i^E, y_i) \quad (7a)$$

$$\dot{\hat{c}}_i^E = f_i^E[\gamma_E(\hat{c}_i^E, y_i), \hat{c}_i^T, \hat{c}_{i-1}^E, \hat{c}_{i-1}^T, \hat{c}_{i+1}^E, \hat{c}_{i+1}^T], \; k = E, T \quad (7b)$$

$$\dot{\hat{c}}_j^k = f_i^k(\hat{c}_j^E, \hat{c}_j^T, \hat{c}_{j-1}^E, \hat{c}_{j-1}^T, \hat{c}_{j+1}^E, \hat{c}_{j+1}^T), \; j \neq i, \; j \in [1,N] \quad (7c)$$

where γ_E (or γ_T) is the solution for c_E (or c_T) of the bubble point measurement equation $c_E = \gamma_E(c_T, y)$ [or $c_T = \gamma_T(c_E, y)$].

4. PASSIVATED ESTIMATOR

According to the constructive-like adjustable-estimation geometric estimation approach (Alvarez, 2000, Lopez and Alvarez 2004), the design of the estimation structure amounts to a suitable compromise between reconstruction rate and robustness, depending on the estimation objectives, the model conditioning of the particular system, and the measurement uncertainty. Low estimation degrees favour robustness and disfavour the reconstruction rate. In a general-purpose estimation structure search procedure [Lopez and Alvarez, 2004]: (i) the passive structure, with maximum robustness, must be seen as the point of departure candidate structure and configurations with more innovated states must be considered to draw the best compromise between robustness and performance, and (ii) the (nominal) detectability structure, with maximum estimation orders equal to the observability indices, constitutes the limit on performance in the absence of modelling errors. The particular (staged, three component, presence of azeotrops, and high separation) features of our ternary distillation column example suggest that the estimation structure should be more on the passive side, and the verification of this conjecture constitutes the scope of the present section.

4.1 Nonpassive structure

Let us recall that the column bottom stage offers the best means of effective data assimilation, the fact that ethanol and terbutanol perform equally well as single-innovated states, consider both the column bottom ethanol and terbutanol concentrations as innovated states with one sensor in the same stage, and write the corresponding PI estimator [Alvarez and Lopez, 1999]:

$$\dot{\hat{c}}_i = f_i(\hat{c}_i,\hat{c}_{i-1},\hat{c}_{i+1}, u) + O^{-1}(\hat{c}_i, \hat{c}_{i-1}, \hat{c}_{i+1}, u)[\pi w + k_p[y_j - \beta(\hat{c}_i)] \quad (8a)$$

$$\dot{w} = k_w[y_i - \beta(\hat{c}_i)], \quad \hat{c}_i = (\hat{c}^E_i, \hat{c}^T_i)' \quad (8b)$$

$$\dot{\hat{c}}^k_j = f^k_i(\hat{c}^E_j, \hat{c}^T_j, \hat{c}^E_{j-1}, \hat{c}^T_{j-1}, \hat{c}^E_{j+1}, \hat{c}^T_{j+1}, u),\ j \neq i,\ j \in [1,N] \quad (8c)$$

$$S^{E\text{-}T}_i = [1/\text{msv}O(c_i, c_{i-1}, c_{i+1}, u)] \quad (9)$$

where

$O(c_i, c_{i-1}, c_{i+1}, u) = \partial_{c_i}\phi(c_i, c_{i-1}, c_{i+1}, u), \quad \pi = (0,1)'$

$\phi(c_i, c_{i-1}, c_{i+1}, u) = \{\beta(c_i), [\partial_{c_i}\beta(c_i)]f_i(c_i,c_{i-1},c_{i+1}, u)\}'$

$k_p = (2\zeta + 1)(\omega, \omega^2)', \ k_w = \omega^3, \ u = (F, R, V, c_{iF})'$

O is the 2x2 observability matrix, $S^{E\text{-}T}_i$ is the error propagation measure for a sensor located in the i-th stage, and msv means "the minimum singular value", and the corresponding plot is presented in Figure 6 (for three times).

Comparing with the same plots (Figures 2 and 4) of the passive structure cases, the two-innovated state error propagation measure is considerably larger, and this is due to: (i) the combination of interactions in O, (ii) the presence of first and second order partial derivatives of the equilibrium (ν_E and ν_T) and bubble point (β) nonlinear functions in O, (iii) stages with close-to-azeotropic compositions, and dependency of O on neighbour tray compositions. This results lead us to disregard nonpassive estimation structures.

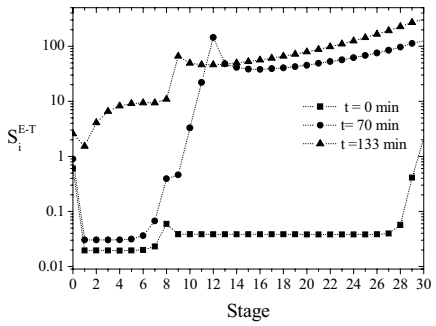


Figure. 6. Singularity measure dependency on (single) sensor location, when ethanol and tertbutanol are innovated state, at three different times.

4.2 Passivated structure

In a way that is analogous to the recursive robust control design via passivation (Krstic *et al*., 1995; Alvarez *et al.,* 2004), and motivated by the decentralized control design for distillation columns (Castellanos-Sahagun *et al*, 2005) as well as by the similar behavior of the passive structures with ethanol or terbutanol as innovated state, let us consider the parallel combination of the two passive estimators, (2) and (5), presented is Subsection 3:

$$\dot{\hat{c}}^E_i = f^E_i(\hat{c}^E_i, \hat{c}^T_i, \hat{c}^E_{i-1}, \hat{c}^T_{i-1}, \hat{c}^E_{i+1}, \hat{c}^T_{i+1}) + [1/\beta_{c_E}(\hat{c}^E_i, \hat{c}^T_i)]\{w_E + 2\zeta_E\omega_E[y_j - \beta(\hat{c}^E_i, \hat{c}^T_i)]\} \quad (10a)$$

$$\dot{w}_E = \omega^2_E[y_i - \beta(\hat{c}^E_i, \hat{c}^T_i)], \quad \dot{w}_T = \omega^2_T[y_i - \beta(\hat{c}^E_i, \hat{c}^T_i)] \quad (10b)$$

$$\dot{\hat{c}}^T_i = f^T_i(\hat{c}^E_i, \hat{c}^T_i, \hat{c}^E_{i-1}, \hat{c}^T_{i-1}, \hat{c}^E_{i+1}, \hat{c}^T_{i+1}) + [1/\beta_{c_T}(\hat{c}^E_i, \hat{c}^T_i)]\{w_E + 2\zeta_T\omega_T[y_j - \beta(\hat{c}^E_i, \hat{c}^T_i)]\} \quad (10c)$$

$$\dot{\hat{c}}^k_j = f^k_i(\hat{c}^E_j, \hat{c}^T_j, \hat{c}^E_{j-1}, \hat{c}^T_{j-1}, \hat{c}^E_{j+1}, \hat{c}^T_{j+1}),\ j \neq i,\ j \in [1,N] \quad (10d)$$

$\omega_E = \omega_T = 0.03\ \text{min}^{-1}$, $\zeta_E = \zeta_T = 1.5$

Note that this estimator has a decentralized error propagation structure, with two passive error propagation mechanisms, one for each innovated state, that have been already displayed (in figures 3 and 5). Consequently, the (same) *sensor location assessment* of the passive cases is inherited by the preceding passivated estimator: (i) the column bottom stage is the best sensor location for bottom (and to a good extent also for distillate) composition estimation purposes, (ii) a sensor in the enriching section (say about tray 22) may be added, in the understanding that such addition may not bring in sufficiently meaningful information. The corresponding single-sensor (column bottom) behavior is presented in Figure 6 (discontinuous plots). Comparing with the two passive structures (Figures 3 and 5), the passivated structure yields a better behavior, or equivalently is a more efficient means to execute the data assimilation task.

4.4 Comparison with EKF

Here the proposed single-sensor passivated estimator is compared with an EKF with two sensors (in the reboiler and stage 27), as it is commonly done in distillation column studies. Since the EKF covariance matrix pair gain tuning is rather complex for the ternary case, the tuning procedure presented in an earlier study was followed (Baratti, et. al, 1995; Baratti, et. al, 1998): (i) a suitable structure of the model error covariance matrix was assumed, on the basis of the column sections and the number of components, (ii) and six matrix parameters were tuned using an optimization scheme. The resulting behavior is presented in Figure 7 (continuous plots), showing that basically the single-sensor passivated estimator (10) and the EKF yield the same behavior.

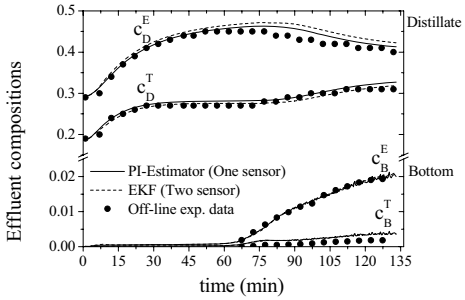


Figure 7. Comparison of the Single-sensor passivated PI-Estimador against the two-sensor EKF.

The advantage of the passivated estimator over the EKF resides in the fact that the construction, implementation and tuning tasks of the passivated are considerably simpler: (i) while the single-sensor passivated estimator has 65 nonlinear ODEs (64 for the model and one for the integral state), the EKF has 2144 nonlinear ODEs (64 for the model and 2080 Riccati equations), (ii) while adequate functioning of the EKF requires a nontrivial tuning via optimization, with parameters devoided of physical meaning, the passive estimator tuning can be performed according to conventional-like guidelines for linear second order filters, with (damping and frequency) parameters that have a clear connection with the column dynamics and the output prediction error response.

5. CONCLUSIONS

The problem of jointly designing the estimation structure and algorithm estimation to infer effluent compositions in ternary distillation columns with temperature measurements has been addressed and the results illustrated with a representative 32-stage experimental column. The constructive estimation approach associated with the adjustable-structure geometric estimation design methodology led to a single sensor (located in the column bottom) two-innovated state passivated estimator with 65 ODE's and a straightforward tuning scheme. The proposed estimator yielded the same behavior than the one obtained by a two-sensor EKF with 2144 nonlinear ODE's, and six tuning parameters chosen with an off-line optimization approach developed before.

Acknowledgment. The experimental data were obtained from University of Padova pilot plant, and the authors are grateful to Prof. Alberto Bertucco for the experimental data.

6. REFERENCES

Alvarez, J., and Lopez, T. (1999). Robust dynamic state estimation of nonlinear plants. *AichE J.,* **45,** 107-123

Alvarez, J. (2000). Nonlinear State Estimation with Robust convergence. *J. Process Control*, 10, 59-71

Alvarez, J.; Zaldo, F.; Oaxaca G. (2004). *Towards a Joint Process and Control Design Framework for Batch Processes: Application to Semibatch Polymer Reactors*. In: The Integration of Process Design and Control; Seferlis, P., Georgiadis, M. C., Eds.; Elsevier: Amsterdam, The Netherlands.

Alvarez, J, Castellanos-Sahagun, E., Fernandez, C. and Aguirre, S. (2005). Optimal Closed-loop Operation of Binary Batch Distillation Columns. Preprints of the 16th IFAC World Congress, Prague, Czech Republic.

Baratti, R., A. Bertucco, A. da Rold, and M. Morbidelli (1995). Development of a composition Estimator for binary distillation columns. Applications to a Pilot Plant. *Chem. Eng. Sci*., 50, 1541-1550

Baratti, R., A. Bertucco, A. da Rold, and M. Morbidelli (1998). A Composition estimator for a multicomponent columns development and experimental test on ternary mixtures. *Chem. Eng. Sci.*, 53, 3601-3612

Castellanos-Sahagun, E., Alvarez, J. and Alvarez-Ramirez, J. (2005). Two-point temperature control structure and algorithm design for binary distillation columns. *Ind. Eng. Chem. Res.* 44, 142-152.

González, P., Alvarez, J. (2005). Combined proportional/integral-inventory control of solution homopolymerization reactors. *Ind. Eng. Chem. Res.* 44, 7147-7163

Krstic, M., Kanellakopoulos, I., Kokotovic, P.V. (1995) *Nonlinear and Adaptive Control Design*. Wiley, New York.

Lopez, T. and Alvarez, J. (2004). On the effect of the estimation structure in the functioning of a nonlinear copolymer reactor estimator. *J. Proc. Control* 14, 99-109

Oisiovici, R.M., and Cruz, S.L. (2000). State estimation of batch distillation columns using an extended Kalman filter. *Chem. Eng. Sci*. 55 , 4667-4680

Quintero-Marmol, E., Luyben, W. L. and Georgakis, C. (1991). Application of an Extended Luenberger Observer to the Control of Multicomponent Batch Distillation. *Ind. Eng. Chem. Res*. 30(8), 1870-1880.

Skogestad, S. (1997). Dinamics and control of distillation columns – A critical survey. *Model Identif. Control*. 18, 177-217.

Tronci, S., Baratti, R., Barolo, M., Bezzo, F. (2005). Geometric Observer for a binary distillation column. Accepted to *IECR*

Venkateswarlu,C., Avantika,S. (2001). Optimal state estimation of multicomponent batch distillation. *Chem. Eng. Sci*. 56, 5771-5786.

Yu, C. C., Luyben, W. L. (1987). Control of multicomponent distillation columns using rigorous composition estimators. Distillation and Absorption. Institute of Chemical Engineers Symposium series no. 104, Institute of Chemical Engineers, London.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

FINITE TIME OBSERVER FOR NONLINEAR SYSTEMS

F. Sauvage [*,1] **M. Guay** [**] **D. Dochain** [*]

[*] *IMAP, Université Catholique de Louvain, Louvain-la-Neuve, Belgium*
[**] *Dept. Chem. Eng., Queen's University, Kingston K7L 3N6, Canada*

Abstract: This paper proposes a nonlinear finite time convergent observer that does not require to compute any inverse coordinate transformation. The finite time estimate is recovered from two asymptotically convergent estimates which have linear error dynamics in transformed coordinates through the transformation Jacobian only. An extended version of this nonlinear finite time observer is next envisaged. The finite time estimate is obtained from two pseudo linear dynamic systems and require to compute only the Jacobian of two functions which are the solutions of two systems of partial derivative equations.

Keywords: State observers, time delay, nonlinear observers, finite time convergence

1. INTRODUCTION

Monitoring of the component concentrations is a key question for productivity and safety in the chemical industry. However it often requires very specific and expensive sensors that cannot be used in practice. Therefore the real-time estimation of component concentrations using a state observer is a very attractive option.

Since the first observers for linear systems have been developed by Kalman and Luenberger several decades ago, several different techniques have been proposed to deal with nonlinearities and model uncertainties. However these techniques give an estimate that reaches the real state asymptotically what may be a limitation for batch and fed-batch processes.

An observer that converges in finite time has been recently proposed (Engel and Kreisselmeier, 2002). The key idea is to use the present and delayed estimates provided by two independent classical observers to compute an estimate that converges exactly to the state after a predefinite time delay. The estimate formulation arises from solving a set of four equations linking the state and each of the two classical estimates at both present time and delayed time. The finite time observer performance relies on the linearity of the estimation error dynamics and its use is therefore restricted to linear time-invariant systems.

The field of application of this technique has been extended to linear time-varying systems (Menold *et al.*, 2003*b*). The use of the transition matrix of the system is introduced to compare the delayed and present estimates. The same authors have also extended the technique to nonlinear systems that can be transformed into the observer canonical form (Menold *et al.*, 2003*a*). Once the nonlinear system is transformed into its normal form, two observers with linear error dynamics can be developed and the finite time estimation can be carried out in these coordinates. The estimate in the original coordinates is then retrieved by the inverse transformation.

In this paper, we propose a finite time observer for nonlinear systems that does not require to

[1] Corresponding author. E-mail: sauvage@imap.ucl.ac.be

compute the inverse coordinates transformation but only its Jacobian. The estimate is computed in transformed coordinates associated to a pseudo linear form of the system (Menold *et al.*, 2003*a*) and its expression comes from the set of equations used to estimate linear time invariant systems (Engel and Kreisselmeier, 2002). The estimate in original coordinates is obtained by differentiating the previous expression introducing the Jacobian of the coordinates transformation. A more general approach using two different changes of variables obtained from two systems of partial derivative equations is then presented. Also in this case, the estimation requires to compute the inverse Jacobian of each of the transformations only.

The paper is structured as follows. In Section 2 we present a finite time observer for nonlinear systems that can be transformed into pseudo linear system with nonlinearities depending on the input and output only. This section ends by an example of a numerical simulation. In Section 3 we propose a finite time observer which require to compute the Jacobian of two functions which are the solutions of two systems of partial derivative equations. This section is ended by an example of a numerical simulation.

2. FINITE TIME OBSERVER

Consider the following observable nonlinear system :

$$\begin{aligned} &\dot{x} = f(x,u),\ x(t_0) = x_0,\ t \geq t_0 \\ &y = h(x) \end{aligned} \tag{1}$$

with state $x \in \mathbb{R}^n$, input $u \in \mathbb{R}^m$ and output $y \in \mathbb{R}$. Assume that there exists a change of coordinates

$$z = \Psi(x) \tag{2}$$

allowing to transform the system (1) into the following observable pseudo linear system (Hou and Pugh, 1999),(Krener and Respondek, 1985) :

$$\begin{aligned} &\dot{z} = Az + \beta(y,u),\ z(t_0) = z_0,\ t \geq t_0 \\ &y = Cz \end{aligned} \tag{3}$$

where β is a known nonlinear function that only depends on the input and output. The observability involves that two gain matrices H_1 and H_2 can be computed so that both following matrices have desired eigenvalues with negative real parts :

$$F_1 = A - H_1 C \tag{4}$$

$$F_2 = A - H_2 C \tag{5}$$

implying that both following systems are observers for system (3):

$$\dot{\hat{z}}_1 = A\hat{z}_1 + \beta(y,u) + H_1\left(y - C\hat{z}_1\right) \tag{6}$$

$$\dot{\hat{z}}_2 = A\hat{z}_2 + \beta(y,u) + H_2\left(y - C\hat{z}_2\right) \tag{7}$$

Each of the reconstruction errors associated with the above observers is governed by a linear dynamics as follows:

$$\dot{\epsilon}_1 = F_1\epsilon_1 \tag{8}$$

$$\dot{\epsilon}_2 = F_2\epsilon_2 \tag{9}$$

This involves that for any time $t \geq D$, the following relations exist between the errors at different time instances:

$$\epsilon_1(t) = e^{F_1 D}\epsilon_1(t-D) \tag{10}$$

$$\epsilon_2(t) = e^{F_2 D}\epsilon_2(t-D) \tag{11}$$

This leads to the following set of four equations :

$$\hat{z}_1(t) = z(t) + \epsilon_1(t) \tag{12}$$

$$\hat{z}_2(t) = z(t) + \epsilon_2(t) \tag{13}$$

$$\hat{z}_1(t-D) = z(t-D) + e^{-F_1 D}\epsilon_1(t) \tag{14}$$

$$\hat{z}_2(t-D) = z(t-D) + e^{-F_2 D}\epsilon_2(t) \tag{15}$$

which, when it is solved for $z(t)$, gives rise to the following observer that converges within the predefinite time D (Engel and Kreisselmeier, 2002), (Menold *et al.*, 2003*a*):

$$\begin{aligned} \hat{z}(t) = \left(e^{-F_1 D} - e^{-F_2 D}\right)^{-1} \big(&e^{-F_1 D}\hat{z}_1(t) \\ &-\hat{z}_1(t-D) - e^{-F_2 D}\hat{z}_2(t) + \hat{z}_2(t-D)\big) \end{aligned}$$

The observer proposed in this paper comes from the same set of equations. However, as the ultimate goal is to estimate the state in the original coordinates, the set of equations will be transformed to depend on x explicitly. This is achieved by differentiating each of the equations, using Equation (2), it becomes:

$$\dot{\hat{z}}_1(t) = \frac{\partial \Psi}{\partial x}\dot{x}(t) + \dot{\epsilon}_1(t) \tag{16}$$

$$\dot{\hat{z}}_2(t) = \frac{\partial \Psi}{\partial x}\dot{x}(t) + \dot{\epsilon}_2(t) \tag{17}$$

$$\dot{\hat{z}}_1(t-D) = \frac{\partial \Psi}{\partial x}\dot{x}(t-D) + e^{-F_1 D}\dot{\epsilon}_1(t) \tag{18}$$

$$\dot{\hat{z}}_2(t-D) = \frac{\partial \Psi}{\partial x}\dot{x}(t-D) + e^{-F_2 D}\dot{\epsilon}_2(t) \tag{19}$$

The expression for the observer proposed in the following theorem arises from solving the above set of equations for $\dot{x}(t)$.

Theorem 1. Assume that the change of coordinates $z = \Psi(x)$ that transforms the nonlinear system (1) into the pseudo linear system (3) exists and that its Jacobian is invertible. Furthermore

assume that the systems (6) and (7) are observers for (3) designed so that for any positive constant D, the following term is invertible :

$$e^{-F_1 D} - e^{-F_2 D} \tag{20}$$

Then the following dynamical system:

$$\begin{aligned}\dot{\hat{\xi}} = &\left(\frac{\partial \Psi}{\partial x}\right)^{-1}_{x=\hat{x}} \left(e^{-F_1 D} - e^{-F_2 D}\right)^{-1} \\ &\Big(e^{-F_1 D}\dot{\hat{z}}_1(t) - \dot{\hat{z}}_1(t-D) - \\ &e^{-F_2 D}\dot{\hat{z}}_2(t) + \dot{\hat{z}}_2(t-D)\Big)\end{aligned} \tag{21}$$

with

$$\Psi(\hat{x}(t_0)) = \hat{z}_1(t_0) = \hat{z}_2(t_0)$$

is a finite time observer for the system (1) in the sense that the estimation $\hat{x}$ converges exactly to the state x after the time delay D.

Proof

Let us choose an arbitrary positive constant D. As the systems (6) and (7) are observers that converge to z with linear error dynamics; furthermore, as they have the same intial conditions; the following system is a finite time observer for (3):

$$\begin{aligned}\hat{z} = &\left(e^{-F_1 D} - e^{-F_2 D}\right)^{-1} \left(e^{-F_1 D}\hat{z}_1(t)\right. \\ &\left.-\hat{z}_1(t-D) - e^{-F_2 D}\hat{z}_2(t) + \hat{z}_2(t-D)\right)\end{aligned} \tag{22}$$

and for any time $t \geq t_0 + D$, we have :

$$\hat{z}(t) = z(t) = \Psi(x(t)) \tag{23}$$

Therefore, integration of the following dynamical system:

$$\begin{aligned}\dot{\hat{z}} = &\left(e^{-F_1 D} - e^{-F_2 D}\right)^{-1} \Big(e^{-F_1 D}\dot{\hat{z}}_1(t)\Big. \\ &\Big.-\dot{\hat{z}}_1(t-D) - e^{-F_2 D}\dot{\hat{z}}_2(t) - \dot{\hat{z}}_2(t-D)\Big)\end{aligned} \tag{24}$$

with the initial conditions:

$$\hat{z}(t_0) = \hat{z}_1(t_0) = \hat{z}_2(t_0) \tag{25}$$

leads to a finite time observer for $z(t)$. Let us define the estimate $\hat{x}$ so that

$$\hat{z} = \Psi(\hat{x}) \tag{26}$$

Then, the following dynamical system

$$\dot{\hat{x}}(t) = \left(\frac{\partial \Psi}{\partial x}\right)^{-1}_{x=\hat{x}} \dot{\hat{z}}(t) \tag{27}$$

with the following initial conditions:

$$\Psi(\hat{x}(t_0)) = \hat{z}(t_0) \tag{28}$$

is an observer for system (1) that converges exactly to the state within the predefined time delay D. □

This one-step approach is different from the two-step one adopted in (Menold *et al.*, 2003*a*). Their technique consists in transforming the system into a linear one by an appropriate transformation and to make a finite time estimation in these coordinates. Then the estimate in the original coordinates is retrieved through the inverse transformation. The observer proposed in this paper provides an estimate in one step only. This is achieved using the transformation Jacobian and therefore it does not require to compute the inverse transformation. This approach is quite similar to that adopted by Kazantzis and Kravaris for their nonlinear observer (Kazantzis and Kravaris, 1998).

Example

Consider the following system

$$\dot{x}_1 = -x_1^2 - x_2 \tag{29}$$

$$\dot{x}_2 = 2x_1 x_2 - x_1^2 \tag{30}$$

$$y = x_1 \tag{31}$$

The system can be transformed into a pseudo linear one with the following coordinates change:

$$z_1 = x_1 \tag{32}$$

$$z_2 = -x_1^2 - x_2 \tag{33}$$

It can then be written as system (3) where

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \beta = \begin{pmatrix} 0 \\ 2y^3 + y^2 \end{pmatrix}$$

Two arbitrary independent high gain observers can be synthetised by taking

$$H_i = \begin{pmatrix} \alpha_1^i/\omega_i \\ \alpha_2^i/\omega_i^2 \end{pmatrix} \quad i = 1, 2$$

with ω_i positive and so that both following polynomials are Hurwitz:

$$s^2 + \alpha_1^i s + \alpha_2^i \quad i = 1, 2$$

The performance of the finite time observer are illustrated by a numerical simulation on Figure 1. The different parameters values used for the simulation are listed in Table 1. It can be seen that the estimate reaches the state exactly after the pre-defined delay D as expected.

The finite time convergence of the estimate provided by the above observer is guaranteed by the linearity of the reconstruction errors dynamics. However in practice, it is not always possible to find a suitable change of variable allowing to build

Table 1. Parameters for the numerical simulation

Variable	Value	Variable	Value
$x_1(0)$	0.5	$x_2(0)$	1
$\hat{x}_1(0)$	0	$\hat{x}_2(0)$	0
Parameter	Value	Parameter	Value
α_1^1	3	α_1^2	8
α_2^1	2	α_2^2	3
ω_1	0.1	ω_2	0.1

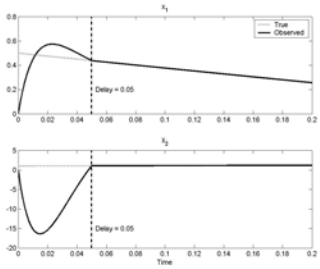


Fig. 1. Simulation results for exemple 1

observers with linear error dynamics. In particular, if the β function of Equation (3) depends on the state, the error dynamics are not linear. In this case, the use of high gain observers allows to converge within a finite time into a neighborhood of the state (Menold, 2004).

3. GENERALIZED FINITE TIME OBSERVER

In the following, we present a general approach of the finite time observation for nonlinear systems.

Assume that the matrices A_1, A_2 are Hurwitz and that the functions β_1, β_2 are such that $[A_i,\ \beta_i]$ form controllable pairs. Furthermore, let Ψ_1 and Ψ_2 be the solutions of the following systems of partial derivative equations :

$$\frac{\partial \Psi_1}{\partial x} f(x) = A_1 \Psi_1 + \beta_1(y) \tag{34}$$

$$\frac{\partial \Psi_2}{\partial x} f(x) = A_2 \Psi_2 + \beta_2(y) \tag{35}$$

Then the following systems are observers for $\Psi_1(x)$ ans $\Psi_2(x)$ respectively:

$$\dot{\hat{z}}_1 = A_1 \hat{z}_1 + \beta_1(y) \tag{36}$$

$$\dot{\hat{z}}_2 = A_2 \hat{z}_2 + \beta_2(y) \tag{37}$$

and we have the following linear dynamics :

$$\frac{d}{dt}\left(\hat{z}_1 - \Psi_1(x)\right) = A_1\left(\hat{z}_1 - \Psi_1(x)\right) \tag{38}$$

$$\frac{d}{dt}\left(\hat{z}_2 - \Psi_2(x)\right) = A_2\left(\hat{z}_2 - \Psi_2(x)\right) \tag{39}$$

Using the following notations :

$$z_1 = \Psi_1(x) \tag{40}$$

$$z_2 = \Psi_2(x) \tag{41}$$

and defining the reconstruction errors as:

$$\epsilon_1 = \hat{z}_1 - z_1 \tag{42}$$

$$\epsilon_2 = \hat{z}_2 - z_2 \tag{43}$$

The error dynamics linearity allows to write the relations between the reconstruction errors at different times instances as follows:

$$\epsilon_1(t) = e^{A_1 D}\epsilon_1(t-D) \tag{44}$$

$$\epsilon_2(t) = e^{A_2 D}\epsilon_2(t-D) \tag{45}$$

This allows to write the same set of four equations as in Section 2, which becomes, after differentiation :

$$\dot{\hat{z}}_1(t) = \nabla\Psi_1 \dot{x}(t) + \dot{\epsilon}_1(t) \tag{46}$$

$$\dot{\hat{z}}_2(t) = \nabla\Psi_2 \dot{x}(t) + \dot{\epsilon}_2(t) \tag{47}$$

$$\dot{\hat{z}}_1(t-D) = \nabla^D\Psi_1 \dot{x}(t-D) + e^{-A_1 D}\dot{\epsilon}_1(t) \tag{48}$$

$$\dot{\hat{z}}_2(t-D) = \nabla^D\Psi_2 \dot{x}(t-D) + e^{-A_2 D}\dot{\epsilon}_2(t) \tag{49}$$

where the following notations are used ($i = 1, 2$):

$$\nabla\Psi_i = \left(\frac{\partial \Psi_i}{\partial x}\right)_{x=\hat{x}(t)} \tag{50}$$

$$\nabla^D\Psi_i = \left(\frac{\partial \Psi_i}{\partial x}\right)_{x=\hat{x}(t-D)} \tag{51}$$

The formulation of the observer proposed in the following theorem arises from solving this set of equations for $\dot{x}(t)$.

Theorem 2. Assume that the matrices A_1, A_2 are Hurwitz and that the functions β_1, β_2 are chosen so that the following systems are controllable:

$$\dot{\hat{z}}_1 = A_1 \hat{z}_1 + \beta_1(y) \tag{52}$$

$$\dot{\hat{z}}_2 = A_2 \hat{z}_2 + \beta_2(y) \tag{53}$$

and let Ψ_1, Ψ_2 be the solutions of the PDE systems (34) and (35). Furthermore, assume that the matrices A_i are chosen so that, for any positive constant D, the following term is invertible:

$$e^{-A_1 D}\nabla\Psi_1 - \nabla^D\Psi_1\left(\nabla^D\Psi_2\right)^{-1} e^{-A_2 D}\nabla\Psi_2 \tag{54}$$

Then if the initial conditions are set as follows:

$$\hat{z}_1(t_0) = \Psi_1(\hat{x}(t_0)) \tag{55}$$

$$\hat{z}_2(t_0) = \Psi_2(\hat{x}(t_0)) \tag{56}$$

and satisfy the following condition:

$$\hat{z}_1(t_0) = \nabla^0\Psi_1\left(\nabla^0\Psi_2\right)^{-1}\hat{z}_2(t_0) \tag{57}$$

the following dynamical system:

$$\dot{\hat{x}}(t) = \Omega^{-1}\left(\Delta(\dot{\hat{z}}_1) - \underline{\nabla}\Delta(\dot{\hat{z}}_2)\right) \quad (58)$$

where

$$\Delta(z) = e^{-A_1 D} z(t) - z(t-D) \quad (59)$$

and

$$\underline{\nabla} = \nabla^D \Psi_1 \left(\nabla^D \Psi_2\right)^{-1} \quad (60)$$

$$\Omega = e^{-A_1 D}\nabla\Psi_1 - \underline{\nabla} e^{-A_2 D}\nabla\Psi_2 \quad (61)$$

is a finite time observer for (1) in the sense that the estimate reaches the states after the predefinite time delay D.

Proof

Let us introduce the reconstruction errors defined by Equations (42) (43). Equation (58) becomes:

$$\dot{\hat{x}}(t) = \Omega^{-1}\left(\Delta(\dot{z}_1) - \underline{\nabla}\Delta(\dot{z}_2) - \Delta(\dot{\epsilon}_1) + \underline{\nabla}\Delta(\dot{\epsilon}_2)\right) \quad (62)$$

By definition of z_1 and z_2 (Equations (40) and (41)), it can be seen that :

$$\Delta(\dot{z}_1) - \underline{\nabla}\Delta(\dot{z}_2) = \Omega\dot{x} \quad (63)$$

Therefore, Equation (62) can be rewritten as follows :

$$\dot{\hat{x}}(t) = \dot{x}(t) - \Omega^{-1}\left(\Delta(\dot{\epsilon}_1) - \underline{\nabla}\Delta(\dot{\epsilon}_2)\right) \quad (64)$$

As for any time greater than $t_0 + D$, Equations (44) (45) hold, both $\Delta(\dot{\epsilon}_1)$, $\Delta(\dot{\epsilon}_2)$ vanish and we have

$$\forall t \geq t_0 + D \ : \ \dot{\hat{x}}(t) = \dot{x}(t) \quad (65)$$

This implies that the estimate dynamics and the state dynamics are the same after the convergence time interval. It remains to show that the estimate reaches the value of the state variables at time $t = t_0 + D$

Let us focus on the time interval $[t_0 \ \ t_0 + D]$. During this time interval, the delayed values for the different variables remain constant and equal to their initial values :

$$\hat{z}_1(t-D) = \hat{z}_1(t_0) \quad (66)$$

$$\hat{z}_2(t-D) = \hat{z}_2(t_0) \quad (67)$$

$$\underline{\nabla} = \underline{\nabla}^0 \quad (68)$$

The observer expression becomes:

$$\dot{\hat{x}}(t) = \Omega^{-1}\left(e^{-A_1 D}\dot{\hat{z}}_1 - \underline{\nabla}^0 e^{-A_1 D}\dot{\hat{z}}_2\right) \quad (69)$$

A simple expression for the integral of the above expression is difficult to compute since in particular Ω is not constant. However, it can be written as follows:

$$\Omega\dot{\hat{x}}(t) = e^{-A_1 D}\dot{\hat{z}}_1 - \underline{\nabla}^0 e^{-A_1 D}\dot{\hat{z}}_2 \quad (70)$$

and using the definition of Ω leads to the following equation:

$$e^{-A_1 D}\dot{\Psi}_1(\hat{x}(t)) - \underline{\nabla}^0 e^{-A_2 D}\dot{\Psi}_2(\hat{x}(t)) = \quad (71)$$

$$e^{-A_1 D}\dot{\hat{z}}_1 - \underline{\nabla}^0 e^{-A_2 D}\dot{\hat{z}}_2 \quad (72)$$

As the initial conditions are set as follows :

$$\Psi_1(\hat{x}(t_0)) = \hat{z}_1(t_0) \quad (73)$$

$$\Psi_2(\hat{x}(t_0)) = \hat{z}_2(t_0) \quad (74)$$

the above equation can be integrated between t_0 and t to give :

$$e^{-A_1 D}\Psi_1(\hat{x}(t)) - \underline{\nabla}^0 e^{-A_2 D}\Psi_2(\hat{x}(t)) = e^{-A_1 D}\hat{z}_1(t) - \underline{\nabla}^0 e^{-A_2 D}\hat{z}_2(t) \quad (75)$$

This equation can be rewritten as follows using the reconstruction errors expressions:

$$e^{-A_1 D}\Psi_1(\hat{x}(t)) - \underline{\nabla}^0 e^{-A_2 D}\Psi_2(\hat{x}(t)) = e^{-A_1 D}\Psi_1(x(t)) - \underline{\nabla}^0 e^{-A_2 D}\Psi_2(x(t)) - e^{-A_1 D}\epsilon_1(t) + \underline{\nabla}^0 e^{-A_2 D}\epsilon_2(t) \quad (76)$$

The evaluation of the above expression at time $t = t_0 + D$ leads to the following expression :

$$e^{-A_1 D}\Psi_1(\hat{x}(t_0+D)) - \underline{\nabla}^0 e^{-A_2 D}\Psi_2(\hat{x}(t_0+D)) = e^{-A_1 D}\Psi_1(x(t_0+D)) - \underline{\nabla}^0 e^{-A_2 D}\Psi_2(x(t_0+D)) - \epsilon_1(t_0) + \underline{\nabla}^0\epsilon_2(t_0) \quad (77)$$

The assumption on the initial conditions (Equation (57)) is such that

$$\hat{x}(t_0 + D) = x(t_0 + D) \quad (78)$$

This shows that the estimate reaches the state at time $t = t_0 + D$. This completes the proof. □

The above procedure is more general than the previous one. It does not require to tune and compute two observers with linear error dynamics. It only requires to compute two functions by solving a set of partial derivative equations. Solving this system can be done by expanding the different functions in Taylor series as proposed in (Kazantzis and Kravaris, 1998). It is worth noting that both functions Ψ_1 and Ψ_2 may be the same provided the matrices A_1, A_2 are such that $\left(e^{-A_1 D} - e^{-A_2 D}\right)$ is invertible. In this case the situation is exactly the same as the one presented in Section 2.

Example

Consider the following Van der Pol oscillator :

Table 2. Parameters for the numerical simulation

Variable	Value	Variable	Value
$x_1(0)$	1	$x_2(0)$	1
$\hat{x}_1(0)$	0.5	$\hat{x}_2(0)$	0.2
Parameter	Value	Parameter	Value
b_1	-7	b_3	14
b_2	-10	b_4	69

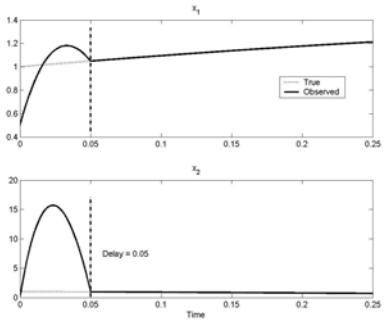


Fig. 2. Simulation results for exemple 2

$$\dot{x}_1 = x_2 \tag{79}$$
$$\dot{x}_2 = -x_1 + x_2 - x_1^2 x_2 \tag{80}$$
$$y = x_1 \tag{81}$$

and the following matrices and functions:

$$A_1 = \begin{pmatrix} b_1 & 1 \\ b_2 - 1 & 1 \end{pmatrix} \beta_1 = \begin{pmatrix} b_1 y + \frac{y^3}{3} \\ b_2 y + \frac{y^3}{3} \end{pmatrix}$$
$$A_2 = \begin{pmatrix} -1 - b_3 & 1 \\ -3 - b_4 & 2 \end{pmatrix} \beta_2 = \begin{pmatrix} b_3 y - \frac{y^3}{3} \\ b_4 y - 2\frac{y^3}{3} \end{pmatrix}$$

where b_1, b_2, b_3 and b_4 are constants to be chosen so that the matrices A_1, A_2 are Hurwitz. Solving the PDE systems (34) and (35) leads to the following expressions

$$\Psi_1 = \begin{pmatrix} x_1 \\ x_2 + \frac{x_1^3}{3} \end{pmatrix}, \Psi_2 = \begin{pmatrix} x_1 \\ x_2 + x_1 + \frac{x_1^3}{3} \end{pmatrix}$$

The implementation of the nonlinear finite time observer (58) requires to define both following systems ($i = 1, 2$)

$$\dot{z}_i = A_i z_i + \beta_i(y) \tag{82}$$
$$y = [1\ 0]\, z_i \tag{83}$$

and to compute the Jacobian of Ψ_1 and Ψ_2 which are respectively given by:

$$\nabla\Psi_1 = \begin{pmatrix} 1 & 0 \\ x_1^2 & 1 \end{pmatrix}, \nabla\Psi_2 = \begin{pmatrix} 1 & 0 \\ 1 + x_1^2 & 1 \end{pmatrix}$$

Simulation results are shown on Figure 2 where it can be seen that the estimate reaches exactly the state after the predefinite time delay D as expected.

4. CONCLUSION

In this paper we have presented a finite time observer for nonlinear systems that proceeds in one step and does not require to compute any inverse coordinates transformation. The estimation only requires to compute the Jacobian of the change of coordinates that transforms the system into a pseudo linear one allowing to build observers with linear error dynamics.

As the change of coordinates that transforms the system into a linear one is not always trivial, a more general approach has been envisaged. It consists in defining two pseudo linear systems allowing to compute two functions by solving a set of partial derivative equations. The estimate is then computed from the integration of a dynamical system using the Jacobian of each of the computed functions.

5. ACKNOWLEDGMENTS

This paper presents research results of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian State, Prime Ministers Office for Science, Technology and Culture. The scientific responsibility rests with its authors. The first author would also like to thank TOTAL S.A. for its financial support.

6. REFERENCES

Engel, R. and G. Kreisselmeier (2002). A continuous-time tbserver which converges in finite time. *IEEE Trans. Aut. Control* **47**, 1202–1204.

Hou, M. and A.C. Pugh (1999). Observer with linear error dynamics for nonlinear multi-output systems. *Syst. Contr. Lett.* **37**, 1–9.

Kazantzis, N. and C. Kravaris (1998). Nonlinear observer design using lyapunov's auxiliary theorem. *Syst. Contr. Lett.* **34**, 241–247.

Krener and Respondek (1985). Nonlinear observer with linearizable error dynamics. *SIAM J. Control Optim.*

Menold, P. (2004). Finite and Asymptotic Time State Estimation for Linear and Nonlinear Systems. PhD thesis. Stuttgart.

Menold, P.H., R. Findeisen and F. Allgöwer (2003*a*). Finite time convergent observers for nonlinear systems. In: *42nd IEEE Conference on Decision and Control, Maui, Hawai USA*.

Menold, P.H., R. Findesein and F. Allgöwer (2003*b*). Finite time convergent observers for linear time-varying systems. In: *11th Mediterranean Conference on Control and Automation*. Rhodos, Greece.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DYNAMIC ESTIMATION AND UNCERTAINTY QUANTIFICATION FOR MODEL-BASED CONTROL OF DISCRETE SYSTEMS

João F.M. Gândara [*], **Belmiro P.M. Duarte** [**], **Nuno M.C. Oliveira** [***]

[*] *Department of Food Science and Technology, ESAC, Polytechnic Institute of Coimbra. Bencanta, 3040-316 Coimbra, Portugal. Tel. +351-239-802940.*
[**] *Department of Chemical Engineering, ISEC, Polytechnic Institute of Coimbra. R. Pedro Nunes, 3030-199 Coimbra, Portugal. Tel. +351-239-790200.*
[***] *Department of Chemical Engineering, University of Coimbra. Pólo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal. Tel. +351-239-798700.*

Abstract: This paper presents an approach to estimate the outputs and the uncertainty associated to the forecast for discrete dynamic systems represented by state-space models. The complete strategy includes three steps: 1. process identification based on a data sample; 2. estimation of the current process state based on the information available during a moving past horizon, which may contain lack of observations; 3. forecast of process states, process outputs and uncertainty along the future horizon. This procedure can be incorporated in control strategies that explicitly consider model uncertainty.

Keywords: Estimation, process monitoring, optimal sampling, quality control.

1. INTRODUCTION

Traditional discrete process control applications assume that the sampling period used for interaction with the process, either through measurements or actuations, is fixed. This parameter is often chosen during the initial design phase of the control system, and before the specification of the control law to be used. However, the recent development of sophisticated control strategies, such as model-based approaches, and the integration of process information acquired from a number of distinct sources, has placed more emphasis on the choice and on-line adjustment of sampling policies, mostly for economical reasons.

The tasks of process and quality control commonly require the use of off-line analytical equipment to measure key product characteristics, such as concentrations and properties of particle systems; this can involve scarce and expensive human and equipment resources. In certain situations, the effective allocation of analytical resources can benefit from an economic performance analysis that simultaneously considers the relative value and costs associated with new information that can be introduced in an optimization problem.

Previous work on the selection of appropriate sampling intervals for process control with basis on economic criteria has been considered by MacGregor (1976), Abraham (1979), and Kramer (1989). The approach followed by these authors

assumed the availability of a linear dynamic model of the process, incorporating a stochastic component, used to predict the average performance of the controlled system when a larger sampling interval, equal to integer multiples of the basic sampling interval, is selected. This requires the use of a cost function that considers the cost of being off-specification, in terms of the variance of the observed errors, and the costs of taking new samples and making further process adjustments.

In this paper we propose a strategy for the forecast of the quality variables and their uncertainty, which are used to predict the probabilities of these variables being outside their quality specifications. Before the forecasts are made, it is necessary to estimate the current process state. For this a procedure is developed which is capable of effectively dealing with incomplete data sets. All these tasks are accomplished using a state-space model with a stochastic component. Finally, the proposed strategy is tested using a simulated continuous fermenter for ethanol production.

2. PROCESS MODEL

The approach described in this paper is applied to state-space models of the family $\mathcal{M}_1$

$$\mathcal{M}_1(A,B,C,D,K,\mathrm{Cov}(e)) = \\ = \begin{cases} x(t_{k+1}) = A\ x(t_k) + B\ u(t_k) + K\ e(t_k) \\ y(t_k) = C\ x(t_k) + D\ u(t_k) + e(t_k) \end{cases} \tag{1}$$

where $x(t_k) \in \mathbf{R}^{n_\mathrm{s}}$ is the vector of states at discrete time t_k, $u(t_k) \in \mathbf{R}^{n_\mathrm{i}}$ is the vector of inputs, $y(t_k) \in \mathbf{R}^{n_\mathrm{o}}$ is the vector of outputs and $e(t_k) \in \mathbf{R}^{n_\mathrm{o}}$ is the vector of stochastic components included in the state variables. The matrices A, B, C, D, K and $\mathrm{Cov}(e)$ are time invariant parameters.

Process identification is performed based on a complete data sample (including all process dynamic features) by employing a subspace projection algorithm (*N4SID*), an approach devoted to discrete systems identification (Van Overschee, 1994). The data represents the open–loop process behavior along the time horizon $N \times \Delta t$, where N is the number of records used for identification and Δt is the sampling interval. The *N4SID* algorithm only requires the knowledge of the system order, thus avoiding the need of *a priori* parametrization, and is non–iterative, avoiding the need of optimization schemes with corresponding problems, such as the convergence rate and the existence of local minima (Ljung, 1999).

The order of the system, n_s, is determined applying an information–based criterion, the Akaike Information Criterion (AIC), to measure the model fitness to process data (Akaike, 1972). The *AIC* metric of the models of the family $\mathcal{M}_1$ with order $n \in \{1, \cdots, {n_\mathrm{s}}^{\max}\}$, $\mathcal{M}_1^n$, is represented as:

$$AIC(\hat{\theta}^n) = \\ \log\left\{\frac{1}{N}\sum_{i=1}^{N}[\epsilon(i,\hat{\theta}^n)]^2\right\} + \frac{\dim(\hat{\theta}^n)}{N} \tag{2}$$

where $\hat{\theta}^n$ is the vector of parameter estimates included in the model $\mathcal{M}_1^n$, $\hat{\theta}^n = \{\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{K}, \mathrm{Cov}(e)\}^n$, and $\epsilon(i,\hat{\theta}^n)$ is the error of estimates of outputs $i \in \{1, \cdots, N\}$. The order of the system is determined as:

$$n_\mathrm{s} = \underset{n\in\{1,{n_\mathrm{s}}^{\max}\}}{\arg\min}\ AIC(\hat{\theta}^n) \tag{3}$$

where ${n_\mathrm{s}}^{\max}$ is the maximum order iterated. Process identification is performed off–line and the model is to be updated whenever process modifications are detected.

3. STATE ESTIMATION

The estimate of the current state process, $\hat{x}(t_0)$, can be obtained from the information available in the form of the inputs and the measurements obtained from sampling the process in the current, t_0, and past sampling times. We consider the receding horizon $\mathcal{H}_r$ comprising the last r discrete sampling times, $\mathcal{H}_r = \{t_{-r+1}, t_{-r+2}, \cdots, t_{-1}, t_0\}$. A possible approach to this problem (Brookner, 1998) consists on obtaining a set of equations in order to $\hat{x}(t_0)$. This is achieved by recursive substitution of all the state variables in the model equations at every sampling time in $\mathcal{H}_r$. This approach involves the use of negative powers of the transition matrix, A, and may lead to ill-conditioned problems, especially for stable systems and large horizons. The approach used in this paper consists on the simultaneous solution of all the model equations in the horizon $\mathcal{H}_r$. Although this leads to larger problems, it is a numerically stable procedure, avoiding ill-conditioning.

In the proposed approach, the problem of estimating the current state process is dealt with by solving the state equations at every sampling time in the horizon $\mathcal{H}_r$, together with the output equations referring to the available measurements

$$\begin{cases} \hat{x}(t_k) = A\ \hat{x}(t_{k-1}) + B\ u(t_{k-1}),\ t_k \in \mathcal{H}_r \\ y(t_k) = C_k\ \hat{x}(t_k) + D_k\ u(t_k),\ t_k \in \mathcal{H}_r, \end{cases} \tag{4}$$

where $y(t_k) \in \mathbf{R}^{n_{\mathrm{o},k}}$, $0 \le n_{\mathrm{o},k} \le n_\mathrm{o}$ is the vector containing the variables measured at sampling time t_k. When no output is measured we have $n_{\mathrm{o},k} = 0$, and when all outputs are measured $n_{\mathrm{o},k} = n_\mathrm{o}$. Matrices $C_k \in \mathbf{R}^{n_{\mathrm{o},k}\times n_\mathrm{s}}$ and $D_k \in \mathbf{R}^{n_{\mathrm{o},k}\times n_\mathrm{i}}$ contain the rows of C and D

corresponding to the measured variables at t_k. We assume that all inputs in the horizon are perfectly known.

Defining the enlarged vectors of states and inputs

$$\hat{\mathcal{X}}_r = \left[\hat{x}(t_0)^{\mathrm{T}}\ \hat{x}(t_{-1})^{\mathrm{T}}\ \cdots\ \hat{x}(t_{-r+1})^{\mathrm{T}}\right]^{\mathrm{T}},$$
$$\mathcal{U}_r = \left[u(t_0)^{\mathrm{T}}\ u(t_{-1})^{\mathrm{T}}\ \cdots\ u(t_{-r+1})^{\mathrm{T}}\right]^{\mathrm{T}},$$

respectively, the system of equations (4) can be formulated as

$$\mathcal{J}_r\ \hat{\mathcal{X}}_r = \mathcal{L}_r, \tag{5}$$

where $\mathcal{L}_r = \mathcal{I}_r + \mathcal{N}_r\mathcal{U}_r$. Vector $\mathcal{I}_r$ contains the output information in the considered horizon:

$$\mathcal{I}_r = \left[y(t_0)^{\mathrm{T}}\ \mathbf{0}_{n_s}^{\mathrm{T}}\ y(t_{-1})^{\mathrm{T}}\ \mathbf{0}_{n_s}^{\mathrm{T}}\ \cdots\ y(t_{-r+1})\right]^{\mathrm{T}},$$

where $\mathbf{0}_{n_s}$ is a vector containing n_s zeroes. Matrix $\mathcal{N}_r$ is formed by the model matrices B and D_k, and $\mathcal{J}_r$ is a sparse and structured matrix formed by matrices C_k, A and I_{n_s} (n_s–dimension identity matrix). For instance, using a generic receding horizon with dimension $r = 4$, we have:

$$\mathcal{J}_r = \begin{pmatrix} C_0 & & & \\ I_{n_s} & -A & & \\ & C_{-1} & & \\ & I_{n_s} & -A & \\ & & C_{-2} & \\ & & I_{n_s} & A \\ & & & C_{-3} \end{pmatrix};$$

$$\mathcal{N}_r = \begin{pmatrix} -D_0 & & & & & \\ & B & & & & \\ & -D_{-1} & & & & \\ & & B & & & \\ & & -D_{-2} & & & \\ & & & & B & \\ & & & & -D_{-3} & \end{pmatrix}.$$

The process state estimate along $\mathcal{H}_r$, $\hat{\mathcal{X}}_r$, which includes the current process state, designated as $\hat{x}(t_0)$, is determined by solving the least-squares problem

$$\min_{\hat{\mathcal{X}}_r} \parallel \mathcal{L}_r - \mathcal{J}_r\ \hat{\mathcal{X}}_r \parallel^2 \tag{6}$$

The optimality of $\hat{\mathcal{X}}_r$ is achieved whenever problem (6) is determined or over-determined. Taking into account that $\mathcal{J}_r \in \mathbf{R}^{m\times n}$, with

$$m = r\ n_s + \sum_{i=0}^{r-1} n_{o,-i}, \quad n = (r+1)\ n_s,$$

the necessary condition that leads to the determination or over-determination of problem (6) is $m \geq n$, thus requiring

$$\sum_{i=0}^{r-1} n_{o,-i} \geq n_s.$$

Therefore, the problem is determined if at least n_s observations are available in $\mathcal{H}_r$.

Problem (6) is solved by performing a QR decomposition of matrix $\mathcal{J}_r$,

$$\mathcal{J}_r = \begin{bmatrix} Q_1\ Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \tag{7}$$

and subsequently applying a back-substitution procedure to determine $\hat{\mathcal{X}}_r$. In closed form, the problem solution is represented as

$$\hat{\mathcal{X}}_r = R_1^{-1} Q_1^{\mathrm{T}} \mathcal{L}_r. \tag{8}$$

The outputs in the horizon $\mathcal{H}_r$ are affected by errors, mainly measurement ones. Since the state estimates are obtained from the available measurements, they also include an error component,

$$\hat{\mathcal{X}}_r = \mathcal{X}_r + e_{\mathcal{X}}.$$

Assuming that $E\{e(t_k)\} = 0, \forall t_k \in \mathcal{H}_r$, then $\hat{\mathcal{X}}_r$ is an unbiased estimator of $\mathcal{X}_r$ (Ikonen, 2002), leading to $E\{e_{\mathcal{X}}\} = 0$. Using (8), the covariance of this error can be computed as

$$\mathrm{Cov}(e_{\mathcal{X}}) = R_1^{-1} Q_1^{\mathrm{T}} \mathrm{Cov}(e_r) Q_1 (R_1^{-1})^{\mathrm{T}},$$

where e_r is a vector containing all the errors in the horizon $\mathcal{H}_r$ (Brookner, 1998), which are due both to modelling and to measurement components.

The structure and sparsity of matrices $\mathcal{J}_r$ and R_1 is exploited by the use of a tailored algorithm for the QR decomposition. This algorithm avoids the decomposition of the full matrix, and performs successive factorizations of smaller blocks, reducing the number of necessary operations. From this point of view, an important feature of this problem is that the the structure of R_1 is not dependent of the number of variables measured in the horizon $\mathcal{H}_r$.

The receding horizon $\mathcal{H}_r$ is updated every time the procedure is used, in order to incorporate the new information, as it becomes available. If the information in $\mathcal{H}_r$ is not sufficient to ensure the optimality of $\hat{x}(t_0)$, then its size can be increased.

4. PREDICTION

The current process estimate, together with the process model, can be used to forecast the quality variables in a future horizon $\mathcal{H}_f$, comprising f discrete sampling times ahead of t_0, $\mathcal{H}_f = \{t_1, \cdots, t_f\}$. This forecast can be split in two components, the expected values of the quality variables, $\hat{y}$ and the uncertainty associated with the forecast, ε:

$$y(t_k) = \hat{y}(t_k) + \varepsilon(t_k),\ t_k \in \mathcal{H}_f. \tag{9}$$

Both these components are used to forecast the probability of a given quality variable being out of a set of predetermined specifications.

The expected values of the quality variables are predicted using the deterministic part of the

model (1). Using the the previously obtained state estimate as the initial condition:

$$\mathcal{M}_2(A,B,C,D,K,\mathrm{Cov}(e)) = \begin{cases} x(t_k) = A\ x(t_{k-1}) + B\ u(t_{k-1}), & t_k \in \mathcal{H}_f \\ \hat{y}(t_k) = C\ x(t_k) + D\ u(t_k), & t_k \in \mathcal{H}_f \\ x(t_0) = \hat{x}(t_0) \end{cases} \tag{10}$$

Again, we assume that the future profile of the input variables has been determined (using, for example, a MPC-type strategy) and is known.

The uncertainty in the obtained forecast of the quality variables can also be predicted, using the stochastic part of the process model (1). This uncertainty is not only due to the error terms, but also to the uncertainty in the value of $\hat{x}(t_0)$. Propagating (1) in the prediction horizon, and including a term due to the error in the initial condition, we obtain:

$$\begin{aligned} &\varepsilon(t_k) = y(t_k) - \hat{y}(t_k) \\ &= C\sum_{j=0}^{k} A^j K e(t_j) + e(t_k) + CA^k e_{x_0},\ t_k \in \mathcal{H}_f. \end{aligned} \tag{11}$$

From the above equation we can conclude that $E\{\varepsilon(t_k)\} = 0$, since $E\{e(t_k)\} = 0, t_k \in \mathcal{H}_f$, and $E\{e_{x_0}\} = 0$.

The result obtained in (11) is useful, not to predict the actual value of the error, but because it allows us to obtain a measure of the uncertainty in the predictions $\hat{y}(t_k)$. Based on the work of Seppala (1998), the variance of the ith element of $\varepsilon(t_k)$ is computed as

$$\begin{aligned} \hat{\sigma}_i(t_k) = \mathrm{var}(\varepsilon_i(t_k)) = \langle \Psi_i^{\mathrm{T}}\ \Psi_i,\ \mathrm{Cov}(e)\rangle + \\ \langle (\Omega_i)^{\mathrm{T}}\Omega_i, \mathrm{Cov}(e_{x_0})\rangle,\ t_k \in \mathcal{H}_f, \end{aligned} \tag{12}$$

where $\langle \bullet, \bullet \rangle$ is the internal product operator, and Ψ_i and Ω_i are the ith rows of matrices Ψ and Ω:

$$\Psi = C\ \left(\sum_{j=0}^{k} A^j\right)\ K + I \tag{13}$$

$$\Omega = CA^k. \tag{14}$$

Note that, since $E\{e(t_k)\} = 0$, the variance of $\hat{y}_i(t_k)$ is the same as that of $\varepsilon_i(t_k)$.

As the prediction instant moves further away from the current sampling time, the second term in (12) continuously increases, due to the accumulation of the model uncertainty. The behavior of the second term, depends on the process stability. If the process is unstable, this term will also increase. If the process is stable, the contribution of the initial error for the forecast variance continuously decreases as t_k mover further away from the current sampling time, since $A^k e_{x_0}$ goes to zero as k increases.

With the obtained information it is possible to predict the probability of a given quality variable being outside it specifications, LS_i and US_i, lower and upper specification limits, respectively. If the random noise, $e(t)$, is well described by a stationary Gaussian distribution, then this probability can be predicted by:

$$\begin{aligned} &pf_i(t) = \\ &\int_{-\infty}^{LS_i} \frac{1}{\sqrt{2\ \pi}\ \hat{\sigma}_i(t_k)}\ \exp\left[-\frac{(z-\hat{y}_i(t_k))^2}{2\ (\hat{\sigma}_i(t_k))^2}\right]\ \mathrm{d}z\ + \\ &+ \int_{US_i}^{+\infty} \frac{1}{\sqrt{2\ \pi}\ \hat{\sigma}_i(t_k)}\ \exp\left[-\frac{(z-\hat{y}_i(t_k))^2}{2\ (\hat{\sigma}_i(t_k))^2}\right]\ \mathrm{d}z, \\ &t_k \in \mathcal{H}_f \end{aligned} \tag{15}$$

A large value of this probability can be due either to a shift in the process or to the increase of the uncertainty in the forecasts. The first can be solved by taking appropriate control measures, in order to drive the process back to the central value of the specifications. If the uncertainty in the forecasts is too large, then a a new measurement of the variable for which pf is too large should be made, before or at the sampling time where this occurs. With this new information, the estimation procedure described earlier is repeated, in order to get a new estimate of the process state, with reduced uncertainty.

5. APPLICATION EXAMPLE

The proposed strategy was tested using a non-linear dynamical model of a continuous fermenter for ethanol production using glucose (Chmúrny, 2000). The measured variables considered in this model are:

(1) biomass concentration, x;
(2) substrate concentration, s;
(3) product concentration, P;
(4) biomass concentration in the output solution, x_v;
(5) carbon dioxide production rate, r_{CO_2};
(6) rate of base consumption, r_z.

The carbon dioxide production and the base consumption rates can be easily measured by online sensors. The concentrations require more complex analytic methods, and thus, are more expensive to obtain. The input variables are D, the ratio between the fermenter feed rate and its volume, and s_0, the substrate concentration in the feed. In this particular model, the temperature and the volume are assumed to be constant.

5.1 Model identification

The fermenter dynamical model was used to generate a complete data set, with random errors

added to the output variables. The *N4SID* algorithm was applied to this data in order to obtain a state-space model, in the form $\mathcal{M}_1(\bullet)$. The substrate concentration variable was not used for identification purposes since its behavior gave rise to much worse models, from the point of view of stability and data fitting. The dimensions of the obtained linear model are $n_\mathrm{s} = 5$, $n_\mathrm{i} = 2$ and $n_\mathrm{o} = 5$ (all of the mentioned above, except the substrate concentration). The sampling interval used is $\Delta t = 0.05\,\mathrm{h}$.

5.2 State estimation

For the current state estimation we have considered an horizon $\mathcal{H}_r$ with dimension $r = 50$. The data for this horizon was generated using the original dynamical model with random error added to the measured variables. This data is different from the one used for identification purposes, but the error has the same characteristics. Not all of the measurements available in $\mathcal{H}_r$ were used. The decision regarding the availability of the measurements was modelled by an independent random binary signal. For the presented results, the number of measurements considered is of 121 out of 250 possible.

The estimated value of $\hat{y}(t_0)$ obtained is:

$$\hat{y}(t_0) = \left[1.03\ 1.03\ 1.01\ 1.01\ 1.03\right]^\mathrm{T},$$

normalised, and the deviation from the value obtained by simulation, in percentage, is

$$\left[-1.39\ -1.51\ -1.44\ -3.23\ -0.91\right]^\mathrm{T}.$$

The profiles of the estimated and measured values of outputs variables x and P, in the horizon $\mathcal{H}_r$, are presented in Figure 1. The difference between all the estimated outputs and all the measurements (both the used and the deleted in the estimation procedure) is presented in Figure 2.

In the estimation step, to obtain $\mathrm{Cov}(e_\mathcal{X})$, we assume that the outputs errors are not correlated with each other and that these errors are mainly due to errors in the measurement methods. Under these assumption, the covariance matrix of the error in $\mathcal{H}_r$, $\mathrm{Cov}(e_r)$, is diagonal with all its elements equal to

$$\sigma^2 = 8.33 \times 10^{-4},$$

the variance of the added random error.

5.3 Prediction

For the prediction phase, we have arbitrarily set the specification limits at $LS = 0.85$ and $US = 1.15$, for all output variables. We have considered a prediction horizon, $\mathcal{H}_f$, with dimension $f = 50$.

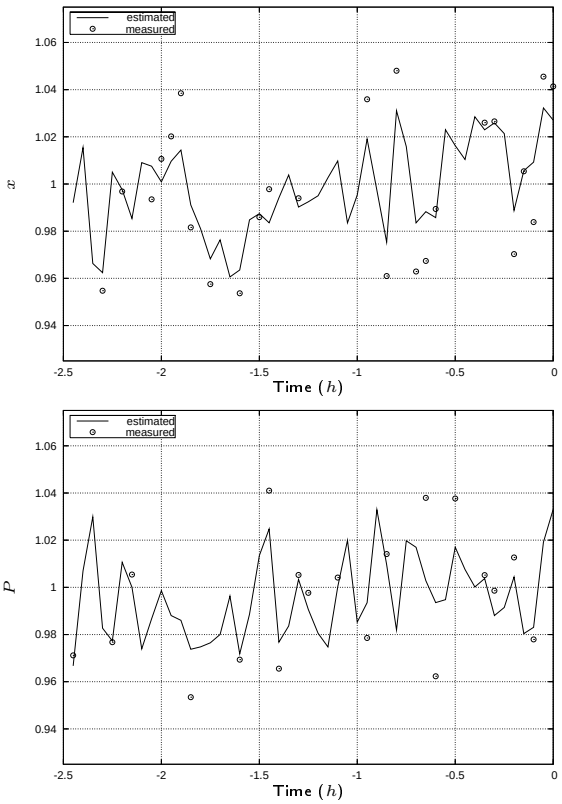


Fig. 1. Real and estimated values of variables x and P in estimation horizon.

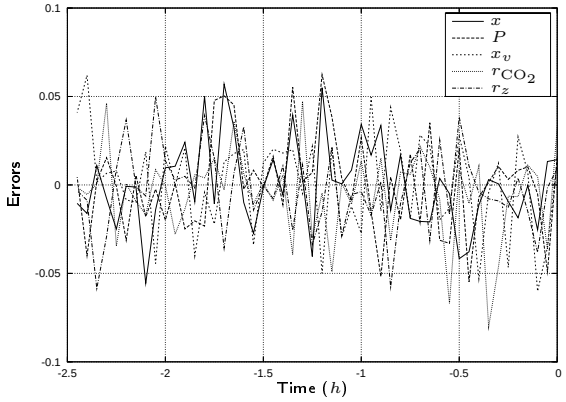


Fig. 2. Estimation errors in the horizon $\mathcal{H}_r$ for all the measured variables.

All the input variables in this horizon are kept at their reference values. The forecasts, for $\mathcal{H}_f$, of all the outputs, their variances and the probability of being out of specifications are presented in Figure 3.

We can see, in Figure 3, than, initially, the variance of the forecasts decreases, due to the the expected decrease of the second term in (12), since the system is stable. As t_k increases, the contribution from the second term becomes dominant, and the variance increases.

In Figure 3 we can see than, without further measurements being made, the probability of all the variables being within their specifications is

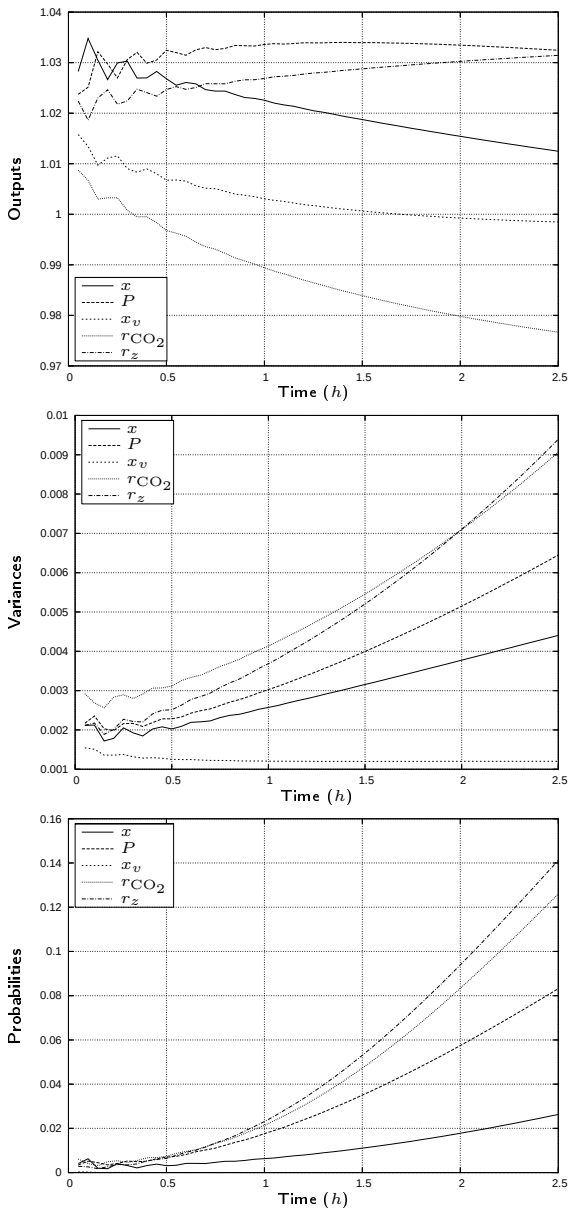


Fig. 3. Forecast of the measured variables, their variance and their probability of being out of specifications for the prediction horizon $\mathcal{H}_f$.

about 98%. From this value it would be reasonable to suppose that no further measurements would be needed up to this time. However, the decision to sample or not would require a greater insight into the system, such as sampling and quality costs.

6. CONCLUSIONS

In this paper we have presented a procedure for the forecast of process measurements and the quantification of their uncertainty. This can be used to predict the probability of a given quality variable being out of its specifications in a future horizon, such as used in model predictive control. The probability can be used to decide if it is possible to take control measures (within the available degrees of freedom) to correct the predicted trajectories, or if instead it is preferable to obtain new information from the process, by performing new types of measurements, in order to decrease the expected operating costs.

This procedure relies on the use of a state-space model which is obtained using process identification techniques. It also includes the estimation of the current process state from the information available in a receding horizon, again using the process model. The problem of missing measurements in the receding horizon is dealt with by considering an outputs vector with variable dimension. Some of the advantages of this procedure are its numerical stability and the capability of dealing with growing or shrinking receding horizons.

The overall procedure can be easily integrated in a Model Predictive Control type strategy, using an objective function that explicitly includes quality and sampling costs.

REFERENCES

Abraham, B., Box G.E.P. (1979). Sampling interval and feedback control. *Technometrics* **21**, 1–8.

Akaike, H. (1972). Information theory and an extension of the maximum likelihood principle. In: *Proc. 2nd. Int. Symp. Information Theory, Supp. to Problems of Control and Information Theory.* pp. 267–281.

Brookner, E. (1998). *Tracking And Kalman Filtering Made Easy.* John Wiley & Sons. New York.

Chmúrny, D., Chmúrny R. (2000). Simulation and control of fermentation complex systems. *Bioprocess Engineering* **23**, 221–227.

Ikonen, E., Najim K. (2002). *Advanced Process Identification and Control.* Marcel Dekker. New York.

Kramer, T. (1989). Process Control from an Economic Point of View. PhD thesis. Univ. Wisconsin–Madison.

Ljung, L. (1999). *System Identification – Theory for the user.* 2nd. ed.. Prentice Hall PTR. New Jersey.

MacGregor, J.F. (1976). Optimal choice of the sampling interval for discrete process control. *Technometrics* **18**, 151–160.

Seppala, C.T. (1998). Dynamic Analysis of Variance Methods for Monitoring Control System Performance. PhD thesis. Queen's University.

Van Overschee, P., DeMoor B. (1994). N4sid: Subspace algorithms for the identification of combined deterministic–stochastic systems. *Automatica* **30**, 75–93.

Keynote 7

Multivariable Controller Performance Monitoring

S. J. Qin and J. Yu,
University of Texas at Austin

Keynote 8

PSE Relevant Issues in Semiconductor Manufacturing: Application to Rapid Thermal Processing

C. C. Yu, A. J. Su, J. C. Jeng, H. P. Huang, S. Y. Hung, and C. K. Chao
National Taiwan University

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

MULTIVARIABLE CONTROLLER PERFORMANCE MONITORING

S. Joe Qin [1] and **Jie Yu**

Department of Chemical Engineering
The University of Texas at Austin
Austin, TX 78712, USA

Abstract: In this paper we give a critical overview of recent development in MIMO control performance monitoring. We discuss a number of MIMO control benchmarks including minimum variance, LQG, and user selected benchmarks. Performance measures are extended from variance based measures in SISO control to covariance based measures in MIMO control. Pros and cons of various benchmarks are discussed. The diagnosis of poor control performance relative to a benchmark is a major focus of the paper. We argue that in the MIMO setting worst performance directions should be analyzed from data to yield meaningful diagnosis information. Therefore, multivariate statistics should be applied for the diagnosis of worst performance directions, much like its use in multivariate process monitoring.

Keywords:
MIMO control performance monitoring, minimum variance, model predictive control, covariance based monitoring, worst performance directions.

1. INTRODUCTION

Control performance monitoring and evaluation can be traced to Åström (Åström, 1970; Åström, 1976) and later Harris (Harris, 1989) who demonstrated that the minimum variance benchmark can be estimated from normal closed-loop operation data. Åström in his CPC-2 paper (Åström, 1976) noted the following:

> In the special case of minimum variance control ... it is known that the covariance function will vanish for lags greater than the sum of the sampling interval and transport delay of the system. It is then sufficient to record output only and to compute its covariance function.

The interest from both academia and industry in control performance monitoring has surged tremendously in the last decade as documented in several review papers and a monograph (Qin, 1998; Harris *et al.*, 1999; Kozub, 1996; Harris and Seppala, 2002; Hoo *et al.*, 2003; Huang and Shah, 1999). The recent survey paper by Jelali (Jelali, 2006) provides a very good collection of recent development in the control performance monitoring area from SISO, MIMO to valve stiction problems. In the application domain, just in HVAC systems alone, Johnson Control has implemented over half a million control monitors in the last ten years based on a pattern recognition technique (Seem, 1998; Seem, 2006). Paulonis and Cox (Paulonis and Cox, 2003) reported the development of a control performance monitoring system spanning over 14,000 PID loops at the Eastman Chemical Company. Industrial case studies (Thornhill *et al.*, 1999; Miller *et al.*, 1998; Harris *et al.*, 1996*b*; Perrier and Roche, 1992; Wein-

[1] Corresponding Author: qin@che.utexas.edu.
Supported by the Texas-Wisconsin Modeling and Control Consortium.

stein, 1992; Desborough and Miller, 2002) have been published on the subject and minimum variance based performance indices are a part of many commercially available control performance monitoring packages.

More recently, academic research interest has shifted to the assessment of MIMO control systems using the minimum variance benchmark (Harris *et al.*, 1996*a*; Huang *et al.*, 1997; Huang, 1997). Harris *et al.* (Harris *et al.*, 1996*a*) reformulated an LQ control solution (Harris and MacGregor, 1987) and show that the optimally controlled process follows a finite $(d-1)th$-order moving average process where d is the maximum delay present in the interactor. They proposed a statistical test of minimum variance based on a cross-correlation of the interactor filtered output vector and past outputs. The minimum variance calculation involves time series modeling of the closed loop system, spectral factorization of the inverse interactor and subsequent solution of a matrix Diophantine equation.

Huang *et al.* (Huang *et al.*, 1997) introduced the unitary interactor as a means of avoiding spectral factorization. The unitary interactor matrix was used to develop an explicit solution to the singular LQ regulation problem by Peng and Kinnaert (Peng and Kinnaert, 1992) and can be used to derive MVC with arbitrary output weighting (Huang, 1997). The need for a process transfer function restricts the practical usefulness of these algorithms. Harris recently (Harris, 2004) established the statistical confidence for the quadratic type of indices like the MVC benchmark.

The MVC benchmark has drawbacks in practice and alternative benchmarks are proposed. One of the limitations is the requirement of the interactor matrix which is essentially a good part of the entire process model. Seppala et al. (Seppala *et al.*, 2002) propose the use of time series analysis to model the control error dynamics and from it to analyze interactions in the multivariable system. No prior information about the process delay structure is required. McNabb and Qin (McNabb and Qin, 2003; McNabb and Qin, 2005) demonstrate that the variance based monitoring index is insufficient for assessing the multivariate covariance of the control performance. As an alternative a covariance based monitoring index is proposed to measure the variance-covariance inflation in terms of the 'volume' of the variability. Another drawback of the existing literature is that little has been done regarding diagnosis. In contrast, a great deal of research has taken place in the area of multivariate process monitoring (MacGregor and Kourti, 1995; Qin, 2003). We argue that in the MIMO setting the worst performance directions should be analyzed from data to yield meaningful diagnosis information. Therefore, multivariate statistics should be applied for the diagnosis of the worst performance directions, much like its use in multivariate process monitoring. Further, the need for the integration of control performance monitoring and process monitoring is pointed out as both problems co-exist in a plant with the same data as the ultimate information source for diagnosis.

In this paper we seek to provide a critical (rather than complete) overview of the MIMO control performance area and point to a new direction of covariance-based monitoring. For a more complete literature review the reader is referred to (Jelali, 2006). This paper is organized as follows. A critical overview of the MIMO control performance monitoring literature is given with some effort to unify some well-known methods. MIMO control performance indices based on the covariance is highlighted. Poor performance diagnosis is conducted by analyzing the worst performance directions using generalized eigenvalue analysis of two covariance matrices. We further propose to have the benchmark covariance as user-defined, rather than from a theoretical calculation. The user-defined benchmark can be a period of operation data that are taken from an exemplary operation. Since the benchmark is not necessarily a lower bound, the diagnosis results from the generalized eigenvector analysis include directions in which the performance deteriorates and those in which the performance improves. The worst performance directions are then analyzed with a proposed contribution analysis that leads to controlled variables or loops most responsible for the performance deterioration. The paper ends with a few concluding remarks.

2. OVERVIEW OF MIMO CONTROL PERFORMANCE MONITORING

2.1 *Minimum Variance Benchmark*

A MIMO process can be represented by the following equation:

$$y(k) = G(q)u(k) + H(q)e(k)$$

where $G(q)$ is the process transfer function matrix which contains possible time delays, $e(k)$ is the white noise innovation and $H(q)$ is the transfer function matrix of the disturbance. For SISO processes $G(q)$ can be represented by

$$G(q) = \tilde{G}(q)q^{-d} \tag{1}$$

where d is the time delay and $\tilde{G}(q)$ is time delay free. If we assume $G(q)$ has no zeros outside the unit circle, $\tilde{G}(q)$ is invertible. For simplicity we assume $G(q)$ has no zeros outside the unit circle except for the time delays.

For MIMO processes the time delays appear in a more complex form. The conventional approach is to find a unitary interactor matrix $D(q)$ such that (Peng and Kinnaert, 1992; Huang and Shah, 1997)

$$\tilde{G}(q) = D(q)G(q) \tag{2}$$

is full rank when $q^{-1} \to 0$, where $D^T(q^{-1})D(q) = I$, that is, $D(q)$ is a unitary matrix. Several methods are available to calculate the interactor matrix from the process model.

By examining the analogy between (1) and (2) we can express $G(q)$ as a product of two parts:

$$G(q) = D^{-1}(q)\tilde{G}(q) = D^T(q^{-1})\tilde{G}(q) \tag{3}$$

where $D^T(q^{-1})$ is analogous to the time delay in (1). Since $H(q)$ is a rational transfer function matrix of the disturbance without time delay, $D(q)H(q)$ should contain some positive factors of q. Denoting

$$D(q)H(q) = \sum_{i=1}^{d} F_i q^i + R(q) \tag{4}$$

where $R(q)$ contains no positive factors of q. We can now express the process output as

$$\begin{aligned} y(k) &= D^T(q^{-1})D(q)[G(q)u(k) + H(q)e(k)] \\ &= D^T(q^{-1})[\tilde{G}(q)u(k) + \sum_{i=1}^{d} F_i e(k+i) + R(q)e(k)] \end{aligned} \tag{5}$$

The innovation sequence $e(k)$ is known up to the current time k once $y(k)$ is measured, but $e(k+i)$ for $i = 1, \ldots, d$ are not known. Therefore, the feedback control $u(k)$ cannot do anything about the $e(k+i)$ terms. $u(k)$ can only be related to $e(k-i) \quad (i \geq 0)$ terms.

By defining the filtered output and using the results in (5),

$$\begin{aligned} \tilde{y}(k+d) &= D(q)y(k) \\ &= \sum_{i=1}^{d} F_i e(k+i) + \underbrace{R(q)e(k) + \tilde{G}(q)u(k)}_{\sum_{j=0}^{\infty} F_{-j}e(k-j)} \end{aligned} \tag{6}$$

For all possible feedback control the second and third terms of (6) can be expressed as past innovations. Further denoting

$$\tilde{y}_{mv}(k+d) = \sum_{i=1}^{d} F_i e(k+i)$$

and using the fact that $\tilde{y}(k)$ is stationary and $e(k)$ is white noise,

$$\begin{aligned} cov\{\tilde{y}(k)\} &= cov\{\tilde{y}_{mv}(k+d)\} + cov\{e(k-j)\, terms\} \\ &\geq cov\{\tilde{y}_{mv}(k)\} \end{aligned} \tag{7}$$

and the MIMO minimum variance control is achieved by

$$u(k) = -\tilde{G}^{+}(q)R(q)e(k) \tag{8}$$

where $\tilde{G}(q)$ is full rank as $q^{-1} \to 0$. Note that pseudo-inverse is used here since $\tilde{G}(q)$ can be non-square.

The above derivation gives the MIMO MVC control law which actually achieves *minimum covariance* in the filtered output. This result, however, has not been widely recognized so far. We make a few remarks about this derivation.

Remark 1. The above MIMO MVC derivation is straightforward and analogous to the SISO MVC derivations (Åström, 1970). The MIMO MVC control law is explicitly expressed in terms of the innovations which correspond to the process output data.

Remark 2. The MIMO MVC law actually achieves minimum covariance in the filtered output, as depicted in (7), which make the difference of the two covariances positive semi-definite. As a consequence, MIMO MVC achieves minimum variance in all possible directions in the filtered output space.

Remark 3. All MVC based performance monitoring methods require the knowledge of $D(q)$ implicitly or explicitly, which is calculated by various means. Huang and Shah (Huang and Shah, 1997) start from the transfer function form, while McNabb and Qin (McNabb and Qin, 2003) start with the state space form. Both methods require only the first d Markov parameter matrices of the process, instead of the entire process model. However, these Markov parameters are difficult to obtain unless some form of identification tests are performed.

Remark 4. Often the sum of the output variances is chosen as a benchmark, which is

$$\begin{aligned} &tr[cov(y_{mv}(k))] = Ey_{mv}^T(k)y_{mv}(k) \\ &= E\left(\tilde{y}^T(k+d)D(q)D^T(q^{-1})\tilde{y}(k+d)\right) \\ &= E\left(\tilde{y}^T(k+d)\tilde{y}(k+d)\right) \\ &= tr[cov(\tilde{y}_{mv}(k))] \\ &= tr\left\{\sum_{i=1}^{d} F_i R_e F_i^T\right\} \end{aligned} \tag{9}$$

where $R_e = cov(e(k))$. We will argue later that the sum of variances is an incomplete measure of the overall output covariance.

The minimum variance parameters F_i can be estimated from routine operational data. The FCOR algorithm (Huang *et al.*, 1997) pre-estimates the innovations $e(k)$ and then performs correlation analysis to estimate F_i. The subspace projection

method of McNabb and Qin (2003) represents the past innovations $e(k-j)$ in terms of past data $y(k-j)$ $(for\ j \geq 0)$ and uses the projection error as $\sum_{i=1}^{d} F_i e(k+i)$. These two algorithms are essentially equivalent.

Harris (Harris, 2004) discusses the issue of the variance of the minimum variance estimated from data, which is an important issue that has not been discussed before. Although many algorithms that calculate the minimum variance are numerically equivalent, the algorithms that estimate the coefficients F_i from closed loop data can differ in terms of statistical efficiency or in the variance of the estimates. The FCOR algorithm, for example, first estimates the innovations sequence and then estimates the coefficients F_i. This procedure resembles the two stage least squares algorithm in (Kashyap and Nashburg, 1974), which is shown to be simple but not efficient in (Mayne and Firoozan, 1982), where an improved efficient algorithm is also proposed.

2.2 *Alternative Performance Benchmarks*

The limitations of the MVC based benchmark are

(1) The benchmark is based solely on time delay restrictions; other restrictions such as hard constraints are not considered.
(2) The minimum variance, although achievable under ideal situations, leads to a non-robust controller. This is characterized by excessive input moves that are usually inherent to MVC.
(3) Only disturbance rejection performance is considered.
(4) The requirement of the interactor is restrictive in practice.

To overcome these limitations, many alternative benchmarks have been studied. Huang and Shah (1998) allow the user to specify the noise decay rate after the interactor order, which has built-in robustness in the benchmark. This approach, however, still requires the interactor matrix. In a similar effect but for the SISO case, Horch and Isaksson (1999) introduced a finite closed-loop pole in the benchmark controller to enhance robustness.

A departure from the use of the interactor matrix is given in the work of Huang et al. (Huang *et al.*, 2005) where, instead of using the exact interactor matrix, only the order of the interactor is used. This method removes the need to estimate the interactor matrix. The time series analysis approach of Seppala et al. (Seppala *et al.*, 2002) does not require any information about the interactor matrix. The control error is analyzed as a time series to detect whether the control loops are interacting or not. Recent work of Harris and Yu (Harris and Yu, 2003) performs degree of freedom analysis to monitor the status of constraints and long run behavior of the control performance.

To address the issue of excessive input moves of MVC, Kadali and Huang (Kadali and Huang, 2002) propose to use LQG as a benchmark. A drawback of this benchmark is the requirement of the entire process model.

As the ultimate multivariable controller in industry is model predictive control (MPC), several attempts have been made to assess the performance of MPC. Loquasto and Seborg (Loquasto and Seborg, 2003) propose the use of similarity factors and pattern recognition to determine the MPC performance is normal or abnormal, and if there is a significant disturbance change. Schaffer and Cinar (Schaffer and Cinar, 2004) propose a knowledge based approach for MPC performance monitoring. Given the complexity of MPC that involves model errors, disturbance changes, optimal target settings, active constraint sets, and controller tuning, the MPC performance monitoring is largely an unsolved problem.

3. COVARIANCE-BASED PERFORMANCE INDEX AND DIAGNOSIS

In MIMO control performance monitoring, the process output variance is an important parameter and the associated performance index may be defined as the ratio of minimum variance to actual variance

$$\eta = \frac{tr\left\{cov(\tilde{y}_{mv}(k))\right\}}{tr\left\{cov(\tilde{y}(k))\right\}} \tag{10}$$

The value of variance index η is between 0 and 1, where the upper bound 1 corresponds to the minimum variance. In the above equation, however, only the diagonal elements of covariance matrix are taken into comparison and the information from the off-diagonal elements is completely ignored (McNabb and Qin, 2003).

To account for the variability that is accurately represented by the covariance matrix, a volume-like performance index is more appropriate, which is defined by the ratio of the determinants as follows,

$$I_v = \frac{\det\left\{cov(\tilde{y}_{mv}(k))\right\}}{\det\left\{cov(\tilde{y}(k))\right\}} \tag{11}$$

Since the determinant is the product of all eigenvalues of the covariance matrix, this index defines exactly the volume ratio.

Denoting the eigenvalues of $cov(\tilde{y}_{mv}(k))$ and $cov(\tilde{y}(k))$ as λ_i^{mv} and λ_i, respectively, the variance based and covariance based performance indices can be rewritten as

$$\eta = \frac{\sum \lambda_i^{mv}}{\sum \lambda_i} \tag{12}$$

$$I_v = \frac{\prod \lambda_i^{mv}}{\prod \lambda_i} \tag{13}$$

Although both indices use information from the eigenvalues, the volume-like index takes into account the covariance information and interactions among variables.

To find a direction in $\tilde{y}(k)$ along which the worst suboptimality occurs, we find the direction p with $\|p\| = 1$ and project $\tilde{y}(k)$ and $\tilde{y}_{mv}(k)$ to this direction:

$$\begin{aligned} \Pi_p \tilde{y}(k) &= p^T \tilde{y}(k)/p^T p &= p^T \tilde{y}(k) \\ \Pi_p \tilde{y}_{mv}(k) &= p^T \tilde{y}_{mv}(k)/p^T p &= p^T \tilde{y}_{mv}(k) \end{aligned}$$

The variance of the projections are, respectively,

$$\begin{aligned} var(\Pi_p \tilde{y}(k) &= p^T cov(\tilde{y}(k))p \\ var(\Pi_p \tilde{y}_{mv}(k) &= p^T cov(\tilde{y}_{mv}(k))p \end{aligned}$$

The direction p along which the largest variance ratio occurs is

$$p = arg\ max \frac{p^T cov(\tilde{y}(k))p}{p^T cov(\tilde{y}_{mv}(k))p} \tag{14}$$

The direction of p after maximization give the direction with the most potential to improve the performance. The solution to this problem is a generalized eigenvector problem,

$$cov(\tilde{y}(k))p_i = \mu_i cov(\tilde{y}_{mv}(k))p_i$$

where p_i is the generalized eigenvector corresponding to the i^{th} largest generalized eigenvalue μ_i. The volume of the suboptimality or variance inflation due to poor control performance is:

$$\prod_{i=1}^{l} \mu_i$$

where l is the number of selected directions. The volume-based performance can be defined as

$$I_v(l) = \prod_{i=1}^{l} {\mu_i}^{-1}$$

It is straight forward to show from (7) that for all possible projections Π,

$$cov\left(\Pi \tilde{y}_{mv}(k)\right) \leq cov\left(\Pi \tilde{y}(k)\right)$$

Therefore, $\mu_i \geq 1$ and I_v is between zero and one. When $\tilde{y}(k)$ achieves the minimum variance performance, I_v approaches one. On the other hand, I_v close to zero indicates poor performance.

4. USER-DEFINED BENCHMARK

The calculation of minimum variance output y_{mv} , however, requires a priori knowledge of the plant and even the model of the system, which is not attractive to implement in practice. Therefore, a user-defined reference is chosen as the benchmark, and the generalized eigenvalue analysis is implemented. The user-defined reference can be a period of "golden" operation data from the process during which desirable control performance was achieved. It could be a period of operation data right after a new controller has been commissioned successfully. It could also used for rolling period monitoring, for instance, benchmarking the performance of the current week against that of last week. Denoting the benchmark data as period I and the monitored data as period II, the direction along which the variance inflation occurs the most is given by

$$p = \arg\ \max \frac{p^T cov(y_{II})p}{p^T cov(y_I)p} \tag{15}$$

The solution is the generalized eigenvector solution,

$$cov(y_{II})p = \mu cov(y_I)p \tag{16}$$

where μ is the generalized eigenvalue and p is the corresponding eigenvector. The direction p is referred to as the worst performance direction (WPD). In addition to the first generalized eigenvector, other subsequent eigenvectors with large enough eigenvalues (especially those much larger than 1) are also of remarkable suboptimality in control performance and should be examined to further improve the control performance.

Since the reference benchmark is not necessarily a minimum variance benchmark, there can be directions along which the monitored period II outperforms the benchmark period I. These directions correspond to the generalized eigenvalues that are significantly less than one, and the corresponding eigenvectors represent the directions with the smallest variance ratio of the monitored period over the benchmark period. These eigendirections constitute the subspace of improved performance over the benchmark. Trying to maintain the loop operations within this subspace will obviously benefit the process control performance.

It is also meaningful to assess the overall variability of the monitored period against the benchmark period by defining a volume-like performance index as follows,

$$I_v = \frac{\det\{cov(y_{II}(k))\}}{\det\{cov(y_I(k))\}} \tag{17}$$

This ratio, while greater than zero, can be greater than or less than one. If it is greater than one, the performance of the monitored is in general worse than the benchmark period and the worst performance directions of the monitored period should be examined. If, on the other hand, this index is significantly less than one, the directions corresponding to the smallest eigenvalues should be

examined to understand where the performance has improved. Denoting μ_i, for $i = 1, 2, \ldots, n_y$ as the generalized eigenvalues in descending order, the volume based index in (17) can be rewritten as

$$I_v = \prod_{i=1}^{n_y} \mu_i^{-1} \tag{18}$$

which is easy to calculate once the generalized eigenvalues are calculated.

5. CASE STUDY

Industrial operating data collected from the DCS system of a wood waste burning power boiler are used here as the example to examine and verify the applicability of the user-defined performance assessment approach. The data set is composed of sample points with the sampling time of five seconds and three subsets of process variables (PV), the corresponding set-point (SP) and controller outputs (OP), respectively. The data processing is applied to the controller error terms, i.e., PV-SP. All these data points are preprocessed by scaling to zero mean and unit variance in every loop. The detailed physical description for these loops is given in Table 1.

The covariance based monitoring is performed on a data set with 150,000 consecutive data points. Here the benchmark period I consists of the first 66,000 samples, while the period II containing 84,000 points is monitored with respect to the benchmark period. It is suspected that the period II has experiences some changes in the performance. The computation results from the proposed monitoring procedure are depicted in Fig.1. The upper-left subplot shows the maximal and minimal eigenvalues, while the lower-right one shows the full spectrum of eigenvalues and their cumulative percentage. It can be easily seen from the plot that the largest eigenvalue is far above one, which implies that the control performance of period II in this eigenvector direction is much worse than that of the benchmark. The loading score plots for the largest and smallest eigenvector directions are given in Fig.1(b) and (c), respectively. It is clear that the variable 4, i.e. loop FC0902, contributes most significantly in the first eigendirection. Thus we may conclude that the control performance of monitored period along the largest eigendirection, especially loop FC0902, deteriorated significantly. In other words, there exists a great margin to improve the performance by re-tuning along this direction as well as the loop FC0902. This can serve as an instructive tool for control engineers to maintain the control system. On the other hand, the smallest eigenvector stands for the direction of improved performance over the benchmark. Fig 1(c) shows that loops 5 and 3 have large contributions to the improved performance.

6. CONCLUDING REMARKS

MIMO control performance monitoring has enjoyed great development recently as it is one of the most important issues in practice after the control design. The minimum variance benchmark is usually considered a good starting point although it requires significant process information. For MIMO performance monitoring we demonstrate in this paper that covariance based monitoring is more appropriate when strong interactions occur among controlled variables. The covariance-based monitoring is extended to benchmarking any two covariance matrices and diagnosis of worse or better performance directions is developed.

Due to limited space several related issues could not be covered in this paper but they are important. One is the deterministic performance loss due to loop oscillations and the need for setpoint tracking. Another is the dual task of control performance monitoring and statistical process monitoring. The current situation is that both issues are studied assuming the other part is problem free. In practice the problems co-exist and only routine operation data are available to tell one problem from another. The integration of control performance monitoring and process monitoring deserves further study.

7. REFERENCES

Åström, K. J. (1976). State of the art and needs in process identification. In: *Proc. of Conf. Process Control-II, AIChE Symposium Series.* pp. 184–194.

Åström, Karl J. (1970). *Introduction to Stochastic Control Theory.* Academic Press. San Diego, California.

Desborough, Lane and Randy Miller (2002). Increasing customer value of industrial control performance monitoring – honeywell's experience. In: *Chemical Process Control - CPC VI.* CACHE. Tuscon, Arizona. pp. 169–189.

Harris, T., F. Boudreau and J.F. Macgregor (1996*a*). Performance assessment of multivariable feedback controllers. *Automatica* **32**(11), 1505–1518.

Harris, T. J. (1989). Assessment of control loop performance. *Can. J. Chem. Eng.* **67**(10), 856–861.

Harris, T. J. and C. T. Seppala (2002). Recent developments in controller performance monitoring and assessment techniques. In: *Chemical Process Control - CPC VI.* CACHE. Tuscon, Arizona. pp. 208–222.

Harris, T.J. (2004). Statistical properties of quadratic-type performance indices. *J. Proc. Cont.* **14**, 899–914.

Harris, T.J. and J.F. MacGregor (1987). Design of multivariable linear-quadratic controllers using transfer functions. *AIChE J.* **33**(9), 1481–1495.

Harris, T.J. and W. Yu (2003). Analysis of multivariable controllers using degree of freedom data. *Int. J. Adaptive Control and Signal Processing* **17**, 569–588.

Harris, T.J., C.T. Seppala and L.D. Desborough (1999). A review of performance monitoring and assessment techniques for univariate and multivariate control systems. *J. Proc. Cont.* **9**, 1–17.

Harris, T.J., C.T. Seppala, P.J. Jofriet and B.W. Surgenor (1996*b*). Plant-wide feedback control performance assessment using an expert system framework. *Control Engineering Practice* **4**(9), 1297–1303.

Hoo, K. A., M. J. Piovoso, P. D. Schnelle and D. A. Rowan (2003). Process and controller performance monitoring: overview with industrial applications. *Int. J. Adaptive Control and Signal Processing* **17**, 635–662.

Horch, A. and R. Isaksson (1999). A modified index for control performance assessment. *J. Proc. Cont.* **9**, 475–483.

Huang, B. (1997). Multivariate Statistical Methods For Control Loop Performance Assessment. PhD thesis. University of Alberta.

Huang, B. and S.L. Shah (1997). Feedback control performance assessment of non-minimum phase MIMO systems. In: *AIChE Annual Meeting*. Los Angeles.

Huang, B. and S.L. Shah (1998). Practical issues in multivariable feedback control performance assessment. *J. Proc. Cont.* **8**, 421–430.

Huang, B. and S.L. Shah (1999). *Performance Assessment of Control Loops: Theory and Applications*. Advances in Industrial Control. Springer-Verlag. London, Great Britain.

Huang, B., S.L. Shah and K.Y. Kwok (1997). Good, bad or optimal? performance assessment of MIMO processes. *Automatica* **33**(6), 1175–1183.

Huang, B., S.X. Ding and N. Thornhill (2005). Practical solutions to multivariate feedback control performance assessment problem: reduced a priori knowledge of interactor matrices. *J. Proc. Cont.* **15**, 573–583.

Jelali, M. (2006). An overview of control performance assessment technology and industrial applications. *Control Eng. Practice* **14**, 441 – C466.

Kadali, R. and B. Huang (2002). Controller performance analysis with lqg benchmark obtained under closed loop conditions. *ISA Transactions* **41**, 521–537.

Kashyap, R.L. and R.E. Nashburg (1974). Parameter estimation in multivariate stochastic difference equations. *IEEE Trans. Auto. Cont.,* **19**, 784.

Kozub, D.J. (1996). Controller performance monitoring and diagnosis: experiences and challenges. In: *Fifth Int. Conf. on Chemical Process Control* (J.C. Kantor, C.E. Garcia and B.C. Carnahan, Eds.). AIChE and CACHE. Tahoe, CA. pp. 83–96.

Loquasto, F. and D. Seborg (2003). Model predictive controller monitoring based on pattern classification and pca. In: *Proc. of ACC*. Vol. 3. pp. 1968 – 1973.

MacGregor, J.F. and T. Kourti (1995). Statistical process control of multivariate processes. *Control Engineering Practice* **3**(3), 403–414.

Mayne, D.Q. and F. Firoozan (1982). Linear identification of arma processes. *Automatica,* **18**, 461–466.

McNabb, C. A. and S. J. Qin (2003). Projection based MIMO control performance monitoring – I. Covariance monitoring in state space. *J. Proc. Cont.* 739-759.

McNabb, C. A. and S. Joe Qin (2005). Projection based MIMO control performance monitoring – II. Measured disturbances. *J. Proc. Cont.* **15**, 89–102.

Miller, R., L. Desborough and C. Timmons (1998). Citgo's experience with controller performance assessment. In: *NPRA 1998 Computer Conference*. San Antonio, Texas.

Paulonis, M.A. and J.W. Cox (2003). A practical approach for large-scale controller performance assessment, diagnosis, and improvement. *J. Proc. Cont.* **13**, 155–168.

Peng, Youbin and Michel Kinnaert (1992). Explicit solution to the singular LQ regulation problem. *IEEE Trans. Auto. Cont.* **37**(5), 633–636.

Perrier, M. and A. Roche (1992). Towards mill-wide evaluation of control loop performance. In: *Control Systems '92*. Whistler, British Columbia.

Qin, S. J. (2003). Statistical process monitoring: Basics and beyond. *J. Chemomotrics* **17**, 480–502.

Qin, S.J. (1998). Control performance monitoring – a review and assessment. *Comput. Chem. Eng.* **23**, 178–186.

Schaffer, J. and A. Cinar (2004). Multivariable mpc system performance assessment, monitoring, and diagnosis. *J. Proc. Cont.* **14**, 113–129.

Seem, J. E. (1998). A new pattern recognition adaptive controller with application to hvac systems. *Automatica,* **34**, 969–982.

Seem, J. E. (2006). An improved pattern recognition adaptive controller. In: *Proceedings of*

the 2006 American Control Conference. Minneapolis, MN.

Seppala, C.T., T.J. Harris and D.W. Bacon (2002). Time series methods for dynamic analysis of multiple controlled variables. *J. Proc. Cont.* **12**, 257–276.

Thornhill, N.F., M. Oettinger and P. Fedenczuk (1999). Refinery-wide control loop performance assessment. *J. Proc. Cont.* **9**, 109–124.

Weinstein, B. (1992). A sequential approach to the evaluation and optimization of control system performance. In: *1992 ACC*. Chicago. pp. 2354–2358.

Table 1. The name tag and description of ten loops from a power boiler unit

Variable No.	Loop Identification	Description
1	FC0400	PB feed water flow control
2	FC0618	Oil burner air flow control
3	FC0620	Bark-air flow control
4	FC0902	Bark feed rate control
5	FC0922	Bark air firing control
6	LC0403	PB drum level control
7	PC0603	Combustion air pressure
8	PC0609	Furnace pressure control
9	PC0622	Over-fire air pressure
10	PC0904	Steam head pressure control

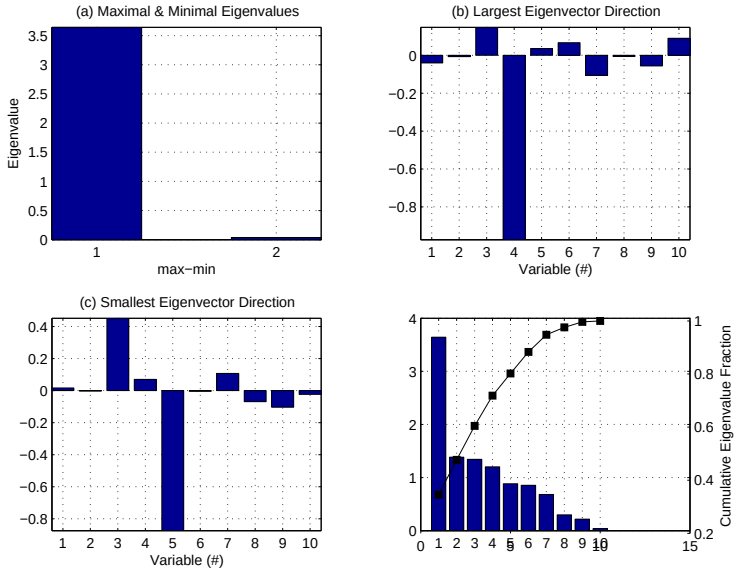


Fig. 1. Generalized eigen-analysis results for the period II against the user-defined benchmark period I with (a) the maximal and minimal eigenvalues; (b) the eigenvector direction corresponding to the maximum eigenvalue; (c) the eigenvector direction corresponding to the minimum eigenvalue; (d) the eigenvalue spectrum and the corresponding cumulative fractions.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PSE Relevant Issues in Semiconductor Manufacturing: Application to Rapid Thermal Processing

Cheng-Ching Yu[1†], An-Jhih Su[1], Jyh-Cheng Jeng[1], H. P. Huang[1], Shih-Yu Hung[2] and Ching-Kong Chao[2]

Dept. of Chem. Eng., National Taiwan University.[1], Taipei 106, TAIWAN
Dept. of Chem. Eng., National Taiwan University of Sci. Tech.[2], Taipei 106, TAIWAN

Abstract: The quality control of the wafer is becoming more and more important as the wafer becomes larger and the feature size shrinks. An advanced IC fabrication process consists of 300+ steps with scarce and usually difficult quality measurements. Thus product yield may not be realized until months into production while in-line measurements are available on the order of a millisecond. The series production nature and measurement setup lead to a unique process control problem. In this work, typical disturbances are explained and possibility for inferential control is explored. This leads to a control architecture with multiple layers in a cascade structure. Next, rapid thermal processing (RTP) is used to illustrate recipe generation and control structure design at the tool level. The resultant multivariable controller gives satisfactory setpoint tracking for a triangular-like temperature program. In order to reduce downtime, process trend monitoring of a tool is essential. Instead of using entire batch data, a key process variable is identified and an index is computed to capture the dynamic behavior of the tool. An RTP example is used to illustrate this approach and results clearly indicate that process trend is well predicted using the index-based time-series model.

1. INTRODUCTION

The continuing miniaturization of integrated circuit (IC) components and the increasing numbers of functions and performance of a single integrated circuit (IC) chip are the trend in the semiconductor industry. The quality control of the wafer is becoming more and more important as the wafer becomes larger (from 200 mm to 300 mm) and the feature size shrinks (from 350 nm to 90 nm). On the corporate level, improved yield is the only solution to remain competitiveness. Thus advanced equipment control and advanced process control (AEC/APC) have become a standard practice in modern semiconductor manufacturing. Edgar et al. (2000) give a comprehensive review in the processes and control issues, Qin et al. (2004) discuss the challenges in the IC industries, and Lewin et al. (2005) explore PSE related issues in IC fabrication. Contrary to general understanding in chemical process industries (CPI), the AEC is generally concerned with keeping the equipment (unit operation in CPI terminology) in working condition and, in so doing, prolonging the time between maintenance and reducing unscheduled downtime. So, the AEC is synonymous with fault detection and classification (FDC) for the individual equipment. However, unlike chemical processes, an advanced IC fabrication process may include 300 steps (or process units), and success in a single step (equipment) certainly does not guarantee an acceptable wafer. The APC addresses the control issue from one step to another. Thus, feedforward (FF) and feedback (FB) control becomes important. The run-to-run (R2R) control is the typical element in the feedback loop, and controllers are integral-only (I-only) or double integrator (PI^2). They are generally termed exponentially weighted moving average (EWMA) and double EWMA algorithms. In chemical process control (CPC) terminology, the AEC can be viewed as the within batch control and fault detection and the APC is similar to batch-to-batch process control. The controllers used rarely go beyond PID types. One may wonder: "Why does such a hi-tech industry use seemingly low-tech control methodology?" The answer is quite simple: "We cannot fix (control) what we cannot detect (measure)." (Wang, 2004) However, the endeavor for yield improvement via improved process control can be seen throughout fabs worldwide. Currently, the AEC/APC symposium (Wang, 2004; Edgar, 2004; Wu et al., 2005) is held in the USA, Europe, and Asia each year with hundreds of attendees to each conference, and they have become the major events for APC division personnel from fabs worldwide. In fact, this is similar to the process control phenomena we witnessed in CPI 20 years ago. However, the approaches taken in the IC industries are quite different from those of the CPI for the following reasons: (1) scarce and sometimes difficult quality measurements, (2) multiple and iterative processing steps, (3) non-straightforward links between processing steps and product specification (e.g., in terms of IC design), and (4) frequent tool

† to whom all correspondence should be addressed.
E-mail: ccyu@ntu.edu.tw
Fax: +886-2-2362-3040

maintenance. In this paper, the process characteristics in IC fabrication are explained in Section 2 and opportunities in process control are explored. In Section 3, a specific tool, rapid thermal processing (RTP), is used to illustrate the tool level control problems. RTP is employed for various single-wafer thermal treatment processes including annealing, oxidation, cleaning, and chemical vapor deposition (Campbell and Knutson, 1992; Huang et al., 2000a,b,c; Chao et al., 2003a,b; Jung et al., 2003; Gunawan et al., 2004). The preventive maintenance problem is studied in Section 4 via an industrial example followed by the conclusion.

2. PROCESS CHARACTERISTIC

2.1 Disturbances

Similar to chemical process control, disturbance rejection is the major concern in semiconductor manufacturing. By disturbance rejection, we mean maintaining the product quality in the face of process changes. Typical sources of process variations in IC fabrication include: (1) tool-induced disturbances which are generally known as process drift and/or process shift, (2) product-induced disturbance which typically comes from the IC foundry where high-mix products are manufactured, and (3) incoming disturbances which are often referred to as the variations which are a direct consequence of proceeding processing steps (Patel et al., 2000; Chen et al., 2005). Generally, some prior knowledge about the quality of the *incoming* wafers is available in semiconductor manufacturing processes. Thus, feedforward control or feed sequence arrangement can be devised to mitigate the incoming disturbance (Chen et al., 2005). A similar approach can be applied to the product-induced disturbance. The tool-induced disturbance is less frequently seen in chemical process control. Nano-scale-based operation generally requires an ultra-clean environment. A small contamination may lead to degraded tool performance. Thus, we have seen almost weekly-based maintenance in fabs as opposed to yearly-based maintenance in chemical plants. It is never the less essential to maintain product quality under gradual degradation using feedback control (Chen and Guo, 2001).

2.2 Measurement

The product nature of IC makes the quality measurement difficult, if not impossible. Unlike the product purity specification in chemical production, the product yield cannot be realized until the end of some 300 processing steps. This implies we may not realize the yield until a *month* into production. The electrical performance of a wafer (die to be specific) cannot be tested till the end of the iteration for each metal layer. The electrical performance of a wafer is generally referred to as the wafer acceptance test (WAT) and the test results are available in the time-scale of a *week* (Fan et al., 2000). The product yield is usually highly correlated to the WAT data. Generally, after each processing step, we have a quality measurement which is often denoted as the "metrology." Nano-scale nature makes the measurement (metrology) difficult and the measuring station (metrology tool) expensive. The cost of a typical metrology tool is in the range of millions of dollars. This leads to a very different measurement setup as compared to chemical plants. That is: the metrology tool is *shared* by similar processing steps and only few of the wafers (1-4 wafers from each lot) are measured. The time-scale for a metrology measurement is in the order of hours to one day. This may result in delay problem if feedback control is installed. Typical metrology measurements include: thickness, resistance, critical dimension (CD), overlay, particles, etch rate etc. Down to the tool level, we have the in-line measurements such as temperature, pressure, flow, current, etc. which are measured in the order of *milli-second to second*. Thus, quality/process variables are available on drastically different time scales and, obviously, the measurement complexity increases as one goes from the tool level to the product level (Figure 1).

2.3 Control Architecture

The ultimate goal of IC production is to improve the yield and, as pointed out earlier, process control is a means to achieve this. However, the process measurement setup in Fig. 1 reveals that effective control cannot be obtained without some type of inferential control (soft sensor in chemical engineering literature). The quality estimation can be further arranged into two tiers. One is at the tool level and the estimator is denoted as *virtual metrology*. The other is at the product level which is generally called *virtual WAT* (Wu et al., 2005) Quality estimation is not unfamiliar to the chemical engineering community and it is often used to estimate product composition in a distillation column, molecular weight distribution in a polymerization reactor etc. with certain degree of success. For example, in distillation, the relationship between product composition and tray temperature is governed by the thermodynamic equilibrium. Thus, a strong correlation between tray temperatures and composition can be established. However, the relationship between in-line measurements (e.g., temperature) and quality variable (e.g., sheet resistance) in semiconductor manufacturing is less obvious, especially when the tool is operated in a batch mode. A successful virtual metrology model relies on identifying key tool indices from the entire batch data. At the product quality level, few attempts have been made to relate end-of-line electrical properties to the metrology data over the entire process (Fan et al., 2000; Wu et al., 2005). Figure 2 shows how the virtual metrology (VM) and virtual WAT (V-WAT) can be incorporated into the control architecture for improved yield management. Here, the estimated quality variable is maintained by changing the recipe (e.g., temperature set point) while

the metrology model is updated when metrology data become available (e.g., via Kalman filtering). The electrical properties of a wafer can also be estimated at the completion of several processing steps using the virtual WAT. The electrical properties of the product are controlled by adjusting metrology set points which subsequently affect the recipes in related tools. Figure 3 gives a detailed description of the control architecture for product quality control. It is clear that quality estimation (VM and V-WAT) plays a vital role in this framework. The series nature of the process flow leads to a feedforward/feedback (FF/FB) structure from a tool perspective provided with multiple layers of cascade control.

3. CONTROL OF RTP

Typically, wafer processing in a tool is described by a recipe which consists of on the order of ten steps. These steps include: warm-up, temperature program, flow manipulation, cool-down etc. Generally, very simple feedback control is used to ensure successful execution of the recipe. We will use rapid thermal processing (RTP) to illustrate the tool level control.

3.1 Process

RTP is an effective tool for various single-wafer thermal treatment processes. It permits processes to be accomplished with minimal dopant redistribution and uniform deposition quality with a smaller thermal budget. However, poor RTP system design can lead to significant temperature differences in the wafer. One of the main shortcoming that RTP must overcome is that of heating (or cooling) the wafers non-uniformly which results in material failure due to an increases in thermal stresses or serious warpage. The damage due to the presence of thermal stresses can represent a limit on the applicability of rapid thermal processing.

The temperature non-uniformity in the wafer is caused by three factors: edge effect, pattern effect, and heat source. The higher heat loss from the wafer edge has been found to result in a radial temperature gradient in the wafer. To improve the wafer temperature non-uniformity produced by the edge effect, several radiative shields can be placed at the edge of the wafer to reduce the heat loss from the wafer edge and reflect the radiative energy back into the wafer during the cooling process. By varying the angle of the shield, an optimal shield configuration can be found to minimize the induced thermal stress (Young and McDonald, 1990). Hebb and Jensen (1998) show that pattern-induced temperature non-uniformity can cause plastic deformation during a RTP cycle and the problem is exacerbated by single-side heating, increased processing temperature and ramp rate. Design and control of RTP to improve temperature uniformity was explored by Huang et al. (2000a,b,c).

A cross-sectional view of the furnace and wafer is shown in Fig. 4. A bank of tungsten halogen lamps provides the thermal radiative energy to the single silicon wafer through a transparent quartz window. Since quartz does not absorb light efficiently within the wavelength band of the lamps, it can be neglected in the thermal system. Let us assume the wafer is 200 mm in diameter held by three quartz pins and enclosed in a cylindrical chamber, where the chamber is axis-symmetric in geometry (Chao et al., 2003a,b). The chamber geometry is described in Huang et al (2000a).

3.2 Recipe Generation

The essential step in the RTP recipe, in addition to preparation steps, is the temperature program. Two types of temperature programs are often used in RTP: soak and spike temperature profiles. Consider the spike annealing of rapid thermal annealing (RTA). The post-implant annealing uses a lamp-based RTA with temperature programs shown in Fig. 5. As pointed out by Jung et al. (2003), the ion-implantation technology is limited in part by transient enhanced diffusion (TED) of dopants during RTA, often leading to significant spreading of the dopant profile. This may lead to defects in extremely shallow pn junctions in electronic devices. Considerable efforts have been put forth to design a temperature program to produce the desired junction depth while maintaining low sheet resistance (Gunawan et al., 2004). A different approach is taken here. We will use the spike annealing to illustrate thermal-stress-based temperature program generation with emphasis on the cooling curve.

Consider the RTP system shown in Fig. 4. The wafer thickness is assumed to be thin as compared to the radius of the wafer r_o, so we can regard this as a one-dimensional plane-stress problem, that is, the temperature T is dependent on r only. The partial differential equations of the present thermoelastic problem can be written as (Nowinski, 1978):

$$k\left(\frac{1}{r}\frac{\partial T}{\partial r}+\frac{\partial^2 T}{\partial r^2}\right)-q^{rad}-q^{conv}=\rho C_p\frac{\partial T}{\partial t} \qquad (1)$$

with boundary conditions given by

$$\frac{\partial T}{\partial r}=0\,, \quad \text{at r} = 0 \qquad (2)$$

$$-k\frac{\partial T}{\partial r}=q_{edge}\,, \quad \text{at r} = r_o \qquad (3)$$

where ρ, C_p and k are the density, specific heat capacity and thermal conductivity of silicon, respectively. q^{rad} and q^{conv} represent the radiative and convective heat flux leaving a wafer surface per unit volume, respectively. The quantity q_{edge} is the heat flux at the wafer edge that includes the heat loss of convection and radiation.

Once the temperature profile has been obtained, the components of stresses are obtained as:

$$\sigma_{rr} = \alpha E\left(\frac{1}{r_o^2}\int_0^{r_o} T(\eta)\eta d\eta - \frac{1}{r^2}\int_0^{r} T(\eta)\eta d\eta \right) \quad (4)$$

$$\sigma_{\theta\theta} = \alpha E\left(-T + \frac{1}{r_o^2}\int_0^{r_o} T(\eta)\eta d\eta + \frac{1}{r^2}\int_0^{r} T(\eta)\eta d\eta \right) \quad (5)$$

$$\sigma_{r\theta} = 0 \quad (6)$$

where σ_{rr} and $\sigma_{\theta\theta}$ are the radial and tangential stress components, respectively. α and E denote the linear thermal expansion coefficient and Young's modulus, respectively. Since the obtained temperature profile is expressed in a discrete manner, the stresses in Eqs (4) and (5) are determined by a trapezoidal integration technique.

In the present study, the maximum shear stress failure criterion is used which assumes that the wafer fails in shear when

$$S = \frac{\tau_{\max} \cdot F_s}{\tau_{yp}} > 1 \quad (7)$$

where S is the normalized maximum resolved stress, F_S is the safety factor which is usually taken to be 2 and the maximum shear stress is calculated using Mohr's circle as:

$$\tau_{\max} = \frac{1}{2}|\sigma_{rr} - \sigma_{\theta\theta}| \quad (8)$$

At high temperature, silicon behaves like a viscous material. The yield stress in shear can be expressed in terms of the temperature and the maximum shear stress rate (Hebb and Jensen, 1998) as:

$$\tau_{yp} = 23.17\exp(16.1 - 0.00916T)\left(\frac{d\tau}{dt}\right)^{0.4} \quad (9)$$

where the stress unit is in Pascal and the temperature unit is in degree Celsius. The stress rate $d\tau/dt$ is taken to be the larger of 2.5×10^5 Pa/s or its calculated value. If the result calculated from Equation (9) exceeds 3.1×10^8 Pa, it is taken to be 3.1×10^8 Pa which means that the wafer is at low temperature. From Equation (9) we know that the yield shear stress will be about 1.5 MPa when T= 1200°C at the beginning of the cooling process which is far less than 310 MPa at the room temperature T= 27°C. This simply indicates that, according to the failure criterion stated in Equation (7), a small temperature non-uniformity may induce material failure at high temperature. Since no analytical solution is available for the present problem, the numerical solutions are sought to the above governing equations. The calculation is carried out using a fully implicit finite difference method (Chao et al., 2003a).

Three scenarios are considered using the lamps radiative cooling condition: (1) fixed temperature-difference control scheme: The maximum temperature difference within a wafer is fixed to 0.7°C (by trial and error such that the normalized maximum resolved stress is less than one during the cooling process), (2) constant cooling-rate control scheme: The lamp's power decreases gradually at a constant rate of 10KW/m^2-s (by trial and error which ensures that the normalized maximum resolved stress is less than one during the cooling process), (3) maximum stress control scheme: The normalized maximum resolved stress is kept close to one until the lamp's power decreases to zero during the cooling process.

Chao et al. (2003a) show that the edge heat loss leads to large temperature gradient toward the wafer edge. Based on the maximum shear stress failure criterion, the results show that material failure always occurs at the edge of the wafer at the beginning of cooling processes. Furthermore, the maximum stress control scheme is shown to be more efficient because it can significantly reduce the required cooling time and thermal budgets. Thus, the conventional constant cooling-rate control scheme or linear temperature ramp-down scheme is not appropriate for the rapid thermal processor.

Fig. 6 shows, for the radiative-only cooling process, the tangential stress at the wafer edge is positive due to thermal shrinkage induced by the edge effect. On the other hand, the compressive tangential stress prevails at the central region of wafer. Since the tangential stress at the central region is far less than the tangential stress at the wafer edge. The wafer failure is dominated by the edge effect in the wafer and yield stress in shear. For the maximum stress control scheme, the lamp's power decreases dramatically during the cooling process. After five seconds have elapsed, the lamp's power for the fixed temperature-difference control scheme decreases gradually with a rate even smaller than the constant cooling-rate control scheme. The required cooling time for the maximum stress control scheme is only 18 sec from 1200°C to 600°C, compared to 30 sec for the constant cooling-rate control scheme, and, moreover, it is only one fifth of the required time for the constant temperature-difference scheme as shown in Fig. 7. This provides an attractive alternative for temperature program generation.

3.3 Control Structure Design

The state-of-the-art RTP typically consists of 7 lamp-heating zones with 7 temperature measurements, in addition to computed emissivity. Here we use a simple RTP model (Huang et al., 2000a) to illustrate the essential steps in the control structure design. This is an RTP system with 3 lamp-heating zones for a 200mm wafer. Once a temperature program becomes available (Fig. 5A), the design procedure consists of the following steps: (1) selection of temperature measurements, (2) controller design, and, possibly, (3) temperature program modification. Spike annealing is considered here. The control objective is to maintain temperature uniformity, especially around the peak temperature. The focus of the program is the temperature range of 1000°C-1050°C with the duration of approximately 2 seconds.

The temperature profile along the radial position plays an important role for the measurement selection. The RTP system uses a linear combination of *three*

lamp powers to match the desired intensity. Notice that each lamp ring has an intensity profile similar to the normal distribution (e.g., Fig. 5). The optimal temperature uniformity corresponds to a unique lamp power combination. The desired temperature profile is a nonlinear function in r and it crosses the temperature set point several times. The profile is similar to a high-order polynomial: $T - T^{set} = \prod(r - z_i)$ where T^{set} is the temperature set point, n is the number of set point crossings and z_i denotes the location of the set point crossing (zero of the polynomial). Therefore, it becomes clear that the best temperature uniformity that can be achieved is the temperature profile minimizing the squares of temperature differences which is termed the *desired* temperature profile. Furthermore, the easiest way to maintain this profile is to keep the temperatures already at (or close to) set point (e.g., Fig. 5) under control. This can be interpreted as retaining the shape of the temperature profile by holding several key positions at the set point. If we have more zero-crossing temperatures than manipulated inputs, The next step is to check system interaction and inherent robustness using the structured singular value (SSV). Therefore, the temperature measurement selection criterion can be summarized as follows (Huang et al., 2000c).

1. Identity the set point crossing locations for the desired temperature profile.
2. Prefer the approximately equal-spaced rule for placing temperature measurements on these locations.
3. Check for system robustness, and if the SSV is not acceptable, go back to step 2.

The procedure suggests control of T_3, T_{17} , and T_{29} out of 30 zones in the radial position.

Once the control structure is determined, the next step is to design a multivariable temperature controller. The conventional PID controller is preferred for its simplicity and transparency. Because almost half of the batch cycle involves ramp-type setpoint trajectory, the IMC design principle of Morari and Zafiriou (1989) is employed (Huang et al., 2000a) and Type-2 system is considered. For the RTP operated at 1050ºC, the model gives the following process transfer function matrix: Note that the sampling rate (0.01 s) is so fast that, a continuous-time model is used here.

$$G(s) = K \cdot \mathrm{diag}(1/\ (\tau_i s + 1)) \tag{10}$$

where K is the steady-state gain matrix and τ_i is the time constant. Following the design procedure of Huang et al. (2000a), it leads to a diagonal PID type of controller with a static decoupler. Moreover, the diagonal controller has *double* integrators.

$$C(s) = K^{-1}\mathrm{diag}(K_{ii}) \tag{11}$$

where K_{ii} is the diagonal PID type of controller.

$$K_{ii} = K_{c,i}(1 + \frac{1}{\tau_{I,i}s} + \tau_{D,i}s)\frac{1}{s} \tag{12}$$

We term this type of controller as PI^{2}D controller hereafter. The controller parameters can be expressed in terms of IMC filter time constant τ_f.

$$K_{c,i} = \frac{\tau_i + 2\tau_f}{\tau_f^2},\ \tau_{I,i} = \tau_i + 2\tau_f,\ \ \tau_{D,i} = \frac{2\tau_i\tau_f}{\tau_i + 2\tau_f} \tag{13}$$

Therefore, once the closed-loop time constant τ_f is set, the tuning constants for the PI^{2}D controller can be determined immediately.

Figure 8 clearly indicates the advantage of PI^{2}D control, derived from type-2 disturbance, over PI control, derived from type-1 disturbance, in which significant offsets are observed in ramp-up and ramp-down periods. Moreover, the two important criteria, peak temperature and duration time over 1000ºC, are completely missed, even with PI^{2}D control. Table 1 summarized the spread of the peak temperature and duration time.

If the peak temperature tracking and duration is the design criteria, the triangular temperature problem in Fig. 5A cannot be achieved with a realizable controller. Thus, a smooth temperature program is used instead as shown in Fig. 5B. The tabulated results in Table 1 also confirm this and the peak temperature spread is reduced to 6.9ºC as compared to 11.8ºC for triangular temperature program. Figure 9 shows the peak-temperature spread across the radial positions is reduced to 6.9ºC using the smooth temperature program for the RTP with 3 heating zones. The trend remains for wafer with different peak temperatures. The results presented here clearly indicate that the advanced control methodology can certainly be applied to semiconductor manufacturing at the tool level.

4. PROCESS MONITORING

Process monitoring and analysis is important in semiconductor manufacturing. Correct trend monitoring can be used to determine appropriate timing for preventive maintenance. In this work, instead of incorporating large number of trajectory data with variable batch time and possibly "missing" data for some process variables using multivariable statistic technique (e.g. MPCA), a key sensitive index (KSI) based approach is proposed for batch process trend monitoring. From process insight or the experience of the process operator, a certain period time within a batch time where the measurements have significant effect on product quality, the key sensitive time-slot (KST), is identified. Next, based on the KST, possible key sensitive process variables (KSV) are chosen. The KSV may not be the measured values themselves in KST, but some quantity, such as area, slope, maximum, etc., computed from the raw measurements. Once a KSV is computed for each batch (wafer-to-wafer) under normal operation, its autocorrelation function is calculated as the batch process progresses. If significant autocorrelation is found, a time-series model is established for the selected KSV, if not, a different KSV is sought. With the time-series model, the process trend can thus be forecasted and then an index for the process operating status (key sensitive index, KSI) is defined and computed. By monitoring the KSI, possible

maintenance action can therefore be called for, whenever necessary. This provides dynamical capability for process trend monitoring while maintaining the simplicity of single-variate analysis. An IC processing example is used to illustrate the KSI-based approach.

In the manufacturing of semiconductor, IC is processed through the recipes which comprise a sequence of different treatments (steps). In general, only some steps are critically related to the product quality so that the processing intervals corresponding to these critical steps are the aforementioned KST. In this example, the recipe comprises 11 steps where the processing time from step 6 to step 10 is identified as KST. Then, three important process variables are selected as possible KSV. From correlation analysis, only the maximum of one variable (say, variable A) in KST shows significant autocorrelation and, hence, this maximum value, $A_{\max}$, is chosen as KSV. However, as shown in Fig. 10(a), $A_{\max}$ for some batches are abnormally greater than the average value. Since different products are usually processed with the same tool, Amax with particularly high values may result from different products. Thus, one product index, as shown in Fig. 10(b), is considered for the modification of $A_{\max}$ values. The result of modified $A_{\max}$, designated as $A'_{\max}$, is shown in Fig. 10(c) where all $A'_{\max}$ values follow the data trend. Consequently, an autoregressive moving average (ARMA) model of the following is built for $A'_{\max}$ based on measurements from 500 wafers.

$$\left(1-1.744\,q^{-1}+0.776\,q^{-2}\right)A'_{\max}(t) = \left(1-1.346\,q^{-1}+0.476\,q^{-2}\right)e(t) \quad (14)$$

where q^{-1} is the backward shift operator and $e(t)$ is white noise. It is found that one root of the autoregressive polynomial is close to unity, which means the time series $A_{\max}(t)$ exhibits nonstationary behavior. For this reason, an autoregressive integrated moving average (ARIMA) model is then built to describe this behavior.

$$\left(1+0.942\,q^{-1}\right)\nabla A'_{\max}(t) = \left(1+0.452\,q^{-1}-0.553\,q^{-2}\right)e(t) \quad (15)$$

where $\nabla=\left(1-q^{-1}\right)$ These two time-series models are then used for forecasting the values of $A'_{\max}$ as the batch process progresses. The result is shown in Fig. 11 where two abrupt changes are observed due to scheduled tool maintenance (PM). Initially, both the forecasts of ARMA and ARIMA models can follow the process trend well. However, as the batch process progresses, the forecast of ARMA model starts to deviate from the actual $A'_{\max}$ more and more, while the forecast of ARIMA model keeps following the actual process. This phenomenon disappears after PM and then can be observed again as the batch process progresses. In order to capture the drifting behavior of this batch process, the KSI is thus defined as the absolute value of difference between residuals of these two models.

$$\text{KSI}=\left|\text{Residual}_{\text{ARMA}}-\text{Residual}_{\text{ARIMA}}\right| \quad (16)$$

The computed KSI is shown in Fig. 12. The results clearly indicate that the process trend can be realized using the proposed KSI and tool maintenance is required once this KSI is greater than a prescribed limit. Therefore, this KSI-based approach not only can be used for batch process trend monitoring, but also it is helpful for the engineers to decide when to call for tool maintenance.

5. CONCLUSION

An advanced IC fabrication consists of 300+ steps with scarce and usually difficult quality measurements. The series production nature and measurement setup lead to a unique process control problem. In this work, typical disturbances in semiconductor manufacturing are explained and the necessity of quality estimation is outlined. This leads to a control architecture with multiple layers in cascade structure. Next, RTP is used to illustrate recipe generation and control structure design at the tool level. The resultant multivariable controller gives satisfactory setpoint tracking for a triangular-like temperature program. In order to prolong the time between maintenance and to reduce unscheduled downtime, process trend monitoring of a tool is essential. Instead of using entire batch data, key process variable is identified and an index is computed to capture dynamic behavior of the tool. An IC processing example is used to illustrate this approach and results clearly indicate that process trend is well predicted using the index-based time-series model.

ACKNOWLEDGMENT

This work was supported in part by the National Science Council of Taiwan. We would like to thank Sunny Wu, B. H. Chen, J. S. Lin, Henry Lo, Jean Wang, C. H. Yu, and M. S. Liang of TSMC for continuing support and discussion. Long-term collaboration with Walters Shen of AMAT is also gratefully acknowledged. We also thank anonymous reviewers for thoughtful comments.

REFERENCES

Chao, C. K.; Hung, S. Y.; Yu, C. C. (2003a). Thermal Stress Analysis for Rapid Thermal Processor, *IEEE Trans. Semi. Manuf.* 13, 335.

Chao, C. K.; Hung, S. Y.; Yu, C. C. (2003b). Effect of Lamp Radius on Thermal Stresses for Rapid Thermal Processing System, *ASME J. Manufac. Sci. Eng.*, 125, 504.

Chen, A.; Guo, R. S. (2001). Age-based double EWMA controller and its application to CMP processes, *IEEE Trans. Semi. Manuf.*, 14, 11.

Chen, Y. H.; Shiu, S. J.; Yu, C. C.; Shen, S. H. (2005). Batch Sequencing for Run-to-Run Control: Application to Chemical Mechanical Polishing, *Ind. Eng. Chem. Research*, 44, 4676.

Edgar, T. F. (2004). Multi-product Run-to-Run Control for High-Mix Fabs, *AEC/APC Symposium Asia,* HsinChu, Dec.

Edgar, T. F.; Butler, S. W.; Campbell, W. J.; Pfeiffer, C.; Bode, C.; Hwang, S. B.; Balakrishnan, K. S.; Hahn, J. (2000). Automatic control of microelectronics manufacturing: practices, challenges and possibilities, *Automatica*, 36, 1567.

Fan, C. M.; Guo, R. S.; Chang, S. C.; Wei, C. S. (2000). SHEWMA: An End-of-Line SPC Scheme Using Wafer Acceptance Test Data, *IEEE Trans. Semi. Manuf.*, 13, 344.

Gunawan, R.; Jung, M. Y. L.; Seebauer, E. G.; Braatz, R. D. (2004) Optimal Control of Rapid Thermal Annealing in a Semiconductor Process, *J. Process Control,* 14, 423.

Hebb, J. P.; Jensen, K. F. (1998). The Effect of Patterns on Thermal Stress during Rapid Thermal Processing of Silicon Wafers, *IEEE Trans Semi. Manuf.*, 11, 99.

Huang, C. J.; Yu, C. C.; Shen, S. H. (2000a). Selection of Measurement Location for the Control of Rapid Thermal Processor, *Automatica*, 36, 705.

Huang, C. J.; Yu, C. C.; Shen, S. H. (2000b). Identification and Nonlinear Control for Rapid Thermal Processor, *J. Chin. Inst. Chem. Eng.*, 31, 585.

Huang, I.; Liu, H. H.; Yu, C. C. (2000c). Design for Control: Temperature Uniformity in Rapid Thermal Processor, *Korean J. Chem. Eng.*, 17, 111.

Jung, M. Y.; Gunawan, R.; Braatz, R. D.; Seebauer, E. G. (2003). Ramp-Rate Effects on Transient Enhanced Diffusion and Dopant Activation, *J. Electrochem. Soc.*, 150, G838.

Lewin, D. R.; Lachman-Shalem, S.; Grosman, B. (2005). More Process System Engineering (PSE) Applications in IC Manufacturing, I*FAC World Congress*, Prague, July.

Morari, M.; Zafiriou, E. (1989). *Robust Process Control*. Prentice-Hall, Englewood Cliff.

Nowinski, J. L. (1978). *Theory of Thermoelasticity with Application.* Sijthoff & Noordhoff.

Patel, N. S.; Miller, G. A.; Guinn, C.; Jenkins, S. T. (2000). Device dependent control of chemical-mechanical polishing of dielectric films, IEEE Trans. Semi. Manuf., 13, 331.

Qin, S. J.; Cherry, G.; Good., R.; Wang, J.; Harrison, C. A. (2004). Control and Monitoring of Semiconductor Manufacturing Processes: Challenges and Opportunities", *DYCOPS-7*, Boston, July.

Wang, T. (2004). Advanced Process Control Road Map and Challenges, *AEC/APC Symposium Asia*, HsinChu, Dec.

Wu, S.; Chen, P. H.; Lin, J. S.; Ko, F.; Lo, H.; Wang, J.; Yu, C. H.; Liang, M. S. (2005). Real-Time Device Performance Prediction for 90nm and Beyond, *AEC/APC Symposium USA*, Palm Spring, CA, Sept.

Young, G. L.; McDonald, K. A. (1990). Effect of Radiation Shield Angle on Temperature and Stress Profiles During Rapid Thermal Annealing, *IEEE Trans Semi. Manuf.,* 3, 176.

Table 1. Control performance of different types of temperature programs

	triangular	Smooth
Mean of peak temp. (°C)	1066.1	1056.9
Range of peak temp. (°C)	11.8	6.9
Std. dev. of peak temp(°C).	3.8	2.2
Mean of duration (s)	2.08	2.08
Range of duration (s)	0.155	0.086
Std. dev. of duration (s)	0.039	0.027

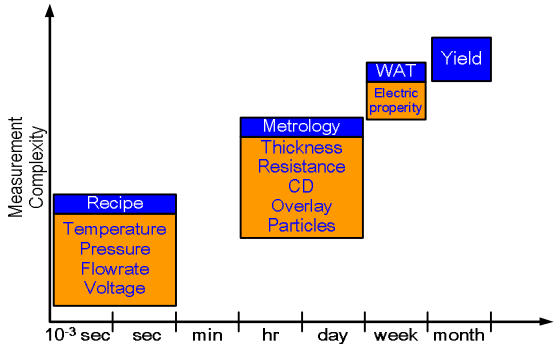


Figure 1. Measurement complexity and frequency

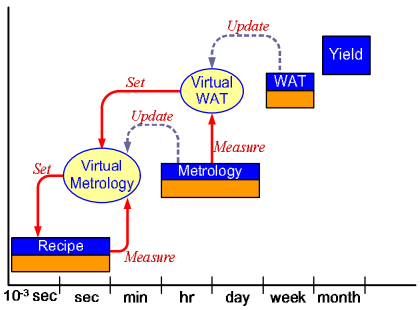


Figure 2. Structure of control action

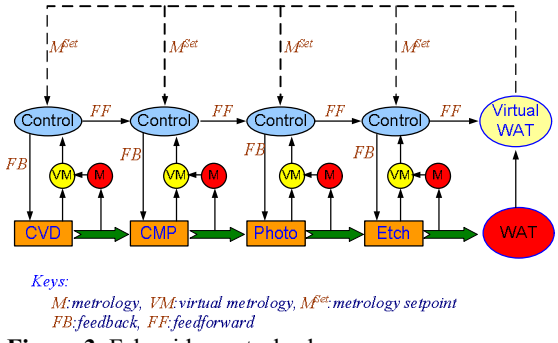


Figure 3. Fab-wide control schema

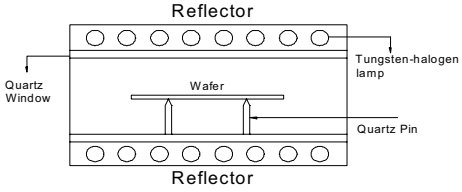


Figure 4. The physical model of RTP

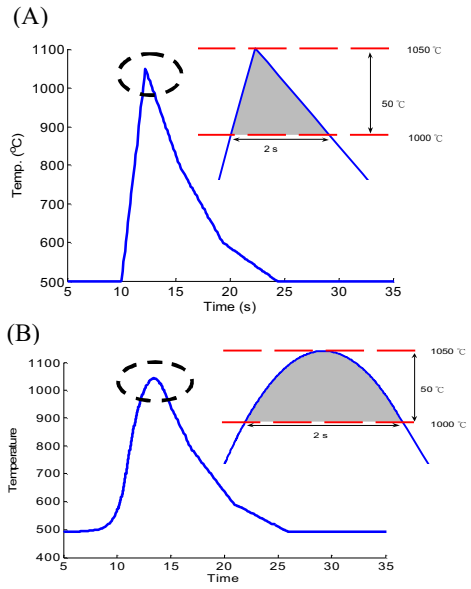


Figure 5. (A) triangular-like temperature program (B)smooth temperature program.

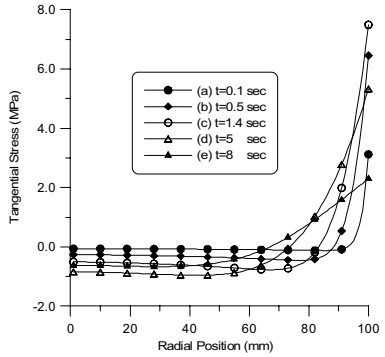


Figure 6. The tangential stress distribution on wafer for the room temperature cooling.

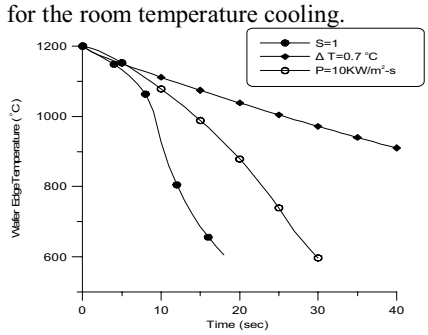


Figure 7. The temperature variation at wafer edge under three different control schemes

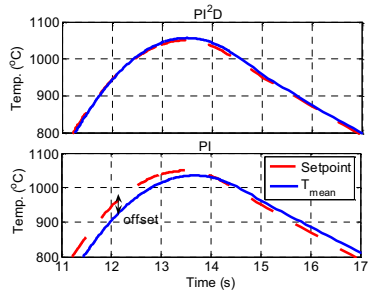


Figure 8. control results of PI ans PI^{2}D for smooth temperature program.

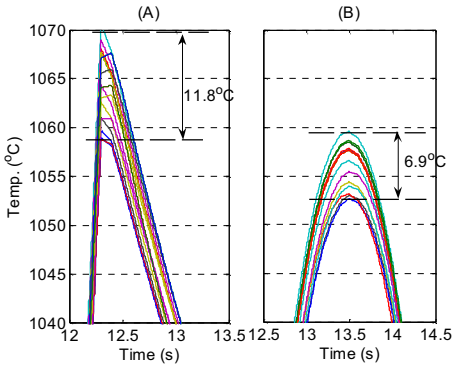


Figure 9. Spread of the peak temperature for (A) triangular-like (B) smooth temperature program.

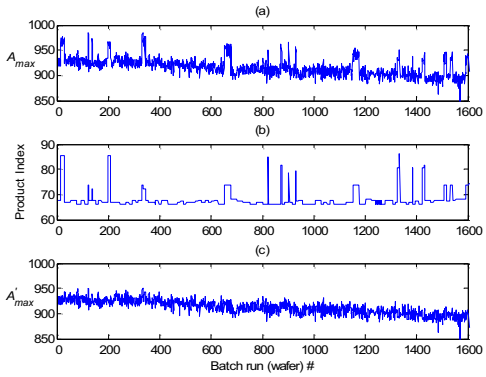


Figure 10. KSV and product index (a) KSV (b) product index (c) modified KSV

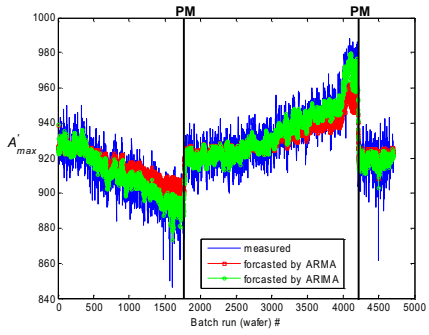


Fig. 11. Comparison of ARMA and ARIMA prediction as compared to the true measurement.

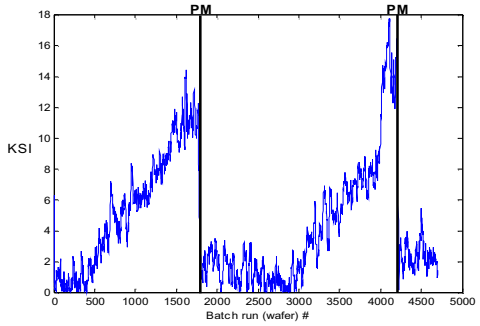


Figure 12. KSI for process trend monitoring

Session 5.1

Analysis and Control of Separation Processes

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PARAMETER AND STATE ESTIMATION IN CHROMATOGRAPHIC SMB PROCESSES WITH INDIVIDUAL COLUMNS AND NONLINEAR ADSORPTION ISOTHERMS

Achim Küpper * **Sebastian Engell** *,1

* *Process Control Laboratory (BCI-AST), Department of Biochemical and Chemical Engineering, Universität Dortmund, Emil-Figge-Str. 70, 44221 Dortmund, Germany.*

Abstract:
In this paper, measurement based parameter and state estimation in Simulated Moving Bed plants with nonuniform columns is investigated. The estimation strategy presented uses the available measurements of the concentrations in the product flows and in one internal flow which is realistic for industrial applications. The estimation task is solved in a decentralized fashion. The correction of the parameters and the state is performed only for the column positioned in front of the respective measurement. Convergence is achieved by the shift of the product concentration measurements. The local estimation problems are solved by Extended Kalman filters. The scheme is validated for a propanolol isomers system with nonlinear adsorption isotherms.

Keywords: Simulated Moving Bed chromatography, parameter estimation, state estimation, Extended Kalman filter, decentralized estimation

1. INTRODUCTION

Preparative chromatographic separation processes are an established separation technology in downstream processing in the pharmaceutical and fine chemicals industries. Most industrial applications are performed discontinuously, leading to low productivity and high solvent consumption. In recent years, continuous Simulated Moving Bed SMB processes are increasingly applied due to their advantages with respect to the utilization of the adsorbent and reduced solvent consumption. The SMB process consists of several chromatographic columns which are interconnected in series to constitute a closed loop. An effective counter current movement of the liquid phase and the solid phase is achieved by periodical and simultaneous switching of the inlet and the outlet ports by one column in the direction of the liquid flow (Figure 1).

Since SMB processes are characterized by mixed discrete and continuous dynamics, spatially distributed state variables with steep slopes, and slow and strongly nonlinear responses of the concentrations profiles to changes of the operating parameters, they are difficult to control. An overview of recent achievements in the optimization and control of chromatographic separations can be found in (Engell and Toumi, 2005). In (Toumi and Engell, 2004*a*) and (Toumi and Engell, 2004*b*), a nonlinear optimizing control scheme was proposed and successfully applied to a 3-zone reac-

[1] Corresponding author: Tel.: +49-231-755-5126; fax: +49-231-755-5129.
E-mail address: s.engell@bci.uni-dortmund.de

tive SMB process for glucose isomerization. In each switching period, the operating parameters are optimized to minimize a cost function. The product purities appear as constraints in the optimization problem. In the optimization, a rigorous model of the general rate type is used. Plant/model mismatch is taken into account by error feedback of the predicted and the measured purities. In addition, the model parameters are regularly updated. In (Toumi *et al.*, 2005), the control concept was extended to the more complex processes Varicol and Powerfeed that offer a larger number of degrees of freedom that can be used for the optimization of the process economics while satisfying the required product purities. A slightly different approach to the control of SMB processes was reported by (Erdem *et al.*, 2004*a*) and (Erdem *et al.*, 2004*b*). Here, the online optimization is based upon a linearized reduced model which is corrected by a Kalman filter that uses the concentration measurements in the product streams. In this work, the switching period is considered as fixed, while in the previously mentioned work it is a parameter in the optimization. In (Toumi and Engell, 2004*a*) and (Toumi and Engell, 2004*b*), the prediction is based on the assumption that the columns are uniform (i.e. they all show the same behavior) and that the modelling errors are small. However, the properties of each individual column differ since they have different effective lengths, different packings with adsorbent and catalyst (for the case of reactive chromatography) and the column temperatures can exhibit some variation. In this paper, an estimation concept is presented for the estimation of parameters and states of chromatographic columns with individual properties. We assume that measurements of the concentrations in one internal and in both product streams are available. The estimation of the column parameters can be used to detect degradations of a column during continuous operation of the plant.

The remainder of this paper is structured as follows: in the next section, the model of the SMB process is introduced. Section 3 reports the observer design for the SMB plant. Simulation results are presented in section 4. Finally, a summary and outlook for future research are given.

2. PROCESS MODEL

The columns of the SMB process can be divided into four different zones according to their relative position with respect to the inlet and the outlet ports as depicted in Figure 1:

(i) Zone I between solvent and extract port: desorption of the more strongly retained component
(ii) Zone II between extract and feed port: desorption of the less retained component
(iii) Zone III between feed and raffinate port: adsorption of the more strongly retained component
(iv) Zone IV between raffinate and solvent port: adsorption of the less retained component.

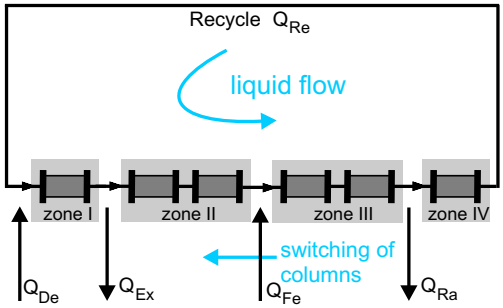


Fig. 1. Schematic diagram of the SMB process

From mass and concentration balances, the relations at the inlet and the outlet nodes can be expressed as

$$\begin{aligned}
\text{Desorbent node}: \quad & Q_{IV} + Q_{De} = Q_I \\
& c^{out}_{i,IV} Q_{IV} = c^{in}_{i,I} Q_I \\
\text{Extract node}: \quad & Q_I - Q_{Ex} = Q_{II} \\
\text{Feed node}: \quad & Q_{II} + Q_{Fe} = Q_{III} \\
& c^{out}_{i,II} Q_{II} + c_{i,Fe} Q_{Fe} = c^{in}_{i,III} Q_{III} \\
\text{Raffinate node}: \quad & Q_{Ra} + Q_{IV} = Q_{III} \\
& i = A, B \ ,
\end{aligned} \tag{1}$$

where $Q_{\mathrm{I,II,III,IV}}$ denote the internal flow rates of the corresponding zones $\mathrm{I, II, III, IV}$, Q_{De}, Q_{Ex}, Q_{Fe}, and Q_{Ra} are the external flow rates of the respective inlet/outlet ports and, $c^{\mathrm{out}}_{i,j}$ and $c^{\mathrm{in}}_{i,j}$ denote the concentrations of the component i in the stream leaving or entering the respective zone j. In this paper, the separation of a racemic mixture of propanolol, a β-blocker, (Toumi *et al.*, 2003) at high purities with a 1/2/2/1 column configuration is investigated. Accurate dynamic models of multi-column continuous chromatographic processes consist of dynamic process models of the single chromatographic columns, the node balances (1) which describe the connection of the columns, and the port switching. The chromatographic columns are described accurately by the *general rate model* (Guichon *et al.*, 1994) which accounts for all important effects of a radially homogeneous column, i.e. mass transfer between the liquid and the solid phase, pore diffusion, and axial dispersion. The concentration of component i is given by c_i in the liquid phase and q_i in the solid phase. D_{ax} is the axial dispersion coefficient, u the interstitial velocity, ϵ_b the void fraction of the bulk phase, $k_{l,i}$ the film mass transfer resistance, and D_p the diffusion coefficient within the particle pores. The concentration within the pores is denoted by $c_{p,i}$. The following set of partial

differential equations can be obtained from a mass balance around an infinitely small cross-section of the column:

$$\frac{\delta c_i}{\delta t} + \left(\frac{1-\epsilon_b}{\epsilon_b}\right)\frac{3k_{l,i}}{r_p}(c_i - c_{p,i|r=r_p}) = D_{ax,i}\frac{\delta^2 c_i}{\delta z^2} - u\frac{\delta c_i}{\delta z} \quad (2)$$

$$(1-\epsilon_b)\frac{\delta q_i}{\delta t} + \epsilon_p\frac{\delta c_{p,i}}{\delta t} - \epsilon_p D_{p,i}\frac{1}{r^2}\frac{\delta}{\delta r}\left(r^2\frac{\delta c_{p,i}}{\delta r}\right) = 0 \quad (3)$$

with appropriate initial and boundary conditions

$$c_{i,t=0} = c_i^{in}; \quad c_{p,i,t=0} = c_{p,i}(0,r,x),$$
$$\left.\frac{\delta c_i}{\delta z}\right|_{z=0} = \frac{u}{D_{ax,i}}\left(c_i - c_i^{in}\right); \quad \left.\frac{\delta c_i}{\delta x}\right|_{z=L} = 0 \quad (4)$$
$$\left.\frac{\delta c_p}{\delta r}\right|_{r=0} = 0; \quad \left.\frac{\delta c_p}{\delta r}\right|_{r=r_p} = \frac{k_i}{\epsilon_p D_{p,i}}\left(c_i - c_{p,i,r=r_p}\right).$$

The adsorption equilibrium behavior and the system parameters for the propanolol isomers investigated here have been determined experimentally by (Toumi *et al.*, 2003). The adsorptive behavior is modelled by a modified competitive Langmuir adsorption isotherm (the components are referred to as A and B):

$$q_i = H_i^1 c_{p,i} + \frac{H_i^2 c_{p,i}}{1 + \sum_j k_j^2 c_{p,j}} \quad i = A, B. \quad (5)$$

For numerical simulation, an efficient discretization (Gu, 1995) is used where a finite element discretization of the bulk phase is combined with orthogonal collocation of the solid phase.

3. STATE AND PARAMETER ESTIMATION

3.1 Estimation strategy

A concept for parameter and state estimation of a SMB process has to take into account the available measurement information as well as the dynamics of the SMB model. We assume here that both concentrations of the species are continuously measured in the two product streams and in the recycle stream. This is the maximum amount of information that is available in a production plant. Thus, the positions of the measurements in the considered six column SMB plant vary within a cycle of operation as indicated by Figure 2. The recycle measurement is permanently located behind the last physical column while the two product measurements move with the product ports in the direction of the liquid flow by one column when a period has passed. Hence, there permanently is a measurement behind the last physical column over the whole cycle, while each of the remaining columns have a product measurement located at their respective outlet for two periods in each cycle (one cycle has six periods). Concerning the dynamics of the SMB model, it

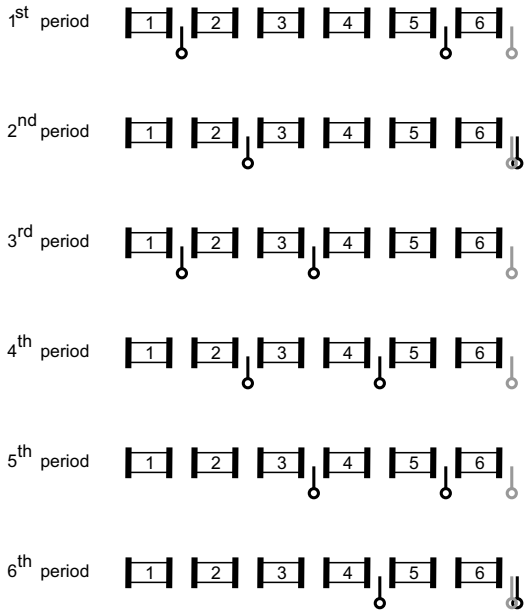


Fig. 2. Measurement positions at the physical columns for a cycle of operation (black: product streams; grey: recycle stream)

can be shown that local column parameters have a greater influence on the concentrations at the outlet of the local column compared to column parameters at a distance to the considered measurement. The influence of distant parameters is subject to a considerable time delay because the liquid flow is the link between the columns and reaches the considered column outlet only after several switching periods.

We therefore propose to perform the estimation by a set of individual, local observers that estimate the states as well as the parameters for one column only, as illustrated by Figure 3. A local observer is activated within the estimation scheme when there is a measurement located at the outlet of its respective column. The columns are of course coupled since a column with an error prone set of parameters that is in front of a column with an active local observer causes a disturbed input flow to the estimated column (indicated by the disturbance d in Figure 3). However, the influence of the disturbed input on the local estimation is reduced by the movement of the product measurements. One period before a local estimator is activated by a product measurement, its corresponding input is corrected by the same measurement sensor that is positioned before the column at that time. Since SMB processes are operated at a periodic steady state with high requirements for the product purities, the dissolved components move with the liquid flow by about one column within one period. Hence, the outlet concentrations of the column are influenced to a large extent by

the inlet concentrations that were measured one period earlier. Therefore, the influence of model errors of the remaining columns on the local estimation of one column is not a dominant factor in the local estimation. However, the estimated input concentration profile of column 6 for which the estimation is performed continuously deviates from the true profile and hence in this column errors of the parameter estimation are induced until a sensor is located in front of this column (see Figure 2).

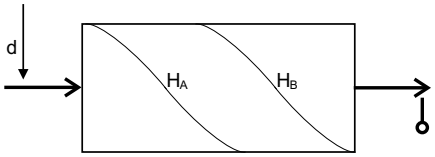


Fig. 3. Column-by-column estimation of parameters and states

3.2 Extended Kalman filter

In the local Extended Kalman filters, the parameters that are estimated are defined as states. At each time step k the dynamics $\hat{f}^i$ of the column model are linearized around the local state estimate $\hat{x}^i$ (each column has 100 states). The algorithm of the Extended Kalman filters is

(1) Prediction:

$$\hat{x}^i_{k+1,k} = \hat{x}^i_{k,k} + \int_{t_k}^{t_{k+1}} \hat{f}^i(\hat{x}^i, \hat{x}^{i-1})dt \quad (6)$$

$$P^i_{k+1,k} = A^i_k P^i_{k,k} A^{iT}_k + Q^i \quad (7)$$

(2) Correction:

$$K^i_k = P^i_{k,k-1} C^{iT} (C^i P^i_{k,k-1} C^{iT} + R)^{-1} \quad (8)$$

$$\hat{x}^i_{k,k} = \hat{x}^i_{k,k-1} + K^i_k (y^i_k - \hat{y}^i_{k,k-1}) s^i_{period} \quad (9)$$

$$P^i_{k,k} = (I - K^i_k C^i) P^i_{k,k-1}, \quad (10)$$

where P^i is the error covariance, K^i the Kalman gain, C^i the output matrix, Q^i the state noise covariance matrix, R the measurement error covariance matrix, y^i are the measurements, and A^i_k the linearized local dynamics of column i. The EKFs were tuned by varying the diagonal values of the matrices P_0, R, and Q^i. s^i_{period} is a binary integer variable that takes the value of one if the local observer of the corresponding column is activated and zero otherwise. The activation trigger s^i_{period} of the local observers is related to Figure 2.

4. RESULTS

For the results presented, the measurements are assumed to be subject to uniformly distributed white noise with a maximum deviation of 5% of the highest concentration value of the SMB profile. Furthermore, it is assumed that the product measurement devices are placed in front of the product pumps, otherwise a deformation of the concentration profiles would be encountered. The Henry coefficients H^1_i are chosen as estimated parameters since they have the strongest influence on the process performance. The chosen operating point achieves high purities but is not optimal with respect to solvent consumption.

Figure 5 shows the estimation of the parameters of column 1 in the case of a step in the Henry coefficient H^1_A (more strongly adsorbed component) of column 3 (scenario 1). The estimator of column 3 converges to the true value. The estimators of columns 1, 2, 4, and 5 are not affected by the error in column 3. However, the input of column 6 is not corrected during one period and the Henry coefficient H^1_A of column 6 is therefore increased. When the estimator of column 3 has converged, the estimator of column 6 converges as well. Figure 6 depicts the estimation of the Henry coefficients of all six physical columns in the presence of considerable model errors (scenario 2). The individual model coefficients are initialized at the nominal values that are given in the appendix while the true values of the individual Henry coefficients of the respective columns differ considerably from the nominal values. The proposed estimation concept manages to estimate all individual Henry coefficients and to reduce the state estimation error at the beginning of a period substantially (see Figure 4), apart from the error caused by the measurement noise. In Figure 7 the estimation of the parameters of column 6 whose local estimator is activated over the whole cycle is compared to the corresponding measurement information for one cycle. For the investigated operating point, the concentration fronts (concentration increases or decreases drastically) are in the 2nd, 4th, and 6th period. The measured

$$J_0 = \sqrt{\sum_{i=1}^{n_{state}} (\hat{x}(i)_{smb,0} - x(i)_{smb,0})^2} \quad (11)$$

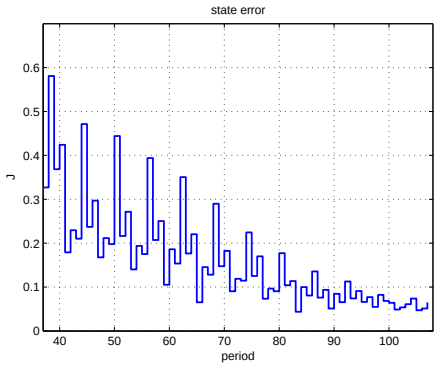


Fig. 4. State error J forscenario 2

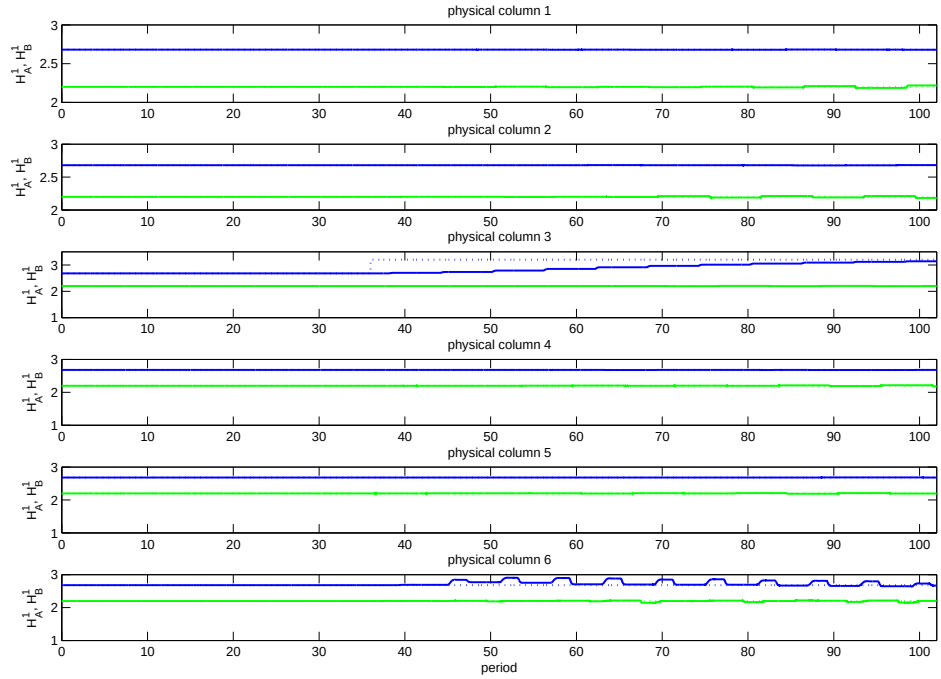


Fig. 5. Parameter pertubation of column 3 (step introduced and estimation started at the 37^{th} period, scenario 1); lines: model, dotted lines: reference plant, black: H_A^1, grey: H_B^1

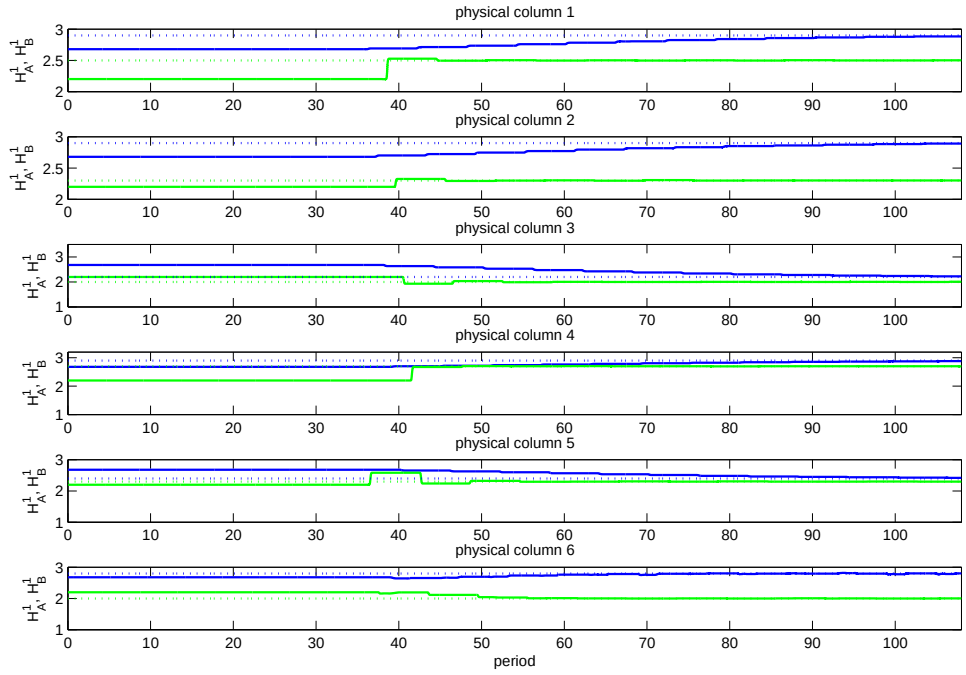


Fig. 6. Parameter estimation for wrong Henry coefficients of all columns (estimation started from the 37^{th} period on, scenario 2); lines: model, dotted lines: reference plant, black: H_A^1, grey: H_B^1

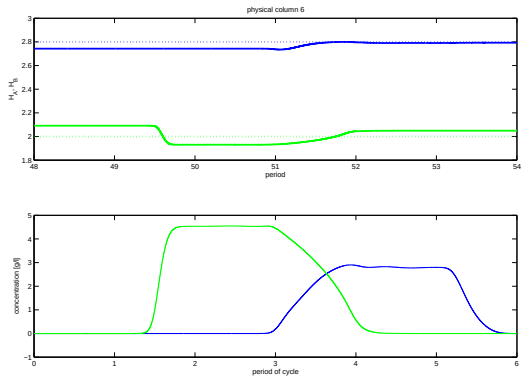


Fig. 7. Column 6 (permanent measurement): parameter estimation and measured concentration profile over one cycle for scenario 2

extract and raffinate profiles correspond to period 2 and 5 modified slightly by the individual column properties. When a concentration front moves over the extract port, the Henry coefficient H_B^1 of the respective column is corrected by the local estimation. However, at the raffinate port (period 5) only a small part of the front is measured. The correction of the Henry coefficient H_A^1 related to the more strongly adsorbed component A is rather slow. This is due to the specific operating point.

5. CONCLUSION

A parameter and state estimation scheme for an SMB process with individual columns applying local Extended Kalman filters based on only three measurement positions has been presented. The observer performs well. The individual column parameters can be reconstructed. It is expected that the performance of a model predictive control scheme can be improved by applying a decentralized state estimation. In future research, the implementation of Moving Horizon Estimator is planned.

SYSTEM PARAMETERS

separator length	$L = 10cm$
separator diameter	$D = 1cm$
adsorption coefficients	$H_A^1 = 2.68$
	$H_A^2 = 0.9412$
	$k_A^2 = 340\frac{cm^3}{g}$
	$k_B^2 = 262\frac{cm^3}{g}$
	$H_B^1 = 2.2$
	$H_B^2 = 0.4153$
film transfer resistance	$k_{l,A} = 0.5610^{-2}\frac{cm}{s}$
	$k_{l,B} = 0.3310^{-2}\frac{cm}{s}$
void fraction	$\epsilon_b = 0.4$
particle void fraction	$\epsilon_p = 0.5$
particle diameter	$d_p = 20\mu m$
particle diffusion coefficient	$D_p = 10^{-5}\frac{cm^2}{s}$
density	$\rho = 1.0\frac{g}{ml}$
viscosity	$\eta = 6.8510^{-4}\frac{g}{cm\,s}$
axial diffusion coefficient	$D_{ax} = 10^{-6}\frac{cm^2}{s}$
feed	$Q_{Fe} = 0.31\frac{ml}{min}$
	$c_{A,Fe} = c_{B,Fe} = 7.5\frac{g}{l}$
period	$\tau = 2.05min$
extract	$Q_{Ex} = 1.94\frac{ml}{min}$
raffinate	$Q_{Ex} = 1.12\frac{ml}{min}$
eluent	$Q_{De} = 2.75\frac{ml}{min}$
recycle	$Q_{Re} = 4.80\frac{ml}{min}$
measurement error covariance	$R = 0.01I$ (I: unity matrix)
initial error covariance	$P_0 = 300I$, $P_{0,6} = 30$
state noise covariance	$Q_x = 100I$,$Q_{x,6} = 0.01I$
	$q_p = 0.05,\ 0.01$

ACKNOWLEDGMENT

This investigation was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant DFG En 152/34. This support is gratefully acknowledged.

REFERENCES

Engell, S. and A. Toumi (2005). Optimisation and control of chromatography. *Computers and Chemical Engineering* **29**, 1243–1252.

Erdem, G., S. Abel, M. Morari, M. Mazzotti and M. Morbidelli (2004*a*). Automatic control of simulated moving beds. *Ind. Eng. Chem. Res.* **43**, 405–421.

Erdem, G., S. Abel, M. Morari, M. Mazzotti and M. Morbidelli (2004*b*). Automatic control of simulated moving beds II: Nonlinear isotherms. *Ind. Eng. Chem. Res.* **43**, 3895–3907.

Gu, T. (1995). Mathematical modelling and scale up of liquid chromatography. *Springer Verlag, New York*.

Guichon, G., S.G. Golshan-Shirazi and A.M. Katti (1994). Fundamentals of preparative and nonlinear chromatography. *Academic Press. Boston*.

Toumi, A. and S. Engell (2004*a*). Optimal operation and control of a reactive simulated moving bed process. *Proc. IFAC Symposium on Advanced Control of Chemical Processes Hong Kong*, 243–248.

Toumi, A. and S. Engell (2004*b*). Optimization-based control of a reactive simulated moving bed process for glucose isomerization. *Chemical Engineering Science* **59**, 3777–3792.

Toumi, A., M. Diehl, S. Engell, H.G. Bock and J.P. Schlöder (2005). Finite horizon optimizing control of advanced SMB chromatographic processes. *IFAC World Congress, Prague*.

Toumi, A., S. Engell, O. Ludemann-Hombourger, R. M. Nicoud and M. Bailey (2003). Optimization of simulated moving bed and Varicol processes. *Journal of Chromatography* **1006**, 15–31.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PARAMETRIC MODEL PREDICTIVE CONTROL OF AIR SEPARATION

Jorge A. Mandler*,
Nikolaos A. Bozinis, Vassilis Sakizlis**[1], Efstratios N. Pistikopoulos**,**
Alan L. Prentice*, Harish Ratna***, Richard Freeman***

*Air Products and Chemicals, Inc., 7201 Hamilton Blvd, Allentown, PA, 18195 USA,
mandleja@airproducts.com
**Centre for Process Systems Engineering, Department of Chemical Engineering,
Imperial College London, London SW7 2AZ, United Kingdom and
Parametric Optimization Solutions (ParOS) Ltd, 90 Fetter Lane
London EC4A 1JP, United Kingdom
n.bozinis@parostech.com, e.pistikopoulos@imperial.ac.uk,
**[1] Currently with Bechtel, Hammersmith Road
London W6 8DP,
vsakizli@bechtel.com
*** Air Products PLC, Hersham Place, Molesey Road, Hersham, Surrey, KT12 4RZ,
United Kingdom

Abstract: This paper describes the application of Parametric Model Predictive Control to small processing units, in particular small Air Separation plants. Multiparametric optimization techniques are used to rigorously solve the MPC problem in two steps: an offline solution which generates a parametric mapping of the optimal control adjustments, and an online solution which reduces to a simple lookup operation. Because of the speed and simplicity of this lookup operation we are able to implement MPC in low-end computing devices such as PLCs, reaping the benefits of model-based control by implementing it at low cost in small plants where otherwise it would not be justified by the cost/benefit ratio.

Keywords: parametric programming, parametric optimization, on-line optimization, predictive control, process control.

1. BACKGROUND

While Model Predictive Control (MPC) is the clear Advanced Control technology of choice in the Process Industries, it has found limited use to date for small processing units, despite its unquestionable superiority in terms of robustness, plant optimization and general control performance. One bottleneck is the complexity and relatively high cost of the controller compared to the unit cost in smaller size plants. This is partially due to the computing hardware and software required for executing on-line, real time optimization in order to determine the appropriate control action for the next time interval.

For the smaller Air Separation plants (single product plants, Nitrogen or Oxygen generators, cryogenic or non-cryogenic) Advanced Control of any kind was in the past an expensive proposition. As a result, the small plants would most often be operated in a conservative manner and suffer from the following operational drawbacks:

- They would consume more energy than required.
- They would be unable to load follow a varying customer demand.
- Venting of product or product backup would be required whenever the customer demand did not match the set production and single point of operation.

In the last few years, academic research on parametric programming has lead to a radically new approach to MPC (Pistikopoulos, *et al.*, 2002a; Pistikopoulos, *et al.*, 2002b; Dua, *et al.*, 2002; Bemporad, *et al.*, 2002). In this approach, the on-line control problem has been recast as a multi-parametric optimization problem where the system state variables act as "parameters". The original MPC problem can now be solved explicitly in an efficient manner, still generating the full control law in a mathematically rigorous fashion. In essence, most of the possible MPC scenarios that are encountered

during the operation of a unit are solved a priori and off-line.

The implementation of the control law is transformed into a simple look-up function operation, where the current values of the state variables determine the control action. The control action taken by such a "parametric" controller is identical to traditional MPC for a given system state representation. The only difference and main advantage of the parametric approach lies in the manner the control action is decided: whereas traditional MPC requires on-line solution of dynamic optimization problems, the new approach just reads the current solution from a complete solution map drawn in advance. This is the concept of on-line control via off-line parametric optimization. Hence, a number of major advantages can be obtained:

1) Lower hardware costs - Simpler hardware, including PLCs and microchips, is completely adequate given the minimal on-line computational requirements;
2) Software costs are nearly eliminated;
3) Simple implementation is possible;
4) Increased control power is obtained, in part due to the very fast sampling now possible given the almost instantaneous solution.

The new parametric control concept has allowed us to extend the applicability of MPC to small Air Separation plants, for example small Nitrogen generators. We call our approach Parametric MPC (pMPC) of Air Separation. We have also called it "MPC on a chip", since the pMPC controller, because of the ease of on-line implementation, can be readily commissioned on a microchip. We should stress at this point that the approach is not limited to small Air Separation plants. It can be equally applied in situations where, even though the Air Separation plant might be very large in terms of its production capacity, because of relative process simplicity there is still a small number of manipulated (MVs), controlled (CVs) and disturbance variables (DVs) in the process. This includes for example very large oxygen generators for GTL (gas-to-liquid) applications.

Parametric MPC is in fact a generic technology, with the only thing specific to a given application being the underlying process model and MPC problem formulation. ParOS Ltd has applied this technology in entirely different sectors, such as in the automotive industry where very fast and accurate control is required (e.g. sampling times of 0.1 millisecond).

2. PROCESS DESCRIPTION, BENEFITS SOUGHT AND CONTROL OBJECTIVES

Fig. 1 is a simplified diagram of a typical Nitrogen generator. Air is compressed, impurities such as CO_2 and water are eliminated in a Pressure Swing Adsorption unit (PSA, not shown), the clean air is then cooled to near its liquefaction temperature in the main heat exchanger and fed to a distillation column. In this distillation column the air is separated into a pure nitrogen fraction in the overhead, and an oxygen rich liquid fraction at the bottom. Part of the pure nitrogen is taken as GAN (gaseous nitrogen) product, and the rest is condensed and returned as reflux to the column. This is done in a reboiler/condenser by heat exchange against the enriched air from the bottom of the column, which boils at a lower pressure on the other side. A small amount of liquid nitrogen (LIN) is fed to the column to provide extra refrigeration. After heat exchange against the incoming feed, the GAN product is compressed and sent to the customer. The customer takes the product through a short pipeline and there may or may not be a buffer tank present.

In typical operation without advanced control, the plant produces GAN at a fixed production rate irrespective of the customer demand. Therefore if the customer demand increases, extra nitrogen may need to be provided by vaporizing LIN from a backup tank. If the customer demand drops, GAN product may need to be vented. While many customers have a constant take pattern that justifies this mode of operation, for other customers the product demand may vary quite frequently. For the latter cases, it is certainly a waste (mainly in terms of power usage) to produce pure gaseous nitrogen and then throw it to vent, and/or to vaporize significantly more expensive LIN when the demand exceeds what the plant is producing. This calls for an advanced control solution allowing the plant to quickly ramp up or down its production to match the customer demand, while maintaining the product purity within specifications. In our work this defined the first control and online optimization objective: To load follow the customer demand.

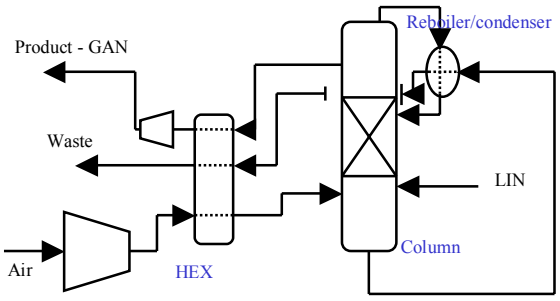


Fig 1. Simplified diagram of a Nitrogen generator.

To run trouble-free and always meet the purity specifications in the face of disturbances, operations personnel may tend to run the plants "fat", namely with extra air fed to the system and with the level of impurities (oxygen and argon in the case of a nitrogen generator) buried down. For example one particular plant had a specification of not more than 5 ppm impurities in the GAN product, but it was observed to be running as low as 0.1 ppm. Of course, this is another source of wasted power. An MPC controller, on the other hand, can be easily set up to run the plant against its true constraints in the face of disturbances. Our second control and optimization objective was thus defined: To operate against the upper impurity limit for the GAN product. This would lead to lower power use for the same production, or equivalently increased production at a given air rate. Although running against the upper impurity limit can be done with standard PID loops, the experience for this particular type of plants was that these loops were hard to tune and too much of an effort was required for the expected benefits.

All the standard benefits of MPC (the multivariable and optimal nature of the solution and the ability to handle constraints) were sought and were achieved, thanks to the parametric control formulation, without a need for expensive online computations.

3. CONTROLLER IMPLEMENTATION DETAILS

The feasibility of pMPC of small ASUs was first demonstrated on a detailed dynamic simulation of a Nitrogen generator. We next proceeded to design and implement a pMPC controller for an actual Nitrogen generator serving a customer. This plant was selected because the customer take pattern was such that it would help us demonstrate load following capabilities. The plant was also conveniently located and it had site personnel resources available as needed for our first prototype. Initially the controller was designed and implemented with only the load following objective, while still maintaining the product purity within constraints. Later a second optimization objective was added as an additional term in the objective function. The second objective led the pMPC controller to reduce air feed whenever possible to drive GAN purity against its upper limit (while not crossing it). Formally, the control and optimization objectives were implemented as follows: 1) Match GAN production to GAN demand (minimizing its difference); 2) Control GAN purity at a setpoint equal to the upper impurity limit. Note that in this plant the GAN demand (product taken by the customer) could not be directly measured but instead it was estimated using the customer valve position and pressure differences.

In order to attain the above control and optimization objectives we designed the following control structure. The pMPC controller was set up to operate as a supervisory controller, with inherently safe fallback mode to the underlying regulatory control

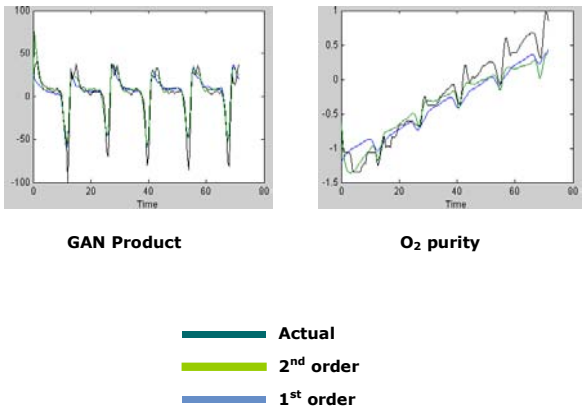


Fig. 2. Model Prediction for Each CV.

loops in case of difficulties. As Manipulated Variables (MVs) we included the air flow setpoint, and the setpoint for the GAN/AIR ratio. There were two disturbance variables (DVs), the first one the deviation in air flow with respect to its setpoint (to account for each switch of the PSA unit), and the second one a measure of the LIN injection to the column. The controlled variables (CVs) were also two: the actual GAN product flow and the ppm O_2 in the GAN product. In summary, the controller had four inputs (2 MVs and 2 DVs) and two outputs.

Data for system identification was obtained via step-change experiments. First and second order ARX models were identified both giving a good fit. Fig. 2 shows a validation set and the prediction by the first and second order models.

As indicated, the first control objective was to follow the customer demand without violating the purity constraints, and further subject to bounds on the MV values and on the maximum allowable rate of change for the MVs. The online MPC controller was obtained off-line via multi-parametric optimization. The techniques and the code employed to conduct this optimization are described elsewhere (Pistikopoulos, *et al.*, 2002a; Pistikopoulos, *et al.*, 2002b; Dua, *et al.*, 2002; Bemporad, *et al.*, 2002). The solution for this first pMPC controller involved 7 parameters, 2 control targets, and 209 piecewise affine control laws (regions of the parametric space). Just as an example, Fig. 3 shows the solution for one of the regions. Notice that the online solution (the value of each MV for the next control move) is obtained explicitly via a linear combination of the parameters. The values of the coefficients a_c and b_c are different for each solution region.

Because of the ease of online computation, we were able to implement the pMPC controller on the existing plant PLC. This PLC (programmable logic controller) was an older type model with very limited computing resources. The pMPC controller was implemented as a C function block and it worked together with the existing ladder logic. Timers ensured that the controller was called every 30 seconds precisely. Since the computation was very fast, the model-based control action could have been

Explicit Control Law:

$$\begin{vmatrix} u_0 = a_c\theta + b_c \\ \text{if} \;\; CR_c^1\theta + CR_c^2 \le 0 \\ c = 1,\ldots N_c \end{vmatrix}$$

Region 117:

MAC(t+1)=5.8971·X_{O2}(t) -4.2801·GAN(t)-3.0664·X_{O2}(t-1) +0.1304·GAN(t-1) +0.0955·MAC(t)+24.7965 ·(1/Ratio)+6.1678·GAN^{set}

1/Ratio(t+1)=0.3063 ·X_{O2}(t)+0.2437·GAN(t)-0.1488 ·X_{O2}(t-1)-0.1554 ·GAN(t-1)- +0.0072·MAC(t)+0.8065·(1/Ratio)-0.0766 ·GAN^{set}

-10.0<MAC(t)<10.0

-71.5< 42.38 ·X_{O2}(t)+33.72·GAN(t) -20.5853·X_{O2}(t-1) +… -10.6 ·GAN^{set} <71.5

Fig . 3. Parametric control solution for Region 117.

done more frequently if this was needed for improved performance. The ability to implement model predictive control on a PLC became in itself a major accomplishment of our work.

Fig. 4 gives a schematic diagram of the function lookup operation constituting the online implementation of pMPC for the small ASU. At every sample the C code read the current plant state and formed a parameter vector consisting of 7 values based on the current and last sample conditions. A search was done to find the corresponding region in parametric space, and the coefficients of the explicit control law for that region were then applied to calculate the next control move for the MVs.

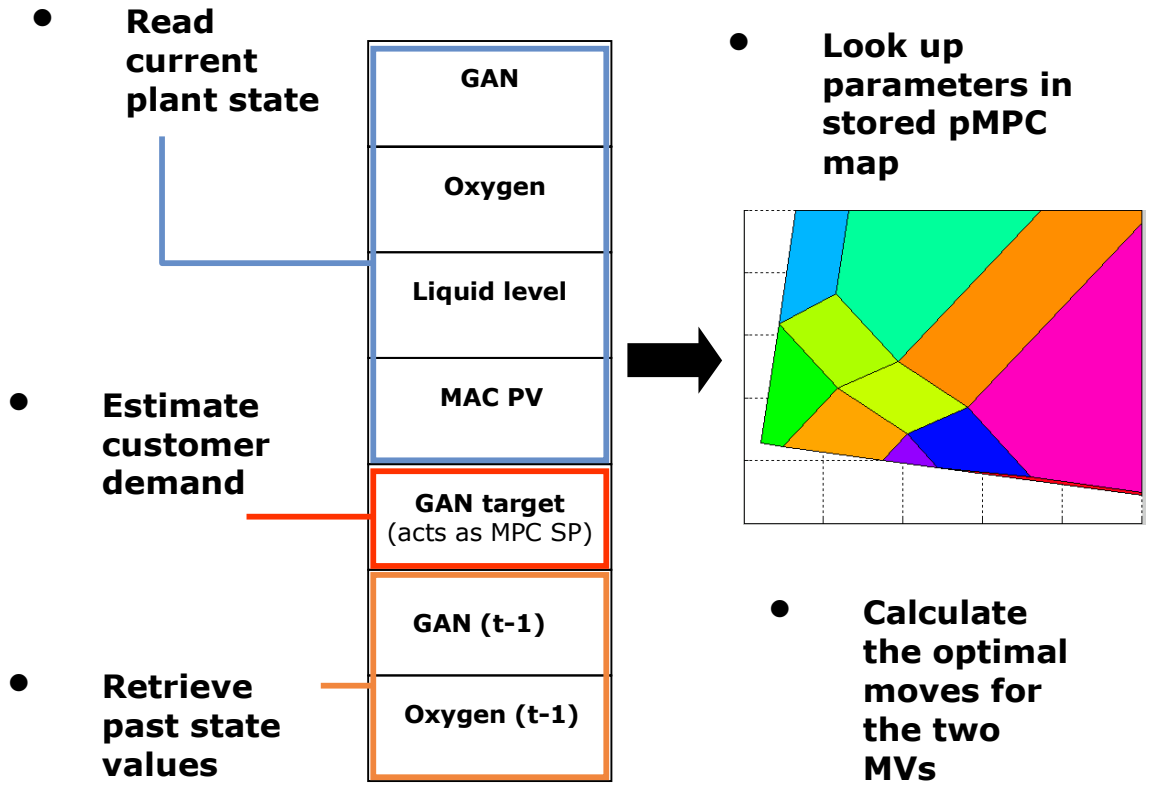


Fig. 4. Implementation of lookup operation of pMPC for a small Nitrogen generator. GAN: Gaseous Nitrogen; MAC PV: Measurement of air flowrate.

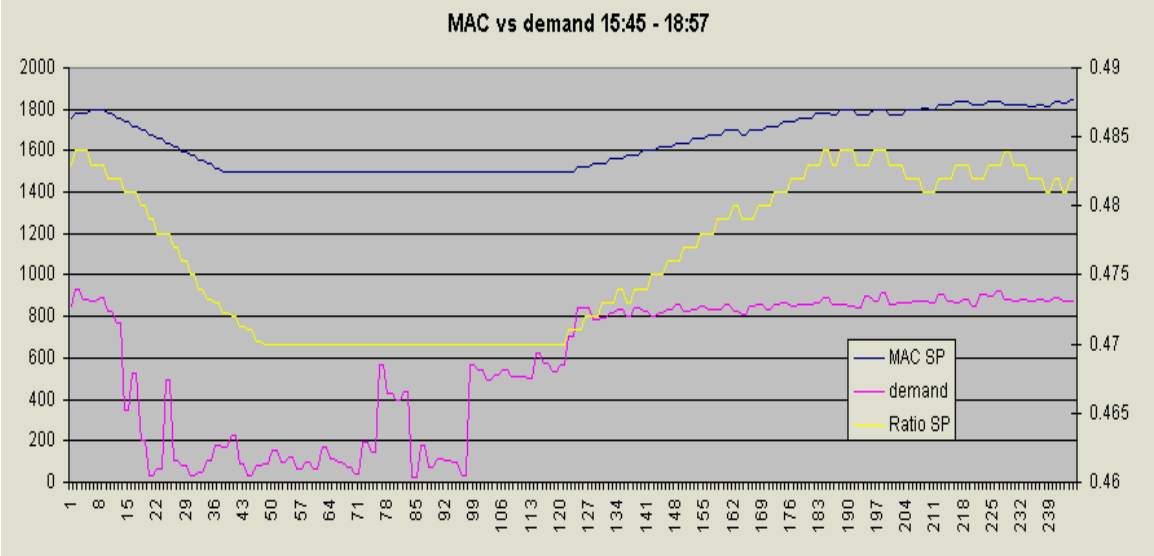


Fig. 5. Load following the customer demand, first controller implementation (left axis, MAC SP and customer demand; right axis: Ratio SP)

4. IMPLEMENTATION RESULTS AND ADDITIONAL STEPS

As mentioned earlier, the first pMPC controller was based on a single objective, i.e. load following. This controller was implemented at the selected Nitrogen plant serving a customer site. Significant benefits resulted from the ability to match the customer demand. The pMPC controller minimized liquid nitrogen (LIN) usage by ramping the plant up to maximum production when the customer demand was high, and it reduced power usage by ramping the plant down during low demand periods. While the speed of the ramp up was observed to be slightly slower than desired, overall the MPC controller was judged to be beneficial, and it was recommended that the operators turn it on every Monday morning after starting the plant up at the beginning of each week (this particular plant was shut down over weekends). The initial load following closed-loop behavior (single objective) is shown in Fig. 5.

The controller was next revised to include a purity control objective. As already mentioned, this second objective was to minimize power by operating against the upper impurity limit. The slow ramp up issue was also corrected. The new controller involved 8 parameters, and its solution 578 piecewise affine control laws. The new controller ramped the plant up or down at 3% of the design flow per minute. This is a significant ramp rate for this type of plants. Fig. 6 shows the load-following performance for this second controller. The plant layout included a buffer tank between the product compressor and the customer. By design, the controller was set up to load the buffer tank whenever possible, and the ramp down was not initiated unless the buffer tank pressure (which is shown in blue – buffer – in Fig. 6) exceeded 12 bar.

Before pMPC, the plant used to run at 0.01 ppm O_2 in the GAN product. By implementing the second optimization objective, the controller instead ran the plant at about 1 ppm, still being able to maintain the purity within specification in the face of disturbances. This led to a 1.5% reduction in air flow, with its consequent power savings. Fig. 7 shows the operation against the upper impurity limit, set at 1 ppm. At around sample 55 in the Figure, the controller started increasing the GAN/AIR ratio from 0.47 to its maximum limit of 0.499. This was effective while the plant was at maximum rates. At around sample 217 the customer suddenly started taking less product, and at that time both MVs, Air flow (not shown) and GAN/AIR ratio dropped to match the new customer demand.

The combined savings, namely from producing more product more efficiently when product was needed, and from ramping down the plant and saving power when less product was needed, were estimated to be in the order of £10000 per year. Without getting into the exact cost details for the controller, we are able to state that based on these results the controller would pay for itself in about half a year or less.

The same controller was duplicated, with almost no change, to a similar nitrogen generator in a different country. This served as an excellent test of the portability and robustness of the pMPC controller. The same model and same controller solution as in

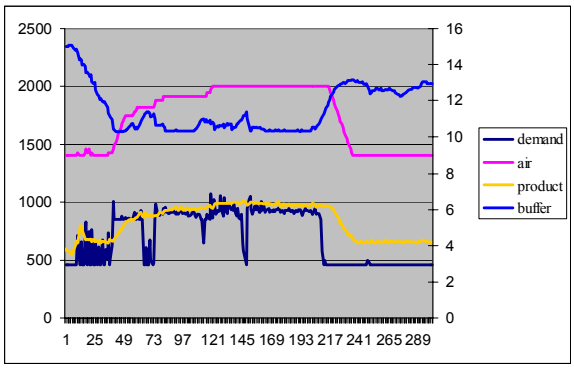


Fig. 6. Load following and operation with buffer tank, second controller implementation.

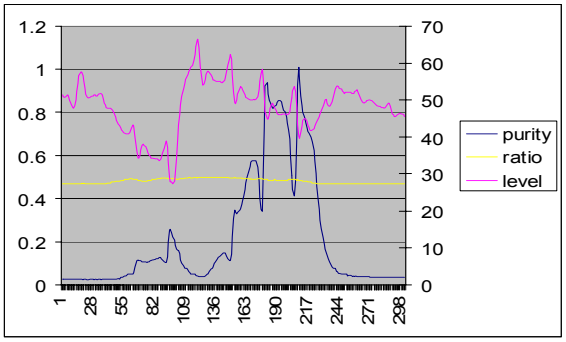


Fig. 7. Operation against upper impurity limit (Left axis, O_2 in GAN, ppm; left axis, GAN/AIR ratio; right axis, column sump level, %).

the previous site were used in spite of some differences in operating limits, customer requirements and customer take pattern. This controller was set up and loaded in just about a day, and ran well from the very beginning. Benefits were quickly obtained because of the increased production rates achievable at the plant with the help of the advanced controller. A third pMPC controller was later commissioned on a larger plant in the USA. Here a side-by-side comparison was done of pMPC operation versus standard operation at site, and the benefits of pMPC were thus demonstrated. Power savings from the use of pMPC were measured to be in the order of 3%. This corresponded well with the original savings estimated in our preliminary, simulation-based evaluation.

Our work also included initial research in the development of a robust parametric controller (Sakizlis, *et al.*, 2004). The idea is to do pMPC plant testing and controller development only once for a generic plant, in much the same way as demonstrated when we copied one controller from one site to another. Then, the robust controller can be simply duplicated to other similar plants, whether of the same size, larger or smaller, and have at most one or two tuning parameters for quick adjustment during commissioning.

CONCLUSION

By implementing MPC via multiparametric optimization (offline solution for online optimization) we can extend the realm and the benefits of model-based, optimizing control to small plants, devices and systems. Parametric MPC of small Nitrogen generators was implemented on existing PLCs and it has been running in several of our plants since 2003. Implementation on new plants is extremely rapid (can be as fast as 1 day). The controller has delivered:

1) Energy savings,
2) Product quality constraint satisfaction,
3) Accurate load following, and
4) Reduced venting of nitrogen.

At the present time, work proceeds for other types of Air Separation plants and in new areas beyond Air Separation.

REFERENCES

Bemporad, A., M. Morari, V. Dua and E.N. Pistikopoulos (2002). The explicit linear quadratic regulator for constrained systems. *Automatica,* **38**, 3-20.

Dua, V., N. A. Bozinis and E. N. Pistikopoulos (2002). A multiparametric programming approach for mixed integer and quadratic engineering problems. *Computers and Chemical Engineering,* **26**, 715-733.

Pistikopoulos, E. N., N. A. Bozinis, V. Dua, J. D. Perkins and V. Sakizlis (2002a). Improved process control. *World Patent Application WO 02/097540 A1*.

Pistikopoulos, E. N., V. Dua, N. A. Bozinis, A. Bemporad and M. Morari (2002b). On-line optimization via off-line parametric optimization tools. *Computers and Chemical Engineering,* **26**, 175-185.

Sakizlis, V., N. M. P. Kakalis, V. Dua, J. D. Perkins, E. N. Pistikopoulos (2004). Design of robust model-based controllers via parametric programming. *Automatica*, **40**, 189-201.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

STABILIZING CONTROL OF AN INTEGRATED 4-PRODUCT KAIBEL COLUMN

Jens Strandberg and Sigurd Skogestad [1]

Department of Chemical Engineering, NTNU, Trondheim, Norway

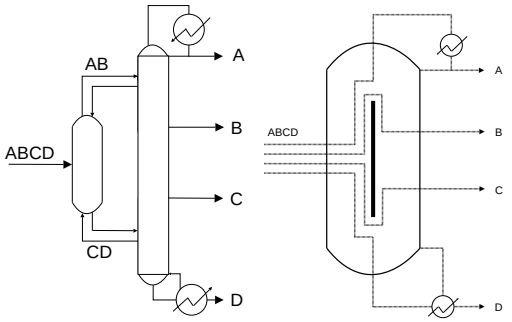
Abstract: This paper considers the Kaibel column, a fully thermally coupled distillation column for the separation of four products in a single column with a single reboiler. The authors of this paper have built a laboratory pilot plant of a Kaibel column with the purpose of investigating its operational performance and control properties. In this paper the requirements for stable operation are discussed, and the location of temperature measurements for optimal operation is investigated.

Keywords: Process control, Distillation, Control structure design, Thermally coupled columns

1. INTRODUCTION

This paper considers the separation of four components in one fully thermally coupled column. The Kaibel column, introduced in 1987 (Kaibel, 1987) separates 4 products in a single column shell with a single reboiler. The main reason for considering the Kaibel column is probably the potential capital savings compared to conventional arrangements with 3 columns in series. The Kaibel column is an extension of the Petlyuk column (Petlyuk et al., 1965). The Petlyuk column and the dividing wall column (DWC) (Wright, 1949) have been extensively investigated in the literature. Even though this research has shown potentially large savings in capital and operational costs, it has taken a long time for the industry to implement the ideas. However, the last 20 years have seen the technology come into use and there are now more than 40 divided wall columns in operation around the world (Adrian et al., 2003). The Petlyuk arrangement can be extended to any number of products with the addition of vertical partitions or column shells. However, a practical realization of a 4-product Petlyuk column would be complex both in construction and operation. The Kaibel arrangement (Figure 1) is easier to implement because one would need only one vertical partition in a one-shell configuration. Both the Petlyuk and the Kaibel columns could be realized in a multi-shell arrangement, still retaining the energy-benefits, but one would of course lose the bonus of reduced capital cost as compared to the conventional three-column sequence. Instead of separating A/D in the prefractionator as in the Petlyuk arrangement, we here have a AB/CD split. This gives a somewhat higher energy requirement, but it is still has the potential to save energy as compared to a three-column sequence (Halvorsen, 2005). Especially, if the B/C separation is easy, the Kaibel configuration ought to be very competitive. BASF, who are the leading industrial company on the dividing wall technology, have at the moment the first Kaibel columns in operation (Kaibel et al., 2004). However, to these author's knowledge there are not reported any lab-scale realizations of the column with thorough investigations of the operational and control properties of the column.

[1] skoge@chemeng.ntnu.no



(a) (b)

Fig. 1. a) Kaibel column with prefractionator arrangement. b) Equivalent one-shell arrangement

2. PILOT PLANT AND COLUMN DATA

A pilot plant of the Kaibel column has been built with the purpose of investigating the controllability of the column arrangement. The column is put together by sections of internal diameter 50 mm. This leads, in effect, to a two-shell implementation (see Figure 2), but it is equivalent to a divided wall column with no heat transfer across the partition wall. The sections are of vacuum-jacketed glass, requiring no further insulation. A kettle reboiler of 3 kW capacity is attached. The product streams (except bottoms) are controlled via solenoid operated swinging funnels built into the glass sections. Another funnel sets the liquid split, R_l, between the prefractionator and the main column. The column will be operated with a constant vapour split, R_v, however a device has been installed that allows for manual adjustment of the split. A total of 24 temperature sensors are distributed inside the column sections and make up the majority of measurements.

The temperature measurements will be used in estimating the composition profile of the column and the purity of the products. Product samples will be analyzed with gas chromatography offline to facilitate tuning and validation of the estimates.

The initial experiments will be run with a mixture of alcohols (methanol, ethanol, propanol and butanol), however, there are also plans to run the column with an alkane-mixture.

2.1 Modelling

The Kaibel column is modelled using a stage-by-stage model with the following simplifying assumptions: Constant pressure, equilibrium stages and constant molar flows. The vapour-liquid equilibrium is modelled using the Wilson equation of state. To model the column we have used 7 column sections with stages (see Figure 2). Section 1 and 2 make up the prefractionator, while the main column consists of sections 3-7. The prefractionator sections have 12 equilibrium stages, while sections 3-7 each have 8 equilibrium stages (our design necessitates the prefractionator having the same height as sections 4-7-5).

Table 1. Nominal operating point. (Flows are scaled with regards to the feed. R_l and R_v are ratios)

Variable	Nominal value
L	2.7864
V	2.5107
$S1$	0.2437
$S2$	0.2530
R_l	0.3013
R_v	0.3233
D	0.2473
B	0.2560

In this study, we separate the same four-component mixture that will initially be used in the pilot plant experiments (methanol, ethanol, propanol and butanol). We use an equimolar feed with partial preheating (q=0.48).

We have specified 4 product purities:

$$\begin{bmatrix} x_D \\ x_{S1} \\ x_{S2} \\ x_B \end{bmatrix} \geq \begin{bmatrix} 0.975 \\ 0.94 \\ 0.94 \\ 0.975 \end{bmatrix}$$

The nominal operating point for the column have been found by optimization. The optimization criterion was to minimize the vapour boil-up, V (minimum energy input), with the model equations as equality constraints and the product purities as inequality constraints.

$$\begin{aligned} \min_{x,u} J &= V \\ \text{s.t.} & \\ f(\dot{x}, \dot{u}) &= 0 \\ h(x) &\leq 0 \end{aligned} \tag{1}$$

Data for the nominal optimum can be seen in Table 1. For the time being we decide to keep the vapour split R_v constant, at its optimal value.

3. CONTROL OF KAIBEL COLUMN

The column has 7 dynamic degrees of freedom (valves): L, V, $S1$, $S2$, R_l, D and B. In addition, it may be possible to adjust the vapour split, R_v, but this is not studied here. Column pressure is self-controlled at atmospheric pressure by the condenser design which has an open vent.

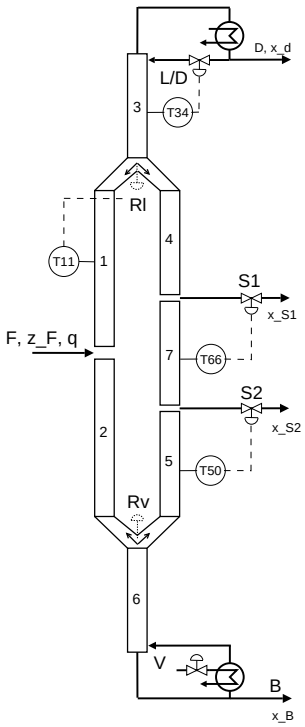


Fig. 2. Stabilizing control scheme (c_3) with four temperature loops (Valves are shown on individual streams for L/D and R_l, but in reality these are implemented as ratios using magnetically operated swinging funnels).

3.1 Stabilizing control of levels

The condenser and reboiler holdups need to be controlled. We choose to use the "L/D V-configuration" where the condenser level is controlled such that L/D remains as a degree of freedom in the top of the column and the reboiler level is controlled (using B) such that V remains as a degree of freedom in the bottom of the column.

We are now left with 5 degrees of freedom (L/D, V, $S1$, $S2$, R_l), and we need to use at least 4 of these to stabilize the column profile.

3.2 Stabilizing control of column profile

In order to avoid "drift" in the column with undesirable breakthrough of impurities in the product we need to stabilize the column profile. First, in the prefractionator we need to maintain the split between components B and C. This may be done using R_l to control some temperature in the prefractionator, probably located in the top for good dynamic response. In the main column, we need to maintain the split between A and B in the top, B and C in the middle, and C and D in the bottom. This requires closing three additional temperature loops, for example using L/D to control a temperature in the top, $S1$ to control a temperature in the middle section, and $S2$ to control a temperature in the bottom (See Figure 2). V and R_v remain unused. These loops need to be relatively fast and since composition measurements are usually slow or not available, we propose to use temperature loops, if possible with composition control in the outer cascade.

3.3 Location of temperature sensors

The objective of the inner loops is mainly to stabilize the column. But, in addition, we would like to keep the column reasonably close to its optimal operation, which is to minimize the energy usage (V), while satisfying the four product purity constraints. One way of achieving this is the "self-optimizing control" approach (Skogestad, 2000). We will use parts of this approach to select the optimal location of the temperature measurements.

We here apply the minimum singular value method (Halvorsen et al., 2003) for selecting the controlled variables. The procedure consists of the following steps.

(1) Obtain a linear model G from the inputs u to the candidate controlled variables y.
(2) Scale the inputs u such that the effect of each input is the same on the objective function.
(3) Obtain the scaled gain G_s of the model by scaling the outputs using sum of their optimal range and their implementation error ("span").
(4) Select controlled variables that maximize the minimum singular value $\underline{\sigma}$ of the scaled gain matrix G_s from u to y.

In the following we will assume that we have an on-line measurement of the bottoms product composition but that the other product streams do not have this feature. The available measurements (candidate controlled variables) are then the temperature at each stage and the bottoms composition, x_B. We also include our 5 inputs as possible outputs in the analysis. We then have a total of 71 candidate variables from which we want to find the best set c of 5 variables to keep at constant set-points.

$$y = \{L/D, V, S1, S2, R_l, T_1 ... T_{65}, x_B\} \quad (2)$$

The implementation errors used to obtain the scaled matrix were: 10% for flow measurments, 0.5 K for temperatures and 0.001 for the mole fraction

Using the exact branch and bound method (Cao et al., 1997) we find that we should pick the following five variables:

$$c_1 = [T_{24} \quad T_{41} \quad T_{56} \quad T_{68} \quad x_B] \quad (3)$$

These five give a minimum singular value $\underline{\sigma} = 113.53$. We see that the method chooses x_B, which one would expect to be a good variable.

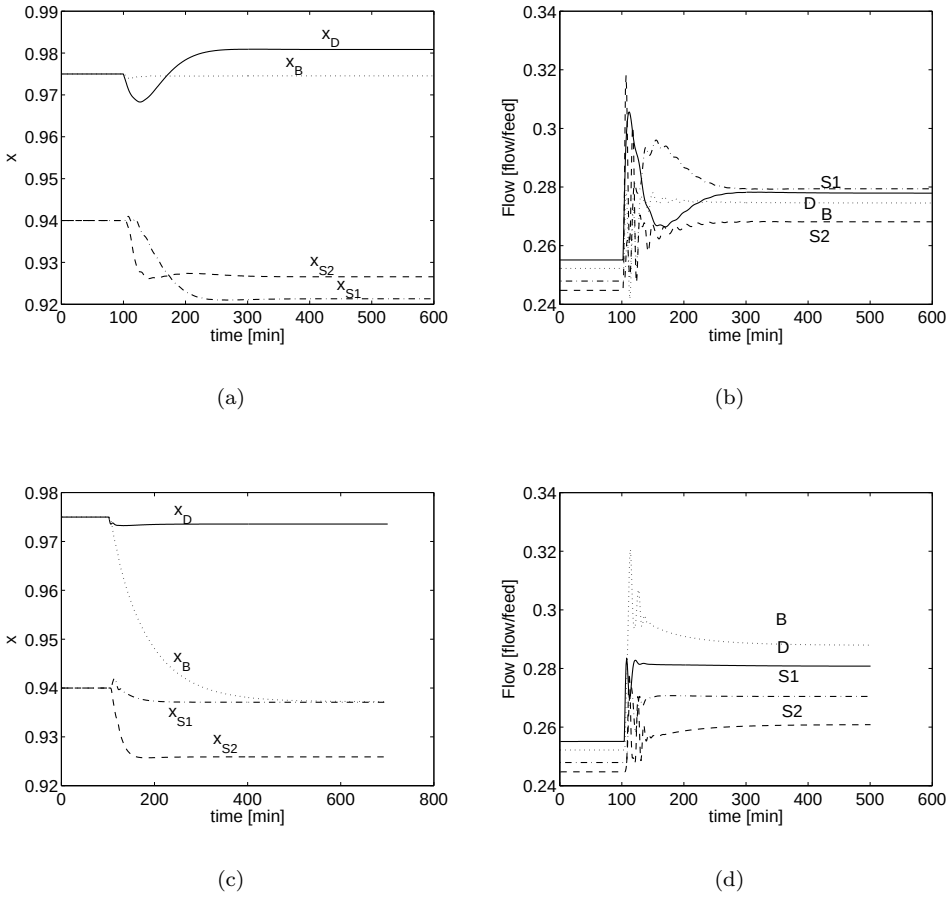


Fig. 4. Disturbance response. 10% increase in feed rate. Top: response for controlled set c_2. Bottom: response for controlled set c_3.

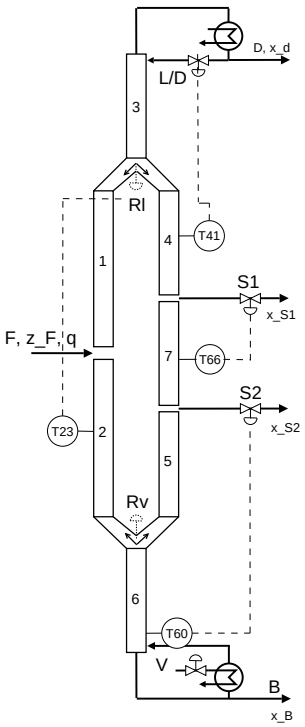


Fig. 3. Control scheme c_2

Now, we will investigate the strategy where we have constant vapour boil-up (we pick V as both input and output) and look for the the best 4 remaining outputs. We then get the following controlled variables:.

$$c_2 = [V \quad T_{23} \quad T_{41} \quad T_{60} \quad T_{66}] \tag{4}$$

The location of the measurements are visualized in Figure 3. We see that the method does not choose the same outputs as before, even though some of them are retained. Here the minimum singular value $\underline{\sigma}$ is reduced to 1.5265, indicating that this it is not an optimal strategy to keep V constant.

If we now go back to our preliminary control structure seen in Figure 2, we can evaluate the minimum singular value for choosing these outputs:

$$c_3 = [V \quad T_{11} \quad T_{34} \quad T_{50} \quad T_{66}] \tag{5}$$

This gives $\underline{\sigma} = 1.200$ which is somewhat lower than for c_2

3.4 Loss calculations

In Table 2 we show the percentage loss in the cost function when disturbances are introduced. We

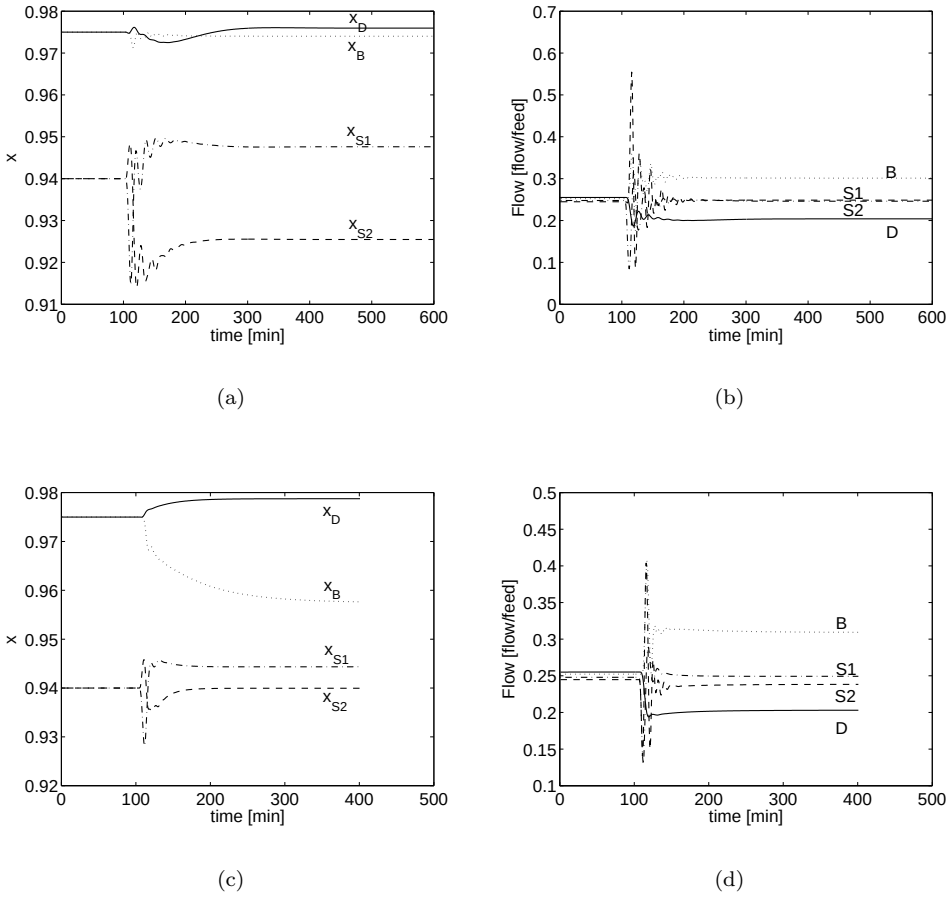


Fig. 5. Disturbance responses. z_A=0.20, z_D=0.30. Top response for controlled set c_2. Bottom: response for controlled set c_3.

Table 2. Loss in cost function due to disturbances when temperature loops are applied.

Disturbance	c_2	c_3
F: 1.0 → 1.1	0.39%	0.39%
z_A: 0.25 → 0.20	12.0%	26.2%
z_B: 0.25 → 0.20	3.32%	10.5%
z_C: 0.25 → 0.20	0.12%	1.04%
z_D: 0.25 → 0.20	0.03%	2.10%

keep the sets (c_2 and c_3) of temperatures at their optimal values and increase V to reach feasibility with respect to the purity constraints. We see that c_2 does indeed produce smaller losses than c_3 as indicated by the singular value.

4. DYNAMIC SIMULATIONS

The two control schemes, c_2 and c_3 were simulated to check the dynamic responses. In each case the four temperature loops were implemented using PI-controllers tuned with Skogestad's IMC tuning rules (Skogestad, 2003). Figure 4 shows the responses from a 10% increase in the feed rate. We note that scheme c_2 does better with regards to the bottoms composition, while c_3 has a somewhat better response in terms of the sidestream compositions.

In Figure 5 the feed composition of component A,(z_A), has been decreased from 0.25 to 0.20, while z_D is increased from 0.25 to 0.30. Again we see that c_3 keeps the sidestream compositions better than c_2, but the bottom composition drifts relatively far away with the set c_3 as compared to the set c_2 found by the singular value method.

To get back to the specified purities we can adjust the boil-up V.

5. EXPERIMENTAL DATA

The experimental work is in progress and will be reported in the conference presentation.

6. CONCLUSIONS

In this paper we have discussed the Kaibel distillation column for the separation 4 products. The Kaibel column is interesting because it has the potential for large capital investment savings as

well as reduced energy consumption when compared to conventional distillation sequences. To stabilize the column operation one has to close a total of 4 control loops. The remaining degrees of freedom (in our case V and potentially R_V) can be used to ensure optimal operation according to some economic objective.

The minimum singular value method has been applied to the problem of finding the optimal location of temperature measurements for stabilizing control. The resulting control scheme has been compared with a predefined scheme using both steady-state and dynamic simulations, showing that the method can be useful in selecting measurements.

REFERENCES

Adrian, T., H. Schoenmakers and M. Boll (2003). Model predictive control of integrated unit operations: Control of a divided wall column. CHEMICAL ENGINEERING AND PROCESSING **43**(3), 347–355.

Cao, Y., D. Rossiter and D.H. Owens (1997). Globally optimal control structure selection using branch and bound method. Submitted for Dycops '98.

Halvorsen, I. J. (2005). Minimum energy for the 4-product kaibel column. Technical report. Dept. of Chemical Engineering, NTNU, Norway.

Halvorsen, I. J., S. Skogestad, J. C. Morud and V. Alstad (2003). Optimal selection of controlled variables. INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH **42**(14), 3273–3284.

Kaibel, G. (1987). Distillation columns with vertical partitions. Chem. Eng. Technol. **10**, 92–98. Petluyk.

Kaibel, G., C. Miller, M. Stroezel, R. von Watzdorf and H. Jansen (2004). Industrial application of dividing wall distillation columns and thermally coupled distillation columns. CHEMIE INGENIEUR TECHNIK **76**(3), 258–+.

Petlyuk, F. B., V. M. Platonov and D.M. Slavinskii (1965). Thermodynamically optimal method for separating multicomponent mixtures. International Chemical Engineering **5**(3), 555–561.

Skogestad, S. (2000). Plantwide control: the search for the self-optimizing control structure. Journal of Process control **10**, 487–507. H2000-10.

Skogestad, S. (2003). Simple analytic rules for model reduction and pid controller tuning. J. of Process Control **13**, 291–309. See C03-1; Nordic Process Control (NPC) Workshop 11, January 9-11, Trondheim, 2003; pages 209-227.

Wright, R.O. (1949). Fractionation apparatus. US Patent No. 2471134.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DYNAMICS AND CONTROL OF HEAT INTEGRATED DISTILLATION COLUMN (HIDIC)

Tomohiro Fukushima * **Manabu Kano** *
Osamu Tonomura * **Shinji Hasebe** *

* *Department of Chemical Engineering, Kyoto University, Kyoto 615-8510, Japan*

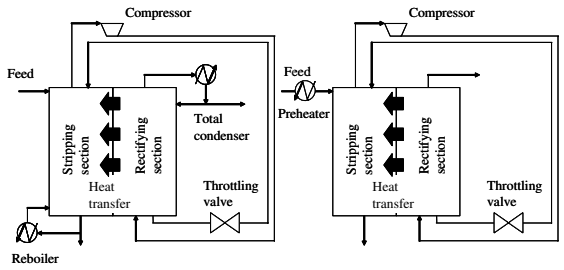
Abstract: A heat integrated distillation column (HIDiC) is a new and highly energy-efficient distillation process. In this work, dynamic simulation models for several types of HIDiCs were developed. The dynamics and controllability of HIDiCs were investigated and compared with those of a conventional distillation column (CDiC). HIDiC has a more complex structure and slower dynamics than CDiC. However, the control performance of HIDiC is comparable to that of CDiC as far as a suitable control system is designed. In addition, an industrial HIDiC plant in Japan was rigorously modeled, and its dynamics and control issues are discussed.

Keywords: Heat integrated distillation column, Energy saving, Process control, Dynamics, Controllability

1. INTRODUCTION

As global warming becomes a more serious problem, the demand to suppress the exhaust of greenhouse gas has increased and technology development to achieve energy saving in industries has been promoted. Since distillation is the most widely-used separation process, quite energy-intensive, and accounts for a large part of energy consumption in industries, the development of an energy-efficient distillation process is crucial.

A heat integrated distillation column (HIDiC) is an energy-efficient distillation column, that has the potential for drastic reduction of energy consumption (Takamatsu et al., 1996). The basic concept of HIDiC is that heat duty needed in a reboiler and a condenser can be reduced simultaneously by enhancing internal heat integration. In the last decade or so, basic characteristics and energy savings of HIDiC have been investigated vigorously. These researches include exergy-based analysis of energy savings of ideal HIDiC (Takamatsu et al., 1997), analysis of energy savings in a multicomponent separation process (Iwakabe et al., 2004), and detailed design in which material transfer rate, heat transfer rate, and pressure drop are taken into account (Noda et al., 2004). In addition, an industrial HIDiC plant has been built and operated to prove its usefulness and also to investigate practical issues. However, past research on HIDiC has focused mainly on its static characteristics. Little research on the dynamics and controllability of HIDiC has been carried out, while internal heat integration and a complex structure of HIDiC might make its operation more difficult than conventional distillation columns. In particular, a new control scheme must be developed for ideal HIDiC, because it does not have both a reboiler and a condenser and thus reflux flow rate and



(a) HIDiC (b) ideal HIDiC

Fig. 1. Schematic diagrams of HIDiCs.

reboiler heat duty cannot be used as manipulated variables.

For practical applications of HIDiC, its controllability needs to be investigated and an appropriate control system needs to be developed. In the present work, a dynamic simulator for HIDiC and ideal HIDiC is developed. By using the developed simulator, static characteristics, especially energy savings, of HIDiCs are investigated, and various multiloop control structures have been studied to clarify a control strategy suitable for HIDiCs. In addition, an industrial HIDiC plant in Japan was rigorously modeled, and its dynamics and control issues are discussed.

2. DYNAMIC MODEL OF HIDIC

In this section, the structures of HIDiC and ideal HIDiC are shown and the developed dynamic models are briefly described.

2.1 Structure of HIDiC

A schematic diagram of HIDiC is shown in Fig. 1(a). HIDiC has a compressor and a throttling valve between the bottom of the rectifying section and the top of the stripping section. Vapor rising from the top of the stripping section is pressurized by the compressor and supplied to the bottom of the rectifying section. Liquid flowing from the bottom of the rectifying section is supplied to the top of the stripping section through a throttling valve. The feed is supplied to the top of the stripping section.

In a conventional column, heat is transferred by a reboiler and a condenser. In HIDiC, on the other hand, the rectifying section and the stripping section are in physical contact, and the pressure in the rectifying section is kept higher than that in the stripping section by using a compressor to enhance heat transfer from the rectifying section to the stripping section through the wall. By this internal heat transfer, vapor in the rectifying section condenses and liquid in the stripping section evaporates. As a result, the heat duty needed in the reboiler and the condenser can be reduced and high energy saving can be achieved. Although a compressor is required in HIDiC, energy consumption by the compressor is less than that reduced in both the reboiler and the condenser. Therefore, HIDiC is more energy-efficient than conventional distillation columns.

Fig. 1(b) shows a schematic diagram of an ideal HIDiC (i-HiDIC) that does not have both a reboiler and a condenser. To achieve an appropriate heat balance, that is, to operate the column without a reboiler and a condenser, the feed needs to be preheated before entering into the column. In the present work, an i-HiDIC, in which the feed is preheated by distillate vapor (product), was also investigated to enhance the energy saving; this type of i-HiDIC is referred to as i-HIDiC (HX). In addition, to improve the controllability of i-HiDIC, i-HiDIC with a condenser is also investigated; this type of i-HiDIC is referred to as i-HIDiC (L). In most cases, cooling water at normal temperature is used in a condenser. Therefore, the use of a condenser does not deteriorate energy-efficiency of i-HiDIC.

2.2 Dynamic Model

A dynamic simulator was developed by using ASPEN Custom Modeler®. In the simulation model, the mass balance, energy balance, and pressure drop are taken into account. Changes in liquid holdups are calculated by using the Francis weir equation. The other assumptions are as follows:

(1) Tray column is used.
(2) Liquid and vapor on each tray are perfectly mixed and in equilibrium.
(3) Vapor holdup is negligible.
(4) In HIDiC, the rectifying section and the stripping section have the same number of trays and exchange heat between the corresponding trays.
(5) Heat transfer is calculated by $UA\Delta T$ where U is overall heat transfer coefficient, A is heat transfer area, and ΔT is temperature difference between the rectifying section and the stripping section.

2.3 Evaluation of Energy Saving

By using the developed simulator, energy savings of a conventional distillation column (CDiC), HIDiC, and i-HIDiC are investigated under the same separation condition. A standard operating condition is summarized in Table 1. In HIDiC

Table 1. Standard operating condition.

No. of stages	30
Feed stage (top of stripping section)	16
Feed flow rate [kmol/h]	100
Feed temperature [°C]	87
Feed composition Benzene/Toluene	0.5/0.5
Distillate composition (Benzene) [mol%]	99.9
Bottoms composition (Toluene) [mol%]	99.9
ΔP [atm]	1.8
UA [kcal/(h K tray)]	7500

and i-HIDiC, the pressure difference between two sections ΔP is kept at 1.8 atm by a compressor.

The results are shown in Table 2. Heat duty reduction in a condenser does not lead to energy saving directly because the heat can be recovered and inexpensive coolant can be used. Therefore, heat duty in a condenser is not included in the comparison of energy consumption. Energy consumption in HIDiC is 32% less than that in CDiC, and i-HIDiC is more energy-efficient than HIDiC even if the heat duty for preheating feed is considered. Furthermore, energy consumption in i-HIDiC (HX) is 70% less than that in CDiC.

3. CONTROLLABILITY ANALYSIS

Multiloop control systems were designed for various column structures and the control performance was evaluated.

3.1 Control System Design

To evaluate the controllability of various column structures, CDiC and HIDiC were regarded as 2×2 processes. The outputs, i.e., controlled variables, are the mole fraction of benzene in the distillate (x_D) and that of toluene in the bottoms (x_B). On the other hand, the inputs, i.e., manipulated variables, depend on structures. In CDiC, there are three manipulated variables: distillate flow rate (D), reflux flow rate (L), and reboiler heat duty (Q_r). Reflux ratio is investigated later in section 4. In HIDiC, compressor duty (C) is added to these three variables. In i-HIDiC, compressor duty (C) and preheater heat duty (Q_f) are manipulated variables. In i-HIDiC (HX), compressor duty (C) and vapor flow rate supplied to the heat exchanger (V_f) are manipulated variables. In i-HIDiC (L), reflux flow rate (L) is added to these two variables. It was assumed that the pressure at the top of a column and liquid levels of both a reflux drum and a column bottom are perfectly controlled and kept constant at their setpoints.

First, step response tests were conducted and multi-input multi-output (MIMO) transfer function models were identified for each column. The processes were approximated to first-order or second-order lag models. In the step response tests, each manipulated variable was changed in both positive and negative directions to evaluate the degree of nonlinearity that is usually observed when high purity distillation is investigated. To suppress the nonlinearity and derive approximated linear models, the following transformed composition (x^*) was used:

$$x^* = \log \frac{1-x}{1-\tilde{x}} \tag{1}$$

where $\tilde{x}$ is setpoint for x.

In the present work, multiloop control systems were designed for various column structures. Multivariable control was not used here because the objective of the analysis is to investigate the characteristics of various HIDiC structures, not to maximize the control performance. To design multiloop control systems, all possible pairings of controlled and manipulated variables were enumerated, and suitable pairings were selected on the basis of relative gain array (RGA). The relative gain is an index which evaluates process interaction and is calculated from process steady state gains. Here, λ_{ij} is the relative gain which relates the jth manipulated variable and the ith controlled variable. In the case of $\lambda_{ij} \approx 1$, the process interaction is weak and thus it is desirable to choose this pairing. The pairing of $\lambda_{ij} < 0$ should not be selected; rather the pairing of $\lambda_{ij} > 1$ should be selected for 2×2 processes. The selected pairings and the corresponding relative gain values are summarized in Table 3.

After control pairing selection, controllers were designed on the basis of the internal model control (IMC) method. For first-order and second-order lag models, PI and PID controllers are derived, respectively. Both integral time and derivative time are determined automatically from a transfer function model. In this work, a derivative mode filter was used to avoid excessive derivative action. On the other hand, to determine proportional gain, it is necessary to tune the IMC filter time constant τ, which corresponds to the closed-loop time constant. The time constant τ was determined by conducting rigorous dynamic simulations so that ISE becomes as small as possible for both a disturbance and a setpoint change.

3.2 Evaluation Measures

To evaluate the control performance, relative disturbance gain (RDG) and integral error under multiloop control (IEML) are used together with integral squared error (ISE). The controllability

Table 2. Comparison of energy consumption.

	Reboiler	Compressor	Preheater	Total [GJ/h]	HIDiC/CDiC [%]
CDiC	4.49			4.49	100
HIDiC	2.16	0.903		3.06	68
ideal HIDiC		1.35	0.816	2.17	48
ideal HIDiC (HX)		1.35		1.35	30

of a multivariable process is usually dominated by RDG, which is given as the product of relative gain and a disturbance factor. The combined effects of inherent process interaction and disturbance type determine the dominant difference between single-loop and multiloop control performance. RDG becomes small when the control performance is good.

To evaluate the control performance for specific disturbances, IEML can be calculated directly from transfer function models of a process and a disturbance (Stanley et al., 1985). IEML of the controlled variable y_1 is given by

$$\begin{aligned} \mathrm{IEML}_1 &\equiv \left[\int_0^\infty E_1(t)dt\right]_{ML} \\ &= \left[\int_0^\infty E_1(t)dt\right]_{SL} f_{1,tune}\mathrm{RDG}_1 \quad (2) \\ \mathrm{RDG}_1 &= \lambda_{11}\left(1-\frac{K_{d2}K_{12}}{K_{d1}K_{22}}\right) \quad (3) \end{aligned}$$

where E_1 denotes an error, K_{di} steady-state gain of a disturbance for the ith controlled variable, respectively. Subscripts SL and ML mean single-loop and multiloop control, respectively. $f_{1,tune}$ is a detuning factor for multiloop control.

3.3 Results and Discussions

To compare the controllability of various column structures, setpoint changes of product compositions and disturbances in feed flow rate and feed composition were investigated. These variables were changed stepwise ±5% from their steady-state values. The results of ISE are summarized in Table 3. Due to space limitation, only ISE results are shown here.

In CDiC, the control performance of the pairing x_D-L and x_B-Q_r is better than the other for the disturbances (a and b), and that of the pairing x_D-D and x_B-Q_r is better for the setpoint changes (c and d). In HIDiC, the control performance of the pairing x_D-L and x_B-Q_r is considerably worse than the others for the disturbance and the setpoint change (a, c, and d). Therefore, it is recommended to use the pairing x_D-D and x_B-Q_r or the pairing x_D-D and x_B-C, the control performance of which is as good as that of CDiC. In addition, the pairing x_D-$Q_r(L)$ and x_B-C can

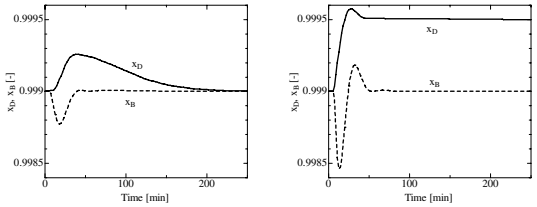


Fig. 2. Control responses of CDiC with (x_D-D, x_B-Q_r) control structure.

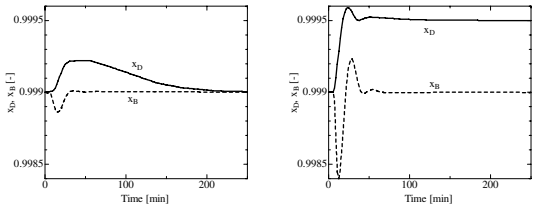


Fig. 3. Control responses of HIDiC with (x_D-D, x_B-Q_r) control structure.

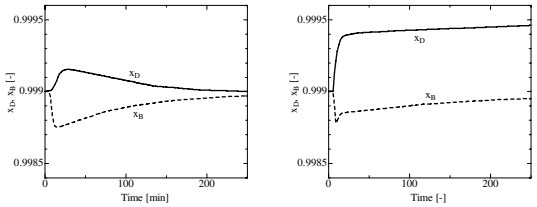


Fig. 4. Control responses of ideal HIDiC (HX) with (x_D-V_f, x_B-C) control structure.

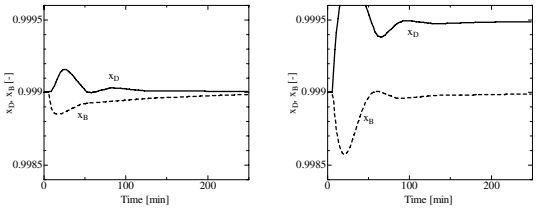


Fig. 5. Control responses of ideal HIDiC (L) with (x_D-L, x_B-C) control structure.

achieve the best control performance of all for the disturbances (a and b); however, this pairing makes the control performance worse for the setpoint changes (c and d). Here, $Q_r(L)$ means the use of Q_r as a manipulated variable under the condition that L is kept constant. Since D cannot be kept constant for satisfying material balance, $Q_r(D)$ cannot be chosen. In i-HIDiC and i-HIDiC (HX), there is only one candidate of pairing. The control performance of both i-HIDiCs is as good as that of CDiC and HIDiC for the disturbances (a and b), but it is worse for the setpoint changes (c and d). In i-HIDiC (L), it is recommended to use the pairing x_D-L and x_B-C or the pairing x_D-D

Table 3. Assessment of control performance. (a: Feed rate change, b: Feed composition change, c: Setpoint change of x_D, d: Setpoint change of x_B)

Type	Sub–Type (A, ΔP)	Pairing x_D	Pairing x_B	Relative gain λ		a	b	c	d
CDiC		L	Q_r	11.4	ISE	0.00908	0.001	0.011	0.004
		D	Q_r	0.569	ISE	0.018	0.016	0.008	0.003
HIDiC	(12, 1.8)	L	Q_r	245	ISE	0.101	0.011	0.094	0.331
		L	C	11.9	ISE	0.019	0.017	0.018	0.035
		$Q_r(L)$	C	12.5	ISE	0.008	0.006	0.017	0.022
		D	Q_r	0.568	ISE	0.015	0.014	0.008	0.003
		D	C	0.582	ISE	0.015	0.014	0.011	0.004
ideal HIDiC	(12, 1.8)	Q_f	C	8.90	ISE	0.022	0.011	0.012	0.023
ideal HIDiC (HX)	(12, 1.8)	V_f	C	14.2	ISE	0.015	0.017	0.033	0.036
ideal HIDiC (L)	(12, 1.8)	L	C	4.67	ISE	0.004	0.005	0.037	0.013
		D	C	0.582	ISE	0.004	0.005	0.037	0.012
		V_f	C	0.582	ISE	0.014	0.015	0.054	0.032
HIDiC	(12, 1.8)	D	C	0.582	ISE	0.015	0.014	0.011	0.004
	(6, 1.8)	D	C	0.576	ISE	0.021	0.013	0.009	0.003
	(15, 1.8)	D	C	0.582	ISE	0.044	0.042	0.012	0.005
	(12, 1.9)	D	C	0.577	ISE	0.018	0.017	0.015	0.009
	(12, 1.7)	D	C	0.582	ISE	0.015	0.014	0.010	0.004

and x_B-C. As expected, the control performance can be improved by using a condenser.

Control responses for the disturbance in feed flow rate (left) and the setpoint change of the distillate product composition (right) are shown in Figs. 2, 3, 4, and 5. The control responses of CDiC and HIDiC are very similar to each other, and the results reveal that HIDiC can be controlled in the same way as CDiC regardless of the complexity of HIDiC. In addition, control responses of x_D are slower than those of x_B in both CDiC and HIDiC, because the composition in the reboiler is affected immediately by changing Q_r, whereas the composition in the reflux drum is not affected directly by changing D. On the other hand, the control response of i-HIDiC (HX) is different from that of CDiC and HIDiC. It takes a long time in i-HIDiC (HX) for both x_D and x_B to settle at the steady state because i-HIDiC (HX) does not have a reboiler and a condenser. By using a condenser, the control performance of i-HIDiC (L) is improved.

The effects of internal heat transfer area (A) and pressure difference (ΔP) of HIDiC on the control performance were also investigated. The internal heat transfer area of each stage changed from 12 m^2 (benchmark) to 6 and 15 m^2. As a result, the ratio of reboiler heat duty in HIDiC to that in CDiC changed from 50% (benchmark) to 75 and 33%. The pressure difference changed from 1.8 atm (benchmark) to 1.7 and 1.9 atm. The pairing is x_D-D and x_B-C through this investigation. The results show that the control performance of HIDiC becomes worse as the heat transfer area increases. In addition, the control performance of HIDiC becomes slightly worse as the pressure difference increases.

Table 4. Feed and product compositions (wt%) of industrial HIDiC.

	Feed	Distillate	Bottoms
n-butane	0.02	0.16	
i-pentane	1.09	8.77	
n-pentane	11.17	89.76	0.02
2,2-dimethlbutane	0.43	0.02	0.49
cyclopentane	39.00	1.30	44.35
2,3-dimethylbutane	2.19		2.50
2-methylpentane	19.18		21.90
3-methylpentane	7.18		8.20
n-hexane	10.81		12.34
methylcyclopentane	8.71		9.95
cyclohexane	0.22		0.25

Table 5. A standard operating condition of industrial HIDiC.

No. of stages	70
Feed stage (top of stripping section)	36
Feed flow rate [kmol/h]	1286
Feed temperature [°C]	288.1
Distillate composition (2,3-DMB) [wt%]	1.3
Bottoms composition (2,2-DMB+CP) [wt%]	0.509
ΔP [atm]	1.8
UA [kcal/(h K tray)]	1533

Table 6. Energy consumption.

	Total [Mcal/h]	HIDiC/CDiC [%]
simulated CDiC	231	100
simulated HIDiC	174	76
simulated i-HIDiC	142	62
simulated i-HIDiC (HX)	135	58
industrial HIDiC	165	71

4. ANALYSIS OF INDUSTRIAL HIDIC

The feed and standard operating condition of the industrial HIDiC is summarized in Tables 4 and 5. It was confirmed that the developed model can describe the steady state of the industrial HIDiC with sufficient accuracy. The energy-efficiency comparison results are shown in Table 6.

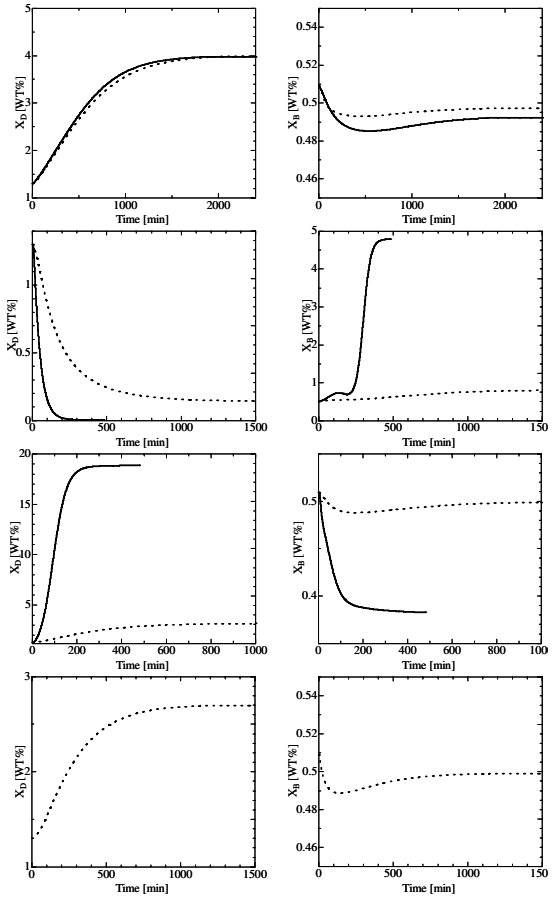


Fig. 6. Simulated step responses of industrial HIDiC and CDiC. (Solid line: CDiC, Dashed line: HIDiC. Step changes in D, L, $Q_r(L)$, and $C(L)$ from the top to the bottom.)

Simulated step responses of the industrial HIDiC are shown in Fig. 6. Although the responses of product compositions in HIDiC from changes in L and $Q_r(L)$ are slower than those in CDiC, the difference between HIDiC and CDiC is small when D is changed. Therefore, control of HIDiC can be easier by using D instead of L as a manipulated variable. In addition, strange responses, which seem to consist of two stages, is observed in x_B when L is changed. Such step responses occurs because physical properties of the key components in the bottoms, i.e., 2,2-dimethlbutane and cyclopentane, are quite different.

By using the developed dynamic simulator, the control performance of various HIDiCs and CDiC is investigated. The control performance of i-HIDiC without a condenser is considerably worse than that of CDiC especially for set-point changes. Therefore, i-HIDiC (L) is strongly recommended. As for control structures, the pairing x_D-V_f and x_B-$C(R)$ is the best of all. Here, R denotes reflux ratio. Although the control performance of i-HIDiC (L) is much better than CDiC for the feed flow rate disturbance, it is worse for the set-point change. To further improve the control performance, decoupling or multivariable control can be used. In the simulation, ISE can be reduced about 40% by using static decouplers.

5. CONCLUSION

In the present work, dynamic models for several types of Heat Integrated Distillation Columns (HIDiCs) were developed, and energy saving, dynamics, and controllability of HIDiC were investigated. In addition, a dynamic model was developed for simulating an existing industrial HIDiC plant to investigate its dynamics and control issues. Although HIDiC has a more complex structure than CDiC, the control performance of HIDiC is comparable to that of CDiC as far as a suitable control system is designed.

ACKNOWLEDGMENT

This research was partially supported by the New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

[1] Iwakabe, K., M. Nakaiwa, T. Nakanishi, K. Huang, Y. Zhu, and A. Rosjorde (2004). Analysis of the Energy Savings by HIDiC for the Multicomponent Separation. *APPChE*, CD-ROM, 0259, Kitakyushu, Japan, Oct. 17-21.

[2] Noda, H., N.Kuratani, T. Mukaida, M. Kaneda, K. Kataoka, H. Yamaji, and M. Nakaiwa (2004). Plate Efficiency and Heat Transfer Characteristics in Heat-Integrated Distillation. *APPChE*, CD-ROM, 0017, Kitakyushu, Japan, Oct. 17-21.

[3] Stanley, G., M. Marino-Galarraga, and T. McAnoy (1985). Short-cut Operability Analysis: 1. The Relative Disturbance Gain. *IEC Proc. Des. Devel.*, **24**, 1181–1188.

[4] Takamatsu, T., M. Nakaiwa, and T. Nananishi (1996). The Concept of an Ideal Heat Integrated Distillation Column (HIDiC) and its Fundamental Properties. *J. Chem. Eng. Japan*, **22**, 985–990.

[5] Takamatsu, T., M. Nakaiwa, T. Nananishi, and K. Aso (1997). Possibility of Energy Saving in the Ideal Heat Integrated Distillation Column (HIDiC). *J. Chem. Eng. Japan*, **23**, 28–36.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

RIGOROUS SIMULATION AND MODEL PREDICTIVE CONTROL OF A CRUDE DISTILLATION UNIT

Gabriele Pannocchia [*,1] **Lorenzo Gallinelli** [*]
Alessandro Brambilla [*] **Gabriele Marchetti** [**]
Filippo Trivella [**]

[*] *Department of Chemical Engineering, Industrial Chemistry and Science of Materials – University of Pisa*
Via Diotisalvi 2, 56126 Pisa (Italy)
[**] *AspenTech S.r.l*
Lungarno Pacinotti 47, 56126 Pisa (Italy)

Abstract: This paper describes the application of a widely-used commercial multivariable predictive controller to a rigorously simulated crude distillation process. After describing the main process and controller features, it is shown how the two simulation and control environments can be interfaced together. A number of simulation results of typical product quality changes and crude switches are presented. The final goal of this paper is to demonstrate how rigorous dynamic simulators can be effectively used to reduce the costs of Advanced Process Control projects by shortening model identification, controller design and commissioning phases.

Keywords: Dynamic simulators, model predictive control, model identification, complex distillation processes.

1. INTRODUCTION

Process industries, such as the petroleum and chemical industries, face very dynamic and unpredictable market conditions, due to world-wide competition, limitation in natural resources, strict national and international regulations. In order to improve the production safety, quality and flexibility, plant automation has become increasingly important and is now recognized as a very effective way to achieve the production goals with satisfaction of safety and quality constraints.

Modern automation control systems for processing plants usually consist of a multi-level hierarchy of control layers. The first layer (starting from the bottom) is usually a distributed control system (DCS) which gathers process measurements, performs simple monitoring and PID-based control of some process variables (such as flow rates, levels, temperatures) to guarantee automatic operation of the plant. The second layer, usually referred to as Advanced Process Control (APC), performs multivariable model-based constrained control to achieve stable unit operation and push the process towards its operational limits for maximum economic benefits. APC regulators typically fall within the class of Model Predictive Control (MPC) algorithms (Morari and Lee, 1999; Mayne *et al.*, 2000; Qin and Badgwell, 2003). On top of APC other layers can be present, such as a Real-Time Optimization (RTO) layer and a Planning and Scheduling layer.

Reduction of costs for application of the second layer is a relevant issue since it would enlarge considerably the range of applicability of APC systems, which at present are mostly limited to capital intensive sectors, such as refinery and petrochemical industries. Qin and Badgwell (2003) reported nearly five thousand MPC applications all over the world, as a snap-shot of the situation in 1999, with a rough increase of about 80% in the subsequent three years. An MPC/APC project typically consists of a number of phases, such as:

[1] Corresponding author. Email: g.pannocchia@ing.unipi.it, Fax: +39 050 511266.

Table 1. TBP of the Zarzaitine crude oil.

Volume %	3.4	12.4	27.5	44.1	56.6	67.5
BP (^{o}C)	20	80	145	225	290	350

(1) A preliminary study comprising selection of manipulated (MV), controlled (CV) and disturbance (DV) variables, check of all instrumentation, and possible re-tuning of regulatory PID controllers.
(2) Plant testing and model identification in which MVs are varied and data of CVs (and DVs) are collected to build a process model, by means of identification techniques.
(3) Controller tuning, simulation and commissioning including selection of MVs and CVs limits and weights, open and closed-loop simulation on the identified plant model and final closed-loop implementation on the plant.

Phase 2 is particularly time-consuming and during plant testing (which may last several weeks) the products may violate some quality specifications. Phase 3 is also time-consuming although most of controller tuning and simulation needs not to be done "on-site".

Rigorous (steady-state and dynamic) simulators, i.e. those based on first-principles/fundamental equations, have become widely-used tools in process analysis, design and control (Yiu *et al.*, 1994; Berber and Coskun, 1996; Luyben and Tyreus, 1998; Huang and Riggs, 2002). In particular dynamic simulators can be useful to simplify Phase 2 and Phase 3 since they "surrogate" the true plant, thus allowing data collection for model identification, controller tuning and simulation. Moreover, it is important to remark that nowadays it is possible to carry out closed-loop model identification using an MPC regulator based on some preliminary model (e.g. one obtained from the data collected using the simulator). This approach can reduce dramatically the required plant testing duration and finally improve the controller performance by means of a more accurate model and more effective tuning.

In the present work, an industrially relevant example of a Crude Distillation Unit is simulated by means of HYSYS™ and controlled by using the commercial MPC algorithm DMCplus™. More details on this study can by found in (Gallinelli, 2005).

2. PROCESS AND CONTROLLER DESCRIPTION

2.1 Crude distillation unit

Crude oil is a mixture of a large number of components (whose exact determination is impossible), ranging from alkanes and iso-alkanes to cycloalkanes and aromatic compounds. Different oils are usually characterized in terms of density, often expressed in API degrees, and in terms of distillation curves, such as True Boiling Point (TBP), Equilibrium Flash Vaporization (EFV) or ASTM curves. In this study a Zarzaitine (Algerian) crude oil is considered, whose TBP data are reported in Table 1.

Crude distillation units (CDU) represent the core plant of any refinery site since most of its products are the

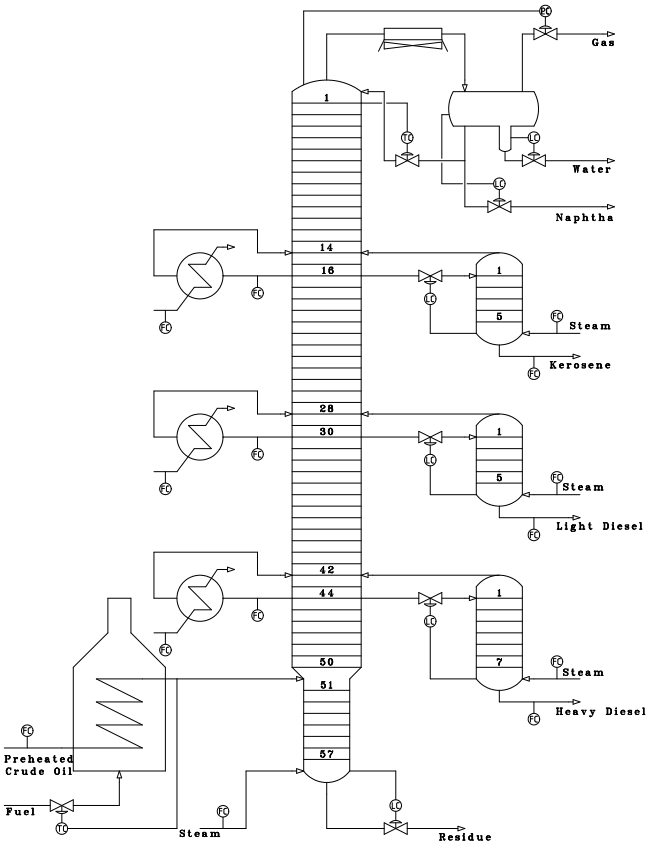


Fig. 1. CDU layout

starting point for a number of subsequent operations and final products. A typical CDU consists, mainly, of four operations:

(1) Desalting and pre-heating: before and after desalting, crude oil is pre-heated at expense of other hot streams available in the plant.
(2) Pre-flash: light components are vaporized in a flash drum to reduce the load at the furnace.
(3) Heating: crude oil is heated at high temperature (350÷390^{o}C) in a furnace.
(4) atmospheric distillation: crude oil is separated into a number of products (such as naphtha, kerosene, light diesel, heavy diesel and a residue) in a complex rectification column featuring side strippers and external coolers (pumparounds).

The CDU layout is depicted in Figure 1, in which only the furnace and the complex distillation column (with strippers and pumparounds) are shown for simplicity of representation. The main column top and bottom pressures are 1.9 atm and 2.8 atm, respectively; the reflux ratio is 1.5. All specifications can be found in (Gallinelli, 2005). Products are themselves mixtures of components and are characterized by given boiling temperature ranges. The boiling temperature ranges of products considered in the present study are reported in Table 2 along with the corresponding expected yield for the chosen crude.

2.2 DMCplus™ algorithm

In order to provide a brief description of the controller algorithm used in this work, it is assumed that a

Table 2. Boiling ranges for CDU products and expected yield.

Product	Boil. Range (°C)	Yield %
Naphtha	35÷150	24.5
Kerosene	150÷240	18.1
Light Diesel	240÷350	21.2
Heavy Diesel	350÷390	7.5
Residue	390÷548	28.7

(stable and proper) convolution model of the process is known. According to this model, the predicted value of the outputs (CVs) at time k given past values of the inputs (MVs) is:

$$\hat{y}_k = \sum_{i=1}^{k} S_i \Delta u_{k-i} \tag{1}$$

in which $\hat{y}_k \in \mathbb{R}^p$ is the predicted output vector at time k, $\Delta u_{k-i} \in \mathbb{R}^m$ is the input variation vector at time $k-i$ and $S_i \in \mathbb{R}^{p\times m}$ is the $i-$th matrix of step response coefficients from each input/output pair. Given the output measurement y_k, a correction term is then computed as:

$$d_k = y_k - \hat{y}_k \tag{2}$$

This term, which is meant to lump different sources of plant/model mismatch (such as disturbances, nonlinearities, noise), is used to guarantee offset-free control (Pannocchia and Rawlings, 2003). Other more general and effective correction terms can be considered (Muske and Badgwell, 2002; Pannocchia and Rawlings, 2003; Pannocchia and Brambilla, 2005).

The DMCplus™ controller, as well as most MPC algorithms, is based on two optimization modules (executed at each sampling time):

- A steady-state target optimizer, which computes optimal targets for inputs and outputs.
- A dynamic optimizer, which computes optimal trajectories for inputs and outputs from their current value towards the computed targets in a fixed-length time window (horizon).

The steady-state optimizer solves a linear program (LP) in the form:

$$\min_{\Delta \bar{u}_k} c^T \Delta \bar{u}_k \tag{3a}$$

subject to

$$u_{\min} \le \bar{u}_{k-1} + \Delta \bar{u}_k \le u_{\max} \tag{3b}$$
$$-\Delta \bar{u}_{\max} \le \Delta \bar{u}_k \le \Delta \bar{u}_{\max} \tag{3c}$$
$$y_{\min} \le G(\bar{u}_{k-1} + \Delta \bar{u}_k) + d_k \le y_{\max} \tag{3d}$$

in which $u_{\min}$ ($u_{\max}$) and $y_{\min}$ ($y_{\max}$) are vectors which contain the minimum (maximum) value for inputs and outputs, $\bar{u}_{k-1}$ is the previous input target vector, $\Delta \bar{u}_k$ is the input target variation vector, $\Delta \bar{u}_{\max}$ is maximum target variation vector, $G \in \mathbb{R}^{p\times m}$ is the model gain matrix and c is a vector of steady-state "costs". Solution of (3) yields the following input and output optimal targets:

$$\bar{u}_k = \bar{u}_{k-1} + \Delta \bar{u}_k, \quad \bar{y}_k = G(\bar{u}_{k-1} + \Delta \bar{u}_k) + d_k \tag{4}$$

It is clear that when it is desirable to maximize (minimize) an input, the corresponding cost should be chosen negative (positive). Possible infeasibility outcomes of (3), due to the output constraints (3d), are handled by iteratively softening output constraints (starting from low priority variables) and penalizing the corresponding variation in the objective function.

Given the optimal targets, the dynamic optimizer computes an optimal sequence of future input variations by solving the following quadratic program (QP):

$$\min_{\Delta u_k, \ldots, \Delta u_{k+N-1}} \sum_{j=k}^{k+N-1} \Delta u_j^T R \Delta u_j + \sum_{j=k+1}^{k+P} \left\{ e_j^T Q e_j + \eta_j^{uT} Q^u \eta_j^u + \eta_j^{lT} Q^l \eta_j^l \right\} \tag{5a}$$

subject to

$$e_j = \bar{y}_k - \hat{y}_{j|k} = \bar{y}_k - \left(\sum_{i=1}^{j} S_i \Delta u_{j-i} + d_k \right) \tag{5b}$$
$$u_{\min} \le u_{k-1} + \sum_{i=k}^{j} \Delta u_i \le u_{\max} \tag{5c}$$
$$-\Delta \bar{u}_{\max} \le \Delta u_j \le \Delta \bar{u}_{\max} \tag{5d}$$
$$y_{\min} - \eta_j^i \le \hat{y}_{j|k} \le y_{\max} + \eta_j^s \tag{5e}$$
$$\eta_j^i \ge 0, \quad \eta_j^s \ge 0 \tag{5f}$$

where: N and P are positive integers referred to as control and prediction horizon, respectively; R, Q, Q^u and Q^l are diagonal matrices with positive entries; u_{k-1} is the previous input vector; e_j is the vector of errors between target and future predicted outputs at time j; $\Delta u_{\max}$ is the maximum input variation vector; η_j^u and η_j^l are non-negative vectors which represent (possible) violations of upper and lower output constraints, respectively. It is important to remark that due to the output soft-constraint approach adopted, problem (5) is always feasible. Moreover, due to patent restrictions, DMCplus™ solves (5) in a suboptimal fashion. Given the "optimal" input sequence only the first "move" is implemented, i.e.

$$u_k = u_{k-1} + \Delta u_k \tag{6}$$

and both modules are re-executed at the next sampling time.

3. SIMULATION AND CONTROL ENVIRONMENT

In this section a description of the process simulation model built using HYSYS™ (version 3.2) and of the commercial controller DMCplus™ (version 6.0) used in this work is given.

3.1 Process simulation model

As remarked, crude oil exact composition is unknown; however this piece of information is necessary to simulate a crude distillation process. It is, therefore, com-

mon practice to represent the crude oil as a mixture of true components (light ends) and a number of pseudo-components. In the present work, seven light components (ranging from methane to n-pentane) and fifty pseudo-components are used.

After defined the crude composition and flow rate, the next step is to build a steady-state flow-sheet. Since the main interest of this work is to focus on the column dynamics and control, a number of "sensible" simplifications are made. First of all, the pre-heat train is simulated as a single heat exchanger; then, each pumparound is considered as a simple heat exchanger in which the hot fluid (column draw) flow rate and the exchanger duty are specified; finally, the cooling system is simulated as a single air cooler with a plant equivalent holdup.

Once a steady-state flow-sheet is defined, a number of steps are necessary to obtain a dynamic simulation model in HYSYS™.

(1) Design and sizing (holdup definition) of each equipment. In particular sieve trays are used throughout the column except for the draw stages where partial chimney trays are chosen.
(2) Specification of pressure and/or flow rate for a number of streams.
(3) Specification of the pressure profiles in each equipment.
(4) Addition of regulatory control loops (flow-rate, level, pressure and temperature controllers) to guarantee automatic operation of the column.

In the present work a further degree of simulation rigor is achieved by considering the static height of each equipment. With regards of the regulatory control loops, shown in Figure 1, PI controllers are used and tuned using IMC-like rules (Skogestad, 2003). It should be remarked that, although not shown, all level, temperature and pressure controllers are actually implemented in cascade on the corresponding flow-rate controllers.

3.2 DMCplus™ controller

In order to implement a predictive controller using the DMCplus™ software a list of manipulated and controlled variables is defined. In particular:

- 17 manipulated variables are considered: these variables are the setpoints of temperature, pressure and (non-cascaded) flow-rate controllers.
- 23 controlled variables are selected: these variables are the ASTM-D86 95% of the four main products (Naphtha, Kerosene, Light Diesel, Heavy Diesel) and the opening percent of all control valves.

Once the controller structure is defined, it is necessary to generate simulation data that can be used for the identification of the dynamic model matrix. This is done by imposing a series of setpoint changes to all the manipulated variables and observing their effect on the controlled variables. The size, direction and duration of the setpoint changes must be chosen so that significant changes are induced in the controlled variables and that all the typical operating conditions of the unit are explored. The sequence of moves can be pre-programmed and then the simulation can run unattended and the results are continuously collected and archived inside the HYSYS™ environment.

The data can then be exported from HYSYS™ and imported in DMCplus™ Model, the identification tool for DMCplus™ controllers. The next step is of course the analysis of the simulation data and the identification of the dynamic model: this can be done exactly as if the data were coming from a non-simulated step-test on a real process unit.

The identified model is then loaded in DMCplus™ Build to prepare the controller configuration file which will then be used by the on-line control engine. The tuning of a DMCplus™ controller is usually performed with the aid of its associated closed-loop simulation tool, Simulate. These simulations are based on the linear model that has been obtained during the model identification, and even if it is possible to introduce some extent of plant/model mismatch, it will always be difficult to closely reproduce the behavior of the controller once it will be applied to a real plant or, as in this case, to a rigorous dynamic simulation.

3.3 Simulator and controller interfacing

The interfacing of the dynamic simulation with the controller is performed with the use of the DMCplus™ block which is available in the HYSYS™ control libraries, together with PID or ratio controllers. This block is connected to the PID controllers which are manipulated variables in a simple master/slave cascade arrangement and is capable of reading the values of the controlled variables even if they are calculated values such as the ASTM qualities of selected streams.

When the HYSYS™ simulation is started with this block in place, the integrator runs until the time specified as the controller execution cycle (1 minute for the current application) has elapsed and pauses the simulation; it then passes the current values of manipulated, controlled and feed-forward variables (if present) to the DMCplus™ online control engine and waits for it to execute and return the new values for the manipulated variables. These values are applied in the HYSYS™ environment and the simulation is then started again for the time corresponding to one controller execution cycle.

4. SIMULATION RESULTS

A number of different closed-loop simulation studies were conducted to test the controller effectiveness in different situations, and to verify the flexibility of the proposed simulation and controller environment (Gallinelli, 2005). In this section some significant examples are presented.

Figure 2 shows the closed-loop results obtained for variations of the products' ASTM-D86 95% limits. In

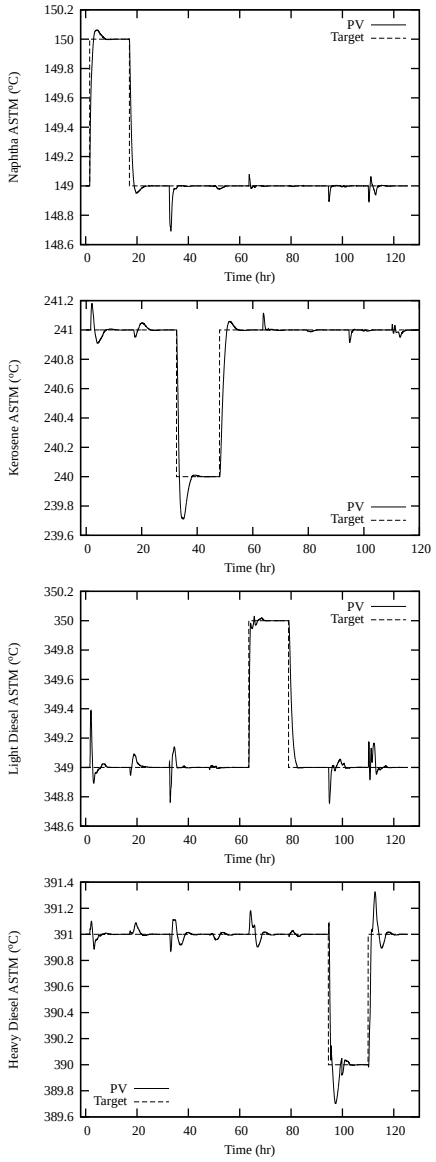


Fig. 2. Closed-loop results for variations of the products' ASTM-D86 95% limits: time behavior of products' ASTM-D86 95%.

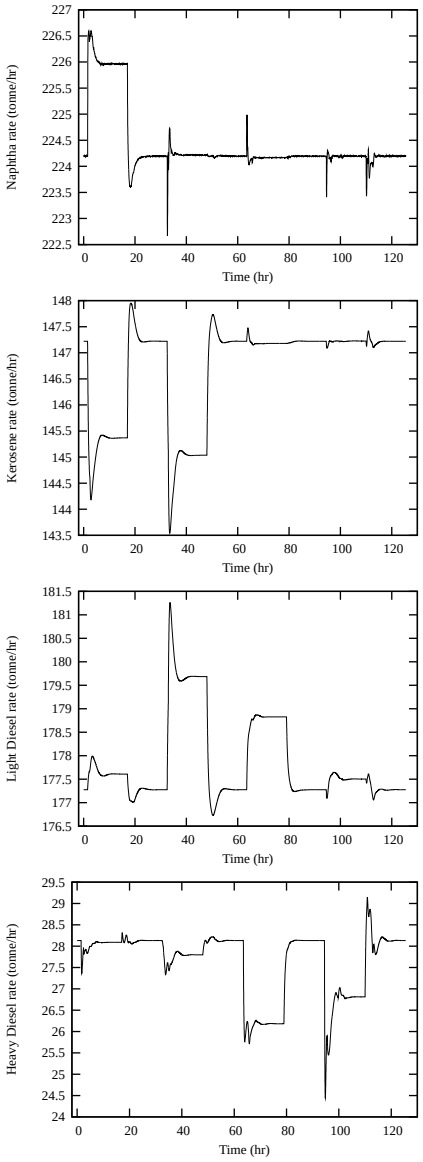


Fig. 3. Closed-loop results for variations of the products' ASTM-D86 95% limits: time behavior of products' flow rates.

each plot the time behavior of the "measured" controlled variables (ASTM-D86 95%) and of the corresponding steady-state targets calculated by the controller are reported. In Figure 3, instead, the corresponding time behavior of the flow rate of each product is reported.

A typical disturbance that occurs in refinery plants is associated to the crude switching. Starting from the original Zarzaitine crude oil, a new crude obtained by mixing the Zarzaitine oil with an Arabian Heavy one is fed to the CDU. For this case, closed-loop results of the products' ASTM-D86 95% are reported in Figure 4, while the corresponding product flow rates are reported in Figure 5.

5. CONCLUSIONS

In this paper, the design and study of a rigorous simulation model of crude distillation unit controlled by a commercial multivariable predictive controller has been presented. A crude distillation unit has been simulated using HYSYS^{TM} with a very high degree of accuracy (equipment design and sizing, pressure profiles, static heads, etc.). This rigorous dynamic model has been interfaced with a commercial controller, DMCplus^{TM}, whose (linear) process model was derived from data collected on the simulated plant. Closed-loop results of common setpoint changes and disturbance rejections showed the effectiveness of the implemented control algorithm.

The main contribution of this work is to emphasize the potential advantages of using rigorous simulators on complex processes of industrial relevance. In particular rigorous simulators can be effectively used to generate data for preliminary model identification and controller tuning. This "initial" controller can be implemented on the actual plant to carry out closed-loop identification tests from which the final process

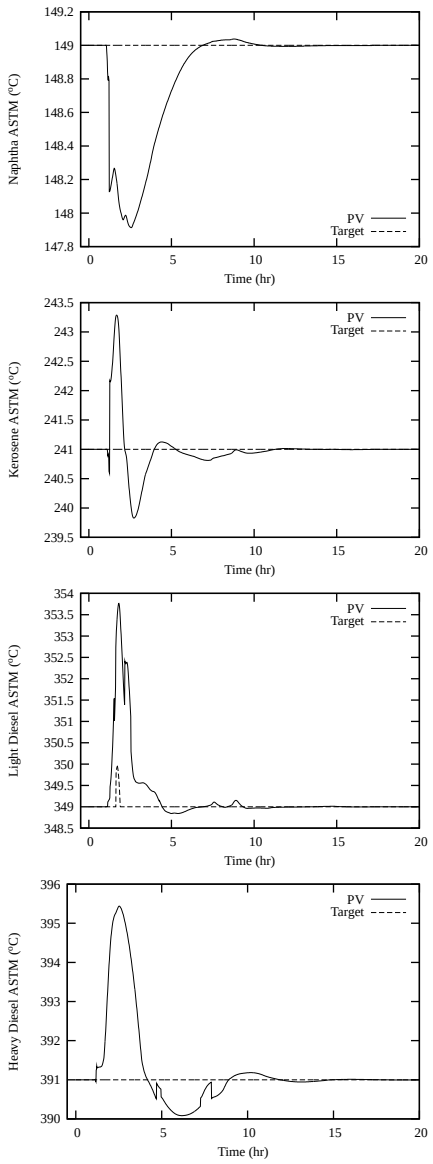


Fig. 4. Closed-loop results for crude switch: time behavior of products' ASTM-D86 95%.

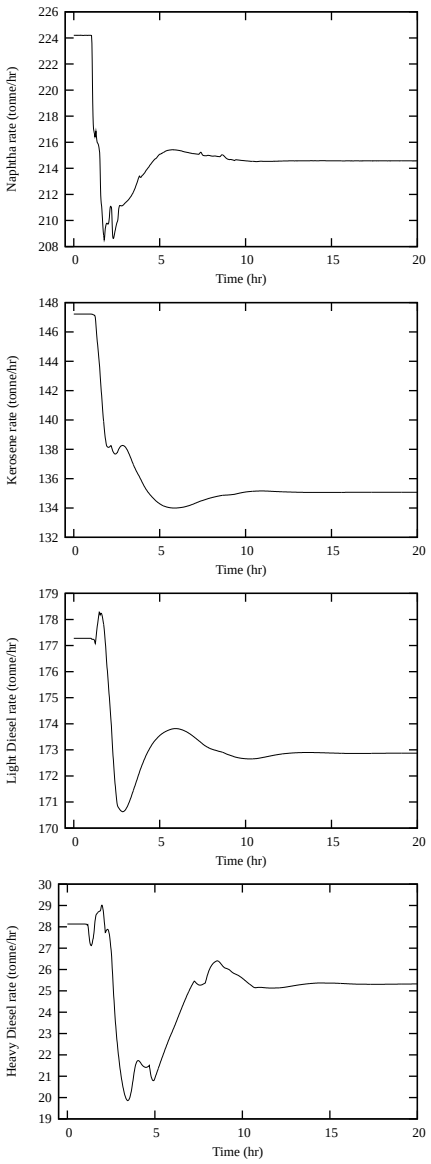


Fig. 5. Closed-loop results for crude switch: time behavior of products' flow rates.

model can be identified and the predictive controller implemented. Simulators can also be used to carry out closed-loop simulations, useful to refine the controller tuning (selection of costs, equal concern errors, limits, etc.) and effectively compare different predictive control algorithms. Therefore, the methodology illustrated in the present paper can potentially lead to reduction of costs for MPC applications, with consequent enlargement of the range of applicability of APC systems.

REFERENCES

Berber, R. and S. Coskun (1996). Dynamic simulation and quadratic dynamic matrix control of an industrial low density polyethylen reactor. *Comput. Chem. Eng.* **20**, S799–S804.

Gallinelli, L. (2005). Studies on dynamics and control issues of complex distillation columns using rigorous simulators (in italian). Master's thesis. Chemical Engineering, University of Pisa.

Huang, H. and J. B. Riggs (2002). Comparison of PI and MPC for control of a gas recovery unit. *J. Proc. Cont.* **12**, 163–173.

Luyben, M. L. and B. D. Tyreus (1998). An industrial design/control study for the vinyl acetate monomer process. *Comput. Chem. Eng.* **22**, 867–877.

Mayne, D. Q., J. B. Rawlings, C. V. Rao and P. O. M. Scokaert (2000). Constrained model predictive control: stability and optimality. *Automatica* **36**, 789–814.

Morari, M. and J. H. Lee (1999). Model predictive control: past, present and future. *Comput. Chem. Eng.* **23**, 667–682.

Muske, K. R. and T. A. Badgwell (2002). Disturbance modeling for offset-free linear model predictive control. *J. Proc. Cont.* **12**, 617–632.

Pannocchia, G. and A. Brambilla (2005). How to use simplified dynamics in model predictive control of superfractionators. *Ind. Eng. Chem. Res.* **44**, 2687–2696.

Pannocchia, G. and J. B. Rawlings (2003). Disturbance models for offset-free model predictive control. *AIChE J.* **49**, 426–437.

Qin, S. J. and T. A. Badgwell (2003). A survey of industrial model predictive control technology. *Cont. Eng. Pract.* **11**, 733–764.

Skogestad, S. (2003). Simple analytic rules for model reduction and PID controller tuning. *J. Proc. Cont.* **13**, 291–309.

Yiu, Y., Y. Fan, L. W. Colwell and M. N. Papadopoulos (1994). Building multivariable predictive control models by process simulation and data regression. *ISA Trans.* **33**, 133–140.

Session 5.2

Modeling of Particulate Systems

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

CHALLENGES OF MODELLING A POPULATION BALANCE USING WAVELET

Johan Utomo, Nicoleta Balliu, Moses O. Tadé[1]

Department of Chemical Engineering, Curtin University of Technology,
GPO Box U 1987, Perth, WA 6845, Australia.

Abstract: Crystallization is one of the oldest separation technologies due to its ability to produce a range of bulk products to high purity chemicals. Aspect of controlling the size distribution is important for downstream operations and characteristics of products. Two cases of population balance problems are considered in this paper to show limitations of some utilized methods. Those cases present the sharp transition phenomena in the particle size distribution. A wavelet-based method by Liu and Cameron (2001) is applied and compared with other conventional methods based on Finite Difference, Orthogonal Collocation and Orthogonal Collocation with Finite Elements. The result show that the wavelet method is faster, more accurate and more efficient in solving the population balance problems. *Copyright © 2005 IFAC*

Keywords: Crystallization, Modelling, Population Balance, Wavelet method

1. BACKGROUND

Crystallization is one of the oldest separation technologies and plays a key role regarding the quality of the products and the economy of a whole plant. This process is used to manufacture large quantities of bulk materials as well as high purity chemicals.

Although crystallization technology has been established for a long time it is difficult to operate and control. Controlling particle size distribution (PSD), shape distribution and crystal purity are challenging due to the complexity and non-linearity of the process and because of lack of reliable on-line instrumentation to measure the key parameters (Rohani, et al. 1999; Braatz 2002). These properties affect downstream operation such as filtration, washing, drying, mixing and formulation (Braatz 2002; Fujiwara, et al. 2005). They also affect the end-usage properties such as entrainment liquid after dewatering, dissolution rate for pharmaceutical products, caking properties, fluidization properties, pneumatic handling properties, bulk density, and esthetic appearance (Randolph and Larson 1988).

According to Braatz (2002), inadequate control of particle size and shape can result in unacceptably long filtration or drying time, or in extra processing steps, such as re-crystallization or milling process. Shekunov et al. (2000) claim that in the pharmaceutical industry, there were advance control systems over drug identity and purity, however control over the physical properties such as form and crystallinity remains inferior. It is indicated that there are many attractive challenges arising from the pharmaceutical processes which can be used for further research directions especially in modelling and building advanced control systems.

Hulburt and Katz (1964) introduced a modelling approach for particulate processes more than 41 years ago. This approach is well known as the concept of population balances. The population balance equation (PBE) can be defined as a mathematical description characterizing particles undergoing the mechanisms of birth, growth, death and leaving a certain particle phase space. In

[1]Corresponding author : Tel: +61-8-9266-4998; Fax: +61-8-9266-2681; M.Tade@exchange.curtin.edu.au

crystallization, those mechanisms can be categorized as nucleation, growth, agglomeration and breakage.

A very sharp transition profiles problem commonly occurs in many chemical engineering cases. For example, concentration profiles in chromatography processes, temperature and activity profiles in solid catalyst, the profiles of reaction in fixed bed reactors as well as the particle size distribution for a well-mixed batch crystallizer in which crystal breakage and agglomeration may be neglected. These processes are represented by parabolic partial differential equations. Effective solutions for these models require certain type of numerical methods to be implemented.

2. NUMERICAL METHODS

2.1 Previous Methods

Most papers addressing population balance problems discussed techniques to solve systems from the unidimensional to multidimensional population balance models. Many numerical methods have been proposed such as method of moment, method of self-preserving distributions, method of weighted residuals, sectional method, and the discretization methods. Other methods, which have been used to solve PB problems, are also based on Monte Carlo method and finite element method. To sum up, the above methods can be categorized into four types i.e. finite difference approach, spectral methods (e.g. orthogonal collocation), finite element and other approach. There are major drawbacks from those methods such as high computationally cost, lack of stability and accuracy of the solution and the in applicability of the solved models for implementation in control based models. Extensive discussion of those methods can be found in the literature (Kostoglou and Karabelas 1994; Ramkrishna 2000; Vanni 2000). In this paper, simulation studies will be conducted to compare the computational efficiency, the accuracy as well as the stability between the finite difference method (FD), the orthogonal collocation (OC), the orthogonal collocation with finite element (OCFE) and the wavelet-based method.

2.2 Finite Difference Methods

Finite difference methods have been commonly used for the solution of all types of partial differential equations (ODEs) systems. FD method approximates the continuous function $f(x)$ with Taylor expansion series (Hangos and Cameron, 2001). They can be a first order or second order approximations. In our case, FD method is used to approximate the first partial derivative of population density over its size ($\partial n/\partial x$) and converts the PDE into a set of ODEs.

2.3 Orthogonal Collocation

This technique was developed more than 70 years ago and applied in various cases of boundary value problems. The trial functions are chosen as sets of orthogonal polynomials and the collocation points are the roots of these polynomials. The solution can be calculated from the collocation points. The use of orthogonal polynomials is to reduce the error as the polynomial order increases (Gupta 1995; Hangos and Cameron 2001).

2.4 Orthogonal Collocation with Finite Elements

The combination of dividing the regions into a number of elements and by applying orthogonal collocation techniques for each element can improve the solution where the profile is very steep. In the region where there is a sharp transition, numbers of small elements can be applied while the remainder utilizes larger size of elements. Selection of the elements size is therefore essential.

2.5 Wavelet-based method

In 2001, Liu and Cameron proposed wavelet based method to solve population balance problems. They developed Wavelet Orthogonal Collocation (WOC) and Adaptive Wavelet Orthogonal Collocation (AWOC) to solve agglomeration in batch vessel. Further information about wavelet method can be found in Liu's papers (Liu and Cameron 2001; Liu and Cameron 2003; Liu and Tade 2004). To our knowledge, wavelet method combined with Galerkin method was first applied in chemical engineering area by Chen et al. (1996) to solve the breakage mechanism in a batch crystallizer.

Significant advantages of using wavelet method are the accuracy in producing solutions in the sharp transition regions, computationally efficient solutions, stable and easily implemented solution that is applicable to another system. These advantages are related to the characteristic of wavelet method such as, localization properties in space and scale, hierarchical organization, sparse coefficients and easy handling of the derivatives as well as non-linear and integral terms. However in Liu and Cameron (2001), there was no comparative study between wavelet method and any other methods.

2.6 Daubechies orthonormal wavelets

Wavelet can be used as a basis function to represent a certain function. In the wavelet function, two-basis functions can be found, the scaling function and the wavelet function. The scaling function coefficient illustrates a local average of the function (coarse illustration) and the wavelet function coefficient describes detailed information of the function (refinements) that cannot be found from the average coefficient. Compared to Fourier expansion, wavelet approximation give smaller error and is highly localized at discontinuity regions (Nielsen 1998). Compared to the traditional trigonometric basis functions which have infinite support, wavelets have compact support, therefore wavelets are able to approximate a function by the placement of the right wavelets at appropriate locations. From Daubechies's

work (1988), scaling function (ϕ) and wavelet function (ψ) can be described by a set of L (an even integer) coefficients (p_k : k = 0,1,…, L-1) through the two-scale relationship:

$$\phi(x)=\sum_{k=0}^{L-1} p_k\,\phi(2x-k) \tag{1}$$

and the wavelet function

$$\psi(x)=\sum_{k=2-L}^{1}(-1)^k\,p_{1-k}\,\phi(2x-k) \tag{2}$$

The support for the scaling function is in the interval 0 to (L-1), whilst for the wavelet function is in the interval (1-L/2) to (L/2). The coefficients p_k are called the wavelet filter coefficients.

Denote $L^2(R)$ as the space of square integrable functions on the real line. Let V_j be the subspace as the L^2-closure of the linear combination of:

$$\phi_{jk}(x)=2^{j/2}\phi(2^j x-k) \tag{3}$$

for $k\in Z=\{..,-1,0,1...\}$. A function $f(x)\in V_j$ can be represented by the wavelet series:

$$f(x)=\sum_{k\in Z} f_{jk}\phi_{jk}(x) \tag{4}$$

The multi-resolution properties of wavelets give another advantage to represent functions in differential equations which can be solved numerically (Motard and Joseph 1994). Detailed information about Daubechies orthonormal wavelets can be found in Daubechies (1988).

2.7 Wavelet Orthogonal Collocation(WOC)

This method was proposed by Betoluzza and Naldi (1996) for solving partial differential equations. In 2001 it was developed and applied for solving population balance problems by Liu and Cameron (2001). The interpolation functions are generated by autocorrelation of the usual compactly supported Daubechies scaling functions $\phi(x)$. Then the function θ called autocorrelation function verifies the interpolation property due to the orthonormality.

$$\theta(0)=\int\phi(x)\phi(x)dx=1 \tag{5}$$

and

$$\theta(n)=\int\phi(x)\,\phi(x-n)dx=0,\ n\neq 0 \tag{6}$$

The approximate solution of our problem will be a function u_j in the term of its dyadic points to obtain the wavelet expression:

$$u_j(x)=\sum u_j(2^{-j}n)\,\theta(2^j x-n) \tag{7}$$

Detailed information can be found in Liu and Cameron (2001, 2003) and Bertoluza and Naldi (1996).

3. CASE STUDIES

We consider two case studies of population balance which have sharp and dramatic transition phenomena in their particle size distribution in the batch crystallizer. Even though the case studies considered here are simple, since the analytical solutions are available for comparison purposes, the more complex models can be solved using the methods described above.

3.1 Case I: Nucleation and size-independent growth

The population balance for nucleation mechanism and size independent growth is described by the partial differential equation:

$$\frac{\partial n(L,t)}{\partial t}+G\frac{\partial n(L,t)}{\partial L}=B_0 \tag{8}$$

where n is number of particle (population density), L is dimensionless particle size, $L\in[0,2]$, G is the growth rate ($G=1$) and B_o is the nucleation rate, $B_0=\exp(-L)$. The initial condition is $n(L,0)=0$ and the boundary condition when $L=0$, $n(0,t)=0$. The analytical solution for this case is:

$$\begin{aligned} n(L,t)&=1-\exp(-L) && ;L-t<0\\ n(L,t)&=\exp(-L)\left[\exp(-t)-1\right] && ;L-t>0 \end{aligned} \tag{9}$$

3.2 Case II: Size-independent growth only

One dimensional population balance for size dependent growth mechanism only is described by the partial differential equation:

$$\frac{\partial n(L,t)}{\partial t}+G\frac{\partial n(L,t)}{\partial L}=0 \tag{10}$$

with:

$$n(0,t)=0;\quad n(L,0)=\exp(-100(L-1)^2) \tag{11}$$

The independent growth rate ($G=1$) is constant. The range of dimensionless particle size, $L\in[0,4]$. The analytical solution for the second case is :

$$n(L,t)=\exp(-((L-G.t-1)\times 10)^2) \tag{12}$$

4. DISCUSSION

All the simulation results presented have been executed on a 3.00 GHz Pentium IV – 1.00 Gigabytes of RAM running under Windows 2000. A MATLAB® version 7.0.1 was used as the computation software to simulate the models.

4.1 Case I: Nucleation and size-independent growth

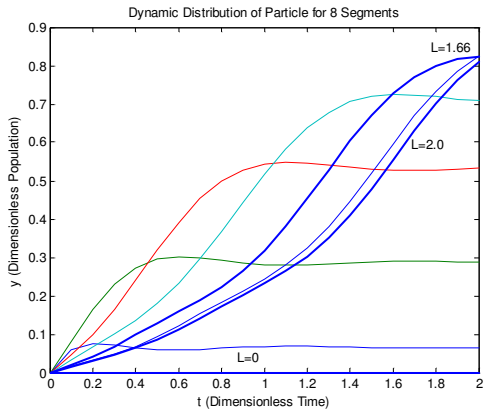


Fig.1. Dynamic distribution of 8 segments of particles in case I

Firstly we divide the size interval range into 8 segments and simulate the dynamic distribution of particles. It is seen from Figure 1 that using $L = 0$ the smallest segment of particles remains zero, while the other increase by following the underdamped mechanism until the overdamped mechanism for the largest segment of particles ($L = 2$). We found that the dynamic particle distribution gives the stable responses.

Finite difference method is employed to solve this problem numerically. The results are shown below in Figure 2 by using 101 discretization points (FD 101). At early time ($t = 0.6$) particles are distributed heavily over the left region (maximum at $L = 0.6$) up to the final time of simulation, particles are mostly distributed on the right segment (maximum at $L = 1.8$). The numerical FD 101 solution is accurate for any segment except the sharp transition region. If we increase the number of discretization point it will increase the accuracy of the solution of the whole region (including the peak region). However, it requires more of computational effort due to the increase of ordinary differential equations (ODEs) needed to be solved.

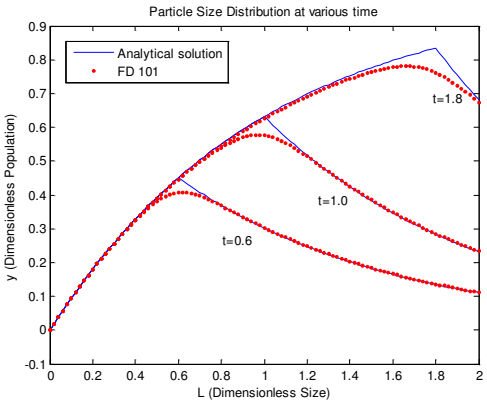


Fig. 2. PSD at various times using FD method

Other methods that can be used to solve this problem are, as previously mentioned Orthogonal Collocation (OC) and Orthogonal Collocation with Finite Element (OCFE). Detailed description of these method can be found in many literature such as Davis (1984), and Finlayson (1980).

Comparisons between the numerical solutions using OC, OCFE and FD are presented in Table 1 and Figure 3. Table 1 shows the comparative error results for the utilized methods, which are SE (Sum of Errors), AE (Average of Errors) and ME (Maximum of Errors), respectively.

In terms of computation time, even though the FD 101 consists of more ODEs than others, it gives reasonable computation time, only 1.37 s. On the other hand, the OCFE 16, which has only 16 ODEs contributes 2.21 s in computation. We can conclude that all methods described have reasonable computation time. The only problem is in the accuracy of the solution in the sharp transition region.

OCFE 31 gives the overpredicted result at the maximum point, whilst the other methods give underpredicted result at that point. OC method cannot be used because it only represents 8 points of solution and it does not cover the entire region proportionally. On the other hand, all OCFE methods perform better than the 101 points of FD method in the sharp transition region. It is indicated by the maximum of errors (ME) results of the OCFE 16 and the OCFE 31 which are less than the FD 101. Even though both methods use less number of collocation points, they successfully present accurate solution and especially in the peak region. It is revealed that the OCFE methods are superior compared to others.

Table 1 Comparative simulation results for case I

Method	Time (s)	SE	AE	ME
FD 101	1.37	8.21e-005	0.0091	0.00200
OC 8	0.90	0.0151	0.1230	0.07060
OCFE 7	0.19	3.3965e-004	0.0184	0.00210
OCFE 16	2.21	1.3513e-004	0.0116	0.00064
OCFE 31	3.34	1.0248e-004	0.0101	0.00110

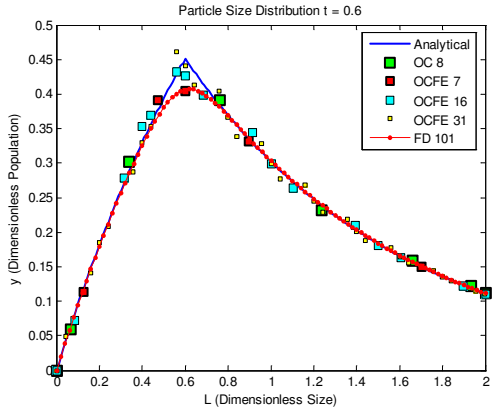


Fig. 3. PSD case I using various methods

At this point we can conclude that for the sharp transition region, the OCFE methods can be used instead of ordinary FD methods with reasonable

computational time and accurate solution. Another question that arose from this study is whether those methods are able to track a very dramatic changing profile as shown in the next case.

4.2 Case II: Size-independent growth only

Figure 4 shows that there is a very steep gradient profile in the particle size distribution. The particles are distributed mostly in region of $L = 1.7 - 2.3$. According to Liu et.al (2000) OCFE method can avoid spurious (unstable) responses under steady state conditions, however, it may fail for the transient model. In our case-II's simulation, unfortunately the OCFE method cannot be applied, since it gives unstable solution. On the other hand, the FD methods even with a very large point of discretization (401 and 801 points) represent inaccurate solutions in terms of the maximum value and also the particle distribution itself. As we can see from the Figure 5, there is a shift phenomena in particle distribution, and it shifts 0.5 unit of dimensionless size compared to analytical solution.

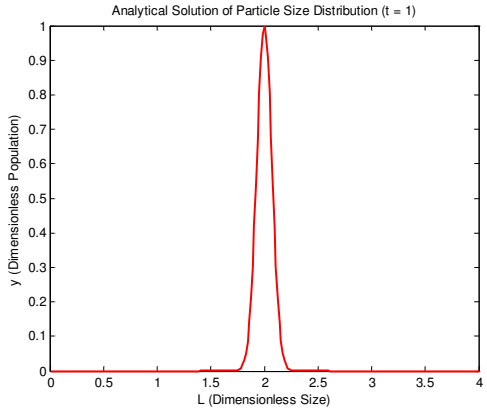


Fig. 4. Analytical solution for case II

Table 2 Comparative simulation results for case II

Method	Time (s)	SE	AE	Max(n)
FD 401	1.42	0.05334	0.2309	0.70717
FD 801	4.62	0.06844	0.2384	0.81654
FD 1201	11.07	0.05842	0.2417	0.86605
Wave 8	0.92	9.91e-6	0.0032	1.00

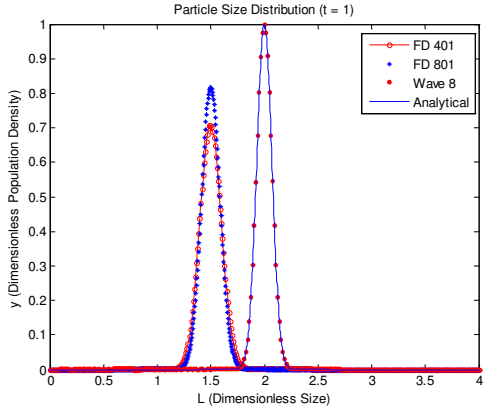


Fig.5. PSD case II using FD and Wavelet method

Other method must be considered here. Wavelet orthogonal collocation method is employed since it can represent a sharp transition region due to its good localization properties both on time and frequency. By using 8-level of wavelet approximation series (resolution), the solution can represent the accurate value at the peak point. Regarding the error parameters, wavelet solution appeared to be superior compared to previous methods. By using 8-level of resolution, we utilize 257 (256+1) wavelet collocation points for the solution and 255 numbers of differential equations. Given that the properties of wavelet which is capable of representing high localization both in space and frequency, allow the preview of the behaviour of the solution at certain time, from localization properties of the solution at previous time-step. Moreover, the computation time needed for FD methods is greater than the wavelet method.

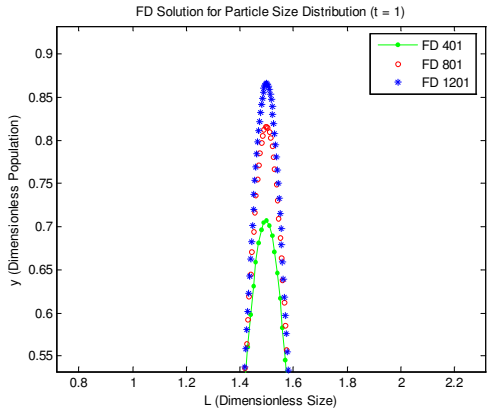


Fig. 6. Comparison of FD solution in case II

In our computational study, we used Runge-Kutta method in MATLAB® to utilize numerical integration of set of ODEs and all the parameters are set to their default value.

The exponential term in the initial condition contributes to the nature of the solution. The numerical solutions become highly non linear. At the early time ($t = 0$), the numerical solution agrees with the analytical solution. As the time increase the shifted solution becomes larger. We can conclude that the use of the FD technique for this case will generate highly inaccurate results. Furthermore we can see that the solution of the FD method still cannot track the dramatic change in particle size distribution even if there was no shift phenomena. As we can observe from Figure 6, the maximum value of $n(L,t) = 0.7$ for FD 401 and $n(L,t) = 0.8$ for FD 801. By increasing the number of discretization points from 801 to 1201 points, the solution still fails to move from 0.5 unit delay however it increased the maximum value of n from 0.816 to 0.866. As can be seen from the Figure 6, there was a small increase in reaching the maximum value by adding 400 points from 801 to 1201 points compared with 401 to 810 points. Comparison between Wavelet method and FD methods; and FD methods with OC and OCFE methods employed in this paper provide us some

insight into the superiority of wavelet in terms of accuracy and computation time.

Further research on wavelet application in chemical engineering field is essentially required. From the computational efficiency result shown, with the wavelet algorithms, the model is suitable to be employed in online control system. Model of population balance with multidimensional properties is necessary for certain cases. Even though from control engineers' perspective, low-order models are needed. The challenges for modelling a population balance with wavelet-based method is to define the suitable complexity for various cases and reduce the appropriate models for design the control strategies. Efforts will be made in future work to validate the results of this study using experimental data from the literature.

5. CONCLUSION

In this work, accurate, fast and general approach by wavelet method is the most efficient way to simulate the case of a very sharp transition phenomenon in population balance system. For case-I, the nucleation and size-independent growth only, the sharp change region can be described effectively by OCFE methods. It, however fail to simulate case-II where there was a very steep gradient in PSD profiles. The FD methods in both cases fail to provide accurate solutions. In case-II where there was a shift of 0.5-unit size they give incorrect prediction of the PSD profiles due to the presence of the highly non-linear term. Wavelet solution gives the fast, stable and accurate solution. The selection of the level of resolution being used depends on the characteristic of the solution itself. If the solution is highly non-linear, the level of resolution must be increased.

REFERENCES

Bertoluzza, S. and G. Naldi (1996). "A Wavelet Collocation Method for the Numerical Solution of Partial Differential Equations." Applied and Computational Harmonic Analysis **3**(1): 1-9.

Braatz, R. D. (2002). "Advanced control of crystallization processes." Annual Reviews in Control **26**(1): 87-99.

Chen, M.-Q., C. Hwang and Y-P. Shih. (1996). "A wavelet-Galerkin method for solving population balance equations." Computers & Chemical Engineering **20**(2): 131-145.

Daubechies, I. (1988). "Orthonormal bases compactly supported wavelets." Communications Pure and Applied Mathematics(41): 909-996.

Davis, M. E. (1984). Numerical Methods and Modeling for Chemical Engineers. New York, John Wiley and Sons.

Finlayson, B. A. (1980). Nonlinear Analysis in Chemical Engineering. New York, McGraw Hill.

Fujiwara, M., Z. K. Nagy, J.W. Chew and R.D. Braatz. (2005). "First-principles and direct design approaches for the control of pharmaceutical crystallization." Journal of Process Control **15**(5): 493-504.

Gupta, S. K. (1995). Numerical Methods for Engineers. New Delhi, New Age International.

Hangos, K. and I. T. Cameron (2001). Process Modelling and Model Analysis. London, Academic Press.

Hulburt, H. M. and S. Katz (1964). "Some problems in particle technology : A statistical mechanical formulation." Chemical Engineering Science **19**(8): 555-574.

Kostoglou, M. and A. J. Karabelas (1994). "Evaluation of Zero Order Methods for Simulating Particle Coagulation." Journal of Colloid and Interface Science **163**(2): 420-431.

Liu, Y., I. T. Cameron and F.Y. Wang. (2000). "The wavelet-collocation method for transient problems with steep gradients." Chemical Engineering Science **55**(9): 1729-1734.

Liu, Y. and I. T. Cameron (2001). "A new wavelet-based method for the solution of the population balance equation." Chemical Engineering Science **56**(18): 5283-5294.

Liu, Y. and I. T. Cameron (2003). "A new wavelet-based adaptive method for solving population balance equations." Powder Technology **130**(1-3): 181-188.

Liu, Y. and M. O. Tade (2004). "New wavelet-based adaptive method for the breakage equation." Powder Technology **139**(1): 61-68.

Motard, R. L. and B. Joseph (1994). Wavelet applications in chemical engineering. London, Kluwer Academic Publisher.

Nielsen, O. M. (1998). Wavelets in Scientific Computing. Department of Mathematical Modelling. Lyngby, Technical University of Denmark**:** 244.

Ramkrishna, D. (2000). Population balance : theory and applications to particulate systems in engineering. London, Academic Press.

Randolph, A. D. and M. A. Larson (1988). Theory of Particulate Processes : Analysis and Techniques of Continuous Crystallization. San Diego, Acadmic Press.

Rohani, S., M. Haeri and H.C. Wood. (1999). "Modeling and control of a continuous crystallization process Part 1. Linear and non-linear modeling." Computers & Chemical Engineering **23**(3): 263-277.

Shekunov, B. Y. and P. York (2000). "Crystallization processes in pharmaceutical technology and drug delivery design." Journal of Crystal Growth 211(1-4): 122-136.

Vanni, M. (2000). "Approximate Population Balance Equations for Aggregation-Breakage Processes." Journal of Colloid and Interface Science **221**(2): 143-16

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DEVELOPMENT OF A DYNAMIC MULTI-COMPARTMENT MODEL FOR THE PREDICTION OF PARTICLE SIZE DISTRIBUTION AND MOLECULAR PROPERTIES IN A CATALYTIC OLEFIN POLYMERIZATION FBR

G. Dompazis, V. Kanellopoulos and C. Kiparissides,

Department of Chemical Engineering and Chemical Process Engineering Research Institute, Aristotle University of Thessaloniki, P.O. Box 472, Thessaloniki, Greece 541 24

Abstract: In the present study a comprehensive multi-compartment model is developed for the prediction of particle size distribution and particle segregation in a catalytic olefin polymerization FBR. To calculate the particle growth and the spatial monomer and temperature profiles in a particle, the random pore polymeric flow model (RPPFM) is utilized. The RPPFM is solved together with a dynamic discretized particle population balance model, to predict the particle size distribution (PSD) in each compartment. In addition, the polymer molecular properties are calculated, in each reactor compartment, by employing a generalized multi-site, Ziegler-Natta, kinetic scheme. The effects of various fluidized bed operating conditions on the morphological and molecular distributed polymer properties are thoroughly analyzed.

Keywords: Polymerization, particle size measurement, modeling

1. INTRODUCTION

High and low density polymers are commercially manufactured in gas phase fluidized bed olefin polymerization reactors using high activity transition metal catalysts such as Ziegler-Natta catalysts, Phillips-Chromium oxide catalysts and supported metallocene catalysts. Although polymer particles are assumed to be very well-mixed, particle segregation may occur in large industrial fluidized bed reactors. This means that the polymer particle size distribution at the reactor exit may differ from the PSDs at different locations along the reactor height. In a fluidized bed reactor strong segregation can occur if the bed contains particles of different densities. Density differences are a common reason for particle segregation but particle size differences can also cause it.

Despite its inherent importance, a limited number of papers have been published on the modeling of the particle-size distribution in gas-phase catalytic olefin polymerization processes. Zacca, et al. (1994), developed a population balance model using the catalyst residence time as the main coordinate, to model particle-size developments in multistage olefin polymerization reactors, including vertical and horizontal stirred beds and fluidized-bed reactors. Choi, et al. (1994), incorporated an isothermal simplified multigrain particle model, by neglecting the external particle mass and heat transfer resistances, into a steady-state PBE to investigate the effect of catalyst deactivation on the PSD and average molecular properties for both uniform and size distributed catalyst feeds. Yiannoulakis, et al. (2001), extended the model of Choi, et al. (1994), to account for the combined effects of internal mass and heat transfer resistances on the PSD for highly active catalysts. In a recent publication Dompazis, et al. (2005), developed a comprehensive integrated model, accounting for the multi-scale phenomena taking place in a continuous gas-phase ethylene copolymerization FBR to describe the molecular and morphological properties of the particulate polymer. Kim and Choi, (2001) presented a steady state multi-compartment population balance model using the concept of size-dependent absorption/spillage model to investigate the effects of fluidization and reaction conditions on the reactor performance.

In what follows, a dynamic multi-compartment model is developed for the prediction of morphological and molecular distributed polymer properties in an FBR.

2. POLYMERIZATION KINETIC MODEL

To describe the molecular weight developments over a heterogeneous Ziegler-Natta catalyst, a generalized two-site kinetic model is employed (Table 1) (Hatzantonis et al., 2000). The kinetic mechanism comprises of a series of elementary reactions, including site activation, propagation, site deactivation and site transfer reactions. The symbol $P_{n,i}^k$ denotes the concentration of "live" copolymer chains of total length 'n' ending in an 'i' monomer unit, formed at the 'k' catalyst active site. P_0^k and D_n^k denote the concentrations of the activated vacant catalyst sites of type 'k' and "dead" copolymer chains of length 'n' produced at the 'k' catalyst active site, respectively.

Table 1 Kinetic mechanism of ethylene-propylene copolymerization over a Ziegler-Natta catalyt.

Activation by aluminum alkyl:	$S_p^k + A \xrightarrow{k_{aA}^k} P_0^k$
Chain initiation:	$P_0^k + M_i \xrightarrow{k_{0,i}^k} P_{1,i}^k$
Propagation:	$P_{n,i}^k + M_j \xrightarrow{k_{p,ij}^k} P_{n+1}^k$
Spontaneous deactivation:	$P_*^k \xrightarrow{k_{dsp}^k} C_d^k + D_n^k$
Chain transfer by hydrogen (H_2):	$P_{n,i}^k + H_2 \xrightarrow{k_{tH,i}^k} P_0^k + D_n^k$

Based on the postulated kinetic mechanism one can define the "live" (λ_ν) and "bulk" (ξ_ν) moments with respect to the corresponding total number chain length distributions (TNCLDs). The average polymer properties of interest (i.e., number and weight average molecular weights) can be calculated.

Number-average molecular weight:

$$M_n = \left(\sum_{k=1}^{N_s} \xi_1^k \Big/ \sum_{k=1}^{N_s} \xi_0^k \right) \sum_{k=1}^{N_s} MW^k \qquad (1)$$

Weight-average molecular weight:

$$M_w = \left(\sum_{k=1}^{N_s} \xi_2^k \Big/ \sum_{k=1}^{N_s} \xi_1^k \right) \sum_{k=1}^{N_s} MW^k \qquad (2)$$

where MW^k is the average molecular weight of the repeating unit in the copolymer chains.

$$MW^k = \sum_{i=1}^{N_m} \Phi_i^k MW_i \qquad (3)$$

In the particle level the number- and weight- average molecular weights are diameter-dependent and are obtained by integrating over the particle volume.

Mean-average molecular weights of a polymer particle of size D:

$$M_n(D) = \frac{1}{r} \int_0^{D/2} M_n(r)dr \,,\; M_w(D) = \frac{1}{r} \int_0^{D/2} M_w(r)dr \quad (4)$$

3. SINGLE PARTICLE GROWTH MODELING

To simulate the growth of a single polymer particle, the random pore polymeric flow model (RPPFM) of Kanellopoulos, et al. (2004), was employed. In the RPPFM, the polymer particle is assumed to be spherical, while the heterogeneous polymer and catalyst phases are treated as a pseudo-homogeneous medium of constant density. Monomer diffusion and heat conduction are assumed to occur only in the radial direction, and diffusion of all the other species (e.g., polymer chains) is considered to be negligible. As a result, the overall monomer transport rate will largely depend on the catalyst/particle morphology, which continuously changes with polymerization time. The equations to be solved for the calculation of spatial ethylene and propylene concentrations and temperature profile in a growing polymer particle, as well as the overall particle polymerization rate are presented elsewhere (Kanellopoulos, et al., 2004).

4. MULTI-COMPARTMENT MODEL

To calculate the dynamic evolution of PSD and particle segregation in a gas-phase fluidized bed reactor a dynamic population balance model needs to be solved together with the system of differential equations describing the radial monomer(s) concentration and temperature profiles in a single particle (Kanellopoulos, et al., 2004). The bed is divided into N equally sized virtual compartments, as illustrated in Figure 1 and each reactor zone consists of a bulk emulsion phase compartment and a wake compartment. Let us assume that the operation of each compartment can be approximated by a perfectly back-mixed, continuous flow reactor. Polymer particles are fed into each compartment from other compartments, while the mass of solids in the compartment is kept constant by controlling the product withdrawal rate.

The dynamic population balance equation and the overall mass balance in each reactor compartment can take the following forms:

Top compartment (n=1)

Bulk Phase:

$$\frac{\partial n_1(D,t)}{\partial t} + \frac{\partial [G(D) n_1(D,t)]}{\partial D} = \frac{1}{W_{b,1}} F_c n_c(D) + \frac{1}{W_{b,1}} \left[F_{tr,1}^{wb} n_{tr,1}^{wb}(D,t) + u_{b,1} A_{w,1} \rho_p n_{w,1}(D,t) \right] - \frac{1}{W_{b,1}} \left[F_{tr,1}^{bw} n_{tr,1}^{bw}(D,t) + F_1 n_1(D,t) + u_{b,1} A_{w,1} \rho_p n_{w,1}(D,t) \right] \qquad (5)$$

$$W_{b,1}\int_{D_{\min}}^{D_{\max}} G(D)n_1(D,t)d\left(\rho_p \pi D^3/6\right) - F_1 + u_{b,1}A_{w,1}\rho_p + F_c = 0 \quad (6)$$

Wake Phase:

$$\frac{\partial n_{w,1}(D,t)}{\partial t} = \frac{1}{W_{w,1}}\left[u_{b,2}A_{w,2}\rho_p n_{w,2}(D,t) + F_{tr,1}^{bw} n_{tr,1}^{bw}(D,t)\right] - \frac{1}{W_{w,1}}\left[u_{b,1}A_{w,1}\rho_p n_{w,1}(D,t) + F_{tr,1}^{wb} n_{tr,1}^{wb}(D,t)\right] \quad (7)$$

$$u_{b,2}A_{w,2}\rho_p - u_{b,1}A_{w,1}\rho_p + F_{tr,1}^{bw} - F_{tr,1}^{wb} = 0 \quad (8)$$

i[th] compartment (n=i)

Bulk Phase:

$$\frac{\partial n_i(D,t)}{\partial t} + \frac{\partial\left[G(D)n_i(D,t)\right]}{\partial D} = \frac{1}{W_{b,i}}\left[F_{i-1}n_{i-1}(D,t)\right] - \frac{1}{W_{b,i}}\left[F_i n_i(D,t) - F_{tr,i}^{wb} n_{tr,i}^{wb}(D,t) + F_{tr,i}^{bw} n_{tr,i}^{bw}(D,t)\right] \quad (9)$$

$$F_{i-1} + W_{b,i}\int_{D_{\min}}^{D_{\max}} G(D)n_i(D,t)d\left(\rho_p \pi D^3/6\right) - F_i = 0 \quad (10)$$

Wake Phase:

$$\frac{\partial n_{w,i}(D,t)}{\partial t} = \frac{1}{W_{w,i}}u_{b,i+1}A_{w,i+1}\rho_p n_{w,i+1}(D,t) - \frac{1}{W_{w,i}}\left[u_{b,i}A_{w,i}\rho_p n_{w,i}(D,t) + F_{tr,i}^{wb} n_{tr,i}^{wb}(D,t) - F_{tr,i}^{bw} n_{tr,i}^{bw}(D,t)\right] \quad (11)$$

$$u_{b,i+1}A_{w,i+1}\rho_p - u_{b,i}A_{w,i}\rho_p + F_{tr,i}^{bw} - F_{tr,i}^{wb} = 0 \quad (12)$$

Bottom compartment (n=i)

Bulk Phase:

$$\frac{\partial n_N(D,t)}{\partial t} + \frac{\partial\left[G(D)n_N(D,t)\right]}{\partial D} = \frac{1}{W_{b,N}}F_{N-1}n_{N-1}(D,t) - \frac{1}{W_{b,N}}\left[F_N n_N(D,t) + F_{tr,N}^{bw} n_{tr,N}^{bw}(D,t) - F_{tr,N}^{wb} n_{tr,N}^{wb}(D,t)\right] - \frac{1}{W_{b,N}}F_{re}n_{re}(D,t) \quad (13)$$

$$W_{N,i}\int_{D_{\min}}^{D_{\max}} G(D)n_N(D,t)d\left(\rho_p \pi D^3/6\right) + F_{N-1} - F_N - F_{re} = 0 \quad (14)$$

Wake Phase:

$$F_{re} - u_{b,N}A_{w,N}\rho_p + F_{tr,N}^{bw} - F_{tr,N}^{wb} = 0 \quad (15)$$

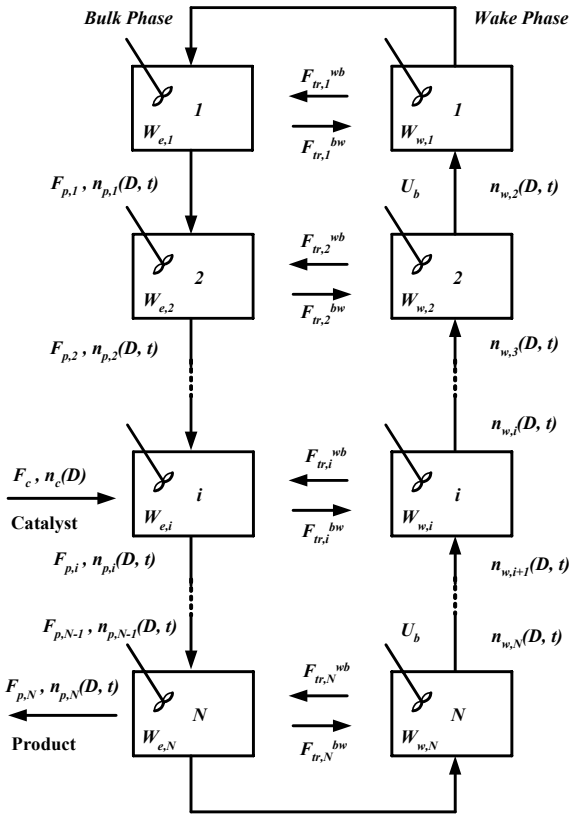


Fig. 1. Schematic representation of the multi-compartment model.

$$\frac{\partial n_{w,N}(D,t)}{\partial t} = \frac{1}{W_{w,N}}\left[F_{re}n_{re}(D,t) + F_{tr,N}^{bw} n_{tr,N}^{bw}(D,t)\right] - \frac{1}{W_{w,N}}\left[u_{b,N}A_{w,N}\rho_p n_{w,N}(D,t) + F_{tr,N}^{wb} n_{tr,N}^{wb}(D,t)\right] \quad (16)$$

where $n_i(D,t)$, $n_{w,i}(D,t)$, $n_{re}(D,t)$, expressed in (#/g/cm), denote the number diameter density functions of particles in the bulk phase, in wake phase of compartment i and in the recycle stream respectively. The term $n_i(D,t)dD$ denotes the number of particles in the size range $(D, D+dD)$ per mass of polymer in the bulk phase of compartment i at time t. u_b is the rising bubble velocity and A_w is the effective cross section area of the wake phase

In order to calculate the individual particle growth rate and temperature profile, one has to solve the monomer and energy balances for each discrete class of particles. According to Hatzantonis, et al. (1998), the particle growth rate, $G(D)$ can be expressed in terms of the overall polymerization rate for each class of particles, $R_p(D)$, as follows:

$$G(D) = 2R_p(D)/\rho_p \pi D^2 \quad (17)$$

The population balance equation has to be solved numerically using an accurate and efficient discretization method. In the present study, the orthogonal collocation on finite elements method

(OCFE) was employed for solving the dynamic PBE (Alexopoulos et al., 2004).

In an FBR solid mixing is induced by fast moving gas bubbles. A rising bubble drags solid particles from the bulk emulsion phase, while solids are moving from wake to the bulk phase at the same time, indicating that there is a continuous interchange of particles between the two phases. It is obvious that the development of a correlation is required to describe correctly the type of particles that are entrained by the bubbles. According to Choi et al. (2001), the following empirical exponential correlation can be applied in order to calculate the particle transfer constant from bulk to wake phase.

$$k_{tr}^{bw}(D) = A\rho_g \exp(-u_t/u_0) \tag{18}$$

where ρ_g is the density of fluidizing gas, u_0 is the superficial gas velocity, u_t is the terminal velocity and A is an adjustable parameter (A = 0.0658). Notice that the particle transfer rate constant from bulk phase to wake becomes size dependent because the terminal velocity is size dependent.

Finally, the number- and weight- average molecular weights for all particles in the reactor will be given by the weighted sum of mean-average molecular weights calculated in the particle level (i.e., $M_n(D)$, $M_w(D)$) with respect to the particle size distribution.

$$M_{n,bed}(t) = \int_{D_{\min}}^{D_{\max}} M_n(D) p_p(D,t) dD \tag{19}$$

$$M_{w,bed}(t) = \int_{D_{\min}}^{D_{\max}} M_w(D) p_p(D,t) dD \tag{20}$$

where $p_p(D,t)dD$ denotes the mass fraction of particles in the size range (D to $D+dD$) at time t per mass of polymer in the bed.

Table 2 Nominal operating conditions and numerical values of the physical and transport properties of the reaction mixture

Reactor Operating Conditions	Physical Properties
W (kg) = 64177	ΔH_r (J/g) = -3832
F_c (g/s) = 0.1	$\rho_{g,1}$ (kg/m^3) = 28
D_c (μm) = 50	$\rho_{g,2}$ (kg/m^3) = 42
T_b (K) = 353.15	$\mu_{g,1}$ (Pa·s) = 1.2×10^{-4}
$[M_1]_b$ (mol/L) = 0.65	$\mu_{g,2}$ (Pa·s) = 10^{-4}
$[M_2]_b$ (mol/L) = 0.15	$D_{b,1}$ (cm^2/s) = 0.006
$[H_2]$ (mol/L) = 0.03	$D_{b,2}$ (cm^2/s) = 0.004
$[Coc]$ (mol/L) = 0.01	

5. RESULTS AND DISCUSSION

Extensive numerical simulations were carried out by using the proposed model (see to Figure 1) to investigate the effects of various reactor operating conditions on the distributed molecular and morphological polymer properties in a catalyzed, gas phase, ethylene propylene copolymerization FBR. The reactor operating conditions and the numerical values of the physical and transport properties of the reaction mixture are reported in Table 2.

In Figure 2, the effect of fluidization gas velocity on the particle size distribution in the reactor compartments is shown. In the present multi-compartment model the number of compartments was set equal to five through all model simulations. For illustration purposes the PSDs in the top, bottom and in an intermediate compartment are shown. As can be seen, at low gas velocities, the PSD is shifted to larger sizes from the top to bottom compartment, according to the general segregation pattern. As the gas velocity increases, the individual PSDs in each compartment collapse into the same distribution, implying that the FBR can be approximated by a single CSTR.

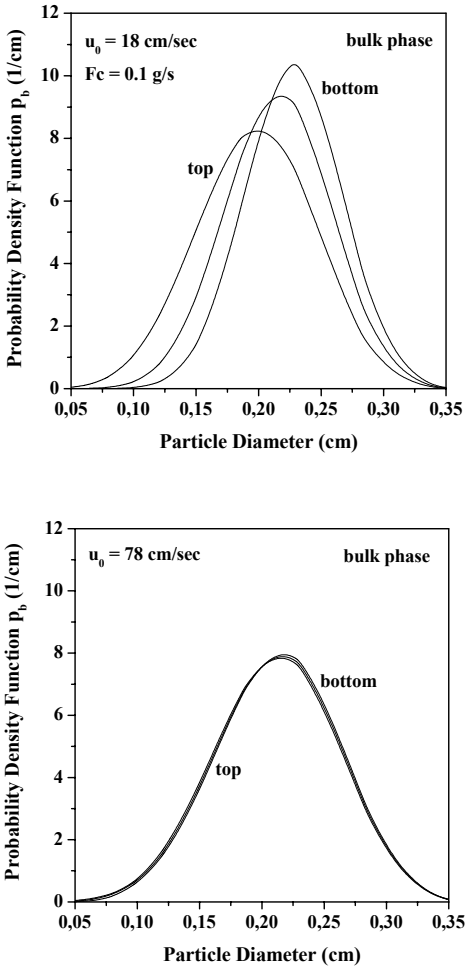


Fig. 2. Effect of fluidization gas velocity

The particle size distributions in the wake phase are shown in Figure 3. As expected, the amount of small particles in the wake phase is substantially larger

than in the bulk phase, which is correct because as a bubble rises in the reactor carries particles small in size.

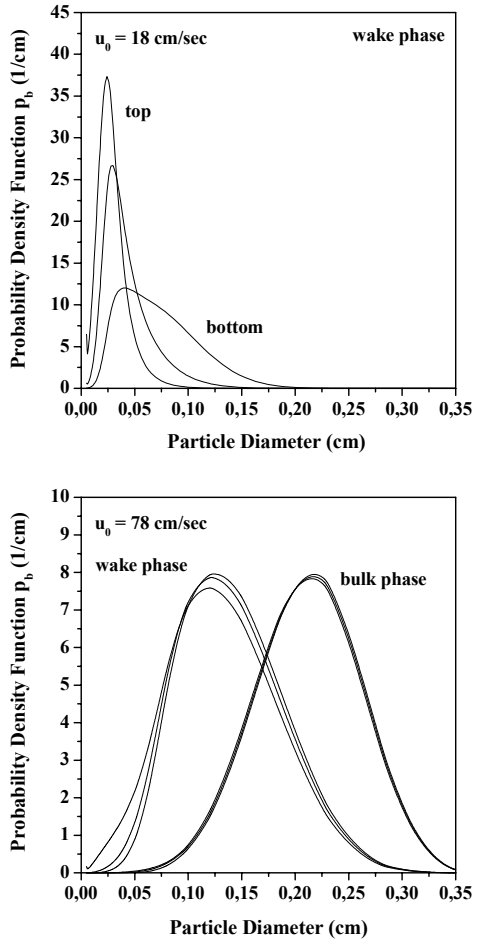


Fig. 3. Effect of fluidization gas velocity.

In Figure 4, the dynamic evolution of ethylene concentration in the reactor is depicted for different fluidization gas velocities. According to this Figure, as the gas velocity increases less monomer is consumed because its conversion is relatively low.

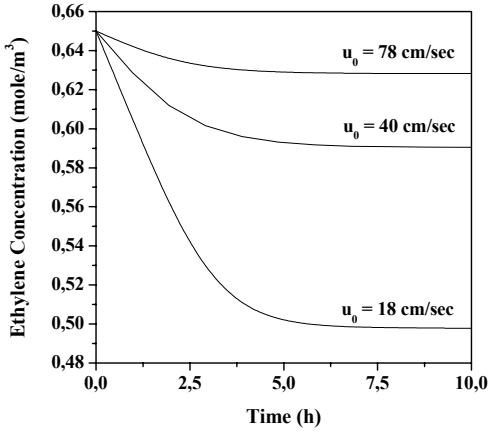


Fig. 4. Effect of fluidization gas velocity on ethylene consumption in the reactor.

In Figure 5, the effect of catalyst feed rate is shown. It is apparent that as the catalyst feed rate increases, the polymer particle size distribution becomes narrower and is shifted to smaller sizes. It is important to point out that as the catalyst feed rate increases, more particles grow in the bed and because the bed weight is kept constant, the residence time of particles in the reactor decreases. As a result the amount of large polymer particles in the reactor decreases.

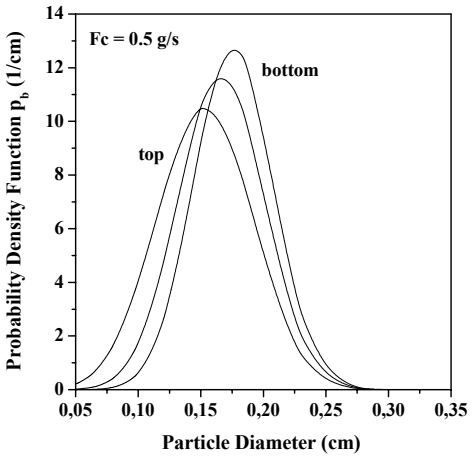


Fig. 5. Effect of catalyst feed rate

In Figure 6, the dynamic evolution of the average particle size of the distribution in the reactor for the two catalyst feed rates studied brfore, is depicted. It is obvious that as the catalyst feed rate decreases the time required for the PSD to reach its final steady-state value increases.

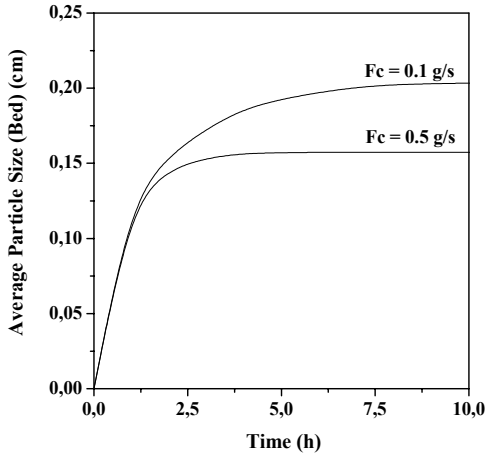


Fig. 6. Effect of catalyst feed rate on the average particle size of the PSD in the reactor.

In Figure 7, the effect of catalyst feed rate on the dynamic evolution of the polymer weight average molecular weight in the reactor compartments is illustrated. According to the results of this Figure, an increase of catalyst feed rate significant affects the WAMW of polymer produced during the dynamic operation of the reactor. Particles in the upper compartment due to their smaller size require less time for their PSD to reach its steady-state value. As a result the time required for their molecular weight to reach its steady state value, is also smaller in

comparison with the corresponding time of particles existed in lower compartments. That is the reason why for a certain time instant more WAMW is produced in the top compartment. Also can be seen, that a steady state WAMW value is achieved after approximately 4 hours, when the catalyst feed rate is equal to 0.5g/s, and 10 hours, when the catalyst feed rate is equal to 0.5g/s, respectively.

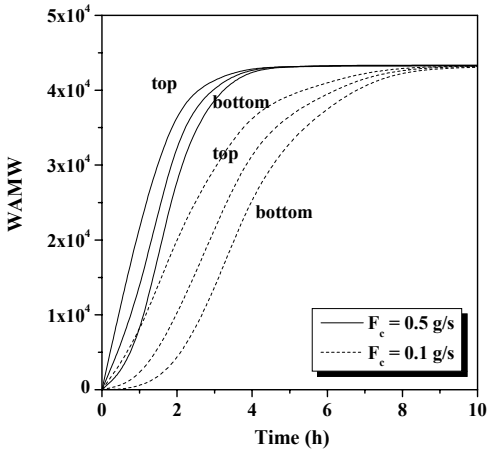


Fig. 7. Effect of catalyst feed rate on the dynamic evolution of the polymer weight average molecular weight in the reactor compartments.

6. CONCLUSIONS

In the present study, a comprehensive multi- scale, multi-compartment dynamic model is developed to analyze the dynamic behavior of fluidized bed reactors for ethylene-propylene copolymerization through mathematical modeling and simulation. The model can also be used for the prediction of morphological (i.e., particle size distribution (PSD) and particle segregation) as well as molecular (i.e., molecular weight distribution (MWD)) distributed polymer properties in a catalytic olefin polymerization FBR. It is illustrated that at low fluidization gas velocities particle segregation phenomena become significant and may influence the morphological and molecular properties of polymer particles in the bed. Although no actual experimental data are reported in the open literature regarding segregation phenomena in industrial fluidized bed reactors, our model results are in qualitative agreement with industrial observations.

REFERENCES

Alexopoulos, H.A., A.I. Rousos and C. Kiparissides (2004). Part 1: Dynamic evolution of the particle size distribution in particulate processes undergoing combined particle growth and aggregation. . *Chemical Engineering Science*, 59, 5751-5769.

Choi, K.Y., X. Zhao and S. Tang (1994). Population balance modelling for a continuous gas phase olefin polymerization reactor. *Journal of Applied Polymer Science*, 53, 1589-1597.

Dompazis, G., V. Kanellopoulos and C. Kiparissides (2005). A multi-scale modeling approach for the prediction of molecular and morphological properties in multi-site catalyst, olefin polymerization reactors. *Macromolecular Materials and Engineering*, 290, 525-536.

Hatzantonis, H., A. Goulas, and C. Kiparissides (1998). A comprehensive model for the prediction of particle-size distribution in catalyzed olefin polymerization fluidized-bed reactors. *Chemical Engineering Science*, 53, 3252.

Hatzantonis, H., A. Yiagopoulos, H. Yiannoulakis and C. Kiparissides (2000). Recent developments in modeling gas-phase catalyzed olefin polymerization fluidized-bed reactors: The effect of bubble size variation on the reactor's performance. *Chemical Engineering Science*, 55, 3237-3259.

Kanellopoulos, V., G. Dompazis, B. Gustafsson and C. Kiparissides (2004). Comprehensive analysis of single-particle growth in heterogeneous olefin polymerization: the random-pore polymeric flow model. *Ind. Eng. Chem. Res.*, 43 (17), 5166-5180.

Kim, J.Y. and K.Y. Choi (2001). Modeling of particle segregation phenomena in a gas phase fluidized bed olefin polymerization reactor. *Chemical Engineering Science*, 56, 4069-4083.

Yiannoulakis, H., A. Yiagopoulos and C. Kiparissides (2001). Recent developments in the particle size distribution modeling of fluidized-bed olefin polymerization reactors. *Chemical Engineering Science*, 56, 917-925.

Zacca, J.J., A.J. Debling and H.W. Ray (1996). Reactor residence time distribution effects on the multistage polymerization of olefins I. Basic principles and illustrative examples, polypropylene. *Chemical Engineering Science*, 51 (21), 4859-4886.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DISTRIBUTIONAL UNCERTAINTY ANALYSIS OF A BATCH CRYSTALLIZATION PROCESS USING POWER SERIES AND POLYNOMIAL CHAOS EXPANSIONS

Z. K. Nagy[†], R. D. Braatz[‡]

[†]*Loughborough University, Department of Chemical Engineering,*
Loughborough, Leics LE11 3TU, U.K.
[‡]*University of Illinois at Urbana-Champaign*
Department of Chemical and Biomolecular Engineering
600 South Mathews Avenue, Urbana, IL 6180, USA

Abstract: Computationally efficient approaches are presented that quantify the influence of parameter uncertainties upon the states and outputs of finite-time control trajectories for nonlinear systems. In the first approach, the worst-case values of the states and outputs due to model parameter uncertainties are computed as a function of time along the control trajectories. The approach uses an efficient contour mapping technique to provide an estimate of the distribution of the states and outputs as a function of time. To increase the estimation accuracy of the shape of the distribution, an approach that uses second order power series expansion in combination with Monte Carlo simulations is proposed. Another approach presented here is based on the approximate representation of the model via polynomial chaos expansion. A quantitative and qualitative assessment of the approaches is performed in comparison to the Monte Carlo simulation technique that uses the nonlinear model. It is shown that the power series and polynomial chaos expansion based approaches require a significantly lower computational burden compared to Monte Carlo approaches, while give good approximation of the shape of the distribution. The techniques are applied to the crystallization of an inorganic chemical with uncertainties in the nucleation and growth parameters.

Keywords: probabilistic analysis, distributional robustness analysis, worst-case analysis, crystallization, optimal control.

1. INTRODUCTION

Comprehensive uncertainty analysis of mechanistic models is crucial especially when these models are used in the optimal control of processes, which normally occurs close to safety and performance constraints. The model-based computation of optimal control policies for batch and semibatch processes is of increasing interest due to industrial interest in improving productivity (Barrera and Evans, 1989; Rippin 1983). However, uncertainty almost always exist in chemical systems in the observed data, in the model parameters, and implemented inputs, and its disregard may easily lead the loss of the benefits of using optimal control (Ma *et al.*, 1999). This motivates the development of techniques to quantify the influence of parameter uncertainties on the process states and outputs. Quantitative estimates obtained from robustness analysis can be used to decide whether more laboratory experiments are needed to provide better parameter estimates (Ma and Braatz, 1999; Miller and Rawlings, 1994). The traditional uncertainty analysis consists of the characterization of uncertainty in model parameters or inputs based on their probability density functions (*pdf*) and then propagating these *pdf*s through the model equations to obtain the *pdf*s of selected model outputs. The propagation of uncertainties via

traditional Monte Carlo methods, based on standard or Latin Hypercube sampling may require performing a large number of simulations, which can be prohibitive in most cases, especially if the propagation has to be performed in real-time. Therefore there is a need to study computationally efficient alternative techniques for uncertainty propagation.

The paper focuses on computationally efficient methods for propagating uncertainty in parameters to the states and outputs of generic batch processes. Two categories of approaches are corroborated. The first is based on first and second order power series approaches (Nagy and Braatz, 2003). The first order approach is based on the analytical computation of the worst-case values of the states and outputs due to the effects of model parameter uncertainties, and then using a contour mapping approach to compute the distribution. The second approach uses polynomial chaos expansion as a functional approximation of the mathematical model (Isukapalli, 1999; Isukapalli *et al.*, 1998; Pan *et al.*, 1997, 1998). Both approaches are suitable for studying the uncertainty propagation in open-loop or closed-loop systems. The techniques are compared with Monte Carlo simulations and applied to compute the distributions for the states and outputs for the batch crystallization of an inorganic chemical subject to uncertainties in the nucleation and growth kinetics.

2. DISTRIBUTIONAL UNCERTAINITY ANALYSIS

2.1 Uncertainty description

In the following we consider the class of finite time (batch) processes. Assume the model is described by the generic ODE vector equation:

$$\dot{x}(t) = f(x(t), u(t); \theta) \tag{1}$$

with $x \in \mathbb{R}^{n_x}$ the state vector, $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$ vector of control inputs, and $\theta \in \mathbb{R}^{n_\theta}$ uncertain parameter vector, and f a vector function that is continuous wit respect to its elements. Characterizing the uncertainties enhances the value of the model by allowing the quantification of the accuracy of its predictions. This information can be used to assess whether the model is adequate for its intended purpose (e.g., for optimal control design) or whether more experimental data are needed to further refine the model.

Define $\hat{\theta}$ as the nominal model parameter vector of dimension ($n \times 1$), and $\delta\theta$ as the perturbation about $\hat{\theta}$. Then, the model parameter vector for the real system is:

$$\theta = \hat{\theta} + \delta\theta \,. \tag{2}$$

We assume that the uncertainty in the parameter θ is characterized by the generalized ball

$$\mathcal{P} \triangleq \{\, \theta \in \mathbb{R}^{n_\theta} \,\big|\, \| \theta - \hat{\theta} \| \le 1 \,\} \tag{3}$$

defined by using appropriate norm $\|\cdot\|$ in $\mathbb{R}^{n_\theta}$. One generic approach is to use a scaled Hölder p-norm ($1 \le p \le \infty$), given by $\| \theta \| = \| \mathbf{W}_\theta^{-1} \theta \|_p$, with the invertible weighting matrix $\mathbf{W}_\theta \in \mathbb{R}^{n_\theta \times n_\theta}$. This generalized description of the uncertainty set includes the case of a confidence hyper-ellipsoid $\varepsilon_\theta = \{\theta : (\theta - \hat{\theta})^T \mathbf{V}_\theta^{-1} (\theta - \hat{\theta}) \le r^2(\alpha)\}$, (Beck and Arnold, 1985), for Gaussian random variable vector θ with expected value $\mathcal{E}(\theta) = \hat{\theta}$, the $(n_\theta \times n_\theta)$ positive definite variance-covariance matrix $\mathbf{V}_\theta$, and the scalar *r* which is the chi-square distribution function with n_θ degrees of freedom ($\chi^2_{n_\theta}(\alpha)$), for a chosen confidence level α:

$$\mathcal{P}_{\text{ellipsoid}}(\alpha) \triangleq \{\, \theta \in \mathbb{R}^{n_\theta} \,\big|\, \| (1/r(\alpha)) \mathbf{V}_\theta^{-1/2} (\theta - \hat{\theta}) \|_2 \le 1 \,\} \tag{4}$$

The generalized ball described by (2) also includes the case of known lower and upper bounds θ_l, θ_u of the uncertain parameters, leading to a general uncertainty hyper-box:

$$\begin{aligned} \mathcal{P}_{\text{box}} &\triangleq \{\theta \in \mathbb{R}^{n_\theta} \,|\, \theta_l \le \theta \le \theta_u\} \\ &= \{\theta \in \mathbb{R}^{n_\theta} \,\big|\, \| \text{diag}(\tfrac{\theta_u - \theta_l}{2})(\theta - \tfrac{\theta_u + \theta_l}{2}) \|_\infty \le 1\} \end{aligned} \tag{5}$$

2.2. Series expansion approaches for Worst-case and Distributional Robustness Analysis

Define $\hat{\psi}$ as the performance (describing an end point property) for the nominal model parameters $\hat{\theta}$, ψ as its value for the perturbed model parameter vector *p*, and the difference $\delta\psi = \psi - \hat{\psi}$. The worst-case robustness approach (Ma *et al.*, 1999; Nagy and Braatz, 2003) writes $\delta\psi$ as a power series in $\delta\theta$:

$$\delta\psi = L\delta\theta + \frac{1}{2}\delta\theta^T \mathbf{M} \delta\theta + \dots \,, \tag{6}$$

where the jacobian $L \in \mathbb{R}^{n_\theta}$, and hessian $\mathbf{M} \in \mathbb{R}^{n_\theta \times n_\theta}$ are:

$$L(t) = \left(\frac{\partial \psi(t)}{\partial \theta} \right)_{\hat{\theta}}, \tag{7}$$

$$\mathbf{M}(t) = \left(\frac{\partial^2 \psi(t)}{\partial \theta^2} \right)_{\hat{\theta}} . \tag{8}$$

The elements of the time-varying sensitivity vector $L(t)$ and matrix $\mathbf{M}(t)$ can be computed using finite differences or by integrating the model's differential-algebraic equations augmented with an additional set of differential equations known as sensitivity equations (Caracotsios and Stewart, 1985):

$$\dot{L} = \mathbf{J}_x L + \mathbf{J}_\theta \tag{9}$$

with the matrixes $\mathbf{J}_x = df/dx \in \mathbb{R}^{n_x \times n_x}$ and $\mathbf{J}_\theta = df/d\theta \in \mathbb{R}^{n_x \times n_\theta}$.

When a first-order series expansion is used, analytical expressions of the worst-case deviation in

the performance index ($\delta\psi_{w.c}$) can be computed and the analysis can be performed with low computational cost (Matthews, 1997). In the case of an ellipsoidal uncertainty description the worst-case deviation is defined by

$$\delta\psi_{w.c.}(t) = \max_{\|\mathbf{W}_\theta \delta\theta\|_2 \le 1} \left| L(t)\delta\theta \right|. \quad (10)$$

The analytical solution of this optimization problem is given by:

$$\delta\psi_{w.c.}(t) = (r(\alpha)L(t)\mathbf{V}_\theta L^T(t))^{1/2}, \quad (11)$$

$$\delta\theta_{w.c.}(t) = \frac{(r(\alpha))^{1/2}}{(L(t)\mathbf{V}_\theta L^T(t))^{1/2}} \mathbf{V}_\theta L^T(t). \quad (12)$$

A probability density function (PDF) for the model parameters is needed to compute the PDF of the performance index. More than 90% of the available algorithms to estimate parameters from experimental data (Beck and Arnold, 1977) produce a multivariate normal distribution:

$$f_{p.d.}(\theta) = \frac{1}{(2\pi)^{n_\theta/2}\det(\mathbf{V}_\theta)^{1/2}} \exp\left(-\frac{1}{2}[(\theta-\hat{\theta})^T \mathbf{V}_\theta^{-1}(\theta-\hat{\theta})] \right). \quad (13)$$

When a first-order series expansion is used to relate $\delta\psi$ and $\delta\theta$, then the estimated PDF of ψ is

$$f_{p.d.}(\psi) = \frac{1}{V_\psi^{1/2}\sqrt{2\pi}} \exp\left(-\left(\psi - \hat{\psi}\right)^2 \Big/ \left(2V_\psi\right) \right) \quad (14)$$

where the variance of ψ is

$$V_\psi = L\mathbf{V}_\theta L^T. \quad (15)$$

The distribution is a function of time since the nominal value for ψ and the vector of sensitivities L is a function of time.

2.3 Propagation of probability distribution

When the first order power series expansion is used the probability density function (*p.d.f.*) can be estimated very efficiently using a contour mapping approach, which instead of mapping the whole parameter space to the output space by performing exhaustive Monte Carlo simulations, maps only the contours of the uncertainty hyperellipsoid obtained for different α levels as shown on Figure 1. The mapping is performed via the worst-case analysis techniques by obtaining the worst-case $\delta\psi_{w.c}$ for different α-levels. Note that with this approach the mapping of the α levels is performed from the n_θ dimensional space characterized by a chi-square distribution of n_θ degrees of freedom ($\chi^2_{n_\theta}(\alpha)$) to an n_ψ dimensional space in which the same α-levels are characterized by a chi-square distribution but with n_ψ degrees of freedom ($\chi^2_{n_\psi}(\alpha)$), with usually $n_\psi = 1$. Hence the probability mapping between the two spaces characterized by different degrees of freedom can be captured by multiplying the obtained worst-case deviations ($\delta\psi_{w.c.}$) with the ratio $(\chi^2_{n_\psi}(\alpha)/\chi^2_{n_\theta}(\alpha))^{1/2}$.

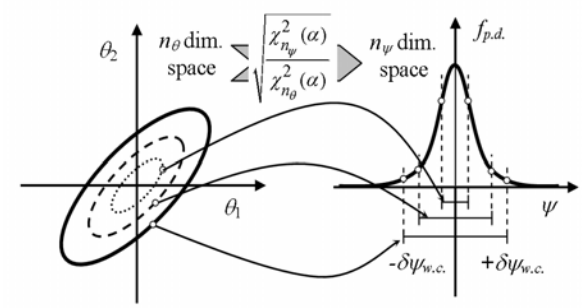


Fig.1. Distributional robustness approach based on a contour mapping technique.

Remark: Since in this approach the sampling is always performed over one parameter (the confidence level α) a significantly smaller number of Monte Carlo simulations are required than using classical sampling of the usually larger dimensional parameter space. Additionally, while the classical sampling procedure requires large number of samples to capture accurately the distribution for values with low probability, in the contour mapping approach samples span over all confidence levels and generate enough data to give insight into the tails of the *pdf* in a similar way as in the Latin Hypercube sampling method.

For increased accuracy of the estimated shape of the *pdf*, the second order series expansion can be used in a classical or Latin Hypercub type Monte Carlo simulation. This Monte Carlo approach is computationally much more efficient than applying the Monte Carlo method to the original nonlinear model. The computational burden of the second order approximate Monte Carlo approach is higher than in the case of the contour mapping approach but provides better estimate of the shape of the output distribution.

2.4 Uncertainty Analysis Using Polynomial Chaos Expansions

An alternative to power series is to use polynomial chaos expansions (PCEs). The PCE (Wiener, 1938) describes the model output ψ as an expansion of multidimensional Hermite polynomial functions of the uncertain parameters θ in the form:

$$\psi = \underbrace{a_0\Gamma_0}_{\text{constant}} + \underbrace{\sum_{i_1=1}^{n_\theta} a_{i_1}\Gamma_1(\theta_{i_1})}_{\text{first order terms}} + \underbrace{\sum_{i_1=1}^{n_\theta}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\Gamma_2(\theta_{i_1},\theta_{i_2})}_{\text{second order terms}} + \underbrace{\sum_{i_1=1}^{n_\theta}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}\Gamma_3(\theta_{i_1},\theta_{i_2},\theta_{i_3}) + \dots}_{\text{third order terms}} \quad (16)$$

where the $\Gamma_\rho(\theta_{i_1},\theta_{i_2},\dots,\theta_{i_\rho})$ are polynomials, n_θ is the number of parameters, and the $a_0, a_1, a_2, \dots, a_{11}, a_{12}, \dots$ are constants in $\mathbb{R}$. Polynomial chaos terms of different order are orthogonal to each other as are polynomial chaos terms of the same order but with a different argument list. The orthogonal polynomials are derived from the

probability distribution of the parameters. In PCE any form of polynomial could be used but the properties of orthogonal polynomials make the uncertainty analysis more efficient. The number of coefficients in the PCE depends on the number of uncertain parameters and the order of expansion. In principle there are two main methods for computing the coefficients of the PCE, (i) the probabilistic collocation method (PCM) (Pan *et al.*, 1997, 1998), and (ii) the regression method with improved sampling (RMIS) (Isukapalli *et al.*, 1998). In both methods the coefficients are calculated from the model at a set of sample points using regression based technique.

3. APLICATION TO A BATCH CRYSTALLIZATION PROCESS

Crystallization from solution is an industrially important unit operation due to its ability to provide high purity separation. The control of the crystal size distribution (CSD) can be critically important for efficient downstream operations (such as filtration or drying) and product quality (e.g., bioavailability, tablet stability, dissolution rate). Most studies on the optimal control of batch crystallizers focus on computing the temperature profile that optimizes some property of the CSD. The problem of computing the optimal temperature profile can be formulated as a nonlinear optimization problem, which is then solved using general-purpose optimization algorithms. A convenient way to describe the temperature trajectory is to discretize the batch time and consider the temperatures at every discrete time k as the optimization variables. In this case the optimal control problem can be written in the following form:

$$\underset{T(k)}{\text{optimize}}\ J \tag{17}$$

subject to:

$$\begin{bmatrix} \dot{\mu}_0 \\ \dot{\mu}_1 \\ \dot{\mu}_2 \\ \dot{\mu}_3 \\ \dot{\mu}_4 \\ \dot{C} \\ \dot{\mu}_{seed,1} \\ \dot{\mu}_{seed,2} \\ \dot{\mu}_{seed,3} \end{bmatrix} = \begin{bmatrix} B \\ G\mu_0 + Br_0 \\ 2G\mu_1 + Br_0^2 \\ 3G\mu_2 + Br_0^3 \\ 4G\mu_3 + Br_0^4 \\ -\rho_c k_v (3G\mu_2 + Br_0^3) \\ G\mu_{seed,0} \\ 2G\mu_{seed,1} \\ 3G\mu_{seed,2} \end{bmatrix} \tag{18}$$

$$\begin{aligned} & T_{\min}(k) \le T(k) \le T_{\max}(k), \\ & R_{\min}(k) \le \frac{dT(k)}{dt} \le R_{\max}(k), \\ & C_{final} \le C_{final,\max}, \end{aligned} \tag{19}$$

where the objective function J is a function of the states, and usually it is a representative property of the final CSD. The equality constraints (18) represent the model equations, with initial conditions given in (Chung *et al.*, 1999), where μ_i is the ith moment ($i = 0, \ldots, 4$) of the total crystal phase (resulted from growth from seed and nucleation) and $\mu_{seed,j}$ is the jth moment ($j = 0, \ldots, 3$) corresponding to the crystals grown from seed, C is the solute concentration, T is the temperature, r_0 is the crystal size at nucleation, k_v is the volumetric shape factor, and ρ_c is the density of the crystal. The rate of crystal growth (G) and the nucleation rate (B), respectively, are given by (Nyvlt, *et al.*, 1985):

$$G = k_g S^g, \tag{20}$$

$$B = k_b S^b \mu_3, \tag{21}$$

where $S = (C\text{-}C_{sat})/C_{sat}$ is the relative supersaturation, and $C_{sat} = C_{sat}(T)$ is the saturation concentration. The model parameter vector consists of the kinetic parameters of growth and nucleation:

$$\theta^T = [g, k_g, b, k_b], \tag{22}$$

with nominal values (Rawlings *et al.*, 1993):

$$\hat{\theta}^T = [1.31, 8.79, 1.84, 17.38], \tag{23}$$

with the uncertainty description of the form (4) characterized by the covariance matrix (Miller and Rawlings, 1994):

$$V_\theta^{-1} = \begin{bmatrix} 102873 & -21960 & -7509 & 1445 \\ -21960 & 4714 & 1809 & -354 \\ -7509 & 1809 & 24225 & -5198 \\ 1445 & -354 & -5198 & 1116 \end{bmatrix}. \tag{24}$$

In the inequality constraints (19) T_{min}, T_{max}, R_{min}, and R_{max} are the minimum and maximum temperatures and temperature ramp rates, respectively, during the batch. The first two inequality constraints ensure that the temperature profile stays within the operating range of the crystallizer. The last inequality constraint ensures that the solute concentration at the end of the batch C_{final} is smaller than a certain maximum value $C_{final,max}$ set by the minimum yield required by economic considerations.

The crystal size distribution (CSD) parameters of interest are: nucleation to seed mass ratio ($J_{n.s.r.}$), coefficient of variations ($J_{c.v.}$), and weight mean size of the crystals ($J_{w.m.s.}$), given by the following expressions:

$$J_{n.s.r.} = (\mu_3 - \mu_{seed,3}) / \mu_{seed,3} \tag{23}$$

$$J_{c.v.} = (\mu_2 \mu_0 / (\mu_1)^2 - 1)^{1/2} \tag{24}$$

$$J_{w.m.s.} = \mu_4 / \mu_3 \tag{25}$$

The optimal temperature trajectory that minimizes the nucleation mass to seed mass ratio at the end of the batch was computed setting $J = J_{n.s.r.}$, and solving the optimal control problem (17)-(19) for the nominal parameter $\hat{\theta}$.

The aforementioned uncertainty analysis approaches are used to assess the effect of parameter uncertainty on the nominal control performance. The distributional uncertainty analysis approaches based on power series approximation and polynomial chaos expansion, respectively, were evaluated in comparison to Monte Carlo (MC) simulations. A

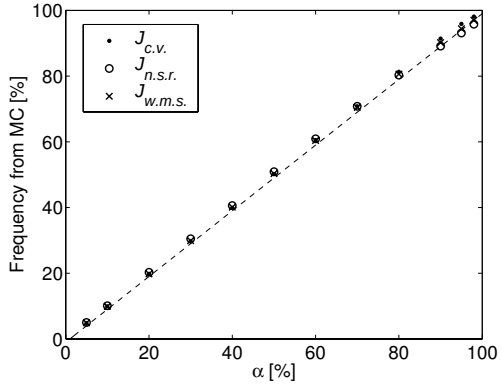


Fig. 2. Comparison between the first-order power series analysis and Monte Carlo simulations. The dashed line corresponds to the ideal fit.

Table 1. Computational burden of the different approaches to compute the distribution for all process outputs during the entire batch (on a P4-1.4 GHz computer)

Method	Computational time
Monte Carlo with dynamic model (80,000 data)	8 hours
First order approach (50 α levels, every 10 min)	1 second
Monte Carlo with 2nd order series expansion model (80,000 data)	4 minutes
Polynomial chaos expansion	2 seconds

number of 80,000 random parameter sets with mean $\hat{\theta}_i$ and covariance $\mathbf{V}_\theta$ were generated from the multivariate distribution using the Cholesky decomposition of the covariance matrix. With the random parameter vectors, Monte Carlo simulation was performed using the dynamic model of the process. Then, the frequency that the simulation output from the Monte Carlo simulation falls in the confidence interval obtained for a certain α level with the power series approach was computed for all states and outputs. The results obtained with the first order analysis approach for the output variables are presented in Figure 2. The accuracy of the first-order approach is very good with a slightly decreasing tendency for large α values. This can be explained by the cumulative effect of the truncation error due to the first-order approach (which increases when α increases) and the tail effect of the Monte Carlo simulation. Although there is a slight decrease in the accuracy of the first order approach for large α in the case of certain process outputs the first-order technique gives good result for practical purposes.

The first-order power series approach approximates the output distributions with a normal distribution. To obtain a more accurate representation of the output PDF, the uncertainties can be propagated through the nonlinear dynamic model via Monte Carlo simulations. To avoid the prohibitively high computational time required in this case, an alternative approach that performs Monte Carlo simulation using a higher order power series expansion or polynomial chaos expansion in place of the dynamic simulation model can be used. Using a second-order power series expansion in the Monte Carlo simulations gives an accurate approximation of the nonlinear distribution with a low computational cost (Table 1). PCE can be used instead of the power series expansion, resulting in similarly good computational efficiency. The advantages of the aforementioned uncertainty analysis approaches compared to classical sensitivity analysis consist in that they provide the variation of the whole distribution for all states and outputs along the entire batch.. Figure 3 shows the results obtained with the two computationally efficient approaches in comparison with the Monte Carlo simulation. In the simulation results presented here a second order PCE was used. Since there are four uncertain parameters the second order PCE requires the determination of 15 coefficients. The probabilistic collocation method was used as described in (Tatang *et al.* 1997, Webster *et al.*, 1996). The obtained PCE provide accurate estimation of the effects of parameter uncertainties at a computational burden similar to the first order power series approach. The power series and PCE based robustness analysis approaches can be used for the assessment and efficient synthesis of robust control approaches. Since the robustness analysis is a convex problem, it is significantly easier to solve it than the direct robust controller synthesis, which usually leads to nonconvex problem formulations. The robust controller synthesis is formulated as a classical minmax optimization or as a multiobjective optimization problem where one of the objective accounts for the nominal term and the other (usually a measure of the variance of the distribution) accounts for robustness. Both uncertainty analysis approaches can be used as an efficient tool of calculating the robustness term.

4. CONCLUSIONS

An overview of several computationally efficient distributional robustness analysis approaches are presented. The approaches are based on the approximate representation of the process model using first or second order power series or polynomial chaos expansions, and provide a qualitative and quantitative estimation of the effect of parameter uncertainties on the states and output variable along the batch time. The computational burden of the robustness analysis approaches is significantly reduced compared to the classical Monte Carlo approach based on the nonlinear model. The algorithms are assessed via a simulated batch cooling crystallization process.

REFERENCES

Barrera, M. D. and L.B. Evans (1989). Optimal design and operation of batch processes, *Chem. Eng. Comm.*, 82, 45-66.

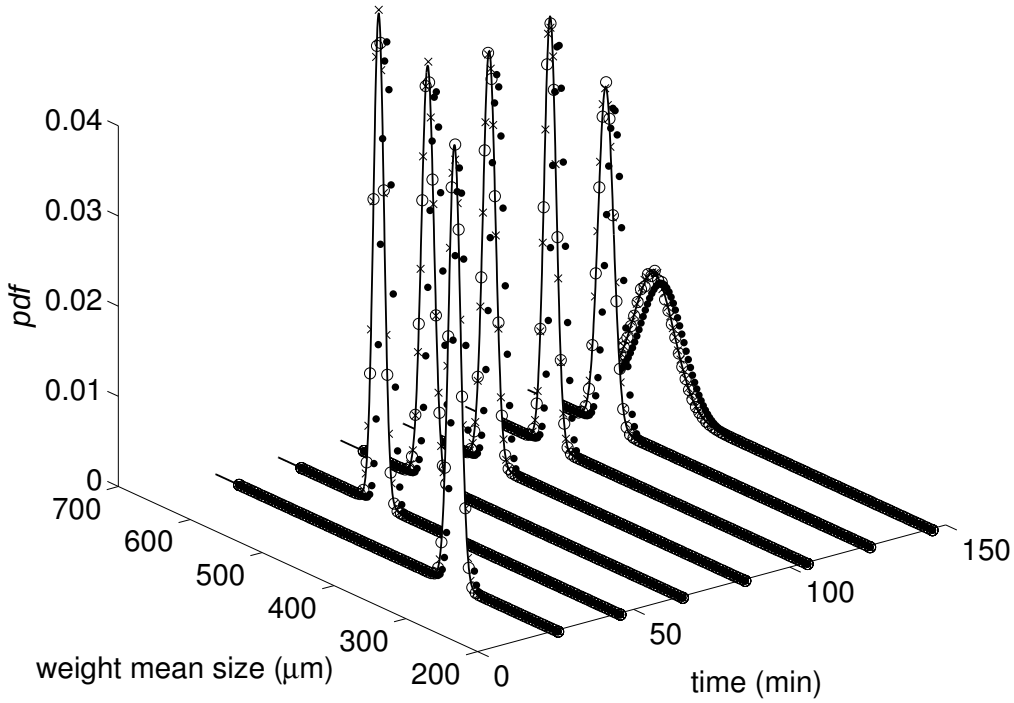


Fig. 3. Variation of the distribution of the weight mean size along the batch, computed using different uncertainty analysis approaches. Solid line is obtained from Monte Carlo simulations using 80,000 parameter sets and the nonlinear model; dots represent the results with first-order power series approach and contour mapping; circles correspond to the result from Monte Carlo simulations based on second order power series expansion; x-marks correspond to the results using the polynomial chaos expansion based approach.

Beck, J.V. and K.J. Arnold (1977). *Parameter Estimation in Engineering and Science*. Wiley, New York.

Caracotsios, M. and W.E. Stewart (1985). Sensitivity analysis of initial value problems with mixed ODEs and algebraic equations, *Computers & Chemical Engineering*, 9, 359-365.

Chung, S. H., D.L. Ma and R. D. Braatz (1999). Optimal seeding in batch crystallization, *The Canadian Journal of Chemical Engineering*, 77, 590-596.

Isukapalli, S.S. (1999). Uncertainty analysis of Transport-Transformation Models, PhD thesis, The State University of New Jersey.

Isukapalli, S.S., A. Roy, and P.G. Georgopoulos (1998). Stochastic Response Surface Methods (SRSMs) for Uncertainty Propagation: Application to Environmental and Biological Systems, *Risk Analysis*, 18(3), 351-363.

Ma, D.L. and R.D. Braatz (2001). Worst-case analysis of finite-time control policies, *IEEE Trans. on Control Systems Technology*, vol. 9, 766-774.

Ma, D.L., S.H. Chung and R. D. Braatz (1999). Worst-case performance analysis of optimal batch control trajectories, *AIChE J.*, 45, 1469-1476.

Miller, S.M. and J.B. Rawlings (1994). Model identification and control strategies for batch cooling crystallizers, *AIChE J.*, vol. 40, pp. 1312-1327, 1994.

Nagy, Z.K. and R.D. Braatz (2003). Worst-case and Distributional Robustness Analysis of Finite-time Control Trajectories for Nonlinear Distributed Parameter Systems, *IEEE Transaction on Control Systems Technology,* 11 (5), 694-704.

Nyvlt, J., O. Sohnel, M. Matuchova and M. Broul (1985). *The Kinetics of Industrial Crystallization*, volume 19 of Chemical Engineering Monographs. Elsevier, Amsterdam.

Pan, W., M.A. Tatang, G.J. McRae, and R.G. Prinn (1997). Uncertainty analysis of direct radiative forcing by anthropogenic sulfate aerosols, *Journal of Geophysical Research*, 102, 21916-21924.

Pan, W., M.A. Tatang, G.J. McRae, and R.G. Prinn (1998). Uncertainty analysis of indirect radiative forcing by anthropogenic sulfate aerosols, *Journal of Geophysical Research*, 103, 3815-3823, 1998.

Rawlings, J.B., S. M. Miller and W.R. Witkowski (1993). Model identification and control of solution crystallization processes: A review, *Ind. Eng. Chem. Res.*, 32, 1275-1296.

Rippin, D.W.T. (1983). Simulation of single- and multiproduct batch chemical plants for optimal design and operation, *Computers & Chemical Engineering*, 7 (3), 137-156.

Wiener, N. (1938). The homogeneous chaos, *Amer. J. Math*, Vol. 60, 897-936.

DYNAMIC EVOLUTION OF THE PARTICLE SIZE DISTRIBUTION IN PARTICULATE PROCESSES.

D. Meimaroglou, A.I. Roussos, and C. Kiparissides

Department of Chemical Engineering, Aristotle University of Thessaloniki and Chemical Process Engineering Research Institute P.O. Box 472, 540 06 Thessaloniki, Greece

Abstract: The present wok provides a comparative study on the numerical solution of the dynamic population balance equation (PBE) in batch particulate processes undergoing simultaneous particle aggregation, growth and nucleation. The general PBE was numerically solved using three different techniques namely, the Galerkin on finite elements method (GFEM), the generalized method of moments (GMOM) and stochastic Monte Carlo simulations (MC). Numerical simulations were carried out over a wide range of variation of particle aggregation and growth rate models.

Keywords: Monte Carlo method, moments method, finite element method, numerical algorithms, distribution

1. INTRODUCTION

The dynamic evolution of the particle size distribution (PSD) in particulate processes is commonly obtained via the solution of the population balance equation (PBE) (Ramkrishna, 2000). In previous publications (Alexopoulos et al., 2004; Alexopoulos and Kiparissides, 2005; Roussos et al., 2005), a comprehensive study on the numerical solution of the dynamic PBE for batch and continuous particulate processes was presented. In general, Galerkin and orthogonal collocation on finite element methods exhibit good numerical performance when applied to processes undergoing simultaneous particle aggregation, growth and nucleation. However, increased computational times and special programming skills are often required for their implementation. Sectional PBE methods are faster and easier to implement but are not sufficiently accurate, especially with strongly size-dependent particle aggregation rate kernels (Kumar and Ramkrishna, 1996[a,b]; Roussos, 2004).

An attractive alternative approach to sectional and finite element (FE) methods is to calculate the leading moments of the distribution instead of the distribution itself. The essential condition for the application of the method of moments (MOM) is that the resulting moment differential equations are in a closed form. Contrary to sectional and FE methods, the computational requirements of the MOM are substantially lower due to the limited number of moment differential equations needed to be solved. However, this results in a less detailed description of the distribution. The reconstruction of a distribution by a finite number of moments (e.g., zero, first, second…), is known in the literature as the inversion or Stieltjes problem.

The dynamic evolution of the PSD in a particulate process can also be obtained via stochastic Monte Carlo (MC) simulations. Spielman and Levenspiel (1965) were the first to employ a MC approach to study the effect of particle coalescence on the reaction progress in two-phase particulate reactive systems in backmix reactors. Later, Shah et al. (1977) developed a general MC algorithm for time varying particulate processes. In 1981, Ramkrishna established the precise mathematical connection between population balances and the MC approach. In MC simulations, the dynamic evolution of the PSD is inferred by the properties of a finite number of particles sampled at appropriate time steps. In order to preserve the statistical accuracy and, at the same time, to keep the computational requirements of a typical MC simulation within reasonable time limits, the number of sampled particles at each time step must be maintained within a specified range (e.g., between 10^3 and 10^6 particles).

In the present study, the general PBE is solved for batch particulate processes using the generalized method of moments (GMOM) and a stochastic MC approach. The performance of the two methods is directly compared with that of the Galerkin on finite elements method (GFEM), for a number of test-problems including processes undergoing simultaneous particle aggregation, growth and nucleation.

2. THE POPULATION BALANCE EQUATION

The general population balance equation for a batch particulate system can be written as follows (Hulburt and Katz, 1964; Ramkrishna, 1985):

$$\frac{\partial n(V,t)}{\partial t} + \frac{\partial \left[G(V)n(V,t)\right]}{\partial V} = B(V) - D(V) + S(V,t) \quad (1)$$

where n(V,t)dV denotes the number of particles per unit volume in the size range [V, V+dV], G(V) is the particle volume growth rate function and S(V,t) is the particle nucleation rate. The terms B(V) and D(V) represent the respective "birth" and "death" rates due

to particle aggregation and are defined by the following expressions:

$$B(V) = \int_0^{V/2} \beta(V-U,U)\, n(V-U,t)\, n(U,t)\, dU \qquad (2)$$

$$D(V) = n(V,t) \int_0^{\infty} \beta(V,U)\, n(U,t)\, dU \qquad (3)$$

$\beta(V,U)$ is the aggregation rate kernel between particles of volumes V and U. In general, Eq. (1) will satisfy the following initial condition:

$$n(V,0) = n_0(V) \qquad (4)$$

where $n_0(V)$ is the initial number density function. If the value of the number density function at the minimum particle volume, $n(V_{min},t)$, is known, the corresponding boundary condition for Eq. (1) takes the following form:

$$n(V_{min},t) = n_1(t) \qquad (5)$$

2.1 The Galerkin on Finite Elements Method.

In the finite elements method, the particle size domain is divided into a number of discrete elements, "*ne*", each containing "*np*" equally spaced nodal points. The number density function, n(V, t), is then approximated over each element "*e*" in terms of its respective values at the nodal points, n_j^e:

$$n^e(V,t) = \sum_{j=1}^{np} \varphi_j^e(V)\, n_j^e(t) \qquad (6)$$

where $\varphi_j^e(V)$ are the well-known Lagrange basis functions. Following the weighted residual formulation of Finlayson (1980), eq. (1) is forced to hold true, in an approximate sense, at each point "*i*" of element "*e*" by satisfying the following orthogonality condition:

$$R_i^e = \int_{V_i^e}^{V_{np}^e} w_i^e(V)\left(\frac{\partial n(V,t)}{\partial t} + \frac{\partial G(V) n(V,t)}{\partial V} - \frac{n_{in}(V,t) - n(V,t)}{\tau} - S(V,t) - B(V) + D(V)\right) dV = 0 \qquad (7)$$

where the indexes "*e*" (= 1, 2,…, *ne*) and "*i*" (= 1, 2,…, *np*) denote the various discrete elements and nodal points, respectively. In the Galerkin approach, the weighting functions $w_i^e(V)$ are identical to the basis functions $\varphi_j^e(V)$. By substituting eq. (6) into eq.(7), the following system of ordinary differential equations is obtained for each element:

$$[A]^e \frac{dn^e}{dt} + \left([E]^e + [C]^e\right) n^e - [S]^e - [B]^e + [D]^e = 0 \qquad (8)$$

A detailed description on the implementation of the GFEM is given in Roussos et al. (2005).

2.2 The Generalized Method of Moments.

According to the method of moments, the general PBE, Eq. (1), is transformed into a system of non-linear integro-differential equations describing the dynamic evolution of the moments of the distribution. In terms of n(V, t), one can easily define the k^{th} dimensionless moment of the distribution, m_k.

$$m_k = \int_0^{\infty} V^k n(V,t)\, dV \Big/ \left(N_0 \cdot V_0^k\right); \; k = 0,1,2... \qquad (9)$$

where N_0 and V_0 are some characteristic values of the distribution. To derive the moment equations, all the terms of Eq. (1) are first multiplied by the quantity $(V/V_0)^k N_0^{-1}$ and the resulting equation is then integrated over the volume domain $[0,\infty]$. It can be easily shown that the dynamic evolution of the k^{th} moment of the distribution will be given by the following integro-differential equation (McGraw, 1997; Wiliams and Loyalka, 1991; Alexiadis et. al., 2004):

$$\dot{m}_k(t) = \left[k \int_0^{\infty} V^{k-1} G(V)\, n(V,t)\, dV + \frac{1}{2} \int_0^{\infty} \int_0^{\infty} \left[(V+U)^k - V^k - U^k\right] \beta(V,U)\, n(V,t)\, n(U,t)\, dU\, dV + \int_0^{\infty} V^k S(V,t)\, dV\right] \left(N_0 V_0^k\right)^{-1} \qquad (10)$$

The main difficulty with the numerical solution of Eq. (10) results from the integral terms that must be expressed in terms of a closed set of moments, so that it can be integrated in time. The closure of moment equations can be achieved either by assuming a specific form for the distribution or using special interpolation techniques.

In the present study, a general formulation, based on an arbitrary choice of the moments, is presented. Let $[M] = [M_1\ M_2]$ be a $(2Nq \times 2Nq)$ matrix with elements defined by the following equations:

$$[M_1]_{ij} = \left(2 - k(i) - 1\right) V_j^{k(i)}; \; i = 1,2,...,2N_q$$
$$[M_2]_{ij} = k(i)\, V_j^{k(i)-1}; \; j = 1,2,...,N_q \qquad (11)$$

where V_i denotes the quadrature rule abscissas and k(i) can take any desired, even negative, values.

The vector $[F] = \left[F_{k(1)}, F_{k(2)}, ..., F_{k(2N_q)}\right]^T$ contains the $2N_q$ contributions of particle growth, aggregation and nucleation mechanisms and its elements will be given by the following equation:

$$[F]_{k(i)} = \Big[\sum_{i=1}^{N_q} k(i)\ V_i^{k(i)-1}\ G(V_i)\ w_i + \frac{1}{2}\sum_{i=1}^{N_q}\sum_{j=1}^{N_q}\Big[\left(V_i+V_j\right)^{k(i)} - V_i^{k(i)} - V_j^{k(i)}\Big]\beta(V_i,V_j)\ w_i\ w_j + \bar{S}_{k(i)}(t)\Big]\left(N_0 V_0^k\right)^{-1} \quad (12)$$

Accordingly, the $2N_q$ elements of the vector $[\mathbf{P}] = \left[P_1, P_2, ..., P_{2N_q}\right]^T$ can be calculated from the solution of the following system of linear algebraic equations:

$$[M][P] = [F] \quad (13)$$

where the quadrature weights and abscissas are directly determined from the solution of the following system of differential equations (Marchisio and Fox (2005)):

$$\frac{dw_j}{dt} = [P]_j \quad ; \quad \frac{d\mathcal{V}_j}{dt} = [P]_{N_q+j} \quad (14)$$

where $\mathcal{V}_j = V_j w_j$; $j = 1, 2, \ldots, N_q$.

Reconstruction of the distribution. The reconstruction of the distribution from a finite set of moments is in general a very difficult problem. A common approach to this problem is to assume a series approximation of the distribution with coefficients expressed in terms of the calculated moments. Different function series have been proposed by various researchers in the past, however, it must be noted that the best results are obtained when some a-priori insight on the form of the distribution is available either from theory or experimentation.

In the present work, it was assumed that the unknown number density function could be approximated by a series of exponential functions:

$$n(V,t) \approx \sum_{i=1}^{N_c} a_i \exp(-b_i V) \quad (15)$$

The value of N_c was selected to be equal to 1or 2. Accordingly, the unknown coefficients a_i and b_i were determined via the minimization of the following objective function:

$$J = \sum_{i=1}^{N_m}\left(\left(m_{k(i)}^{model} - m_{k(i)}^{num.}\right)\Big/ m_{k(i)}^{model}\right)^2 \quad (16)$$

using an appropriate non-linear parameter estimator (e.g., NPSOL). The terms $m_{k(i)}^{num.}$ and $m_{k(i)}^{model}$ in the above equation denote the numerical (i.e., calculated by the GMOM) and the model values of the k(i) moment, respectively. From Eq. (15), one can easily show that the values of $m_{k(i)}^{model}$ moments will be given by the following analytical equation:

$$m_{k(i)}^{model} = \sum_{i=1}^{N_c} a_i \frac{\Gamma[k(i)+1]}{b_i^{k(i)+1}} \quad (17)$$

where $\Gamma(x)$ is the gamma function. To estimate the unknown parameters (i.e., a_1, a_2, b_1 and b_2) in Eq. (15), a set of four target moments were selected (i.e., $N_m = 4$ and k(i)=0, 0.5, 1 and 2). For particle growth systems, Eq. (15) was multiplied by a Heaviside step function, $\mathcal{H} = (V - V_{min}(t))$, to account for the time-varying minimum particle volume, where $V_{min}(t)$ is the minimum particle volume at time, t. Furthermore, for processes undergoing particle nucleation, the number density function was assumed to exhibit a bimodal form. The first mode of the distribution represented the new nucleated particles while the second mode accounted for the dynamic evolution of the distribution due to particle growth and aggregation. As a result, an additional exponential function, $aV\exp(-bV)$, was added into Eq. (15) to account for the first mode of the distribution. Thus, the total number of estimated parameters was increased by two.

2.3 Monte-Carlo Simulations.

The stochastic Monte Carlo (MC) method is based on the principle that the dynamic evolution of an extremely large population of particles (e.g., 10^8) can be followed by tracking down the corresponding changes or events (i.e., growth, aggregation, nucleation) occurring in a smaller number of sample particles, (e.g., 10^4). Initially, the particle volume domain is divided into a number of discrete volume intervals using a logarithmic discretization rule. Subsequently, each particle in the sample population is assigned to an appropriately selected volume, V_i, so that the particle array at time zero, $N_s(0)$, closely represents the initial distribution, according to the inverse transform method (Rubinstein, 1981). Once all the particles in the sample population have been assigned to randomly selected volumes, the MC algorithm is initiated and the effects of particle aggregation, growth and nucleation mechanisms on the dynamic evolution of the particle population are stochastically simulated in a consecutive series of variable-duration time steps.

In problems involving particle aggregation, the time step can be determined in terms of the number of aggregation events, N_{agg}, that take place (Gooch et al., 1996). According to the above procedure, the time required for the occurrence of the duration of N_{agg} events, Δt, will be given by the following equation:

$$\Delta t = \int_{m_0}^{m_0-\Delta m_0}\left[\int_0^{\infty}\left[B(V) - D(V)\right]dV\right]^{-1} dm_0 \quad (18)$$

$$\begin{aligned}\Delta m_0(t) &= \left|m_0(t) - m_0(t+\Delta t)\right| \\ &= \left|N_p(t) - N_p(t+\Delta t)\right| \\ &= \left|N_s(t) - N_s(t+\Delta t)\right|\left(N_p(t)/N_s(t)\right)\end{aligned} \quad (19)$$

where $m_0(t)$ and $\Delta m_0(t)$ denote the total number and the change in the total number of particles due to aggregation. Similarly, $N_s(t)$ and $\Delta N_S(t)=|N_S(t)-N_S(t+\Delta t)|$ denote the number and the change in the number of particles in the sample population due to the occurrence of N_{agg} aggregation events in the time interval, Δt. In the absence of particle aggregation, the time step does not need to be explicitly calculated via Eq. (18) and, therefore, it can be arbitrarily selected in the MC algorithm.

To simulate the occurrence of a particle aggregation event, two particles of volumes V and U are randomly selected from the sample population. Following the developments of Garcia et al. (1987), an aggregation event is assumed to be successful if the following condition is satisfied:

$$\beta(V,U)/\beta_{max} \geq \kappa_i \tag{20}$$

where β_{max} is the maximum value of the particle aggregation kernel and κ_j is a randomly generated number in the range [0,1]. If the above probability criterion is met, the two randomly selected particles are removed from the sample population and a new particle with volume equal to (V+U) appears while the number of particles in the sample, $N_s(t)$, is reduced by one. In the opposite case, two new particles are randomly selected and the whole procedure is repeated till all the specified aggregation events, N_{agg}, have been completed.

In the presence of a particle growth mechanism, the volume of each particle in the sample population is subsequently increased from V_i to V_i' by taking into account the integral of the particle growth rate function, G(V), over the time interval, Δt.

$$V_i' = V_i + \int_t^{t+\Delta t} G(V)\,dt \tag{21}$$

Finally, in the presence of particle nucleation mechanism, a procedure similar to that employed for the reconstruction of the initial distribution is applied. Thus, at each time step, known numbers of new particles, having a specified distribution, are added to both total and sample populations.

In processes involving particle aggregation, as the MC simulation advances in time, the number of particles in the sample is constantly reduced. As a consequence, the statistical accuracy of the simulation is gradually lost. In order to deal with this problem the number of particles in the sample needs to be restored to its initial number, $N_s(0)$. Thus, when the particle number reaches a predetermined lower bound (e.g., $N_s(t)=f_A\,N_s(0)$), new particles of appropriate sizes are introduced into the various discrete volume intervals in such a way so that the sample distribution is preserved. This is achieved by the following procedure.

Let $V_{j,tot}$ be the total particle volume in the sample bin $[V_j,\ V_{j+1}]$ at time t. That is, $V_{j,tot}=\sum_{i=1}^{N_j} V_{j,i}\,N_{j,i}$ where $V_{j,i}$ and $N_{j,i}$ are the volume and the number of the "i" particles in the interval $[V_j,\ V_{j+1}]$, respectively. Let $N_{j,tot}$ be the total number of particles in the sample bin $[V_j,\ V_{j+1}]$ at time t (i.e., $N_{j,tot}=\sum_{i=1}^{N_j} N_{j,i}$) and f_A a number fraction parameter varying from 1 to 0. To ensure that the form of the distribution does not change during the particle refreshing procedure, the volumes assigned to the new particles, added to the interval $[V_j,\ V_{j+1}]$, must satisfy the following condition:

$$V_{j,ref} = V_{j,tot}\,(1-f_A)/\left(f_A\,N_{j,ref}\right) \tag{22}$$

where $N_{j,ref}=INT\left\{N_{j,tot}\left[(1-f_A)/f_A\right]\right\}$ is the number of particles added to the volume interval (where the symbol INT denotes the integer part of the result). The above refreshing procedure does not alter the information gathered from the precedent particle events and allows the simulation to carry on, theoretically, for an infinite period of time while the statistical error is maintained within acceptable limits.

In processes involving particle nucleation, the number of particles in the sample population is constantly increased. This increase in the number of particles raises the computational demands of the MC simulation and, therefore, the number of particles in the sample needs to be kept below a predetermined upper limit (i.e., f_N % of the initial number $N_s(0)$). Thus, when the number of particles in the sample reaches the specified upper limit, particles are randomly removed from the sample, so that the total number of particles, $N_s(t)$, is restored down to its initial value, $N_s(0)$, while the current form of the sample distribution is preserved.

3. RESULTS

Detailed numerical simulations were carried out for several particulate processes undergoing particle aggregation, growth and nucleation. Several particle aggregation rate functions (i.e., constant, sum and Brownian aggregation kernels) and particle growth rate functions (i.e., size independent and size dependent) were considered. The particle nucleation rate function was assumed to follow an exponential, size-dependent model (i.e., $S(V,t)=(N_{0,s}/V_{0,s})\exp(-V/V_{0,s})$, where N_{0s} and V_{0s} are some characteristic values of the distribution). Finally, in most cases studied, the initial number density function, n(V, 0), was assumed to have an exponential dependence with respect to particle volume,

$$n(V,0)=(N_0/V_0)\exp(-V/V_0) \tag{23}$$

In one case, it was assumed that n(V, 0) followed a Gaussian-distribution of the form:

$$n(V,0)=\left(\sigma\sqrt{2\pi}\right)^{-1}\exp\left(-(V-V_0)^2/2\sigma^2\right) \tag{24}$$

As in previous publications (i.e., Roussos et al. 2005), the following dimensionless aggregation, τ_a, and growth, τ_g, time constants were defined:

$$\tau_a = \beta_0 V_0^{\gamma} N_0 t \quad ; \quad \tau_g = G_v(V_0) t / V_0 \qquad (25)$$

where β_0, N_0 and V_0, are some characteristic values of the aggregation rate constant, particle number and particle volume, respectively.

It should be noted that in a previous publication (Roussos et al. 2005), it was shown that the GFEM results (i.e., the calculated distributions and their respective moments) were in excellent agreement with the analytical solutions.

3.1 Pure Aggregation Processes

In Fig. 1, the distributions calculated by MC and GFEM, for the case of constant particle aggregation (i.e., $\beta(V,U) = \beta_0$), are depicted for two different values of the dimensionless aggregation time (i.e., τ_a=1 and τ_a=10), and an initial Gaussian density function: $n(V,0) \approx 2\exp\left(-(V-1)^2/0.08\right)$

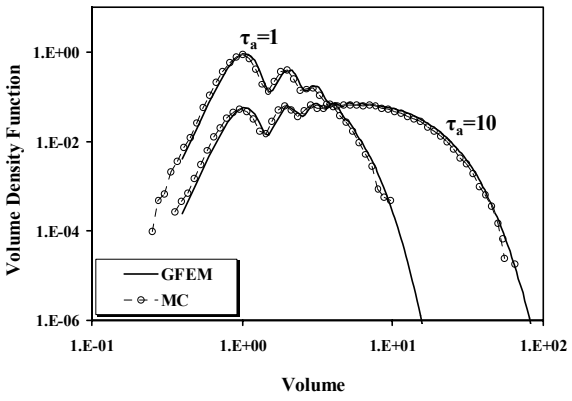


Figure 1. Comparison of dynamic PSDs for constant particle aggregation.

It is interesting to note that both methods are capable of predicting very accurately the multiple picks appearing in the small-volume part of the distribution.

In Fig. 2, the calculated distributions by the MC method and the GFEM, for the case of a Brownian particle aggregation kernel (i.e., $\beta(V,U)=\beta_0/4\{(V/U)^{1/3}+(U/V)^{1/3}+2\}$), are plotted for three different values of the dimensionless aggregation time (i.e., τ_a=1, τ_a= 10and $\tau_a=10^2$).

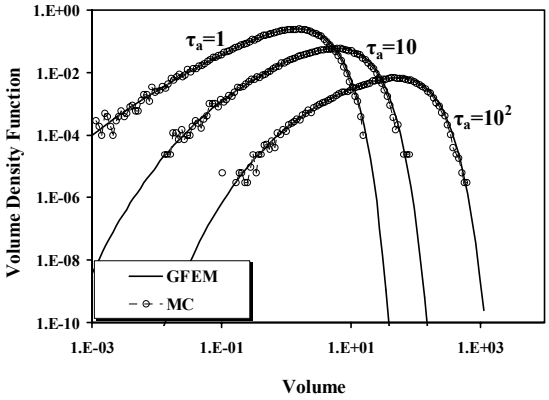


Figure 2. Comparison of dynamic PSDs for brownian particle aggregation

3.2 Combined Aggregation and Growth Processes

The calculated distributions by the two methods, for the case of a sum particle aggregation kernel and a linear particle growth rate function, for two different sets of dimensionless times (i.e., τ_a=3, τ_g=1 and τ_a=3, τ_g=2), are shown in Fig. 3.

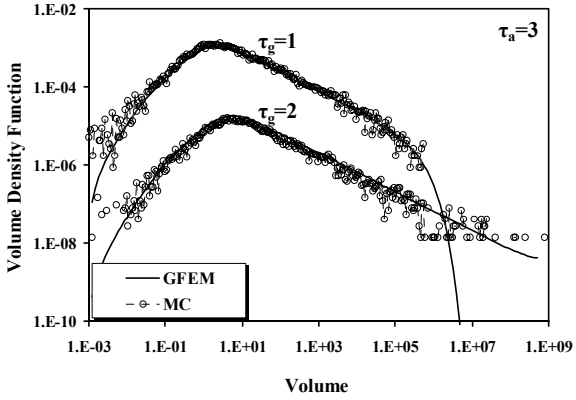


Figure 3. Comparison of dynamic PSDs for sum particle aggregation and linear particle growth

The calculated distributions by the two methods are in very good agreement despite the large oscillations displayed by the MC method at the low and high volume range.

3.3 Combined Aggregation, Growth and Nucleation Processes

In this case, all three mechanisms (i.e., constant particle aggregation and growth, exponential nucleation function) were assumed to take place simultaneously. The distributions calculated by the two methods are compared in Fig. 4 for two sets of the dimensionless parameters (i.e., τ_a=1, τ_g=1 and τ_a=1, τ_g=10).

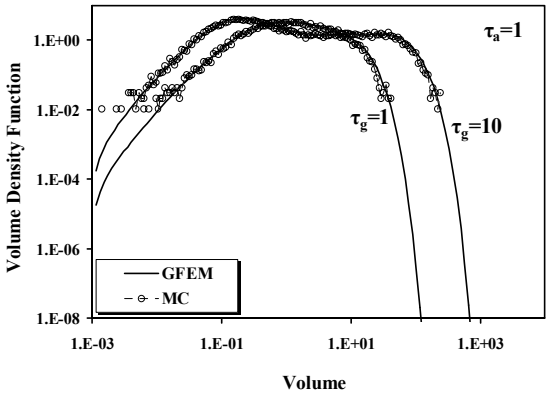


Figure 4. Comparison of dynamic PSDs for constant particle aggregation, constant particle growth and exponential particle nucleation

There is a very good agreement between the distributions calculated by the GFEM and the MC method.

3.4 Reconstruction of the distribution from its moments

In Fig. 5, the GMOM reconstructed distributions for various cases (i.e., constant particle aggregation,

combined constant particle aggregation and constant particle growth, combined constant particle aggregation and linear particle growth and combined constant particle aggregation, constant particle growth and exponential particle nucleation) are plotted. In all cases, the distributions were reconstructed using a set of four moments (i.e., m_0, $m_{0.5}$, m_1 and m_2) following the procedure described in detail in section 2.2.

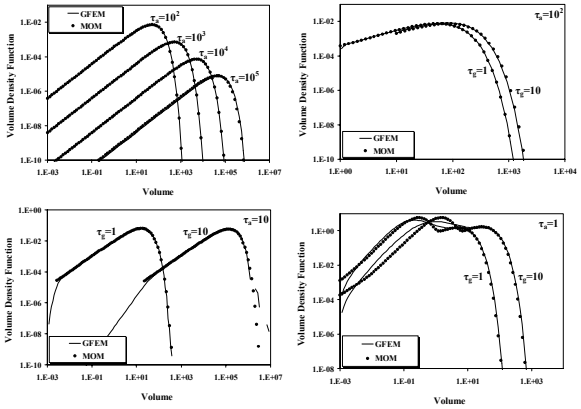


Figure 5. Comparison of dynamic PSDs calculated with the use of the GMOM

The reconstructed distributions are compared with the ones calculated by the GFEM. Apparently there is a good agreement for the cases in which no particle nucleation takes place). In the last case, the reconstructed distribution displays a satisfactory agreement with the distribution calculated by the GFEM. However, there is a notable deviation in the first part of the distribution, representing the contribution of the newly generated particles. This discrepancy is due to the fact that the first part of the distribution cannot be represented accurately by the selected form of the density function (see Eq.(15)).

REFERENCES

Alexiadis, A., M. Vanni. and P. Guardin (2004). Extension of the method of moments for population balances involving fractional moments and application to a typical agglomeration problem. Journal of Colloid and Interface Science, 276, 106-112.

Alexopoulos, A.H., A.I. Roussos and C. Kiparissides (2004). Part I: Dynamic Evolution of the Particle Size Distribution in Particulate Processes Undergoing Combined Particle Growth and Aggregation. Chemical Engineering Science, 59, 5751-5769.

Alexopoulos, A.H. and C. Kiparissides (2005). Part II: Dynamic Evolution of the Particle Size Distribution in Particulate Processes Undergoing Simultaneous Particle Nucleation, Growth and Aggregation. Chemical Engineering Science. 60, 4157-4169.

Finlayson, B.A. (1980). Nonlinear analysis in chemical engineering. McGraw-Hill, NY, New York.

Garcia, A.L., C. van de Broek, M. Aertsens and R. Serneels (1987). A Monte Carlo Simulation of Coagulation.. Physica, 143(3), 535-546.

Gooch, J.R. and M.J. Hounslow (1996). Monte Carlo Simulation of Size-Enlargement Mechanisms in Crystallization. AIChE J., 42(7), 1864-1874.

Hulburt, H.M. and S. Katz (1964). Some Problems in Particle technology. A Statistical Mechanical Formulation. Chemical Engineering Science, 19, 555-574.

Kumar, S. and D. Ramkrishna (1996). On the Solution of Population Balance Equations by Discretization-I. A Fixed Pivot Technique. Chemical Engineering Science, 51(8), 1311-1332.

Kumar, S. and D. Ramkrishna (1996). On the Solution of Population Balance Equations by Discretization-II. A Moving Pivot Technique. Chemical Engineering Science, 51(8), 1333-1342.

Marchisio, D.L., and R.O. Fox (2005). Solution of population balance equations using the direct quadrature method of moments. Chemical Engineering Science, 36(1), 43-73.

McGraw, R. (1997). Description of aerosol dynamics by the quadrature method of moments. Aerosol Science and Technology, 27, 255-265.

Ramkrishna, D., (1981). Analysis of Population Balance IV. The Precise Connection Between Monte Carlo and Population Balances. Chemical Engineering Science, 36, 1203-1209.

Ramkrishna, D. (1985). The status of population balances. Reviews Chemical Engineering, 3(1), 49-95.

Ramkrishna, D. (2000). Population Balances: Theory and Applications to Particulate Systems in Engineering. San Diego, California, Academic Press.

Roussos, A.I. (2004). Development of numerical methods for the solution of population balances: Application to batch and continuous particulate processes. PhD Thesis, Aristotle University of Thessaloniki, Greece.

Roussos, A.I.,A.H. Alexopoulos and C. Kiparissides (2005). Part III: Dynamic evolution of the particle size distribution in batch and continuous particulate processes: A Galerkin on finite elements approach. Chemical Engineering Science, 60, 6998-7010.

Rubinstein, R.Y. (1981), Simulation and the Monte Carlo Method, ch. 3, Wiley, New York.

Shah, B.H., J.D. Borwanker and D. Ramkrishna, (1977). Simulation of Particulate Systems Using the Concept of the Interval of Quiescence. AIChE J., 23, 897-904.

Spielman, L.A. and O. Levenspiel (1965). A Monte Carlo Treatment of Reacting and Coalescing Dispersed Phase Systems, Chemical Engineering Science, 20, 247-254.

Williams, N.M.R. and S.K. Loyalka (1991). Aerosol science: Theory and practice (with special application to the nuclear industry). New York, Pergamon Press.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

NON LINEAR OBSERVER FOR THE RECONSTRUCTION OF CRYSTAL SIZE DISTRIBUTIONS IN POLYMORPHIC CRYSTALLIZATION PROCESSES

Toufik Bakir *,[1] **Sami Othman** * **Gilles Fevotte** *
Hassan Hammouri *

* *LAGEP, UMR CNRS 5007/ CPE Université Claude Bernard Lyon 1, 43 bd du 11 Nov. 1918, 69622 Villeurbanne Cedex France*

Abstract: This paper deals with the problem of estimating CSD (Crystal Size Distributions) during polymorphic crystallization processes. The proposed approach is based on the high gain observer using the discretization of the PBE's (Population Balance Equations). First, in the growth phase, the monitoring of the nuclei production permits the estimation of the CSD using an adequate observer. In the dissolution of the metastable phase, the lack of on line sensors doesn't allow to synthesize a classical observer. However, the stability of the PDE allows to design an open loop observer. In fact, this stability is necessary to guarantee the convergence of the proposed observer. The performance of the given observer is discussed in the presence of noise measurements.

Keywords: Chemical industry, Observers, Partial differential equations, Finite difference method, Nonlinear systems.

1. INTRODUCTION

In biotechnology and pharmaceutical industries, polymorphism may occur for many products, which means that a given compound may exhibit several crystal structures. Such industrial products are obtained generally in a solid state through a batch crystallization process. Polymorphic structures exhibit different physical and chemical properties such as crystal morphology, solubility, and color, which affect the performance of the ingredients. Concerning drugs, and from a safety point of view, the polymorphic structure has to be controlled to keep the proper product performance. To do so, on line measurements have to be realized. The monitoring of crystallization processes has been object of number of publications. In (Ono *et al.*, 2004), Raman spectroscopy was used in order to measure polymorphic composition, a simulation of the process was also developed. In (Starbuck *et al.*, 2002) and (Caillet *et al.*, 2006), the aim was also the monitoring of different transitions using in-situ Raman spectroscopy. In the current work, is designed an asymptotic observer to estimate the CSD yield by polymorphism in crystallization process. The performances of this technique are discussed through simulation results. The stability of the PDE (partial differential equation) describing the PBE of the CSD is basically used, it justifies the possibility of the estimation of distribution without specific measurements (dissolution phase), this part will be developed below.

The paper is organized as follows, the polymorphism in batch crystallization is briefly described

[1] bakir@lagep.univ-lyon1.fr

(section 2). The principle of discretization of the PBEs (Population Balance Equations) is then exposed in section 3. Section 4 is devoted to the observer synthesis. In section 5, the estimation technique is validated through simulation.

2. MODEL DEVELOPMENT

Polymorphism in crystallization process can be defined as a set of N crystalline forms produced in the same stirred reactor. The dynamical model of such process is described by a set of population balances, a material balance relating the solute concentration and the different solid concentration of the crystalline forms, and an energy balance for all the components of the reactor. In the case of two crystalline forms, the following population balance approach is applied for both CSDs (crystal size distributions), it yields the following partial differential equations (PDEs):

$$\frac{\partial n_1(x_1,t)}{\partial t} + G_1(t)\frac{\partial n_1(x_1,t)}{\partial x_1} = 0 \qquad (1)$$

$$\frac{\partial n_2(x_2,t)}{\partial t} + G_2(t)\frac{\partial n_2(x_2,t)}{\partial x_2} = 0 \qquad (2)$$

$n_1(x_1,t)$ and $n_2(x_2,t)$ are the number population density function for the two crystal forms respectively (stable and metastable forms). each function represents the number of crystals of size x_1 or x_2 per unit volume of suspension and per unit of size. In equations (1) and (2), only nucleation and growth are be considered, agglomeration and breakage are not taken into account. The growth kinetics $G_1(t)$ and $G_2(t)$ are assumed to be size independent.
The solute concentration balance describing the mass transfer from the liquid to the solid phase is:

$$\frac{dV_t(t)C(t)}{dt} + \frac{dV_T C_{S1}(t)}{dt} + \frac{dV_T C_{S2}(t)}{dt} = 0 \qquad (3)$$

$C(t)$ represents the solute concentration, V_T is the suspension volume, variations of this volume, due to solute mass transfer can be neglected. $C_{S1}(t)$ and $C_{S2}(t)$ being the solid concentration of the two phases, they can be deduced from the crystal size distributions (CSDs) :

$$C_{S1}(t) = \frac{K_{V1}\rho_s}{M_s}\int_0^\infty x_1^3 n_1(x_1,t)dx_1 \qquad (4)$$

$$C_{S2}(t) = \frac{K_{V2}\rho_s}{M_s}\int_0^\infty x_2^3 n_2(x_2,t)dx_2 \qquad (5)$$

where K_{V1} and K_{V2} are the shape factors for the two forms (for sphere $K_V = \frac{\pi}{6}$), M_s is the molecular weight of solid of density ρ_s, and $V_t(t)$ is the solution volume (i.e. the continuous phase), which is calculated as :

$$V_t(t) = V_T(1 - \frac{M_s}{\rho_s}C_{ST}(t)) \qquad (6)$$

with :

$$C_{ST}(t) = C_{S1}(t) + C_{S2}(t) \qquad (7)$$

The crystallizer temperature is described by the energy balance :

$$\sum_{i=1}^{3} Cp_i n_i \frac{\partial T_{cr}}{\partial t} = -\triangle H_c V_T \frac{dC_{ST}}{dt} - UA(T_{cr} - T_j) \qquad (8)$$

where Cp_i and n_i represent respectively the molar heat capacities and the number of moles of the different components in the crystallizer. T_{cr} and T_j are respectively the crystallizer and jacket temperatures. $\triangle H_c$ is the crystallization enthalpy. U and A_c are respectively the overall heat transfer coefficient and contact surface through the jacket wall. The solubility, which refers to the solute concentration under saturated conditions, is assumed to obey Van't Hoff equation and is given by :

$$C_{sat1}(T) = A_1 \exp(\frac{-\Delta H_c}{RT}) \qquad (9)$$

$$C_{sat2}(T) = A_2 \exp(\frac{-\Delta H_c}{RT}) \qquad (10)$$

$C_{sat1}(T)$ and $C_{sat2}(T)$ represent respectively the solubility for stable and metastable form ($A_1 < A_2$), the absolute supersaturation ($C - C_{sat}$) is the driving force of the crystallization process. When this value is positive, the overall growth rate, including possible diffusive limitations, is assumed to be represented by the following model.

$$G_1(t) = K_{c1}\frac{M_s}{2\rho_s}\eta_1(C(t) - C_{sat1}(t))^g \qquad (11)$$

$$G_2(t) = K_{c2}\frac{M_s}{2\rho_s}\eta_2(C(t) - C_{sat2}(t))^g \qquad (12)$$

where K_{c1} and K_{c2} represent the kinetic growth rate coefficients, η_1 and η_2 represent the effectiveness factors. For example, for the first population, η_1 is the solution of the following equation :

$$\frac{K_{c1}}{K_{d1}}(C(t) - C_{sat1}(t))^{g-1}\eta_1 + \eta_1^{\frac{1}{g}} - 1 = 0 \qquad (13)$$

K_{d1} represents the mass transfer coefficient through diffusion which will be assumed to be the same for all crystal sizes. In the literature, values of exponent g were generally assumed to lie between 1 and 2, Analytical solution of equation (13) is available if g is equal to 1 or 2, a numerical solution can be considered in the other cases.

In the case of a negative supersaturation, the growth kinetic is replaced by a dissolution kinetic, it takes the following form :

$$D_1(t) = -K_{dis1}\frac{M_s}{2\rho_s}(C(t) - C_{sat1}(t))^g \qquad (14)$$

$$D_2(t) = -K_{dis2}\frac{M_s}{2\rho_s}(C(t) - C_{sat2}(t))^g \qquad (15)$$

In this case, g is assumed to be equal to 2. K_{dis1} and K_{dis1} are the dissolution coefficients for stable and metastable population. Concerning the two populations, the nucleation rate B is the result of two competitive nucleation mechanisms. Primary nucleation takes place in the absence of any crystal in the solution :

$$B_{11} = A_{11}\exp\left(\frac{B_{11}}{\ln^2(\frac{C(t)}{C_{sat1}(t)})}\right) \qquad (16)$$

$$B_{21} = A_{21}\exp\left(\frac{B_{21}}{\ln^2(\frac{C(t)}{C_{sat2}(t)})}\right) \qquad (17)$$

and secondary nucleation, which may occur at lower supersaturation level, is favored by the presence of solid in suspension (i.e. added in the crystallizer through seeding or generated through primary nucleation) :

$$B_{12} = A_{12}M_{T1}^i(C(t) - C_{sat1}(t))^j \qquad (18)$$

$$B_{22} = A_{22}M_{T2}^i(C(t) - C_{sat2}(t))^j \qquad (19)$$

A_{11}, A_{21}, B_{11} and B_{21} are the primary nucleation parameters, A_{12} and A_{22} are the secondary nucleation parameters, M_{T1} and M_{T2} are respectively the crystal mass of the stable and metastable crystal form in the solution. In the case of positive supersaturation, the boundary condition for equations (1) and (2) are usually set as follows :

$$n_1(x_1^*, t) = \frac{B_1(x_1^*)}{G_1(x_1^*)} \simeq \frac{B_1}{G_1} \qquad (20)$$

$$n_2(x_2^*, t) = \frac{B_2(x_2^*)}{G_2(x_2^*)} \simeq \frac{B_2}{G_2} \qquad (21)$$

Where only small crystal nuclei of critical size x_1^* and x_2^* are assumed to grow.

At first, clear solution is prepared, the two forms being undersaturated. The solution is cooled until nuclei of the metastable form are produced. The production of metastable nuclei yields a decrease of the solute concentration. The decrease of the temperature generates more supersaturation, and thus, a growth of the two forms. The process behavior for the metastable form changes when metastable concentration crosses the metastable solubility curve. Dissolution of this form begins, and the polymorphic fraction of the metastable form decreases. At the same time, the growth of stable form continues until the consumption of the solute concentration.

This description is done in order to analyze the observability of both forms. Concerning the stable form, the nuclei production guarantees the observability during all of the process. It is the case for the metastable form until the behavior changes

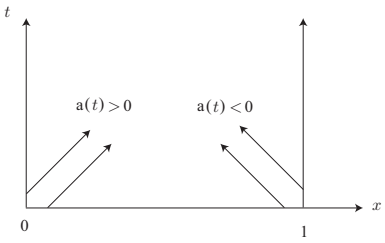


Fig. 1. a(t) conditions for the PDE stability

(dissolution phase). The stability of the equation system during the dissolution phase permits the estimation of the remaining part. The PDE describing the metastable CSD is an hyperbolic equation, it has the following form :

$$\frac{\partial F(x,t)}{\partial t} + a(t)\frac{\partial F(x,t)}{\partial x} = 0 \qquad (22)$$

Figure (1) gives the condition for the stability of the equation (22), the variable x is normalized ($0 \leq x \leq 1$). The stability is guaranteed if the condition concerning $a(x)$ is respected. In our case, when the metastable form is in dissolution phase, $G_2(t)$ is replaced by $D_2(t)$, this dissolution kinetic gives negative values, thus, the system is stable, which allows the boundedness of the error between the model and the estimated values.

3. DISCRETIZATION OF THE PBE "POPULATION BALANCE EQUATION"

Finite difference method is applied in the current study for the discretization of the PBEs. This choice is motivated by the structure obtained by this method which corresponds exactly to the observer one. Indeed, the state matrix involved exhibits tri-diagonal form. Moreover, the method concurs with the physical behavior of the system. The principle of the discretization for (1) and (2) is exactly the same. The system resulting for one of the two PDEs from the discretization turns out to be :

$$\begin{cases} \dot{n_x} = \alpha(t)An_x \\ y = Cn_x \end{cases} \qquad (23)$$

With:

$$\alpha(t) = \frac{G(t)}{\Delta x} \qquad (24)$$

In the case of dissolution :

$$\alpha(t) = \frac{D(t)}{\Delta x} \qquad (25)$$

$$n_x = \begin{pmatrix} n_{x_1} \\ n_{x_2} \\ n_{x_3} \\ \vdots \\ n_{x_{N-1}} \\ n_{x_N} \end{pmatrix}, A = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & \dots & 0 & 0 & 0 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \end{pmatrix},$$

Where $n_x \in \mathbb{R}^N$, $A \in \mathbb{R}^N \times \mathbb{R}^N$ and $C \in \mathbb{R}^N$

4. HIGH GAIN OBSERVER SYNTHESIS

As far as the crystal forms are supersaturated, system (23) associated to the output $\frac{Rn(t)}{G(t)}$ is observable. The production of nuclei allows to synthesize a high gain observer. In the undersaturation case, measurements are not available. However, an open loop observer may be applied to both forms to estimate the CSDs. The convergence of the proposed observer is dependent on the stability of the PBEs. In the following, the observer is applied for both PBEs. In the case of single output systems, the high gain observer is dedicated to the uniformly observable systems class of the following form :

$$\begin{cases} \dot{z} = f(z) + \sum_{i=1}^{N} u_i g_i(z) \\ y = h(z) \end{cases} \quad (26)$$

where $z(t) \in \mathbb{R}^N, y \in \mathbb{R}, u \in \mathbb{R}^p$ System (26) is said to be uniformly observable if for any two initial states $z \neq \bar{z}$ and every admissible inputs defined on any $[0, T]$, there exists $t \in [0, T]$ such that $y(z, u, t) \neq \bar{y}(\bar{z}, u, t)$, where y(z,u,t) is the output associated to the initial state z and the input u. In our case, system(26) takes the particular form of system (23) which is clearly observable due to its triangular form. The canonical form may be used to construct an exponential observer for system (23) under the following assumption :
$0 < \gamma \leq \alpha(t) \leq \xi \ \forall t \geq 0$
for some constants γ and ξ.
With continuous measurements, a candidate exponential observer for this system is given by (Farza *et al.*, 1997) and (Gauthier *et al.*, 1992):

$$\dot{\hat{z}}(t) = \alpha(t) A \hat{z}(t) - \alpha(t) S_\theta^{-1} C^T (C\hat{z}(t) - Y(t)), \quad (27)$$

where S is symmetric positive definite matrix given by the following equation :

$$\dot{S}_\theta(t) = -\theta S_\theta(t) - A^T S_\theta(t) - S_\theta(t) A + C^T C \quad (28)$$

If $\alpha(t)$ is negative for any time $t > 0$, the sign of the correction term should be changed :

$$\begin{cases} \dot{\hat{z}}(t) = \alpha(t) A(t) \hat{z}(t) + \\ \qquad \alpha(t) S_\theta^{-1} C^T (C\hat{z}(t) - Y(t)) \end{cases} \quad (29)$$

An other alternative is to use the following diffeomorphism $\phi : \mathbb{R}^N \to \mathbb{R}^N$
$z \to \phi(z) = [h, L_f(h), \dots, L_f^{n-1}(h)]$
Such diffeomorphism transforms the system (23) into the observable canonical form with :

$$A_1 = \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}, C_2 = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix},$$

The resulting observer has the following form :

$$\dot{\hat{z}}(t) = \alpha(t) A \hat{z} - \alpha(t) (\frac{\partial \phi}{\partial z}(\hat{z}, t))^{-1} S_\theta^{-1} C_2^T (C_2 \hat{x}(t) - Y(t)) \quad (30)$$

Where S_θ is given by the following Lyapunov equation :

$$\theta S_\theta(t) + A^T S_\theta(t) + S_\theta(t) A = C^T C \quad (31)$$

The terms of this matrix $S_\theta = [S_\theta(l,k)]_{1 \leq l,k \leq N}$ have the following form :

$$S_\theta(l,k) = \frac{(-1)^{l+k} D_{l+k-2}^{k-l}}{\theta^{l+k-1}} \quad (32)$$

With :

$$D_n^k = \frac{n!}{(n-k)!k!} \quad (33)$$

5. SIMULATION RESULTS AND DISCUSSION

5.1 Simulation conditions

The parameters used in this simulation are taken from the crystallization of adipic acid in water by (Marchal, 1989). Predictive models of homogenous primary nucleation A_1 were also taken from (Mersmann *et al.*, 2000). Concerning the jacket, the cooling fluid is assumed to be brine at $0°C$. Figure 2 summarizes the parameters values which were used during the simulation.

5.2 Simulation discussion

Figure (3) represents the solute concentration profile and the saturation concentration for both crystal forms. It can be seen that between the two temperatures (322K and 312K), the nuclei production of both forms is very small. For lower temperature, nucleation and crystal growth begin and stay until a temperature of about 300K. Below 300K, the metastable form is undersaturated while the stable form is supersatured. The metastable form therefore begins to dissolve. Nucleation and growth of the stable form still go on.

Figures (4) and (5)represent some examples of simulated classes of crystal sizes and the corresponding estimated ones. The choice of these crystal sizes is arbitrary. The same performances can

parameter	definition	unit	value
A_{11}	homogeneous primary nucleation parameter for CSD 1	$nb.m^{-3}.s^{-1}$	1 10^{10}
B_{11}	homogeneous primary nucleation parameter for CSD 1	$nb.m^{-3}.s^{-1}$	0.63
A_{21}	homogeneous primary nucleation parameter for CSD 2	$nb.m^{-3}.s^{-1}$	1 10^{12}
B_{21}	homogeneous primary nucleation parameter for CSD 2	$nb.m^{-3}.s^{-1}$	0.63
A_{12}	secondary nucleation parameter for CSD 1	$nb.m^{3(i+j-1)}mol^{-i-j}.s^{-1}$	1440
K_{c1}	growth constant for CSD 1	$mol^{(1-g)}.m^{(3g-2)}.s^{-1}$	0.0157
A_{22}	secondary nucleation parameter for CSD 2	$nb.m^{3(i+j-1)}mol^{-i-j}.s^{-1}$	1440
K_{c2}	growth constant for CSD 2	$mol^{(1-g)}.m^{(3g-2)}.s^{-1}$	0.0170
K_{dis1}	growth constant for CSD 1	$mol^{(1-g)}.m^{(3g-2)}.s^{-1}$	2 10^{-8}
K_{dis2}	growth constant for CSD 2	$mol^{(1-g)}.m^{(3g-2)}.s^{-1}$	2.5 10^{-8}
i	exponent	no dimension	1.968
j	exponent	no dimension	1
g	exponent	no dimension	2
M_s	molar mass	$Kg.mol^{-1}$	146.14 10^{-3}
ρ_s	volume mass	$Kg.m^{-3}$	1360
K_{V1}	shape factor for CSD 1	no dimension	$\frac{\pi}{6}$
K_{V2}	shape factor for CSD 2	no dimension	$\frac{\pi}{10}$
Cp_1	solute molar heat capacity	$J.K^{-1}.mol^{-1}$	3.72
Cp_2	solid molar heat capacity	$J.K^{-1}.mol^{-1}$	7.44
Cp_1	water molar heat capacity	$J.K^{-1}.mol^{-1}$	75.33
ΔH_c	crystallization enthalpy	$J.mol^{-1}$	-48000
U	overall heat transfer coefficient	$J.m^{-2}.K^{-1}.s^{-1}$	1000
A_c	contact surface through jacket wall	m^2	0.022

Fig. 2. simulation parameters values

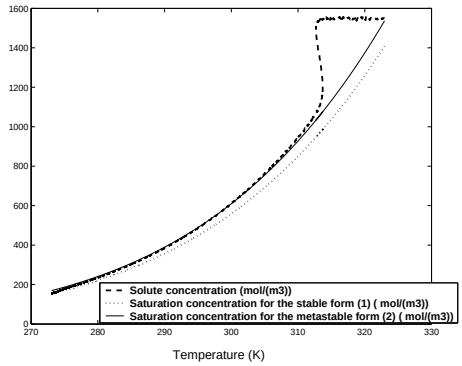


Fig. 3. solubility for stable and metastable forms

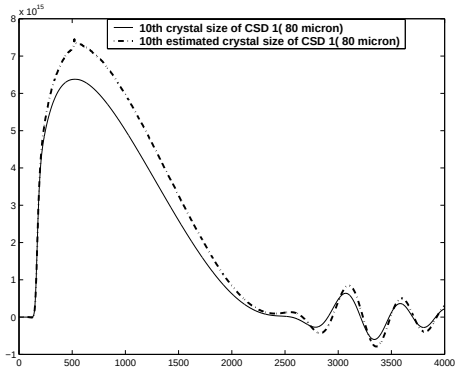


Fig. 4. 1Oth size crystal of the stable form (Model and estimate (dotted))

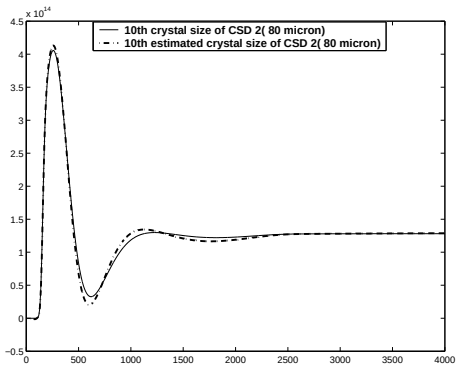


Fig. 5. 1Oth size crystal of the metastable form (Model and estimate (dotted))

be shown for the other sizes. Figure (4) represents the time variation of the number of stable crystals in the 10th size class (i.e. size around 80 μ m) and its estimation. Figure (5) represents the crystal size of the metastable form and its estimation. Before 500 s, the metastable form is supersaturated. Then, the metastable form becomes undersaturated (this time corresponds to the temperature of 300K), the open loop observer is then used. As mentioned above, the stability of the system yield by the PBE discretization implies the assymptotic convergence of the observer. The crystal size and its estimate tend to the same final value with an acceptable error, as shown on the figure.

Figures (6) and (7) represent the CSD of the stable form and its estimation. Figures (8) and (9) represent the CSD of metastable form and its es-

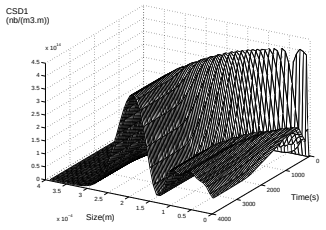


Fig. 6. crystal size distribution based on stable form model

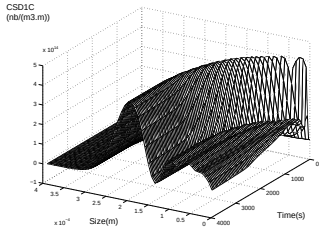


Fig. 7. estimation of stable form crystal size distribution

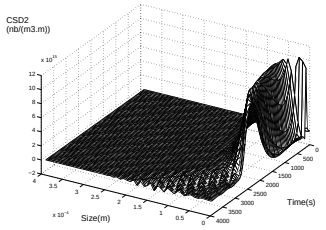


Fig. 8. crystal size distribution based on metastable form model

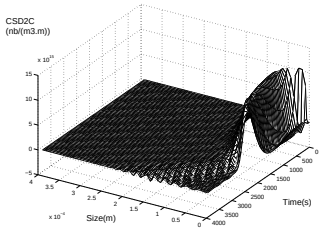


Fig. 9. estimation of metastable form crystal size distribution

timation. These figures summarize the comments made after the preceding results. The estimation of both CSD's is acceptable.

6. CONCLUSION

In this work, a methodology to estimate CSDs in polymorphic crystallization process has been presented. This methodology is based on a model for each crystal form. This model in obtained by the solute concentration balance, the energy balance in the crystallizer in addition to population balance equations for both crystal forms. A high gain observer is applied to estimate the CSDs of stable and metastable forms in nucleation phase. In the dissolution phase, the metastable form can be estimated using open loop estimation. This estimation gives good results because of the stability of the PBEs. Additional simulations have shown that modelling errors in primary nucleation parameters don't affect considerably the observer robustness.The observer can be used for process supervision to prevent any variation of crystals population which may affect the product quality. In the case of control applications, the estimated crystal size distribution could be used to ensure advanced quality control objectives such as reproducible crystal number mean sizes and variances.

REFERENCES

Caillet, A., F. Puel and G. Fevotte (2006). In-line monitoring of partial and overall solid concentration during solvent-mediated phase transition using raman spectroscopy. *International journal of pharmaceutics* **307(2)**, 201–208.

Farza, M., H. Hammouri, S. Othman and K. Busawon (1997). Non linear observer for parameter estimation in bioprocesses. *Chemical engeneering Science* **52**, 4251–4267.

Gauthier, J.P., H. Hammouri and S. Othman (1992). A simple observer for non linear systems application to bioreactors. *IEEE Trans. Automat. Control* **37**, 875–880.

Marchal, P. (1989). *Génie de la cristallisation: application l'acide adipique*. Thesis,Institut National Polytechnique de Lorraine. Nancy.

Mersmann, A., B. Braun and M. Löfflmann (2000). Prediction of crystallization coefficients of the population balance. *Chemical engeneering science* **57**, 4267–4275.

Ono, T., H. J. M. Kramer, J. H. ter Horst and P. J. Jansens (2004). Process modeling of the polymorphic transformation of l-glutamic acid. *Crystal growth and design* **4**, 1161–1167.

Starbuck, C., A. Lindemann, L. Wai, J. Wang, P. Fernandez, C. Lindemann, G. Zhou and Z. Ge (2002). Process optimization of a complex pharmaceutical polymorphic system via in situ raman spectroscopy. *Crystal growth and design* **2**, 515–522.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

CALCULATION OF THE MOLECULAR WEIGHT – LONG CHAIN BRANCHING DISTRIBUTION IN BRANCHED POLYMERS

Krallis Apostolos; Kiparissides Costas

Department of Chemical Engineering and Chemical Process Engineering Research Institute, Aristotle University of Thessaloniki, P.O. Box 472, 54006 Thessaloniki, Greece

Abstract: In the present study a population balance approach is described to follow the time evolution of molecular polymer properties in free-radical polymerizations. The model formulation is based on the fixed pivot technique (FPT) which was properly adapted to calculate the combined molecular weight - long chain branching distribution. At first the predictive capabilities of the proposed model were tested against experimental measurements and simulation results taken from the open literature, on molecular weight distribution (MWD) of branched polymers. Then the MWD calculated by the FPT was compared with the MWD calculated by the method of classes. However the FPT proved to be a faster method for the calculation of the MWD.

Keywords: Computing elements, Distributions, Mathematical model, Numerical algorithms, Polymerization

1. INTRODUCTION

The molecular properties (e.g., molecular weight distribution, MWD, copolymer composition distribution, CCD, long chain branching distribution, LCBD, etc.) of polymers are directly related to their end-use properties (e.g., mechanical, rheological, etc.). Hence, the ability to control accurately the molecular architecture of polymer chains in a polymerization reactor is of profound interest to the polymer industry. This presupposes a thorough knowledge of the polymerization kinetics and the availability of advanced mathematical models to quantify the effects of process operating conditions on the molecular polymer properties.

Branched polymers are characterized by the presence of long or/and short branches attached to the main backbone of a polymer chain. Thus, the end-use properties of branched polymers will also depend on the number, the type and the distribution of the branches. Long chain branching has a strong impact on the rheological behavior of the polymer. In fact, it affects the flow properties of the polymer melt (e.g., extensional viscosity, shear viscosity and elasticity) as well as the polymer solid state properties (e.g., orientation effects and stress induced crystallization). Thus, the elucidation of the LCB formation and its correlation with the various rheological and physical polymer properties are two subjects of significant research interest.

The free-radical polymerization of vinyl acetate (VAc) is a typical system that leads to the formation of long chain branching that largely affect the MWD and thus, the polymer rheological properties. In this system, transfer to monomer and to polymer reactions as well as terminal double bond polymerization largely control the molecular weight developments via the formation of highly branched polymer chains.

In the past twenty years, several mathematical models dealing with the calculation of the MWD of branched polymers have been published (Lorenzini et. al., 1992; Tobita and Hatanaka, 1996; Nordhus et. al., 1997; Thomas, 1998; Pladis and Kiparissides, 1998, Iedema et. al., 2000). A variety of numerical methods have been employed to calculate the MWD of branched polymers, including 'numerical fractionation' (Teymour and Campbell, 1992, 1994), Monte-Carlo simulations (Tobita, 1996; Tobita and Hatanaka, 1996), global orthogonal collocation (Canu and Ray, 1991; Nele et. al., 1999) and discrete weighted Galerkin (Wulkow, 1995). In general, the above numerical attempts suffer from two key kinetic limitations, (e.g., the use of the quasi steady state approximation (QSSA) for 'live' radical chains and the absence of gel and glass effect).

The 'numerical fractionation' method can provide information on the full MWD of branched polymers by dividing the total polymer chain population into a finite number of classes of polymer chains having narrow MWDs. The method assumes that the transition from one class of polymer chains to a higher one occurs exclusively by a geometric growth mechanism (e.g., termination by combination polymer and terminal double bond reactions are

reactions). However, in systems in which transfer to important, this assumption will not be sufficient. The use of global orthogonal collocation methods for the prediction of the MWD in free-radical polymerization systems is partially successful because a single interpolation polynomial is only employed for the entire collocation domain. As a result, prior knowledge on the type of the approximated distribution is required. Furthermore, global collocation schemes have been proved inadequate in accommodating complex MWDs (e.g., bimodal distributions, MWDs for branched polymers, etc.). Pladis and Kiparissides (1998) employed a polymer chain fractionation approach to calculate the molecular weight – long chain branching bivariate distribution for branched polymers. The total population of the polymer chains was divided into a number of classes with respect to the number of long chain branches. However, in addition to the well-known problem of closure of the 'higher order' moments, the reconstruction of the overall MWD at high monomer conversions and high LCB content, requires a very large number of classes to reduce the approximation errors associated with the high molecular weight fractions of the distribution. Monte Carlo simulations are straightforward techniques that can generally handle complex kinetic mechanisms but usually require significant computational effort for the determination of the MWD. Finally, the discrete weighted Galerkin formulation, even though is computationally demanding, provides a powerful tool for the prediction of the MWD in complex polymerization systems. However, the approximation of the infinite summation terms (e.g., resulting from termination by combination reactions) requires special treatment.

The present study deals with the numerical solution of the dynamic bivariate population balance equations (PBEs) for 'live' and 'dead' polymer chains, arising in highly branched polymer systems. The fixed pivot technique (Kumar and Ramkrishna, 1996) is employed to solve the resulting system of bivariate population balance equations. The validity of the proposed numerical method is tested by a direct comparison of model predictions with experimental data on the number average degree of branching, the number and weight average molecular weights for the free-radical polymerization of VAc (Thomas, 1998). The calculated bivariate MW-LCB distribution is also compared with simulations obtained by an improved method of classes (Pladis and Kiparissides, 1998) as well as with predictions of MWD obtained by Monte Carlo Simulations (Tobita and Hatanaka, 1996).

2. KINETIC MECHANISM AND RATE FUNCTIONS

In the present study, the following kinetic mechanism was employed to describe the formation of highly branched polymers:

Initiator decomposition:

$$I \xrightarrow{k_d} 2PR^{\bullet}$$

Chain initiation:

$$PR^{\bullet} + M \xrightarrow{k_I} P_{0,1}$$

Propagation:

$$P_{b,n} + M \xrightarrow{k_p} P_{b,n+1}$$

Chain transfer to monomer:

$$P_{b,n} + M \xrightarrow{k_{fm}} D_{b,n} + P_{0,1}$$

Chain transfer to solvent:

$$P_{b,n} + S \xrightarrow{k_{fs}} D_{b,n} + P_{0,1}$$

Reaction with terminal double bond:

$$P_{b,n} + D_{r,m} \xrightarrow{k_{db}} P_{b+r+1,n+m}$$

Chain transfer to polymer:

$$P_{b,n} + D_{r,m} \xrightarrow{k_{fp}} D_{b,n} + P_{r+1,m}$$

Termination by combination:

$$P_{b,n} + P_{r,m} \xrightarrow{k_{tc}} D_{b+r,n+m}$$

Termination by disproportionation:

$$P_{b,n} + P_{r,m} \xrightarrow{k_{td}} D_{b,n} + D_{r,m}$$

The symbols $P_{b,n}$ and $D_{b,n}$ denote the respective 'live' and 'dead' polymer chains with 'b' long chain branches and a chain length equal to 'n'. The above kinetic mechanism includes initiation and propagation reactions, termination by combination and disproportionation, molecular weight control reactions via transfer to monomer and chain transfer agent (solvent) and long chain branching formation via transfer to polymer and terminal double bond reactions. Polymer chains with terminal double bonds, formed via termination by disproportionation and transfer to monomer reactions, can react with 'live' polymer chains to produce long chain branches. Transfer to polymer reactions involve the transfer of reactivity from a growing polymer chain to a 'dead' polymer chain. More specifically, a hydrogen atom abstracted from the backbone of a 'dead' polymer chain leads to the formation of a new 'live' polymer chain with an internal radical center and a 'dead' polymer chain.

In the present study, to reduce the number of bivariate population balances to be numerically solved, it was assumed that the concentration of the 'dead' polymer chains having a terminal double bond, was some known fraction of the 'dead' polymer chains (Baltsas et. al., 1996). Based on the above kinetic mechanism and assumptions, the following dynamic population balance equations for the 'live', $P(b,n,t)$, and 'dead', $D(b,n,t)$, polymer chains can be derived:

$$\frac{1}{V}\frac{\partial P(b,n,t)}{\partial t} = r_{P(b,n,t)} \quad (1)$$

$$\frac{1}{V}\frac{\partial D(b,n,t)}{\partial t} = r_{D(b,n,t)} \quad (2)$$

The net formation rates for the 'live' and 'dead' polymer chains are given by the following equations:

Net formation rate of 'live' polymer chains of length 'n' with 'b' branches :

$$r_{P(b,n,t)} = k_I\left[PR^{\bullet}\right][M]\delta(n-1)\delta(b) + k_{fm}[M]\sum_{z=0}^{N_b}\sum_{x=1}^{N_n}P(z,x,t)\delta(n-1)\delta(b) +$$
$$k_{fs}[S]\sum_{z=0}^{N_b}\sum_{x=1}^{N_n}P(z,x,t)\delta(n-1)\delta(b) + k_p[M]\left[P(b,n-1,t) - P(b,n,t)\right] -$$
$$\left(k_{fm}[M] + k_{fs}[S]\right)P(b,n,t) + k_{fp}nD(b-1,n,t)\sum_{z=0}^{N_b}\sum_{x=1}^{N_n}P(z,x,t) -$$
$$k_{fp}P(b,n,t)\sum_{z=0}^{N_b}\sum_{x=2}^{N_n}xD(z,x,t) - k_{tc}P(b,n,t)\sum_{z=0}^{N_b}\sum_{x=1}^{N_n}P(z,x,t) -$$
$$k_{td}P(b,n,t)\sum_{z=0}^{N_b}\sum_{x=1}^{N_n}P(z,x,t) - k_{db}P(b,n,t)\sum_{z=0}^{N_b}\sum_{x=2}^{N_n}D(z,x,t) +$$
$$k_{db}\sum_{z=0}^{b-1}\sum_{x=1}^{n-1}P(z,n-x,t)D(b-z-1,x,t) \qquad (3)$$

Net formation rate of 'dead' polymer chains of length 'n' with 'b' branches:

$$r_{D(b,n,t)} = \left(k_{fm}[M] + k_{fs}[S]\right)P(b,n,t) + k_{fp}P(b,n,t)\sum_{z=0}^{N_b}\sum_{x=2}^{N_n}xD(z,x,t)$$
$$-k_{fp}nD(b,n,t)\sum_{z=0}^{N_b}\sum_{x=1}^{N_n}P(z,x,t) - k_{db}D(b,n,t)\sum_{z=0}^{N_b}\sum_{x=1}^{N_n}P(z,x,t)$$
$$+k_{td}P(b,n,t)\sum_{z=0}^{N_b}\sum_{x=1}^{N_n}P(z,x,t) + \frac{1}{2}k_{tc}\sum_{z=0}^{b}\sum_{x=1}^{n-1}P(z,x,t)P(b-z,n-x,t) \qquad (4)$$

where $\delta(n)$ is the Kronecker's delta function [e.g., $\delta(n)=1$ if $n=0$ and $\delta(n)=0$ if $n\neq 0$]. N_b and N_n denote the maximum number of branches and the maximum chain length, respectively. It should be pointed out that the actual number of rate equations for the 'live' and 'dead' polymer chains will depend on the total degree of polymerization, that may be of the order of hundreds or/and thousands monomer units. Consequently, the computational effort associated with the solution of the complete set of differential equations becomes prohibitively high for the most cases of interest and makes the on-line application of such a model unrealistic. To deal with the above high-dimensionality problem, several methods have been proposed to reduce the infinite system of differential equations into a low-order system of DAEs.

In the present work the fixed pivot technique was applied for the solution of the bivariate PBEs [see eqs. (3) and (4)] to predict the joint MW-LCB distribution of branched polymers.

3. FIXED PIVOT TECHNIQUE

The fixed pivot technique was properly adapted for solving the bivariate population balance equations for the 'live' and 'dead' polymer chains [see eqs (3) and (4)]. The method assumes that the overall polymer chain population can be assigned to selected discrete points, also called 'grid' points. The bivariate PBEs which are derived from the application of the proposed method are then solved at the discrete points. Thus, the initial infinite system of PBEs, is reduced to a system of discrete-continuous differential equations. Since the chain populations in various chain lengths and number of branches are assumed to exist only at the representative discrete points, specific reaction steps (i.e., termination by combination, propagation, chain transfer to polymer and terminal double bond), involving such chain populations, can result in the formation of new polymer chains whose chain lengths and/or number of branches do not correspond to the representative grid points. According to the 2-D FPT, the polymer chains that do not correspond to specific grid points are incorporated in the set of discrete-continuous dynamic PBEs in such a way that any four moments (two in each dimension), of the joint MW-LCB distribution, are exactly preserved.

In the bivariate PBEs, the distribution of polymer chains with a specific number of branches is considered to be continuous over the chain length domain and the number of long chain branches domain. Based on the original developments of Kumar and Ramkrishna (1996), the total branch and chain length domains, are divided into a number of finite elements $N_{e,b}$ and $N_{e,n}$ respectively. Let P(j,i,t) and D(j,i,t), be the concentrations of the 'live' and 'dead' polymer domain, which correspond to the discrete point u(j,i) of the 2-D domain (see Fig. 1).

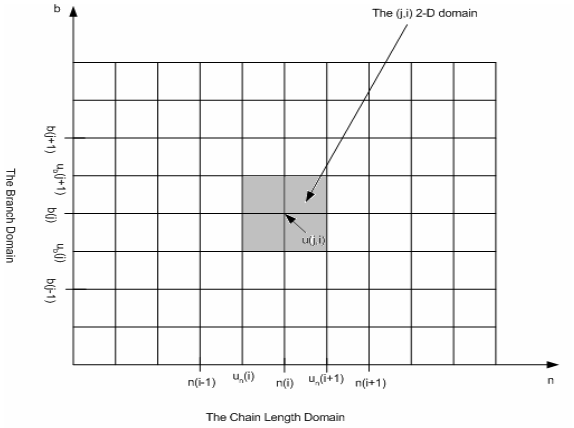


Fig. 1: The two-dimensional grid which can be used with the FPT.

Let n(i) and b(j) be the corresponding middle points in the i^{th} element ($u_n(i)$, $u_n(i+1)$) and j^{th} element ($u_b(j)$, $u_b(j+1)$, respectively. When a new polymer chain is formed within the 2-D discrete element (e.g., due to termination by combination or transfer to polymer reactions), its concentration is assigned to the four neighboring grid points in such a way so that selected moments of the MWD are exactly preserved. On the other hand, polymer chains formed via initiation, transfer to monomer, transfer to solvent or termination by disproportionation reactions, are always assigned to the existing grid points. From the application of the FPT to the bivariate PBEs of the 'live' and 'dead' polymer chains [see eqs (3) and (4)] we obtain the following system of continuous-discrete differential equations:

Continuous-discrete differential equations for linear 'live' polymer chains:

$$\frac{1}{V}\frac{d\left[VP(0,i,t)\right]}{dt}=2fk_d[I]\delta(i-1)+k_{fs}[S]\left(\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)\right)\delta(i-1)$$
$$+k_{fm}[M]\left(\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)\right)\delta(i-1)+k_p[M]\sum_{k=1}^{i}A(i,k)P(0,k,t)$$
$$-k_p[M]P(0,i,t)-k_{fs}[S]P(0,i,t)-k_{fm}[M]P(0,i,t)$$
$$-k_{fp}P(0,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}n(k)D(l,k,t)-k_{db}P(0,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}D(l,k,t)$$
$$-k_{td}P(0,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)-k_{tc}P(0,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t) \qquad (5)$$

Continuous-discrete differential equations for branched 'live' polymer chains:

$$\frac{1}{V}\frac{d\left[VP(j,i,t)\right]}{dt}=k_p[M]\sum_{k=1}^{i}A(i,k)P(j,k,t)-k_p[M]P(j,i,t)$$
$$-k_{fs}[S]P(j,i,t)-k_{fm}[M]P(j,i,t)-k_{fp}P(j,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}n(k)D(l,k,t)$$
$$+k_{fp}n(i)\left(\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)\right)\left(\sum_{l=0}^{j}C(j,l)D(l,i,t)\right)$$
$$-k_{db}P(j,i,t)\left(\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}D(l,k,t)\right)-k_{td}P(j,i,t)\left(\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)\right)$$
$$+k_{db}\sum_{l=0}^{j}\sum_{q=0}^{j}\sum_{k=1}^{i}\sum_{m=1}^{i}B(i,k,m)\,T(j,l,q)\,P(l,k,t)\,D(q,m,t)$$
$$-k_{tc}P(j,i,t)\left(\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)\right) \qquad (6)$$

Continuous-discrete differential equations for linear 'dead' polymer chains:

$$\frac{1}{V}\frac{d\left[VD(0,i,t)\right]}{dt}=k_{fs}[S]P(0,i,t)+k_{fp}P(0,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}n(k)D(l,k,t)$$
$$+k_{fm}[M]P(0,i,t)+k_{td}P(0,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)$$
$$+k_{tc}\sum_{k=1}^{i}\sum_{m=k}^{i}B(i,k,m)\,P(0,k,t)P(0,m,t)$$
$$-k_{fp}\,n(i)D(0,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)-k_{db}\,D(0,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t) \qquad (7)$$

Continuous-discrete differential equations for branched 'dead' polymer chains:

$$\frac{1}{V}\frac{d\left[V\,D(j,i,t)\right]}{dt}=k_{fs}[S]P(j,i,t)+k_{fp}P(j,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}n(i)\,D(l,k,t)$$
$$+k_{fm}[M]P(j,i,t)+k_{td}P(j,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)$$
$$-k_{fp}\,n(i)D(j,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)-k_{db}D(j,i,t)\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}P(l,k,t)$$
$$+k_{tc}\sum_{l=0}^{j}\sum_{q=l}^{j}\sum_{k=1}^{i}\sum_{m=k}^{i}B(i,k,m)\,O(j,l,q)\,P(l,k,t)P(q,m,t) \qquad (8)$$

for j = 1, 2,…, $N_{e,b}$ and i = 1, 2,…, $N_{e,n}$

Assuming that the zero and first moment of the MWD are preserved, the matrices $A(i,k)$, $B(i,k,m)$, $C(j,l)$, $T(j,l,q)$ and $O(j,l,q)$, can be calculated by the following expressions:

$$A(i,k)=\begin{cases}\dfrac{n(i+1)-n}{n(i+1)-n(i)} & n(i)\le n\le n(i+1)\\ \dfrac{n-n(i-1)}{n(i)-n(i-1)} & n(i-1)\le n\le n(i)\end{cases} \quad \text{and} \quad n=n(k)+ \qquad (9)$$

$$B(i,k,m)=\begin{cases}\dfrac{n(i+1)-n}{n(i+1)-n(i)} & n(i)\le n\le n(i+1)\\ \dfrac{n-n(i-1)}{n(i)-n(i-1)} & n(i-1)\le n\le n(i)\end{cases} \quad \text{and} \quad n=n(k)+n(m \qquad (10)$$

$$C(j,l)=\begin{cases}\dfrac{b(j+1)-b}{b(j+1)-b(j)} & b(j)\le b\le b(j+1)\\ \dfrac{b-b(j-1)}{b(j)-b(j-1)} & b(j-1)\le b\le b(j)\end{cases} \quad \text{and} \quad b=b(l)+1 \qquad (11)$$

$$T(j,l,q)=\begin{cases}\dfrac{b(j+1)-b}{b(j+1)-b(j)} & b(j)\le b\le b(j+1)\\ \dfrac{b-b(j-1)}{b(j)-b(j-1)} & b(j-1)\le b\le b(j)\end{cases} \quad \text{and} \quad b=b(l)+b(q)+1 \qquad (12)$$

$$O(j,l,q)=\begin{cases}\dfrac{b(j+1)-b}{b(j+1)-b(j)} & b(j)\le b\le b(j+1)\\ \dfrac{b-b(j-1)}{b(j)-b(j-1)} & b(j-1)\le b\le b(j)\end{cases} \quad \text{and} \quad b=b(l)+b(q) \qquad (13)$$

where δ is the Kronecker's delta function.

The resulting differential-discrete equations were integrated in time to calculate the dynamic behavior of the 'live' and 'dead' bivariate number chain length distributions. The concentrations of the 'dead' polymer chains at the grid points were then used to reconstruct the weight chain length distribution (WCLD) that corresponds to a specific grid point of the branch domain:

$$W(j,i,t)=n(i)\;\;D(j,i,t)\big/\left(u_n(i+1)-u_n(i)\right) \qquad (14)$$

The overall WCLD was then calculated by the weighted sum of all polymer branch distributions:

$$W_{total}(i,t)=\frac{\sum_{l=0}^{N_{e,b}}\left[n(i)D(l,i,t)\big/\left(u_n(i+1)-u_n(i)\right)\right]}{\sum_{l=0}^{N_{e,b}}\sum_{k=1}^{N_{e,n}}n(k)D(l,k,t)} \qquad (15)$$

For the discretization of the chain length and branch domains a logarithmic discretization rule was employed. Typically, the chain length and branch domains were partitioned into 50 and 8 finite elements, respectively, leading to a total number of 800 discrete-continuous differential equations.

To ensure that the selected number of elements was sufficient for the accurate reconstruction of the

MWD, the following convergence criterion was established:

$$\left(\mu_1 - \sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} n(k) D(l,k,t) \right) \Big/ \mu_1 \le \varepsilon \qquad (16)$$

ε is a convergence parameter with typical values in the range of (0, 0.03).

Finally, the number and weight average molecular weights and the number and weight average degrees of branching are calculated using the following equations:

Number average molecular weight:

$$M_n = \left(\sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} n(k) D(l,k,t) \Big/ \sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} D(l,k,t) \right) MW_m \qquad (17)$$

Weight average molecular weight:

$$M_n = \left(\sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} n(k)^2 D(l,k,t) \Big/ \sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} n(k) D(l,k,t) \right) MW_m \qquad (18)$$

Number average degree of branching:

$$B_n = \sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} b(l) D(l,k,t) \Big/ \sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} D(l,k,t) \qquad (19)$$

Weight average degree of branching:

$$B_w = \sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} n(k) b(l) D(l,k,t) \Big/ \sum_{l=0}^{N_{e,b}} \sum_{k=1}^{N_{e,n}} n(k) D(l,k,t) \qquad (20)$$

where, $D(j,i,t)$, denotes the concentration of polymer chains with chain length $n(i)$ and number of branches $b(j)$.

4. RESULTS AND DISCUSSION

The free-radical polymerization of VAc was selected as a representative example for the production of branched polymers. In this system, transfer to monomer and polymer largely control the MWD of the poly(vinyl acetate) produced. Furthermore, the monomer radicals that are produced from the transfer to monomer reaction, propagate giving 'live' and 'dead' polymer chains with a terminal double bond. Thus, terminal double bond polymerization is an important reaction for this system, producing highly branched polymer chains.

It is well known that for the VAc polymerization system, the termination kinetic rate constant becomes gradually controlled by the diffusion phenomena as the monomer conversion and hence the viscosity of the mixture increases (Hamer and Ray, 1986). In order to account for this variation the termination rate constant was expressed as the sum of two terms, one taking into account the effect of the diffusion of polymer chains, k_t^{dif}, and the other describing the so-called 'residual termination', k_t^{res}:

$$k_t = k_t^{dif} + k_t^{res} \qquad (21)$$

The analytical calculation of the diffusion controlled termination rate constant is provided in the work of Keramopoulos and Kiparissides (2002).

The numerical performance of the FPT was first tested by a direct comparison of numerical results with experimental measurements on number average degree of branching B_n, for the free-radical polymerization of VAc (Thomas, 1998). Two temperatures (i.e., 60^0C and 80^0C) and different initiator concentrations (i.e., 2,2'-azobis(2-methylpropionitrile, AIBN) were used in the comparison analysis (see Fig.2).

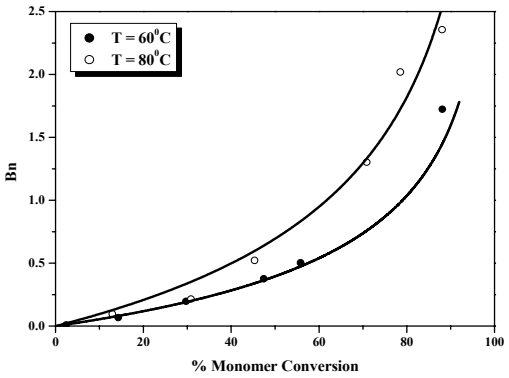


Fig. 2: Predicted and experimental number average degree of branching with respect to monomer conversion (T=60^0C and $[I_0] = 5 \times 10^{-5}$ mol/L; T=80 ^{0}C and $[I_0] = 1 \times 10^{-4}$ mol/L).

Figure 3 shows a comparison between the MWDs calculated by the FPT and the method of classes, at different monomer conversions. In both cases, the AIBN initial concentration was equal to 1.6×10^{-3} mol/L while the polymerization temperature was 60^0C. Notice that both methods are capable of predicting the MWDs up to very high monomer conversions. It was found that a number of 160 classes, leading to a total number of 960 differential equations, was sufficient for the convergence of the method of classes.

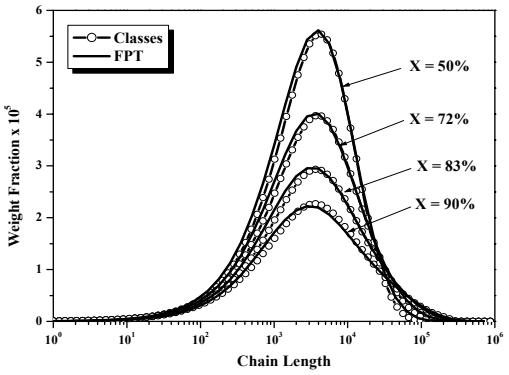


Fig. 3: Predicted MWDs via the application of the FPT and the method of classes, at different monomer conversions

In Figure 4 the MWD calculated by the FPT is compared with the distribution obtained by Monte

Carlo simulation (Tobita and Hatanaka, 1996). The reactor temperature was 60^0C, and the kinetic rate constants for the free-radical polymerization of VAc were taken from the original work of Tobita and Hatanaka (1996). The comparison was made for a specific value of monomer conversion equal to 85%. It can be seen that the MWD calculated by the FPT is in good agreement with the one obtained by the Monte Carlo simulation. The observed discrepancy in the tail of the distribution can be attributed to the use of the QSSA in the Monte Carlo simulation and to the inherent statistical difficulties of Monte Carlo simulations associated with the sampling of chains placed at the tail of the distribution

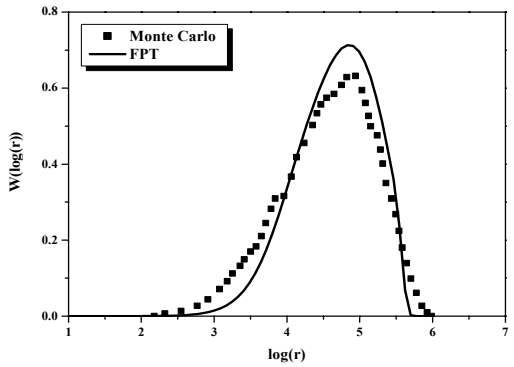


Fig. 4: Comparison of the calculated total weight fraction distribution at monomer conversion equal to 85%. The discrete points are the calculated results by Tobita and Hatanaka (1996). The continue line represents the simulated results using the FPT.

The FPT is capable of predicting the entire joint molecular weight - long chain branching distribution. The calculated combined MW-LCB distributions are depicted at 60^0C at a specific value of monomer conversion (i.e., 90%) as it can be seen in Figure 5.

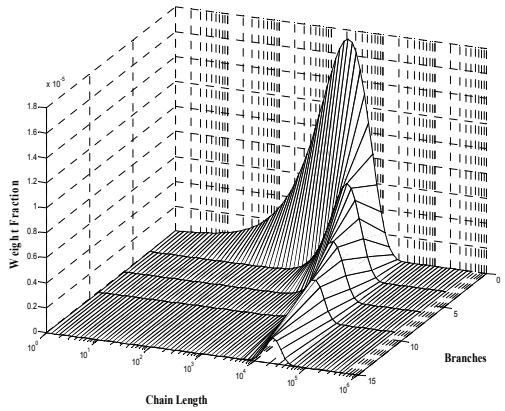


Fig. 5: Predicted combined MW-LCB distribution at 90% monomer conversion (simulation conditions same as in Figure 3).

REFERENCES

Baltsas, A., A.D. Achilias and C. Kiparissides (1996). A theoretical investigation of the production of branched copolymers in continuous stirred tank reactors. *Macromolecular Theory and Simulation*. 5, 477-497.

Canu, P., and W.H. Ray (1991). Discrete weighted residual methods applied to polymerization reactions. *Comput. Chem. Engin.* 15, 549 - 558.

Hamer, J.W. and W.H. Ray (1986). Continuous tubular polymerization reactors - I. A detailed model. *Chemical Engineering Science*. 41, 3083-3093.

Iedema, P.D., M. Wulkow and C.J. Hoefsloot (2000). Modeling Molecular Weight and Degree of Branching Distribution of Low-Density Polyethylene. *Macromolecules*. 33(19), 7173-7189.

Keramopoulos, A. and C. Kiparissides (2002). Development of a comprehensive model for diffusion-controlled free-radical copolymerization reactions. *Macromolecules*. 35, 4255-4272.

Kumar, S. and D. Ramkrishna (1996). On the solution of population balance equations by discretization – I. A fixed pivot technique. *Chemical Engineering Science*. 51, 1311-1332.

Lorenzini P., M. Pons and J. Villermaux (1992) Free-radical polymerization engineering - III. Modelling homogeneous polymerization of ethylene: mathematical model and new method for obtaining molecular - weight distribution. *Chemical Engineering Science.* 47, 3969-3980.

Nele, M., C. Sayer and J.C. Pinto (1999). Computation of molecular weight distributions by polynomial approximation with complete adaptation procedures. *Macromolecular Theory and Simulations*. 8, 199-213.

Nordhus, H., O. Moen and P. Singstad (1997). Prediction of molecular weight distribution and long - chain branching distribution of low-density polyethylene from a kinetic model. *J.M.S.-Pure Appl. Chem*. A34, 1017-1028.

Pladis, P. and C. Kiparissides (1998). A comprehensive model for the calculation of molecular weight – long chain branching distribution in free-radical polymerizations. *Chemical Engineering Science*. 53**,** 3315-3333.

Teymour, F. and J.D. Campbell, (1992). *In Proceedings of the 4th International Workshop on Polymer Reaction Engineering,* eds K.H. Reichert, H.U. Moritz, DECHEMA Vol. 127, p. 149. VCH, Weinheim.

Thomas, S. (1998). Measurement and modeling of long chain branching in chain growth polymerization. *PhD Thesis*, Mc Master University.

Tobita, H. (1996). Random degradation of branched polymers. 1. Multiple branches. *Macromolecules*. 29, 3010-3023.

Tobita, H. and K. Hatanaka (1996). Branched structure formation in free radical polymerization of vinyl acetate. *Journal of Polymer Science*. 34, 671-681.

Wulkow, M. (1995). The simulation of molecular weight distributions in polyreaction kinetics by discrete Galerkin methods. *Macromol. Theory Simul.,*5, 393-407.

Session 5.3

Process Monitoring

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

A DATA-BASED MEASURE FOR INTERACTIONS IN MULTIVARIATE SYSTEMS

Rossi M. * **Tangirala A.K.** ** **Shah S.L.** ***,[1] **Scali C.** *

* *Computer Process Control Lab, Dept. of Chemical Engineering, University of Pisa, Pisa, Italy*
** *Dept. of Chemical Engineering, IIT Madras, Chennai, India*
*** *Department of Chemical & Materials Engineering - University of Alberta, Edmonton, Canada*

Abstract: This article focusses on the control loop performance diagnosis of a multivariate system with emphasis on the presence of interactions and poor performance of control loops. The paper provides a data-driven technique to determine if a decentralized PI(D)$^+$ controller will suffice or if an advanced controller (*e.g.*, MPC) is necessary to handle the control interactions and improve the loop performance. Two different techniques are proposed: the first one, based on the Power Spectrum of the error, analyzes interactions in the frequency domain, while the second one, based on the evaluation of a modified IAE (Integral of Absolute Error), analyzes interactions in the time domain. A performance index for the controller is also proposed for the case of set-point tracking. Simulation and experimental case studies are presented to highlight the applicability of the proposed techniques.

Keywords: Interactions, MIMO systems, frequency domain, time domain

1. INTRODUCTION

Over the last two decades, monitoring control loop performance has been addressed in several ways and several performance indices have been proposed (see Hoo *et al.* (2003) for a good survey). Different causes for low loop performance such as improper controller tuning, sensor faults, valve non-linearities have been identified (Bialkowski, 1993; Kozub, 1997). An important cause that demands attention in addition to these causes is the presence of interactions among loops. A key impact of the interaction on the loop performance is the propagation of the effects of other causes that deteriorate the loop performance, thereby corrupting other loops.

The schematic of a multivariate (MIMO) system under discussion in this sequel is shown in figure 1: the process **P**, not necessarily square; the controller **C** initially considered as a decentralized $PI(D)^+$ type. A disturbance through **Pd** and white noise passing through a first order filter **F** are included for completeness. In a routine operation, the set point array **r**, the control action array **u** and the controlled variables array **y** are measured quantities.

Diagonal elements of the matrix **P** represent the process transfer functions, while the off-diagonal elements (P_{ij}, $i \neq j$) represent the interaction transfer functions. When an excitation affects a loop *i*, some effect is also present on another loop *j* depending on the interaction transfer function P_{ij}.

The Relative Gain Array (RGA) is often used to describe the level of interaction among loops, for instance in (Persechini *et al.*, 2004). However, it has two key limitations: (*i*) a model of the process must be known and consequentially the RGA measure depends on the model uncertainty (Chen and Seborg, 2002) and (*ii*) RGA gives only a measure of stability once loops are closed and no indication on the real interaction among them.

Therefore, a novel approach is proposed, which does not use an explicit process model, but instead di-

[1] sirish.shah@ualberta.ca

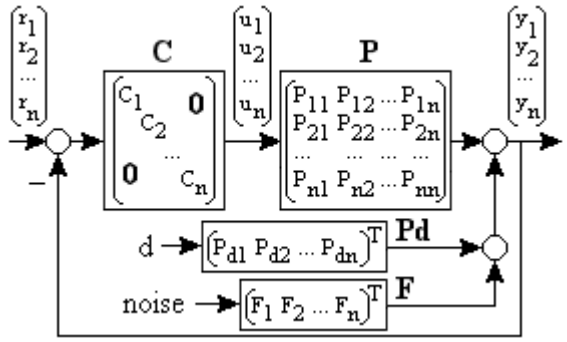


Fig. 1. The reference setup of a MIMO system

rectly uses routine operating data. Once the loops that are suspected to interact are selected, the proposed method can be used to assess the level of interaction.

Among all types of excitations, the set-point excitation is most preferred since it allows us to obtain **y** as a function only of **P** and **C**. The effect of the disturbance transfer function **Pd**, which may cloud the interaction measure, is thus avoided.

The outline of the paper is as follows. Two interaction measures are described in section 2, followed by an analysis of the controller performance in section 3. Application of the techniques to simulated, industrial and experimental setups are presented in Sections 4 and 5 respectively. The paper ends with a few concluding remarks in Section 6.

2. INTERACTION MEASURES

Depending on the nature of process excitation, interaction can be analyzed in either time- or a frequency-domain as may be deemed appropriate. For instance, in rotary machines, it is necessary to evaluate the interaction in a defined range of frequencies by exciting loops with oscillatory set-points. A frequency domain analysis, named in the sequel *Power Spectrum Analysis*, is more suited to such situations. On the other hand, oscillatory set-point changes (steps, ramps, etc.) are not a commonplace in chemical industries and therefore, a time domain technique, named in the sequel as *IAE technique*, may be chosen.

For both cases a comparison between controlled variables belonging to different loops has to be performed. For this reason a normalization factor is chosen so as to make the task independent of the measuring scale. Denoting as $CR_{UP,i}$ and $CR_{LW,i}$ the upper and lower limit of the control range for the loop *i* respectively, the normalization factor NF_i can be evaluated as described in equation 1:

$$NF_i = min\{CR_{UP,i} - \overline{r_i};\quad \overline{r_i} - CR_{LW,i}\} \quad (1)$$

where $\overline{r_i}$ is the mean value of the set-point of loop *i*. All controlled variables are divided by their respective normalization factors for subsequent analysis.

2.1 *Power Spectrum Analysis*

Detection of interacting loop is performed by the use of Power Spectral Correlation Index (PSCI) (Tangirala *et al.*, 2005). Its application allows one to exclude loops characterized by different frequencies due to other oscillating sources. The PSCI between loop *i* and *j* is calculated as:

$$PSCI_{i,j} = \frac{\sum_{\omega} PS_{y_i}(\omega) \cdot PS_{y_j}(\omega)}{\sqrt{\sum_{\omega} PS_{y_i}(\omega)^2 \cdot \sum_{\omega} PS_{y_j}(\omega)^2}} \quad (2)$$

where $PS_{y_i}(\omega)$ is the raw power spectrum of the controlled variable of the loop *i* evaluated at the frequency ω. This index (Tangirala *et al.*, 2005), lies in the range [0 1]: with similar shapes of power spectra its value is near one, indicating the presence of interaction.

Once an interacting loop is detected, the amount of the interaction is calculated as:

$$IFD_{i,j} = 1 - \frac{\sqrt{max(PS_{SP})}}{\sqrt{max(PS_{SP})} + \sqrt{max(PS_I)}} \quad (3)$$

where SP and I indicate respectively the loop affected by the set point change and the interacting loop. IFD lies in the range $[0\ 1]$, the larger the interaction, the higher is the index.

Equations 2 and 3 can be used in combination to assess the interaction in the frequency domain.

2.2 *IAE technique*

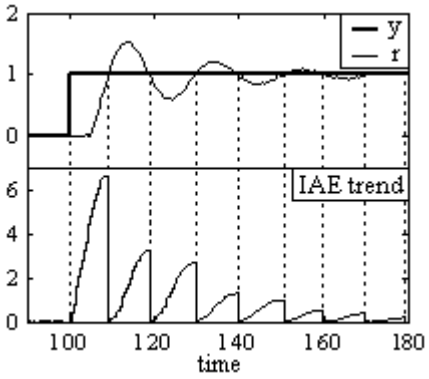


Fig. 2. Example of IAE trend for a set-point change

For the time domain analysis, the error signal **e=r-y** is considered: if the error does not change its sign from the sample $k-1$ to the sample k a modified IAE (Integral of Absolute Error) is evaluated as given in equation 4:

$$IAE(k) = IAE(k-1) + |e(k)| \cdot h \quad (4)$$

where h is the sampling interval. If a change in the error sign occurs, $IAE(k)$ is reset to zero (Hägglund, 1995). The trend of IAE is composed of peaks that coincide with the zero crossing of the error signal as shown in figure 2.

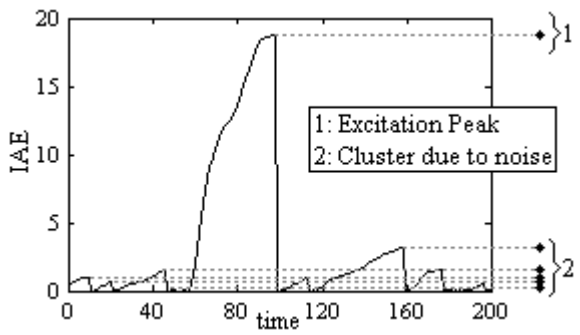


Fig. 3. Example of Excitation Peak and Cluster of Noise peaks

The use of this index allows one to magnify the difference between excitations (bigger peaks) and noise (smaller peaks) taking into account both the amplitude of the error and the duration between two consecutive zero crossings. Furthermore, the comparison of IAE peaks for different variables can be directly used to evaluate the amount of interaction between loops. To analyze only peaks due to excitations, a technique for the detection of outliers is applied (Daszykowski *et al.*, 2001). This technique analyzes maxima of each peak: maxima of noise peaks generate a cluster from which maxima of excitation peaks are excluded (figure 3).

In the presence of interactions, a set-point change will generate a peak in IAE trends in the two examined loops almost at the same time. Considering the time delay of the interaction as unknown, it is impossible to establish the exact gap which occurs between the two peaks. To overcome this problem a time window is chosen according to the duration of the source peak: defining t_0 the time in which the set point change starts and t_1 the time in which the time trend reaches its maximum, the time horizon t_h for the time window can then be evaluated as:

$$t_h = t_0 + a \cdot (t_1 - t_0) \tag{5}$$

where the parameter a is adjustable and set to 4 in this work. This setting of a provides an over-estimate of the interaction delay θ_I). A few remarks follow:

- $t_1 - t_0$, the time gap in which the controlled variable reaches the set-point value for the first time, is an overestimation of θ_P.
- Under the hypothesis of similar values of θ_I and θ_P, the choice $a = 4$ allows to obtain for most cases a time horizon bigger than θ_I.

However the value of a can be changed easily by the operator to analyze the effect of the time window horizon on the interaction measure.

An interaction index Interaction in Time Domain (ITD) is thus proposed based on the IAE of the windowed trend:

$$ITD_{i,j} = 1 - \frac{\sum_{t_0}^{t_h} IAE_{SP}}{\sum_{t_0}^{t_h} IAE_{SP} + \sum_{t_0}^{t_h} IAE_I} \tag{6}$$

with the same formalism used in equation 3. Similar to IFD, ITD lies in $[0\ 1]$ and a strong interaction is associated with a high value.

A heuristic interpretation of the proposed interaction indices is given in table 1. Of particular importance is the limit of 0.5, over which the value implies that the set-point change in a loop *i* affect more other loops than loop *i* itself.

Table 1. Interpretation of the index values

ITD (/IFD)	Interpretation
$[0 \quad 0.125]$	No Interaction
$[0.125 \quad 0.25]$	Low Interaction
$[0.25 \quad 0.375]$	Medium Interaction
$[0.375 \quad 0.5]$	High Interaction
$[0.5 \quad 1]$	Very High Interaction

It is remarked that a limitation of this method is that it can not correctly estimate the interaction when set-point activity in a loop and a disturbance in another loop coincide. However, the presence of other set-point changes in the data set can help overcome this limitation to a large extent.

3. CONTROLLER PERFORMANCE INDEX

Information from the interaction measure can be used to establish if a retuning is sufficient to improve the performance or if an advanced controller is required. For this purpose, a new Controller Performance Index (CPI) is defined.

The CPI is proposed on the basis of the response to a set-point change. Given a set-point change, under *minimum variance control*, after $\theta_P + t_0$, the error immediately reaches zero. Suppose a minimum error e_{min} is associated with this case. Otherwise a residual error is still present until the controlled variable reaches the settling time. Denote the error in such a case by e_{tot}. The CPI is then defined as,

$$\text{CPI} = \frac{e_{tot} - e_{min}}{e_{tot} + e_{min}} \tag{7}$$

If e_{tot} is near to the minimum achievable, the controller has a good performance and CPI is near zero. If $e_{tot} \gg e_{min}$, the controller has a poor performance and CPI is near to one. Given the fact that the minimum variance controller is an idealistic case and of little practical use (Huang and Shah, 1999) and considering the presence of interaction, a threshold value of $CPI = 0.5$ is chosen. Below this value of CPI, retuning would be practically of little benefit. Furthermore, a high value of the CPI with a high value of ITD/IFD implies that the present controller configuration yields good performance but unable to handle interactions. Therefore, a structural change may be necessary.

To evaluate the CPI, the time delay of the process θ_P must be estimated. The recorded response y to the set-point change in closed loop can be approximated

by a open loop response to a step-test $\tilde{y}$. Choosing a second order model $\tilde{P}$ and varying its parameters, it is possible to find the best approximation in the least square sense. The obtained model will not have any physical meaning: it is used only to generate a good estimate of the time-delay (for the same reason the order of the model is not critical). Assuming a fixed value for the time delay $q = \theta_P/h$ with h sampling time and defining n the length of the data set, it is possible to generate the best approximation of y in the least square sense:

$$y(z^{-1}) = \frac{b_1 z^{-1-q} + b_2 z^{-2-q}}{a_1 z^{-1} + a_2 z^{-2} + 1} \cdot r(z^{-}1) \quad (8)$$

$$y_k = b_1 r_{k-1-q} + b_2 r_{k-2-q} - a_1 y_{k-1} - a_2 y_{k-2} \quad (9)$$

$$\underbrace{\begin{bmatrix} y_{q+3} \\ y_{q+4} \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} -y_{q+2} & -y_{q+1} & r_2 & r_1 \\ -y_{q+3} & -y_{q+2} & r_3 & r_2 \\ \vdots & \vdots & \vdots & \vdots \\ -y_{n-1} & -y_{n-2} & r_{n-1-q} & r_{n-2-q} \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix}}_{\mathbf{p}} \quad (10)$$

$$\mathbf{p} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T \cdot \mathbf{y} \quad (11)$$

Among all models $\tilde{P} = f(q, \mathbf{p})$, the one that generates the lowest error in the least square sense is associated with the best estimation of θ_P.

4. CASE STUDIES

This section presents two of the several MIMO systems that were successfully analyzed with the proposed techniques. The first system is the *Marlin Column*, known to contain a high level of interaction. The process transfer functions are reported below:

$$\mathbf{y} = \begin{bmatrix} \frac{0.0747e^{-3s}}{12s+1} & \frac{-0.0667e^{-2s}}{15s+1} \\ \frac{0.1173e^{-3.3s}}{11.7s+1} & \frac{-0.1253e^{-2s}}{10.2s+1} \end{bmatrix} \cdot \mathbf{u} \quad (12)$$

It can be observed that the gains, time constant and time delay for diagonal and off-diagonal elements are similar, indicating a strongly interacting system. A more detailed description of the process together with the definition of a decentralized PI controlled are reported in Marlin (2000). Pre-specified set-point changes were performed to analyze the presence of interaction as depicted in figure 4a); the corresponding IAE trends are reported in 4b). The presence of interaction is indicated by the high values of $\text{ITD}_{1,2} = 0.47$ and $\text{ITD}_{2,1} = 0.43$, which confirms with the earlier discussion. The CPI is over 0.9 for both the controllers indicating that a retuning will improve the performance but, considering the high values of ITD in this case, a different structure is suggested for the controller (e.g. MPC).

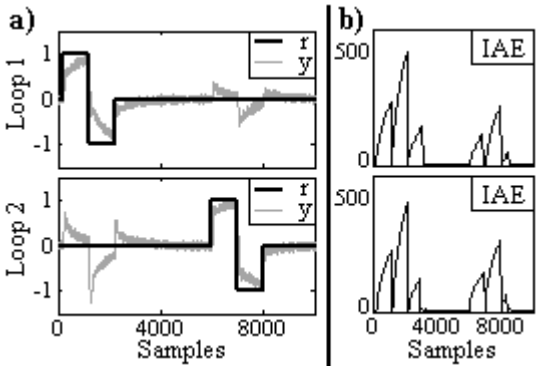


Fig. 4. Marlin Column: a) set-point (r) and controlled variable (y) values; b) IAE trends

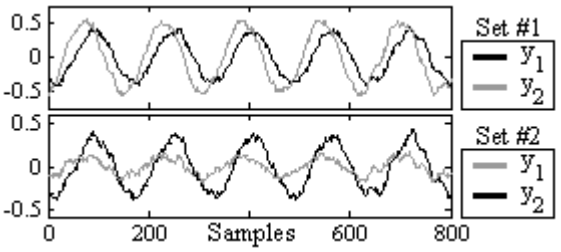


Fig. 5. Marlin Column: controlled variables for the loop affected by the oscillatory set-point (black) and the interacting loop (gray)

A frequency-domain analysis was also performed using oscillatory set-points. In figure 5 the controlled variable for the loop affected by the set-point change (black) and the controlled variable for the interacting loop (gray) are shown. The value of PSCI is over 0.98 for both the set of data indicating that the interaction is present: for this case the interaction from loop 1 to loop 2 is higher (IFD = 0.57) than the one from loop 2 to loop 1 (IFD = 0.26). It is important to recall that this analysis is suited to set-point changes that are well-localized in frequency while the time domain analysis, on the contrary, is well-suited to set-points that contain a range of frequencies.

The second system under study is the *Shell problem*; the process transfer functions are reported below:

$$\mathbf{y} = \begin{bmatrix} \frac{4.5e^{-27s}}{50s+1} & \frac{1.77e^{-28s}}{60s+1} & \frac{5.88e^{-27s}}{50s+1} \\ \frac{5.39e^{-18s}}{50s+1} & \frac{5.62e^{-14s}}{60s+1} & \frac{6.9e^{-15s}}{50s+1} \\ \frac{4.38e^{-20s}}{33s+1} & \frac{4.42e^{-22s}}{44s+1} & \frac{7.2}{19s+1} \end{bmatrix} \cdot \mathbf{u} \quad (13)$$

For this problem two solutions have been analyzed: firstly a decentralized PI controller has been implemented and secondly it has been compared with the MPC proposed in (Patwardhan and Shah, 2004). It is noted that, as explained in (Patwardhan and Shah, 2004), y_3 can be considered as a "slack" variable. The response for the two cases to the same set-point changes are shown in figure 6 and figure 7 respectively. The two different situations are well explained by the values of ITD measure:

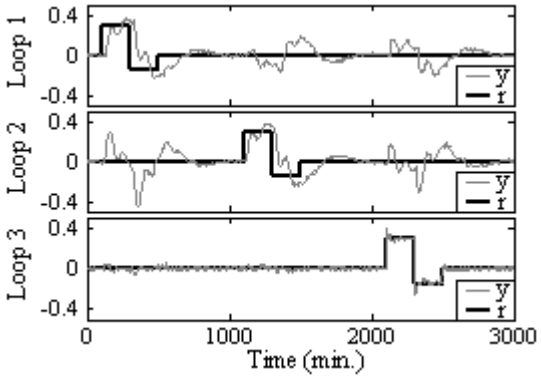


Fig. 6. Shell Problem with decentralized PI controllers; set-points (black) and controlled variables (gray)

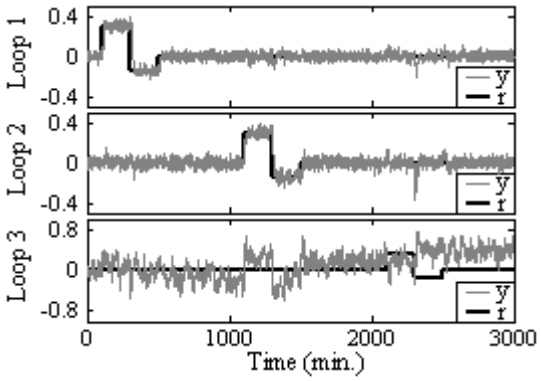


Fig. 7. Shell Problem with MPC; set-points (black) and controlled variables (gray)

$$\underbrace{ITD = \begin{bmatrix} 1 & .47 & .01 \\ .46 & 1 & .01 \\ .98 & .97 & 1 \end{bmatrix}}_{dec.PI} ; \quad \underbrace{ITD = \begin{bmatrix} 1 & .05 & .6 \\ .16 & 1 & .92 \\ .0 & .0 & 1 \end{bmatrix}}_{MPC} \tag{14}$$

It is clear that the decentralized PI controllers do not yield a satisfactory performance; the first two loops are strongly interacting (ITD = 0.46 and 0.47); and loop 3 is affecting them (ITD = 0.97 and 0.98) without being affected (ITD < 0.1). The CPI values for the three PI controllers are respectively 0.39, 0.59 and 0.53 indicating that a new tuning cannot be expected to improve the performance. Therefore, an advanced control scheme such as MPC is required. With such a scheme, the first two loops are no more interacting because the third loop is absorbing all excitations; the only residual interaction from loop 2 to loop 1 ($\text{ITD}_{2,1} = 0.16$) has low importance. In both cases, the ITD measure is able to rightly explain the interacting behaviour.

5. EXPERIMENTAL SETUP

The IAE technique was applied to an experimental setup consisting of the four-tank system depicted in figure 8. Two combinations were considered - the first comprising tank #1 and tank #2 (this is a minimum phase system) and the second one comprising tank #3 and tank #4 (this is a non-minimum phase system) (for more details see Johansson (2000)). In the first

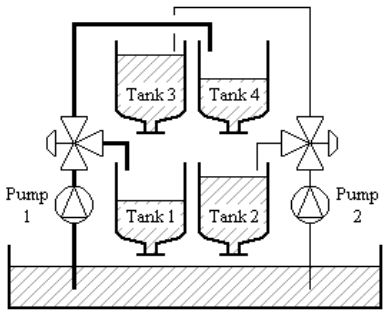


Fig. 8. Simple schematic for the four tank problem

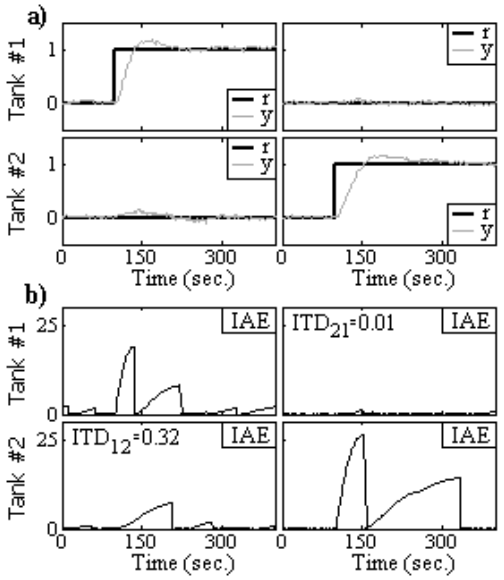


Fig. 9. Minimum Phase System: a) Set point (black) and controller variables (gray); b) IAE trends and interaction measures

case, pump #1 is feeding the left tank and pump #2 is feeding the right tank, while in the second case it is the converse. For each case, the target is the control of the levels in the two tanks by manipulating the inlet flowrates to the tanks. The set-point was changed for each of the levels and the interaction measure was evaluated for the two cases.

Set-point changes and controller variables are shown in figure 9a while IAE trends and interaction measures are shown in figure 9b for the minimum phase system. Analyzing the trend of the controlled variables, it is very difficult to establish properly the level of interaction: an excitation with small amplitude is shown in tank #2 for a set-point change in tank #1, but it appears as a weak interaction. Compare this with the IAE of the trends which exhibits a significant peak similar to peaks showed in tank #1. On the other hand, a set point change in tank #2 does not generate excitations in tank #1. Both of these phenomena are captured by the corresponding interaction measures, $\text{ITD}_{1,2} = 0.32$ reveals the presence of a moderate interaction from tank #1 to tank #2; while $\text{ITD}_{2,1} = 0.01$ implies that tank #2 is not affecting tank #1.

For the second combination, the set-point changes and controller variables are shown in figure 10a) while the IAE trends and interaction measure are shown

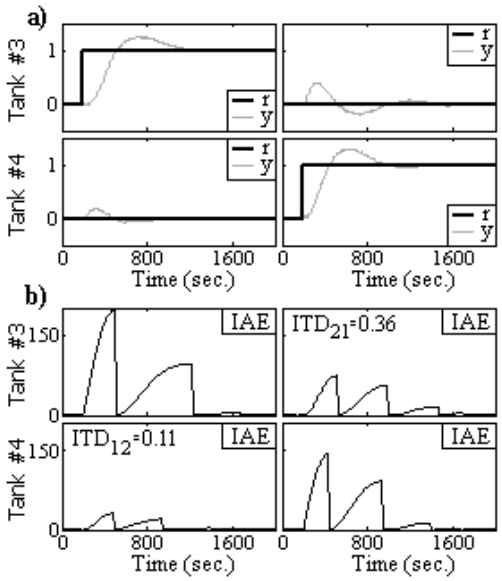


Fig. 10. Non-Minimum Phase System: a) Set point (black) and controller variables (gray); b) IAE trends and interaction measures

in figure 10b) for the non-minimum phase system. Again the analysis of IAE peaks reveals the presence of interaction which is stronger from tank #4 to tank #3: the higher ITD is now on the right tank while in the previous case it was on the left tank. These results are in agreement with the switch in the position of the feed for the two cases, once again indicating that the IAE technique is successful in highlighting the interaction among the loops. The level of interaction for this pair is larger than the earlier one due to the time delay.

The CPI values for the two cases are $[0.58 \;\; 0.71]$, indicating that the used controllers have a sufficiently good performance; the benefit obtained with a new tuning would be marginal. In retrospection, the IAE technique is able to capture the interaction, which otherwise appeared insignificant by visual inspection.

6. CONCLUSIONS

Two different techniques have been proposed to detect and quantify control loop interactions in MIMO processes. The IFD and PSCI measures used in the frequency domain are well-suited to analyze process excitations that are well-localized in the frequency domain; and the ITD, which analyzes data in the time-domain, is suited to process excitations that are spread over a range of frequencies. An important feature of these indices is that they can be computed from measured data, without any need for an explicit knowledge of the process model.

To avoid the effect of the disturbance transfer functions, which may cloud the interaction measures, set-point changes have been analyzed. In the presence of disturbance effects in the given data set, the reliability depends on the time coincidence of the process excitations caused by these two different sources.

The interaction measure in the time domain is completed by an analysis of the performance of the controller and an index CPI has been defined: it serves as an indicator to determine whether retuning of the controller is beneficial or if it is better to use an advanced MIMO (e.g. an MPC) controller.

The application of the proposed techniques have been demonstrated on two industrial simulation case studies and on an experimental setup comprising the four (interacting) tanks. In all cases, the proposed methods successfully revealed and quantified the interaction. which in one case appeared insignificant from a direct observation of the time-domain trends.

7. REFERENCES

Bialkowski, W.L. (1993). Dreams versus reality: a view from both sides of the gap. *Pulp and Paper Canada* **94**, 19–27.

Chen, D. and D.E. Seborg (2002). Relative gain array analysis for uncertain process models. *AIChE Journal* **48(2)**, 302–310.

Daszykowski, M., B. Walczak and D.L. Massart (2001). Looking for natural patterns in data. part 1: Density-based approach. *Chemometrics and Intelligent Laboratory System* **56**, 83–92.

Hägglund, T. (1995). A control loop performance monitor. *Control Eng. Practice* **3(11)**, 1543–1551.

Hoo, K.A., M.J. Piovoso, P.D. Schnelle and D.A. Rowan (2003). Process and controller performance monitoring: overview with industrial applications. *International Journal of Adaptive Control and Signal Processing* **17**, 635–662.

Huang, B. and S.L. Shah (1999). *Performance Assessment of Control Loops, Theory and Applications*. Springer Verlag. London.

Johansson, K.H. (2000). The quadruple tank process: a multivariable laboratory process with an adjustable zero. *IEEE Trans. on Control System Technology* **8(3)**, 456–465.

Kozub, D.J. (1997). Controller performance monitoring and diagnosis: experiences and challenges. *In: AIChE Symposium Series* **94**, 83–96.

Marlin, T.E. (2000). *Process Control: Designing Processes and Control Systems for Dynamic Performance*. McGraw Hill. Boston.

Patwardhan, S.C. and S.L. Shah (2005). From data to diagnosis and control using generalized orthonormal basis filters. part 1: Development of state observers. *J. of Process Control* **15(7)**, 819–835.

Persechini, M.A.M., A.E.C. Peres and F.G. Jota (2004). Control strategy for a column flotation process. *Control Eng. Practice* **12(8)**, 963–976.

Tangirala, A.K., S.L.Shah and N.F. Thornhill (2005). PSCMAP: a new tool for plant-wide oscillation detection. *Journal of Process Control* **15(8)**, 931–941.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ISSUES IN ON-LINE IMPLEMENTATION OF A CLOSED LOOP PERFORMANCE MONITORING SYSTEM

Claudio Scali ([1]), Fabio Ulivari([1]), Antonio Farina([2])

(1) Laboratory of Chemical Process Control (CPCLab)
Department of Chemical Engineering (DICCISM)
University of Pisa, Italy
(2) ENI Refining & Marketing
Refinery of Livorno, Italy

Abstract: The paper illustrates main features and implementation issues of a performance monitoring system which, on the basis of data recorded during normal operation, is able to detect the presence of anomalies, to investigate causes and to propose strategies of action. The off-line architecture of the system, successfully applied to industrial plant data, is briefly recalled. Continuous monitoring of a multi-loop refinery section finds hard constraints in heavy computation load and excessive traffic on the communication bus. A mixed structure, featuring on-line detection of anomalies, followed by research of their causes performed on an external computer, is studied. Effects of key factors as: sampling time, number of data, supervision time, loss of initial data, are analyzed and a supervision strategy, compatible with plant DCS characteristics, is proposed. © *IFAC'06*

Keywords: Performance monitoring, Control loops, Automatic recognition, Valves, Friction.

1. INTRODUCTION

The importance of closed-loop performance monitoring (CLPM), as a means of improving product quality and hence the overall economy of industrial plants, has recently led to a large interest in academic research and industrial applications (Huang and Shah, 1999).
The possibility of detecting the onset of anomalies and determining causes of performance deterioration in base control loops is certainly of vital importance, as the success of advanced control layers (Multivariable, Optimization) depends on the correct operation of them.

In industrial-scale processes, typically involving thousands of variables and hundreds of control loops, a monitoring system needs to operate automatically, leaving only key decisions to the operators. Furthermore, for a wider acceptability, it is desirable that the monitoring system operate on the basis of data made available from the data-acquisition system, without need of introducing additional perturbations in the plant. It is also highly desirable that process monitoring is able to account for various causes of performance deterioration, such as incorrect design or tuning of controllers, anomalies and failures of sensors, presence of friction in actuators, external perturbations, and deteriorations in the process itself. Whatever the causes, the monitoring system should be able to detect them and to indicate actions to perform, ranging from retuning of controllers, to substitution of faulty sensors, compensation or maintenance of valves, or operations on upstream equipment.

A number of issues and problems still remain unresolved in the theory (e.g., significance and reliability of proposed performance indexes, their applicability in the case of multivariable control, simple and reliable technique for automatic detection of causes) and this explains efforts and research activity in the academy (Qin, 1998). Issues coming from applications seem less severe (e.g., the recommended degree of automation and interaction with the operator, off-line versus on-line architectures), but the success of a performance monitoring system depends strongly on them. In addition, there is a feedback from the field, in the sense that, depending on the characteristics of plant and control system, the most suitable architecture can

be chosen and customized according to operators' needs.

This paper focuses on implementation issues for on-line monitoring. In the first part the system architecture and the adopted techniques are briefly illustrated; in the second part constraints coming from available computation power and allowed data transfer traffic are faced and necessary changes in the system architecture to design a flexible supervision strategy, compatible with plant DCS, are analyzed.

2. FEATURES OF THE CLPM SYSTEM

Referring to Figure 1, available data from the acquisition system are: controlled variable (PV), set point (SP), controller output (OP); in addition also controller parameters and control ranges are known; in general the manipulated variable (MV) is not recorded.

Figure 2 provides a schematic illustration of the structure of the Closed Loop Performance Monitoring system.

The first module detects the onset of anomalies, i.e. is able to separate good performing loops from poor ones (oscillating, slow); tests are based on techniques firstly proposed by Hägglund (1995, 1999) and modified, during previous activity, in order to improve their efficiency (Ulivari et al., 2005). These modifications are briefly illustrated in the sequel.

To detect oscillations (Hägglund, 1995), the integral of absolute error (*IAE*) is computed for every half-cycle and compared with a limit value IAE_{lim}:

$$IAE = \int_{t_i}^{t_{i+1}} |e(t)| \cdot dt \qquad (1)$$

$$IAE_{lim} = f(a, \tau_I) \qquad (2)$$

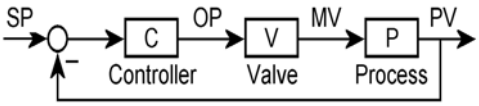


Fig. 1. The reference scheme.

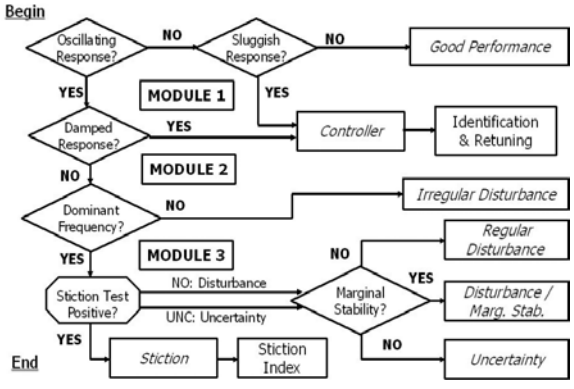


Fig. 2. Architecture of the CLPM system.

Where: t_i and t_{i+1} are two consecutive zero crossing of error $e = SP - PV$, a is a parameter that must be chosen (suggested value: 1%) and τ_I is the controller integral time constant. An oscillation is considered significant when the value of *IAE* exceeds IAE_{lim}. If the number N of detected oscillations exceeds a fixed value N_{lim} (for instance N_{lim}=10), during a supervision time T_{sup}, the oscillation is considered persistent. The suggested value of T_{sup} (Hägglund, 1995), depends on the loop ultimate period and can be correlated to the controller integral time constant τ_I ($T_{sup} = \approx 50\ \tau_I$). This is certainly reasonable when the main objective of the analysis is to detect tuning problems. In presence of stiction the frequency of oscillations can change largely, according to stiction characteristics while keeping a constant tuning.

This is shown in Figure 3 where simulation results, obtained by adopting the data driven model proposed by Choudury et al. (2005), are reported. Similar results are given by the analytical model proposed by Karnhopp (1985).
Therefore, for stiction detection purposes, it is more convenient to use a mobile supervision window T_{sup}, constantly updated on the basis of duration of last anomalous half-cycle.
For every anomalous half-cycle, the two zero-crossing times T_1 and T_0 are defined (Figure 4) and the supervision time is updated as:

$$T_{sup} = T_1 + \beta (T_1 - T_0) \qquad (3)$$

The parameter β is generally taken equal to 1.1. In the case of half-cycles not complete before T_{sup}, the analysis is extended to the end of the cycle.

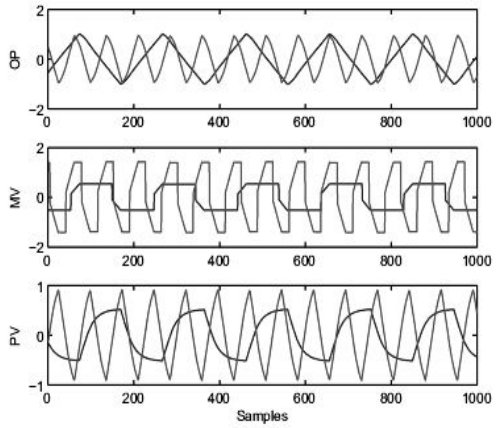


Fig. 3. Different trends of OP, MV and PV with stiction parameters (constant tuning).

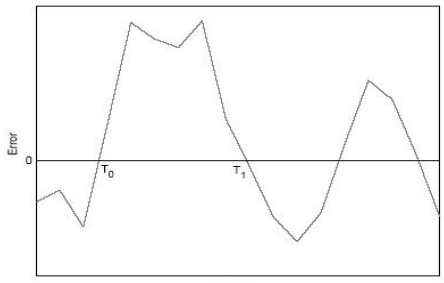


Fig. 4. Zero-crossing time for a half-cycle.

To detect slow responses, an Idle Index is proposed (Hägglund, 1999); this parameter is a function of time periods when the correlation between OP and PV signal increments is positive or negative. In the package, to avoid sensitivity to plant noise, also slow responses are identified on the basis of *IAE*, calculating when the error of a single deviation become too large, that is greater then an assigned limit value L_{min} (for instance L_{min}=10):

$$\frac{IAE}{IAE_{\lim}} > L_{\min} \tag{4}$$

The second module investigates the frequency behaviour of oscillating loops; if a damping response is detected, then the controller is indicated as cause of poor performance (as well as for the case of slow response). In this case a procedure for identification of process and disturbance dynamics and controller retuning (Rossi et al., 2003) is started; the performance improvement with the new tuning is shown to the operator who takes the final decision of changing controller settings.
The dominant frequency of the oscillation is also evaluated and not regular disturbances are isolated.

Oscillating signals are sent to the third module which allows to detect the presence of stiction in actuators, distinguishing this phenomenon from the presence of disturbances or from marginal stability conditions. The presence of stiction can be hidden by variations of process parameters and by stiction characteristics and a region of uncertainty may remain, where no decision can be taken (Rossi and Scali, 2004).
For this reason, different techniques recently proposed in literature, are applied in sequence, in order to reduce the number of uncertain cases. Among them: the Cross-Correlation (Horch, 1999), the Bicoherence (Choudury et al., 2004), the Relay technique (Rossi and Scali, 2005).
A stiction index is also evaluated in order to quantify its extent and to permit scheduling of valve maintenance from few analysis repeated in time.
A direct comparison of MV(OP) plots is also shown to the operator, for cases when MV is available (for instance in flow control), thus confirming/excluding the presence of stiction.

The efficiency of the CLPM system, firstly analyzed by intensive simulations, has been validated in several applications to industrial data, obtained from refineries and petrochemical plants. In particular, in Rossi et al. (2003), problems of frequent retuning of temperature controllers for a polymerization reactor undergoing surface fouling are reported. In Rossi et al. (2005), monitoring of refinery loops, mainly affected by valve stiction, is successfully carried out.

To conclude, off-line applications to industrial data have confirmed several positive features of the system: (a) complete automation of the procedure (after calibration of few parameters on the plant), (b) no perturbations need to be introduced in the plant, (c) open architecture (with easy adoption of new or updated techniques), and (d) flexibility to incorporate operator's knowledge.

About point (c), techniques for automatic recognition of the presence of stiction when MV is available, as proposed by Yamashita (2006), are currently under experimentation.

3. ON-LINE IMPLEMENTATION

Off-line applications are limited to "*una tantum*" analysis of closed loop performance (for instance before deciding the adoption of advanced control) or to periodic check (for instance to evaluate the current status of friction in valves and to schedule maintenance operation). Evident advantages would be given by a continuous on-line monitoring of plant loops. Taking into account the heavy computation load required for assessment of causes of anomalies, this operation must be done in an external computer.
Therefore a mixed structure, partly on-line and partly off-line is proposed.
The detection of onset of anomalies can be performed on-line, in order to discriminate good performing loops directly on DCS and limiting data acquisition to bad ones. In fact, the proposed indexes require few parameters and bring a limited increase of the computation load, as they consist only in few program lines (summation and comparison with constant values).
This is in agreement with Hägglund (2002), who proposes a DCS implementation of detection indexes, describing an application oriented only to detection of anomalies, with indications to operators (flashing alarms), without automatic detection of causes.

In more details, the original technique for oscillation detection required 11 parameters, while the modified technique needs 5 more parameters: T_0, T_1, β (already defined), plus $T_{sup\text{-}old}$ (observation time at previous step) and ΔT_{old}, duration of previous half-cycle. No additional parameters are required for detection of slow responses.

Some further considerations are worth for a complete picture of problems and possible solutions, as illustrated in the sequel.

1. Data acquisition with small sampling time and their transfer to the external computer where the CLPM systems performs a check of loops conditions would generate a too intense traffic, with consequent overload on the communication bus. A drastic reduction of amount of acquired data can be obtained by increasing the sampling time from the present applications value (T_s=10 seconds), to the value of the DCS archive (typically, T_s=60 seconds); in this case, the CLPM system would analyze the same amount of data already acquired for the DCS archive, without any additional traffic. Not surprisingly, that will bring a deterioration in information on loops status: a quantitative evaluation of this phenomenon and its effect on the quality of results in the plant under current analysis can be interesting.

2. A consistent saving of traffic can be obtained by acquiring only data belonging to anomalous loops to detect causes, without transferring data of good performing loops (in general, in previous off-line applications, they represent about 50% of total). A possible problem may arise from the fact that, as data acquisition starts once the anomaly is detected, there is a loss of data corresponding to the first time interval (where anomaly is detected): it can be of interest to evaluate its effect on the efficiency of the monitoring system.

3. In addition, a continuous supervision of all plant loops could be not necessary and could be limited to some of them, with time windows and strategies to be decided, according to tasks and priorities assigned to the CLPM system. Possible solutions to be investigated are illustrated below.

The first two points will be investigated in the sequel and then considerations about the third point (supervision strategy) will follow.

3.1 Effect of sampling time on results.

A total of 38 loops, referring to data coming from refinery plants and already used in off-line analysis, have been investigated and results are reported in Table 1.
The first row contains original verdicts obtained with a sampling time T_s=10 seconds (considered the right one); in rows 2 and 3 contain indications with T_s=30 and 60 seconds: the first number indicates loops maintaining the original verdict, while the second number indicates loops having different verdicts in the original classification.
It can be noted that increasing the sampling time the number of good performing loops remains almost constant: for T_s=60 seconds, 1 missed alarm appears. The number of loops tagged as affected by stiction decreases from 18 to 14 to 12. The number of Uncertain verdicts and Irregular Disturbance increases.
As expected, data sampled at the same rate as the DCS archive cannot be used, because the deterioration of information with the increase of the sampling time affects the quality of the analysis.
Smaller sampling times (less than 10 seconds) would increase the accuracy in signal reconstruction, but the consequent improvement in the quality of analysis results does not seem to justify the more intense traffic generated, as shown by a specific experimentation (with T_s=1 second) on a fewer number of loops.

Table 1 Influence of Sampling Time
(total of 38 loops)

T_s [s]	Good Perf.	Stic.	Unc.	Slow Resp.	Irr. Dist.
10	13	18	4	1	2
30	11	14	4+5	1	2+1
60	10+1	12	3+8	1	2+1

Table 2 Influence of loss of initial data
(total of 24 loops)

Data	Stiction	Uncertain	Irregular Disturbance
All	18	4	2
No Initial	17+3	3	0+1

3.2 Effect of loss of initial data on results.

For this evaluation, 24 oscillating loops in the previous set of 38 (18 tagged as Stiction, 4 as Uncertain and 2 as Irregular Disturbance), have been analyzed. From the global set, initial data, corresponding to the first appearance of the anomaly, have been eliminated.
From Table 2, it can be seen that the number of Stiction loops increases, passing from 18 to 20 (including 3 False Alarm), Uncertain loops change from 4 to 3, Irregular Disturbances from 2 to 1. Thus, the loss of initial data seems to cause less severe errors in stiction detection; this can be explained by considering that the stiction phenomenon, once started, continues to show up for long times, with persistent oscillations (the extent will increase after days or weeks).

3.3 A possible supervision strategy.

In theory, detection indexes could be able to perform a continuous supervision of all plant loops; Hagglund (2002), reports an application including over 90% of total loops; however, depending on the number of loops and on capacity of the DCS, it is quite reasonable to expect some limitations.

In the case under study, the system is a Honeywell TDC3000 with HPM's and Basic Controllers and is in charge of about 600 control loops over a total of 13 thousands of configured variables.
Actual constraints on the computation load and on traffic of the communication bus between DCS and external computer do not allow a supervision of a number of loops (N_{loop}) larger than 10÷15 at the same time.

The following supervision strategy can be proposed:

- A fixed number of loops ($N_{loop} < N_{tot}$, total) is maintained under observation for a fixed time (T_{obs}), considered sufficient to detect the anomaly onset,
- The generic N_i loop not showing anomaly in T_{obs} is tagged as Good Performing loop; for this loop monitoring lasts up to $T_{end}= T_{obs}$,
- The generic N_j loop showing anomaly in T_{obs} is tagged as Bad Performing loop; for this loop, at time T_{det} (when anomaly appears) data acquisition starts for a total number of N_{sam} data ($T_{acq}= N_{sam}*T_s$) and ends at time $T_{end}= T_{det} + T_{acq}$,
- At the end of the cycle, lasting $T_{end}=T_{obs}$ (for GP loops) and $T_{end} =T_{det} + T_{acq}$ (for BP loops), monitoring of loops belonging to a new set starts.

Adopting this strategy, all plant loops (N_{tot}) are monitored in a time equal to T_{plant}. A quantitative evaluation of these factors have been performed referring to a subset of data of the same plant, already available from previous analysis.

By analyzing Table 3, it is evident that the number of verdicts changing with a decrease of the number of data (N_{sam}) increases: a value of N_{sam}=700, corresponding to an acquisition time of T_{acq}= N_{sam}*T_s= 7000 seconds (≈ 2 hours), can be considered sufficient to obtain reliable results about causes detection.

Table 3 Influence of number of analyzed data (total of 24 loops)

N_{sam}	Modified verdict
800	2
700	3
600	3
500	4
400	5
300	5
200	14
100	18

Table 4 Time of occurrence of anomalies (total of 24 loops)

N° Loop	T_{det}	N° Loop	T_{det}
1	44’	13	5h 1’
2	35’	14	24’
3	34’	15	2h 59’
4	36’	16	11’
5	2h 58’	17	20’
6	1h 31’	18	36’
7	15’	19	57’
8	12’	20	32’
9	10’	21	25’
10	10’	22	4h 1’
11	18’	23	5h 53’
12	7h 11’	24	1h 4’

From Table 4, the time of occurrence of anomalies (T_{det}), is always less than 8 hours for all loops and then a value of T_{obs}=8h can be safely proposed. In more details: T_{det} shows to be less than 30 minutes for 9 loops, 60 minutes for 16 loops, 120 minutes for 18 loops and 240 minutes for 21 loops; therefore the choice of T_{obs}=8h is largely conservative and could be reduced to 4h.

At this point, an estimation of the total time required to supervise the complete plant (T_{plant}) can be done. Assuming a total number of plant loops N_{tot}=50, and a number of loops under supervision at the same time N_{loop}=10, the total time depends on the total monitoring time T_{end} and is easily computed as:
T_{plant}= (N_{tot}/N_{loop}) * T_{end}= 5* T_{end}

The duration of a supervision cycle for each loop is estimated under different hypotheses regarding the occurrence of anomalies (more significant parameters are illustrated in Figure 5).

For instance:
Hypothesis #1. All loops are tagged as Bad Performing and show the first occurrence of anomalies at the end of the observation period: T_{det}= T_{obs}= 8h; T_{end}= T_{det} + T_{acq} = 10h
→ T_{plant} = 5* T_{end} = 50h.
Hypothesis #2. All loops are tagged as Good Performing; in this case: T_{end}= T_{obs}= 8h;
→ T_{plant} = 5* T_{end} = 40h.
Hypothesis #3. All BP loops, with average value of T_{det}=4h; T_{end}= T_{det} + T_{acq} = 6h
→ T_{plant} = 5* T_{end} = 30h.

Clearly the first and second hypothesis are very conservative; the third one seems more appropriate.

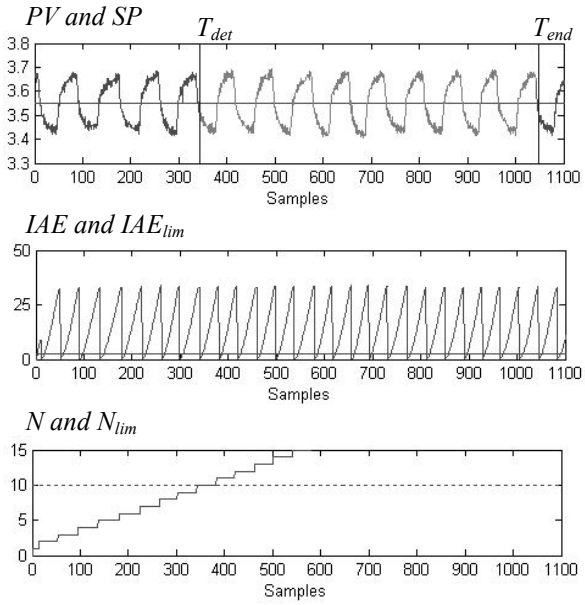


Fig. 5. Main parameters in the analysis of an anomalous loop.

A more realistic scenario should take into account the fact that the situation of each individual loop will be different: therefore the supervision of loops where anomalies shows up at short time (Table 4) will end in much shorter times, thus allowing a faster supervision of the whole plant.

Some conclusive remarks can be drawn:
- Supervision times are considered quite acceptable in the plant under analysis;
- The proposed strategy has large flexibility in order to take into account results from first analysis and to incorporate operator experience or specific needs; for instance:
  - Loop priority and frequency in the monitoring procedure can be changed according to causes indicated from off-line analysis;
  - GP loops maintain high priority in order to detect as soon as possible the onset of an anomaly;
  - Loops affected by stiction can be monitored with larger period (and lower priority), by

considering the slow evolution of this phenomenon; in the case of valve without bypass, the monitoring can be also suspended up to time of the first plant shut down.

4. CONCLUSIONS

The Closed Loop Performance Monitoring system presented in this paper has the global objective of detecting anomalies and tracing back also causes, in order to indicate more appropriate actions to perform. For these reasons, constraints on the computation load and excessive traffic on the communication bus, force to split the two tasks.

Indexes to detect the onset of anomalous responses can be implemented on the DCS, while the more demanding analysis of causes must be hosted on an external computer.

The analysis of loops data of the refinery plant under study has allowed an evaluation of the effect of key factors as: sampling time, loss of initial data, time of occurrence of anomalies, number of data and duration of acquisition period.

The proposed supervision strategy, which allows a monitoring of a subset of total control loops of the plant at the same time, is fully compatible with the present DCS characteristics.

Under different hypotheses about the occurrence of anomalies in the plant, the time required for a complete supervision of all plant loops is considered quite acceptable.

Finally, the proposed strategy has large flexibility in order to incorporate operator experience and to modify priorities and supervision time of each loop, taking into account results from off-line analysis.

REFERENCES

Choudhury M.A.A.S, Shah S.L., and Thornhill N.F. (2004): "Diagnosis of Poor Control Loop Performance using High Order Statistics", *Automatica*, **40**(10), 1719-1728.

Choudhury M.A.A.S, Shah S.L., and Thornhill N.F. (2005): "Modelling Valve Stiction", *Control Eng. Practice*, **13**, 641-658.

Hägglund T. (1995): "A Control Loop Performance Monitor", *Control Eng. Practice*, **3**, 1543-1551.

Hägglund T. (1999): "Automatic Detection of Sluggish Control Loops", *Control Eng. Practice*, **7**, 1505-1511.

Hägglund T. (2002): "Industrial Applications of Automatic Performance Monitoring Tools", *Proc. of IFAC - World Congress;* Barcelona (E)**,** pap#717**.**

Horch A. (1999): 'A Simple Method for Detection of Stiction in Control Valves', *Control Eng. Practice*, **7**, 1221-1231.

Huang B. and Shah S. (1999): Control Loop Performance Assessment: Theory and Applications, Springer-Verlag London.

Karnhopp D. (1985): "Computer Simulation of Stic-Slip Friction in Mechanical Dynamic Systems", *ASME, Journal of Dynamic Systems, Measurement and Control*, Vol. **10**, 100-103.

Qin S. J. (1998): "Control performance monitoring: a review and assessment", *Comp. & Chem. Engineerin*g, **23**, 173--186.

Rossi M. and Scali C. (2004): "Automatic Detection of Stiction in Actuators: A Technique to Reduce the Number of Uncertain Cases", *Proc. (on CD) 7th IFAC - DYCOPS'7 - Int. Symp;* Cambridge (USA), pap#157.

Rossi M. and Scali C. (2005): "A Comparison of Techniques for Automatic Detection of Stiction: Simulation and Application to Industrial Data", *J. Proc. Control*, **15** (5), 505-514.

Rossi M., Scali C. and Amadei M. (2003): "Development of a Technique for Performance Evaluation of Industrial Controllers", *Proc. of IFAC - ADCHEM'03 Congress;* Hong Kong (PRC)**; 1***, 183-188.*

Rossi M., Scali C. and Farina A. (2005): "Monitoraggio e Individuazione del Malfunzionamento degli Attuatori", *Automazione e Strumentazione,* **53** (4), 98-104.

Ulivari F., Scali C. and Farina A. (2005): "Applicazione ad un Impianto di Raffineria di un Sistema di Monitoraggio delle Prestazioni", *Proc. (on CD) ANIPLA'05 Int. Congress*, Napoli (I), pap#30.

Yamashita, Y. (2006): "An automatic method for detection of valve stiction in process control loops", *Control Eng. Practice,* (in press).

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

STEADY-STATE DETECTION FOR MULTIVARIATE SYSTEMS BASED ON PCA AND WAVELETS

L. Caumo[1], A. O. Kempf[2], J. O. Trierweiler[1]

Group of Integration, Modeling, Simulation, Control and Optimization of Processes (GIMSCOP)
Department of Chemical Engineering, Federal University of Rio Grande do Sul (UFRGS)
R. Eng. Luiz Englert, s/n°, 90040-040 – Porto Alegre – RS – Brazil
[1] {leti,jorge}@enq.ufrgs.br
[2] ariel@trisolutions.com.br

Abstract: Steady-state detection has been an important tool in data processing, for nonlinear model identification, real time optimization, variability analysis, and so on. In this article, it is proposed a new methodology applied to multivariate systems for steady-state detection based on PCA and wavelets. The proposed approach is applied to an industrial distillation column. The combination of PCA and wavelets allows quantifying the steady-state considering a single variable generated by a PCA projection.

Keywords: waves, signal analysis, multivariate systems, Principal Component Analysis, steady-state.

1. INTRODUCTION

An efficient method for steady-state detection is of great importance for process analysis, optimization, model identification, and data reconciliation. These applications require data under steady-state or very close to it.

With this aim, several methods have been developed. Most methods are based on statistical tests. Narasimhan *et al.* (1986) presented a Composite Statistical Test - CST (1986) and a Mathematical Test of Evidence - MTE (1987). In CST method, successive time periods are defined and evaluated according to covariance matrices and sample mean. In MTE method, differences in averages are compared to the variability within the periods. More recently, Cao and Rhinehart (1995) proposed a method based on moving average or conventional first-order filter which is used to replace the sample mean.

But these approaches evaluate the process status over a period of time, instead of a point in time. This is an important detail for on-line applications. Besides these techniques consider only the presence of random errors, and it is known that nonrandom errors are present in form of spikes for example (Jiang *et al.*, 2003).

The wavelet transform (WT) has been widely applied, in signal and image processing, singularity detection, fractals, trend extraction, denoising, data suppression and compression, due to its simple mathematical application and because it provides time-frequency localization simultaneously.

The WT is a tool that cuts up data or functions into different frequency components, and then studies each component with a resolution matched to its scale (Daubechies, 1992). In other words, WT consists of scaled and shifted versions of a mother-wavelet (the original wavelet). The process of multiplying the signal by scaled and shifted wavelets over all time produces wavelet coefficients that are function of scale and position. It is like a resemblance or correlation index between the section of the analyzed signal and the wavelet. One advantage of wavelets is to work with global or local analysis. Other advantages are to denoise a signal without degradation of the original signal (without losing information), to choose the resolution level, to obtain signal derivatives and to process unsteady signals.

Hence, in this work wavelets are used as a tool for steady-state detection of process signals. The methodology is based on a fast algorithm of two channel subband coder using conjugate quadrature

filters or quadrature mirror filters. Process trends are extracted from raw measurements via wavelet-based multi-scale processing by eliminating random noise and nonrandom errors. This "clean" signal still preserves the nuances of the original signal. Then the process status is measured using an index with value ranging from 0 to 1 according to the wavelet transform modulus of the extracted process signal and historical data. This index has a great application since it can be used for data compression and determination of optimal operating points for example.

Since most chemical processes are multivariable, it is necessary to have a procedure which makes possible to quantify how close it is to the steady-state. Therefore, it is necessary a way to deal with multivariable systems. Usually, a unique index for the whole process would be recommended, since it is easier to analyze. Jiang *et al.* (2003) suggest selecting key variables and combining them through the Dempster's balance rule. Thus, it is necessary to calculate a status index for each key variable and, by the balance rule, it is necessary to attribute a weight for each variable. Instead, in this work it is proposed to use the PCA (Principal Component Analysis) approach to combine all variables of a multivariable process into a single steady state measurement index, which would be representative of the whole process.

2. WAVELET TRANSFORM APPLIED TO STEADY-STATE DETECTION

2.1. Background of Wavelet Transform Concepts

Wavelet Transform (WT) is a tool for non-stationary signal analysis, and it is applied to steady-state detection in this work.

The Discrete Wavelet Transform (DWT) represents a signal as successive approximations of the original signal and it can be considered as the convolution of the input signal f with a wavelet function y, as seen in Eq. (1), according to the decomposition level.

$$W_{2^j} f(x) = f * y_{2^j}(x) \quad (1)$$

The wavelet function $y_{2^j}(x)$ is related to the high frequency components and so there is a scaling function f_{2^j} related to the low frequency components at each scale j. Therefore, the signals could be considered as a composition of approximations (identity or low-frequency content) and details (nuances or high-frequency components). Thus, if an abnormal sudden change occurs in the signal, the detail coefficients will be affected (Jiang *et al.*, 2000). Then, for any j = 0,

$$a_j[n] = \left\langle f(x), f_{2^j}(x-n) \right\rangle \quad (2)$$

$$d_j[n] = Wf(n, 2^j) = \left\langle f(x), y_{2^j}(x-n) \right\rangle \quad (3)$$

where a_j are the approximation coefficients and d_j (or $W_{2^j}f$) are the detail coefficients or WT modulus.

It is specially attractive if the y is the first-order wavelet, i.e., the first-order derivative of the scaling function $y(x) = df(x)/dx$, so thus Eq. (1) can be written as:

$$W_{2^j} f(x) = f\left(2^j \frac{df_{2^j}}{dx}\right) = 2^j \frac{d}{dx}\left(f f_{2^j}\right)(x) \quad (4)$$

where $f_{2^j}(x) = 1/\sqrt{2^j}\, f\left(x/2^j\right)$.

However, there is a fast algorithm to compute the DWT, computed as presented in Eq. (5).

$$a_{j+1}[n] = a_j * h_j[n], \quad d_{j+1}[n] = a_j * g_j[n] \quad (5)$$

The output a_{j+1} of a FIR filter to any given input may be calculated by convolving the input signal a_j with the impulse response expressed by the coefficients of the filter h_j. For a given filter x with coefficients $x[n]$, $x_j[n]$ denotes the filter obtained by inserting 2^j-1 zeros between every x coefficient.

The process of synthesizing or reconstructing the signal is mathematically computed by the Inverse Discrete Wavelet Transform. Hence, the process of reconstruction can be expressed as the sum of the details, or modulus maxima, and the coarser approximations.

2.2. Procedure for steady-state detection

The proposed technique consists of a process trends extraction of raw data using wavelet-based multi-scale analysis and after detection of the process status with extracted process trends at various scales. The process status is measured using a status index with value ranging from 0 to 1 according to the WT modulus of the extracted process signal. This methodology is based on Jiang *et al.* (2000, 2003).

The process begins with a decomposition of the original signal (WT on process data) generating a_j and d_j at each scale j. The algorithm is based on two quadrature mirror filters h and g proposed by Mallat and Zhong (1992), where h_j and g_j are filters with 2^j-1 zeros interpolated between two successive coefficients of h and g respectively. The wavelet function used is a quadratic spline.

In the next step, soft-thresholding is applied on d_j for scales $1 < j < J$, obtaining d_j'. The threshold for the first scale is assigned as the average of the modulus maxima of historical data, because at scale j = 1 the WT modulus is completely dominated by noise.

Afterwards, abnormal peaks, such as spikes, are detected and treated with symmetric extension technique for scales $2 < j < J$, resulting in new d_j' and a_j'. Spikes are identified if a couple of maximum WT modulus with opposite sign occurs, which duration is less than a time interval t_p considered from historical data. This corresponds to a sudden change in the process data. The threshold for identification of a spike p is computed by the variance of WT modulus of historical data at a defined scale. The duration $p_2 - p_1$ of the spike is determined from the average of WT modulus of historical data attributed a weight.

Later the signal is reconstructed using the threshold coefficients a_j' and d_j', from scale $j = J$ to 2. Jiang *et al.* (2003) suggest reconstructing up to $j = 1$, but as level 1 is dominated by noise it was removed from the reconstruction step.

Another WT is applied on the reconstructed signal, and the extracted trend fs is obtained at the characteristic scale $j = s$, determined by the response time constant and the sample time. The detail coefficients of this last decomposition will indicate the process status.

The status index B is basically determined by the derivatives of the extracted trend fs, expressed as WT_1 and WT_2.

Equation (6) expresses the estimation of the status index, where T_s, T_w and T_u are thresholds estimated from historical data.

$$B(t) = \begin{cases} 0 & ,q(t) \geq T_u \\ x[q(t)] & ,T_s \leq q(t) \geq T_u \\ 1 & ,q(t) \leq T_s \end{cases} \qquad (6)$$

For more details, refer to Jiang *et al.* (2003).

3. APPLICATION OF PCA

3.1. Steady-state detection based on key variables

As mentioned before, the original methodology for steady-state detection (Jiang *et al.*, 2003) is essentially developed for one process variable. For multivariate systems, the author suggests selecting key variables, calculating the status index for each one and then combining them using the Dempster's combination rule (Shafer, 1976). But this is an off-line methodology and it has some drawbacks considering its implementation.

The first drawback is related to the selection of the key variables (i), which requires good process knowledge. The key variables must be uncorrelated and should cover the whole system. Another drawback is that in the Dempster's combination rule some weights w_i must be established, as shown in Eq. (7).

$$B_m(t) = \prod_{i=1}^{N} [B_i(t)]^{w_i / \sum w_i} \qquad (7)$$

In this work, we are proposing to eliminate these drawbacks through the Principal Component Analysis (PCA) discussed as follows.

3.2 Steady-state detection based on principal components

Principal Component Analysis (PCA) is a linear dimensionality reduction technique, optimal in terms of capturing the variability of the data. It determines a set of orthogonal vectors (loading vectors) ordered by the amount of variance explained in the loading vector directions (Chiang *et al.*, 2001). The loading vectors are calculated by solving the stationary points of the optimization problem shown in Eq. (8).

$$\max_{\mathbf{v} \neq 0} \frac{\mathbf{v}^T X^T X \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \qquad (8)$$

where $\mathbf{v}$ are the loading vectors and X is the data matrix. The stationary points are computed via singular value decomposition.

The proposed methodology based on PCA has some advantages. It can be easily applied to multivariate systems. The process variables are combined in a new orthogonal variable so that there is no need of choosing key variables and weighting them. So the combination rule is different and it is simpler to be applied.

The steady-state detection based on principal components begins with a dimensional reduction by using PCA. Once the variables are chosen, they are transformed into new variables which are linear combinations of the original variables.

These new variables are then individually computed with WT for steady-state identification, as described in section 2.

4. INDUSTRIAL APLICATION

The industrial plant consists of a toluene column which is fed by the bottom stream of a benzene column. The toluene column has 60 valve plates and the feed plates are 30 and 36. The temperature of stage 20 is controlled through the reboiler steam flow rate. There are 5 flow measurements, 9 temperature measurements throughout the column and a top pressure measurement, as shown in Fig. 1.

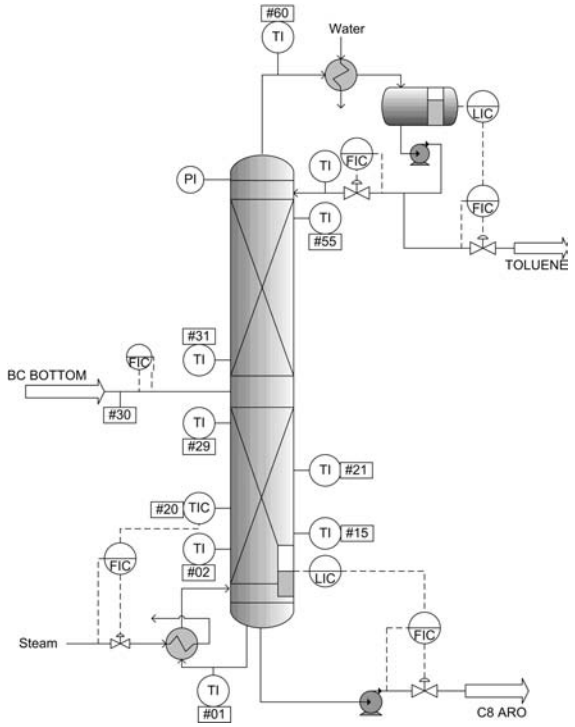


Fig. 1 – Measurements of the toluene column.

A time period was selected and its temperature profile is shown in Fig. 2.

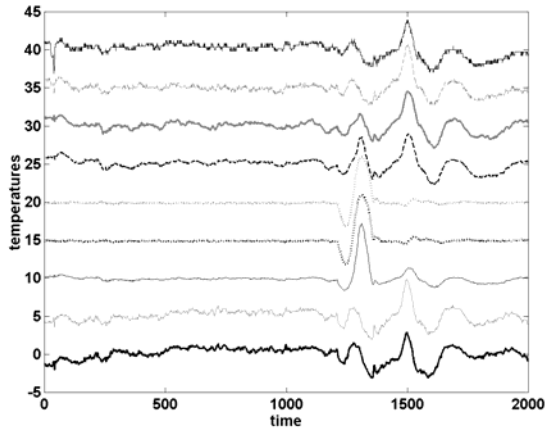


Fig. 2 – Temperature profile.

5. RESULTS

In this section, the PCA and Dempster's approaches are compared using the temperature profile of the industrial toluene distillation column. In this study, all selected variables are considered with same importance, what is translated into the following Dempster's combination rule:

$$B_m(t) = \prod_{i=1}^{N} B_i(t) \qquad (8)$$

Equation (8) implies that the column will be considered in steady-state if all variables are in steady-state at the same instant of time.

5.1. Setting the algorithm parameters

To initialize the algorithm, it is necessary to inform the typical process time constant t. The time constant used in the case study is t = 30 min. This value was estimated through the approximation of the step response of a 10-order ARX identified model obtained with the Matlab® System Identification Toolbox. The corresponding step responses were approximated through the SK method (Sundaresan and Hrishnaswamy, 1977), which delivered the time constant.

As a consequence, the parameter that represents the time interval over which a change usually persists, t_p, is estimated as 1/3-1/5 of t. This parameter is used for identification of abnormal peaks, as cited in section 2.

5.2. Status index by key variables

In Dempster's approach, the decision variables should be non-correlated. For the case study, these variables were chosen through a correlation analysis, which selected the following variables: the temperatures TI02 and TI21, the top pressure PI18 and the bottom level LIC09. The plant data of these variables are shown in Fig. 3.

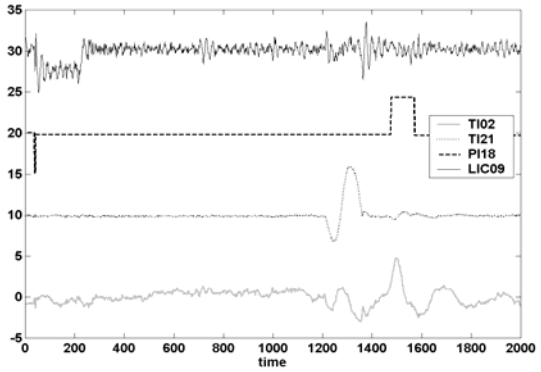


Fig. 3 – Selected key variables for the steady-state determination of the distillation column.

The results obtained for each key variable are presented in Figs. 4 to 7.

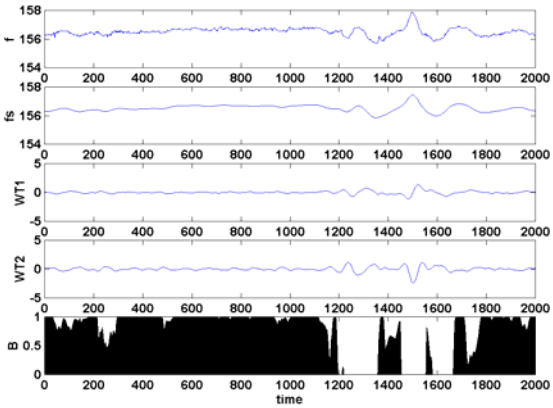


Fig. 4 – Representation of the steady-state detection using the WT for the temperature TI02.

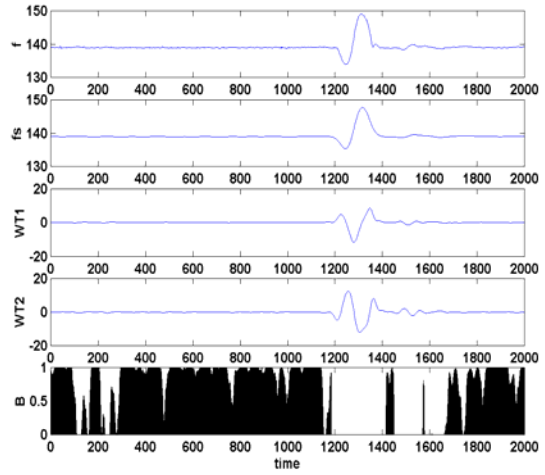


Fig. 5 – Representation of the steady-state detection using the WT for the temperature TI21.

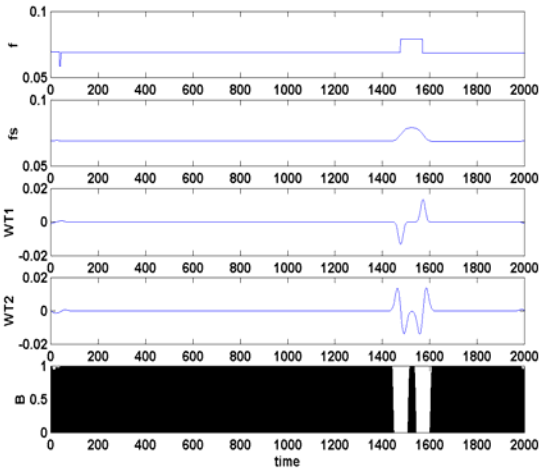


Fig. 6 – Representation of the steady-state detection using the WT for top pressure PI18.

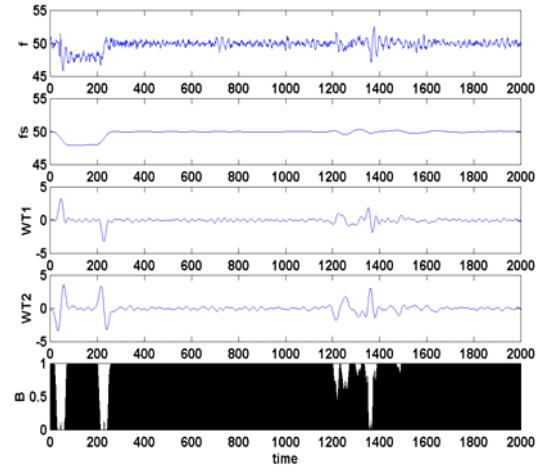


Fig. 7 – Representation of the steady-state detection using the WT for bottom level LIC09.

The status is computed for each variable and the overall status is computed by a combination as the one expressed in Eq. (8).

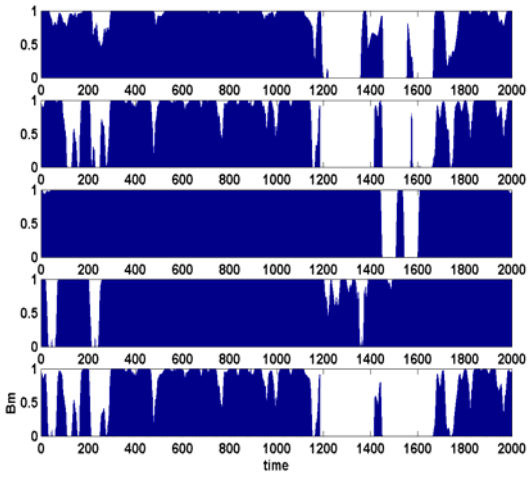


Fig. 8 – Combination of the status indexes of the key variables results in a unique status index B_m for the whole distillation column.

5.3. Status index by principal component analysis

The temperature profile (Fig. 2) is composed by the 9 temperature measurements indicated in Fig. 1. The analysis of the temperature profile by PCA results in two new variables, expressed here as t_1 and t_2, as shown in Fig. 9.

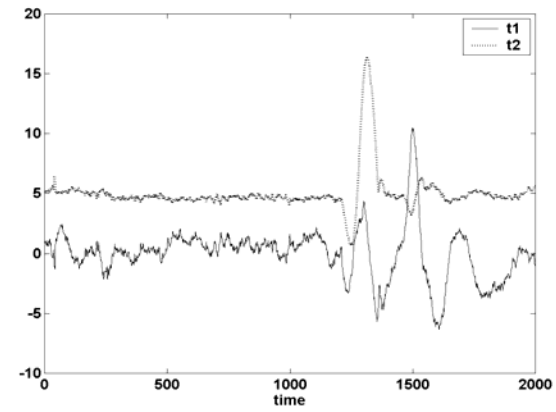


Fig. 9 – Resulting variables by PCA: t_2 and t_1.

For each variable t_1 and t_2, it was made a steady-state analysis and a status index was computed for each one. The input parameters were the response time constant and the historical data period. This period was considered as the first 600 points of t_1 and t_2. It is important here to emphasize the adequate historical period selection. This is an important point for the correct status index estimation. Historical data must bring representative features of the process variable, but without periods of unsteady conditions.

Figures 10 and 11 show the steady-state analysis, where f is the original signal (for t_1 or t_2), fs is the extracted trend, WT_1 is the first-order wavelet transform, WT_2 is the second-order wavelet transform, and B is the status index.

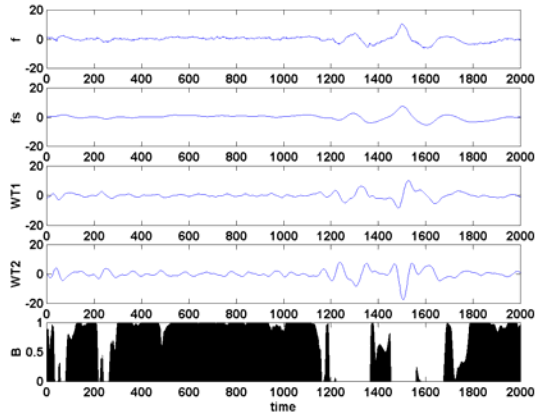


Fig. 10 – Representation of the steady-state detection using the WT for the first orthogonal variable t_1.

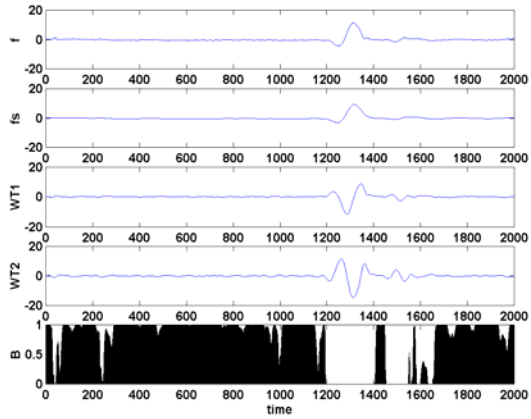


Fig. 11 – Representation of the steady-state detection using the WT for the variable t_2.

The combination of the two indexes B according Eq. (8) results in a unique index B_m, shown in Fig. 12, which represents the column status.

As seen in Fig. 12, the variables t_1 and t_2 have the same results, i.e., the same steady and non-steady time periods. This is a general observation that indicates it is not necessary to analyze the status of all orthogonal variables. Analyzing only the first variable, t_1 in this example, already brings enough information for the column status determination.

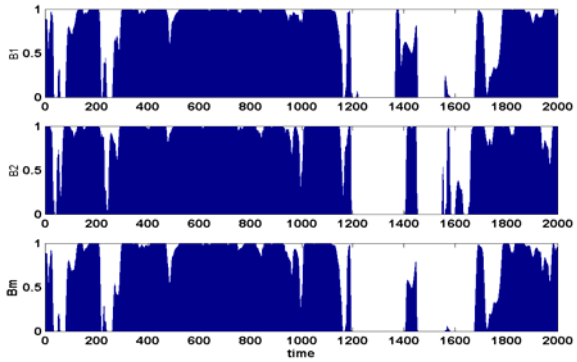


Fig. 12 – Combination of the status indexes B_1 and B_2 results in a unique status index B_m for the whole distillation column.

6. CONCLUSIONS

The results shown in Fig. 8 and 12 are very similar for the discussed industrial case study. Both approaches practically lead the same conclusion. However, the PCA approach is much easier and simpler for dealing with variables and does not require weighting attribution. The variables are selected and linearly combined by PCA without need of knowing what are the principal variables and what are exactly their influences in the process. This is an important point for practical applications. Another advantage is the status estimation of only one variable instead of all key variables, what considerably reduces computational effort.

ACKNOWLEDGMENTS

The authors thank PETROBRAS and FINEP for the financial support and Vanessa Conz and COPESUL for providing the industrial data used in this work.

REFERENCES

Cao, S. and R. R. Rhinehart (1995). An efficient method for on-line identification of steady state. *J. Process Control*, **5** (6), 363-374.

Chiang, L. H., E. L. Russell and R. D. Braatz (2001). *Fault detection and diagnosis in industrial systems*. Springer-Verlag London, Great Britain.

Daubechies, I. (1992). *Ten Lectures in Wavelets* (CBMS-NSF Series Appl. Math.), SIAM, PE.

Jiang, T., B. Chen and X. He (2000). Industrial application of wavelet transform to the on-line prediction of side draw qualities of crude unit. *Computers and Chemical Engineering,* **24**, 507-512.

Jiang, T., B. Chen, X. He and P. Stuart (2003). Application of steady-state detection method based on wavelet transform. *Computers and Chemical Engineering*, **27**, 569-578.

Mallat, S. and S. Zhong (1992). Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14** (7), 710-732.

Narasimhan, S., R. S. H. Mah and A. C. Tamhane (1986). *AIChE Journal*, **32** (9), 1409-1418.

Shafer, G. (1976). *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ.

Sundaresan, K. R. and P. R. Hrishnaswamy (1977). Estimation of time delay time constant parameters in time, frequency, and Laplace domains. *Can. J. Chem. Eng.*, **56**, 257.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

FAULT DETECTION USING PROJECTION PURSUIT REGRESSION (PPR): A CLASSIFICATION VERSUS AN ESTIMATION BASED APPROACH

Shijin Lou, Thomas Duever and Hector Budman

Department of Chemical Engineering, University of Waterloo, Waterloo, ON, Canada

Abstract: Two fault detection approaches are compared using a Projection Pursuit Regression (PPR) algorithm: i- a classification approach where the fault detection PPR model is trained based on the class numbers and ii- an estimation approach where the PPR model is trained to predict the value of the process variable that define the class boundaries and then the corresponding class is identified by comparing the estimated value versus the limits of the fault classes. The comparison is carried on for simple illustration examples, to elucidate the main issues, and for a copolymerization process. The classification approach is found superior provided that the training data closest to the boundaries are located at equidistant locations from these boundaries.

Keywords: Fault Detection, Regression Algorithm

1. INTRODUCTION

One of the goals in fault detection or classification problems is to establish, from measurements, that a specific variable value lies within a certain class defined by a range of values of that variable. Thus, for fault detection problems, the outcomes are a set of discrete values for the variable in question. On the other hand the values of the variables that define the boundary of a class can be continuously predicted from measurements using for example a state estimator. The prediction surface for the variable to be estimated is usually continuous. However, such an estimation model could be easily used for fault detection by comparing the predicted value of a variable to the boundaries defining the different classes or faults and then assessing the class or fault. The additional benefit of having continuous estimates of certain variables is that they could be used for feedback or feed-forward control. This paper is addressing the differences and relative advantages and disadvantages between these two approaches, i.e. the classical fault detection approach where a model is trained to predict directly a class versus the estimation based approach where the model is trained to predict the value of a variable and then the corresponding class is identified from that value. For clarity, the former will be referred to as the classification model whereas the later will be referred to as the estimation-based detection model. Intuitively, it is possible to expect that as the range of values, defining a class for the purpose of fault detection, becomes smaller and smaller, a fault detection model will eventually converge to an estimation model. Does this imply that fault detection is a "rough" version of estimation? Furthermore, will it be always true that a fault detection model will require less experiments for training as compared to the estimation based approach? It will be shown in this manuscript that the answers to these questions is not always affirmative and they are directly related to the level of noise, the linearity of the problem and the specific modelling methodology utilized to obtain the detection or estimation models. Many different modelling techniques have been proposed for estimation or fault detection problems. For example, Kalman filters have been often used for estimation or detection when a mechanistic model is available. Also, a number of empirical techniques have been investigated ranging from Neural Networks (Bakshi, 1999) to multivariate statistical modelling methods such as Partial Least Squares (Yoon and MacGregor, 2004). Projection Pursuit Regression (PPR) is an

additional multivariate modelling technique (Friedman, 1981) based on basis functions that are tailored specifically to the particular set of data to be modelled resulting generally in less parsimonious models with lower sensitivity to noise. The authors of this work have conducted extensive research on the use of PPR for class detection and have compared this modelling fault detection methodology with other techniques such as back propagation neural networks and Radial Basis Functions Neural networks. They have found in these studies that PPR provides a good tradeoff between sensitivity to noise and generalization accuracy as compared to other neural network based methodologies. (Lou, 2003). For instance, for a 2-dimensional problem, Lou found that PPR results in approximately 50% of the classification error obtained with a Haar Wavenet-based model and 35% of the classification error obtained with a Backpropagation Neural Network model. Therefore, this study will conduct the comparison between direct detection and estimation-based detection using specifically PPR based models.

This paper will be organized as follows. Section 2 will briefly summarize the PPR algorithm and its application to detection and estimation problems. Section 3 will discuss simple examples that were specifically tailored to elucidate some of the issues discussed in the introduction regarding the comparison between detection and estimation problems using PPR. Section 4 will discuss a more involved chemical engineering problem, the estimation of impurities in a copolymerization process from conversion and temperature measurements. Finally, conclusions are presented in Section 5.

2. PROJECTION PURSUIT REGRESSION (PPR): BRIEF SUMMARY

PPR is a multivariate statistical technique originally proposed by Friedman and Stuetzle (1981). The technique can be viewed as a 3 layer-neural network composed of an input layer, one hidden layer and an output layer. The input layer operates on inputs or independent variables x whereas the output layer produces the outputs or dependent variables y.

Three sets of parameters: projection directions given by weights between the input and the hidden layers $\alpha_k^T = [\alpha_{k1},...,\alpha_{kp}]$, projection 'strengths' given by weights between the hidden and the output layers) $\beta_k = [\beta_{1k},...,\beta_{qk}]$, and the a priori unknown activation functions in the hidden layer $\{f_k\}$, are estimated via the least squares criteria by minimizing the squared error cost function: (Utojo and Bakshi (1999))

$$L = \sum_{i=1}^{q}\left[(y_i - \sum_{k=1}^{m}\beta_{ik} f_k(\alpha_k^T x)\right]^2 \qquad (1)$$

Each response variable, y_i $(i = 1,2,...q)$, is modeled as a weighted linear combination of the activation function $\{f_k\}$. Each of these functions is a nonlinear function or 'look-up' table, of a weighted linear combination of the weighted independent variables. The output of a hidden function f_k is decided according to the nearest neighbor or neighbors in the 'look-up' table. Projection Pursuit Regression learns function by function and layer-by-layer cyclically after all the training patterns are presented. Specifically, it applies linear Least Squares to estimate the output-layer weights and the Gauss-Newton nonlinear Least Squares method to estimate the input-layer weights. The optimization algorithm grows the model step-wise as in the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm used for the Partial Least Squares (PLS) method. The main difference between PPR and PLS is that the later uses fixed-shape basis functions either linear or polynomial, while PPR uses adaptive basis functions, which are decided by the training data. The PPR basis functions are computed by smoothing the projected data versus the output by using a variable-span smoother such as the supersmoother (Friedman (1984)). The adaptability of the basis functions in PPR allows it to determine more parsimonious models, i.e., using less basis functions than those modeling tools using fixed basis functions, for the same approximation error. A detailed mathematical description is given by Utojo and Bakshi (1999).

Finally, in the introduction, two different forms of constructing a fault detection algorithm have been discussed, i.e. direct detection of the class or fault versus estimation of the variable value and then testing this value versus the ranges of values that define the classes or faults. The difference between the two methodologies is that for the first case, the output data y is discrete and it is typically given in terms of integer numbers whereas for the second method continuous values of y are used for training. In section 3 and 4 examples are given to compare these two methodologies based on the PPR regression algorithm.

3. SIMPLE ILLUSTRATION EXAMPLES

In this section two simple examples are presented to address the comparison between the direct-classification approach versus the estimation-based detection approach. The examples have been specifically tailored to elucidate the issues especially with regards to sensitivity to measurement noise and nonlinearity of the underlying process for which faults are to be detected.

3.1 Linear example

In this example, a linear process model is represented by the following equation:

$$x = p \qquad (2)$$

Where, x is the process measurement; p is the process variable. The objective of a classification model is to find a specific class or fault based on a

measurement x. The classes are defined by the value of p as follows:

$$\text{Class 1: } 0<p\leq 0.5 \quad \text{Class 2: } 0.5<p<1 \quad (3)$$

Unlike the classification model, the goal of the estimation-based model is to establish a direct mapping from x to p, i.e., to predict the true value of p, according to the measurements, x. The estimate of p is then used to decide which class x belongs to. Thus the inputs to the PPR model, referred to as the network input, are the measured values of x and the output from the PPR model y, referred to as the network output, are equal to the class number for the classification model or to the estimated values of p for the estimation-based model.

For this example it is assumed that 3 measurements of x are available x=[0.2 0.6 0.95]. Correspondingly, for the training of a classification model, the PPR model is trained on a data pattern given by y= [1 2 2]. Based on this training data the PPR model is tested for different values of x providing the results shown in Figure 1. Clearly, the PPR model locates the class boundary at x=0.4 instead of x=0.5 that is the actual location of the boundary according to (3) resulting in misclassification of all the point in the range 0.4<x<0.5. The explanation for this misclassification is that the two training data on the two sides of class boundary (x=0.2 and x=0.6) are not symmetric with respect to the actual class boundary $x=0.5$. Since the PPR output calculation is based on the nearest neighborhood concept, the PPR model locates the class boundary at the midpoint between the rightmost point of class 1 (x=0.2) and the leftmost point from class 2 (x=0.6) locating the boundary at x=0.4 with resulting misclassification of testing data.

On the other hand the training data for a PPR estimation-based model are the actual measured values as follows y=[0.2 0.6 0.9] instead of y=[1 2 2] used for the classification model. The estimation-based model finds correctly the straight line relation described by (2) passing through all three training data. Consequently for this case, the estimation model can make accurate prediction, even though the training data on two side of the class boundary are not symmetric with respect to it. In this experiment, the estimation model predicts the testing data accurately, and the classification based on the estimation-based model does not produce any misclassification. This example show that in a noise-free linear problem, a PPR estimation-based model trained with the absolute values of the measurements works better than a PPR classification model trained with the class number values, especially when the training data in the two classes are not symmetric with respect to the actual class boundary.

3.2 Linear Example using Training Data Corrupted by Noise

The system in this example is the same as described by (2) above. In this example, there are also three training data, as in the previous example, with one training data in Class 1 and two in Class 2. The training data pattern for the estimation model is plotted in Figure 2. In this case the training data is corrupted by noise, and consequently is biased from the actual process model represented by the solid straight line in figure 2.

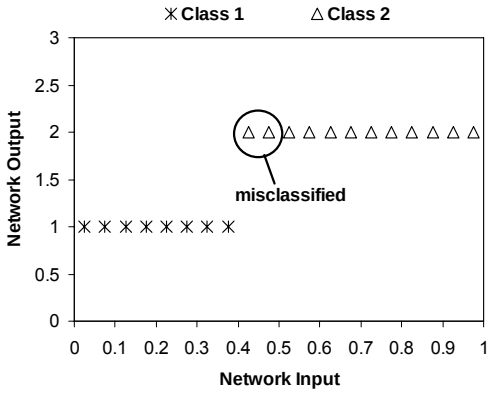


Figure 1. Testing results by a PPR classification model, for a 1-dimensional linear example with noise-free training data.

In the training of a PPR classification model, the inputs are x=[0.2 0.8 0.9] and the outputs for training are the corresponding class numbers as follows y=[1 2 2]. In this case the PPR classification model correctly locates the class boundary at x=0.5 because the rightmost data point from class 1 (x=0.2) and the leftmost data point from class 2 (x=0.8) are now symmetric with respect to the class boundary at x=0.5. Then, the PPR classification model correctly predicts all faults by assigning class one to all measured x<0.5 and class 2 to all measured x>0.5.

For the estimation based model the training data is given by x=[0.2 0.8 0.9] whereas the output data is y=[0.3 0.88 0.9]. In this case, due to noise and the sparseness of the training data a poor PPR estimation-based model is obtained. The prediction of the testing data for different values of x is quite different from their true value as shown in figure 3, resulting in misclassification of 10% of the tested points as illustrated in that figure.

Thus, in a classification problem, the noise in the training data will not affect the classification accuracy, unless the noise level is so significant that it causes data to be assigned to the wrong class. Thus, the noise has no harmful effect on a classification model, if it is small enough such as the training data are still located in the correct classes. This is exactly the situation in this example. Therefore, the classification model makes no misclassification in the testing. This example shows that, due to the discretization of the network outputs, a classification model may be less sensitive to the noise in the training data, as compared to an estimation-based model.

3.3 Nonlinear Example using Noise-free Training Data

This example assumes a nonlinear model, and there is no noise in the training data. The process model can be described by the following model.

$$x = \log_{10}(p) \qquad (4)$$

The classification is decided as follows.

Class 1: $p \leq 3.16$ Class 2: $p > 3.16$

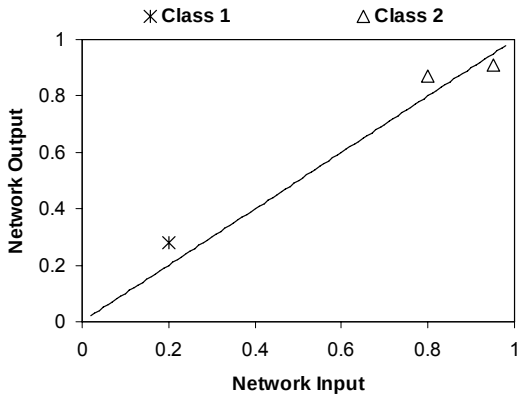


Figure 2. Training data with noise for a PPR estimation model, 1-dimensional linear example

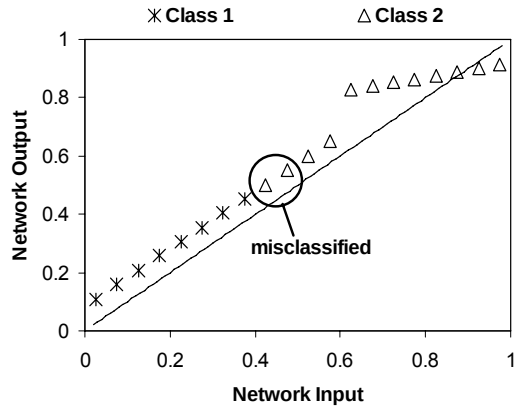


Figure 3. Testing result by a PPR estimation model, 1-dimensional linear example with noisy training data

The training data for the estimation model is presented in Figure 4. The training data in Class 1 and Class 2 are represented by *star* and *triangle* symbols, respectively. The class boundary is located at $x=0.5$ and $p= 3.16$. For these data, the *star* and the *triangle* closest to the class boundary are symmetric with respect to it, in terms of the network input, x. Consequently the PPR classification model accurately predicts the testing data without any misclassification. On the other hand, due to the nonlinearity of the problem and the sparseness of the training data, the estimation-based PPR model misclassifies testing data as shown in Figure 5. The sudden change in the output around the input $x=0.6$ is a consequence of the particular basis functions that the PPR algorithm found for this problem and for the given training data. The training of the estimation model has been done to obtain a training error of zero for the 3 data points in Figure 4. The difference between the estimation and the actual value results in 10% misclassification out of the total data tested. Thus, although PPR is a suitable algorithm to describe nonlinear systems, the resulting estimation model is not accurate due to the sparseness of the training data.

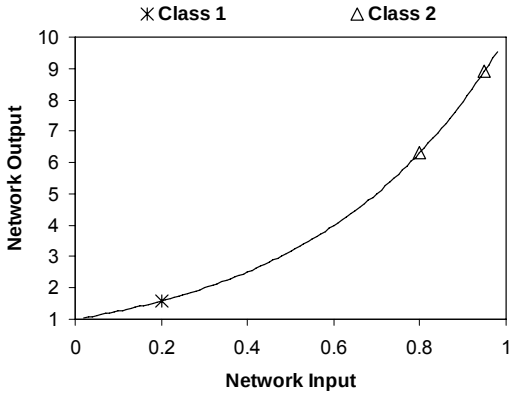


Figure 4. Noise-free training data for a PPR estimation model, 1-dimensional nonlinear example

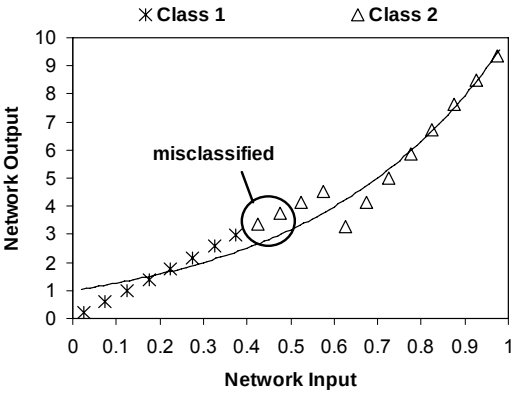


Figure 5. Testing result by a PPR estimation model, 1-dimensional nonlinear example with noise-free training data

This example shows that, in a nonlinear problem, a classification model may need less training data to reach desirable classification accuracy, as compared to an estimation-based model.

3.4 Nonlinear two-dimensional example

The process model investigated here can be mathematically expressed as follows:

$$(5 \cdot x_1)^{0.5} \cdot (1.5 \cdot x_2)^5 = p + v \qquad (5)$$

First all the data are noise free, i.e. v is set to zero. The classification is decided by the value of the process variable, p.

Class 1: $p \leq 0.4$
Class 2: $p > 0.4$

The function is geometrically illustrated in Figure 6. A complete grid is sampled and plotted in the measurement domain in Figure 7. A data point is either represented by a *star* or a *plus*, according to its corresponding class. The boundary between the classes shown in Figure 7 is not straight as in the previous one-dimensional examples. For training, the

following four process measurements are sampled: [0.025 0.025], [0.025 0.975], [0.975 0.025], and [0.975 0.975]. The corresponding output to train the classification model are y= [1 2 1 2] and their corresponding output values for training of the estimation-based model are y=[2.622×10^{-S}, 2.366, 1.637×10^{-7}, 14.773] respectively. All, the points in Figure 7 are used for testing of the resulting PPR regression model.

It is possible to show from figure 7 that the selected training data is located approximately symmetrically with respect to the class boundary corresponding to p=0.4, i.e. the training data in class 1 and class 2 are located at similar distances to the class boundary in terms of their x coordinates. The missclassification on the testing data are 16.8% for the estimation model, and 3.2% for the classification model, as summarized in Table 1.

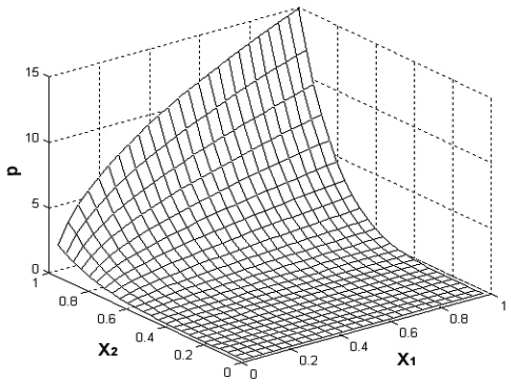


Figure 6. Function surface, 2-dimensional nonlinear example with noise-free data

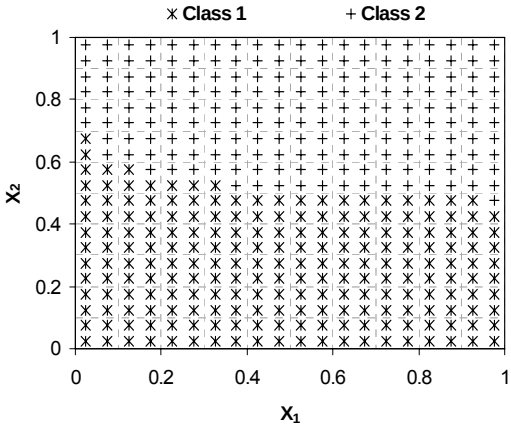


Figure 7. Testing data set for the 2-dimensional nonlinear example with noise-free data.

Subsequently, the training data is corrupted by random noise. The noise is assumed to be of a magnitude, so that the difference between each measurement (x_1 or x_2) and its true value is smaller than 0.5 times its sampling rate. The training data for the estimation model are: [0.0320 0.0117], [0.0300 0.9579], [0.9301 -0.0556], and [0.9883 0.9791] with the corresponding desired outputs for training the estimation-model y=[0.1489, 2.0961, -0.89646, and 15.499]. The classification model is trained with y=[0 1 0 1].

The misclassification for the testing data is summarized in Table 1. The results in Table 1 verify for the 2 dimensional case the following conclusion: for nonlinear systems and in the presence of measurement noise, a PPR classification model can outperform a PPR estimation-based model, when training data is located on the two sides of class boundary and in symmetrical locations with respect to it. This conclusion is consistent with the results obtained in the one-dimensional case.

Table 1. Comparison of classification and estimation technique in two-dimensional examples

	estimation	**classification**
Misclassification percentage in noise-free data	*0.1675*	*0.0325*
Misclassification percentage in noisy data	*0.215*	*0.04*

	estimation	**classification**
Misclassification percentage in noise-free data	*0.1675*	*0.0325*
Misclassification percentage in noisy data	*0.215*	*0.04*

4. EXAMPLE OF A COPOLYMERIZATION PROCESS

Finally, the comparison between a pure classification model to an estimation-based PPR models is carried out for a fault detection task in a polymerization process. The process is a batch copolymerization of STY/MMA. A detailed mathematical model proposed by Landry (1996) has been used. The model is given by six 1st order ODE's derived from energy, mass and component balances. Reactive impurities are commonly encountered in industrial polymerization processes. Consequently, the objective of the fault detection algorithm is to identify the impurity in ranges of values defining classes as follows:

Class 1: $0 \le y < 100\,ppm$
Class 2: $100\,ppm \le y < 300\,ppm$
Class 3: $300\,ppm \le y < 500\,ppm$
Class 4: $500\,ppm \le y < 700\,ppm$

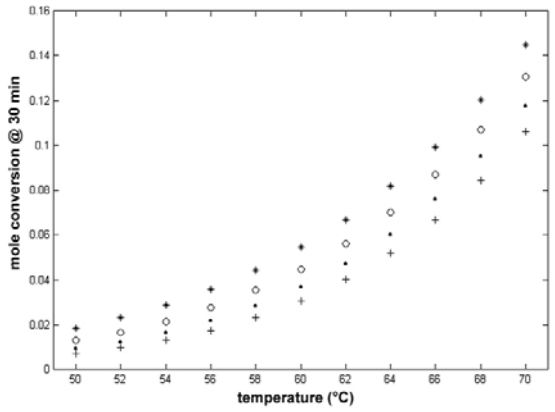
The impurities are detected based on two measurements: the temperature and the mole conversion after 30 minutes of operation. The authors of this work have theoretically shown that the impurity concentration is observable from these two measurements (Lou, 2005). The training data for the case without noise is shown in Figure 8. Based on the observations made for the examples shown in Section 3, the training data points were selected at equidistant locations from the two sides of the class boundaries defined above. The simulation results are

summarized in Table 2. The results show that the classification model gives better performance than the estimation model for both the noise free data and data corrupted by Gaussian noise. However, the difference is not as large as expected. To clarify further, the simulated data have been investigated in graphic form. Figure 9 presents the noise-free testing data. In this diagram, the impurity is plotted with respect to the temperature and the mole conversion. Although the overall data pattern is obviously nonlinear, the nonlinearity is not very large.

In general the observations from this more complex example confirms that the PPR classification model tends to outperform a PPR estimation model, when the problem is nonlinear and in the presence of measurement noise. It is expected based on the simple examples shown above, that the improvement of the classification model versus the estimation-based model could be especially significant when the nonlinearity is more pronounced.

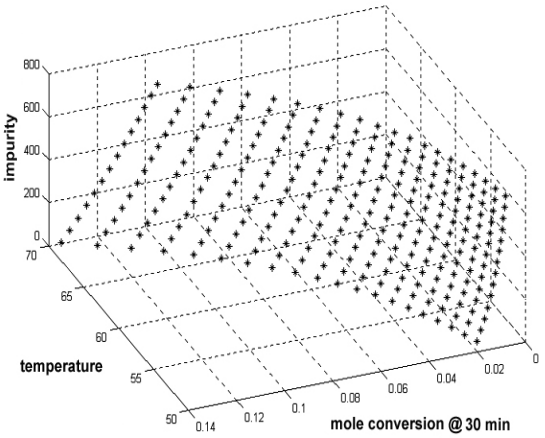
Table 2. Comparison of classification and estimation technique in polymerization examples

	Noise-free data	**Data with noise**
Estimation	*0.32*	*0.49*
Classification	*0.27*	*0.46*



✱ Class 1; ○ Class 2; • Class 3; + Class 4

Figure 8: Original training data in the process measurement space, isothermal copolymerization example



✱ testing data (not symbolized according to classification)

Figure 9. Noise-free Testing data, 2-dimensional polymerization example

5. CONCLUSIONS

In this work two modelling approaches for fault detection are compared using a PPR algorithm: a classification approach where the model is trained based on the class number versus an estimation based approach where the value of the process variable defining the fault is identified and then the class is identified based on that value. It was found that the classification approach generally outperforms the estimation based approach for nonlinear systems, when the data is sparse and in the presence of measurement noise. This result holds provided that the training data is distributed approximately symmetrically with respect to the class boundaries. Then, the PPR algorithm based on such data correctly locates the class boundary since it uses the nearest neighbourhood concept to calculate the output.

References

[1] Bakshi, B. R.; Stephanopoulos, G.; WaveNet: a Multiresolution Hierarchical Neural Network with Localized Learning, AIChE Journal, vol. 39, no. 1, pp: 57-81, 1993

[2] Bakshi, B. R.; Utojo, U.; A common framework for the unification of neural, chemometric and statistical modeling methods, Analytica Chimica Acta, 384, 227-247, 1999.

[3] Friedman, J. H.; Stuetzle, Werner; Projection Pursuit Regression, Journal of the American Statistical Association, vol. 76, no. 376, p: 817-823, Dec. 1981

[4] Landry, R.; Model discrimination issues in polymer reaction engineering, Master Thesis, University of Waterloo, Canada, 1997

[5] Lou, S. J., Budman, H., Duever, T. A., Comparison of fault detection techniques, Journal of Process Control, v 13, n 5, 2003, p 451-464.

[6] Yoon, S., MacGregor, J. F.; Principal Component Analysis of Multiscale Data for Process Monitoring and Fault Detection, AIChE Journal, 50, 2891-2903, 2004.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

FAULT DETECTION USING CORRESPONDENCE ANALYSIS: APPLICATION TO TENNESSEE EASTMAN CHALLENGE PROBLEM

Detroja K. P.[1], Gudi R. D.[2*], Patwardhan S. C.[2]

[1] Interdisciplinary Programme on Systems and Control Engineering,
[2] Department of Chemical Engineering
Indian Institute of Technology Bombay, Powai, Mumbai – 400076, India.
* Email: ravigudi@che.iitb.ac.in

Abstract: This paper presents an approach based on the use of correspondence analysis (CA) for the task of fault detection and diagnosis. Unlike other tools (PCA / DPCA) that are used for this latter task, CA is shown to use a different metric to represent the information content in the data matrix **X**. Decomposition of the information represented in the metric is shown to yield superior performance from the viewpoints of data compression, discrimination and classification as well as early detection of faults. We demonstrate these performance improvements over PCA and DPCA on the Tennessee Eastman problem, which is a representative benchmark problem used in the literature. CA is shown to yield vastly superior performance for the monitoring of the TE problem, when compared with PCA and DPCA.

Keywords: Correspondence Analysis (CA), Principal Components Analysis (PCA), Dynamic PCA (DPCA), Fault Detection (FD)

1. INTRODUCTION

Early detection of the occurrence of an abnormal event in an operating plant is very important for plant safety and maintaining product quality. Tremendous advancements in the area of advanced instrumentation have made it possible to measure hundreds of variables every few seconds. These measurements bring in useful signatures about the status of the plant operation. A wide variety of techniques, for detecting faults, have been proposed in the literature. These techniques can be broadly classified into model based methods and historical data based methods. While model based methods can be used to detect and isolate signals indicating abnormal operation, such quantitative (or qualitative) cause-effect models may be difficult to develop from the first principles.

Historical data based methods for fault detection attempt to extract maximum information out of the archived data and require minimum physical knowledge of the plant. Due to the high dimensionality and correlation amongst the variables of the plant data, multivariate statistical tools, which take correlation amongst variables into account, are better suited for this task. Dimensionality reduction is also a very important aspect of historical data based methods.

Generally, the information content in a data matrix $\mathbf{X}$ can be quantified in terms of a number of criteria or metrics. The most commonly used metric, the variance or the multivariate analysis of the variance (MANOVA), usually yields a wealth of knowledge from the information embedded in the matrix $\mathbf{X}$. Multivariate statistical tools, such as PCA, are based on decomposition of the variances and address issues related to correlation along the column *or* the row spaces. PCA determines the lower dimensional representation of the data, in terms of capturing the data directions that have the most variance. This is done via singular value decomposition (SVD) of a suitably scaled (mean centered and variance scaled) data matrix ($\mathbf{X}$) and retaining those principal components that have significant singular values. PCA achieves dimensionality reduction in the column space by considering the correlation amongst the variables. The statistical model thus built, characterizes the normal plant operation. PCA has been used for fault detection using statistical control limits Q (Squared Prediction Error) and/ or T^2 statistics (Nomikos and MacGregor, 1995). Once a fault is detected using either Q or T^2 statistics,

contribution plots (Miller *et al.*, 1998) have been used to help fault isolation. One of the drawbacks of PCA, however, is that it is representation-oriented and not discrimination-oriented. As shown in Chiang *et al.* (2000), there are other algorithms such as multiple discriminant analysis that can better discriminate between the normal and abnormal operating regions in the data and hence yield smaller misclassification rates during on-line monitoring.

An important aspect that also needs to be considered is that the variance need not be the best metric for capturing cause and effect relationships. Usually, such cause and effect relationships are dynamic and can be more effectively analyzed by assessing the row (sample) versus column (variable) associations. In PCA or in the multiple discriminant analysis (MDA) approach, such dynamic relationships require expanding of the column space to generate a static map of the dynamic relationships. This latter strategy has drawbacks in terms of larger matrix and data sizes and increasing computational intensity.

This paper proposes to address the above problems using an approach that is based on CA for the task of FDD. Correspondence Analysis (CA) (Greenacre, 1984; Greenacre, 1993; Hardle and Simar, 2003) is a powerful multivariate statistical tool, which is based on generalized SVD (GSVD). CA is a dual analysis, as it simultaneously analyzes dependencies in column, row and the joint row-column space in a dual lower dimensional space. Thus, dynamic correlation can be represented relatively easily without having to expand and deal with larger data sizes. CA primarily uses a measure of the row-column association and decomposes it to obtain directions in the lower dimension space which discriminate as well as compress information. Unlike its earlier counterparts such as PCA and MDA, it represents the cause-effect relationships in terms of a chi-square (χ^2) value, that measures row-column associations. Since, decomposition of the χ^2 value takes joint row-column association into account; it can be expected to perform better than conventionally used variance decomposition based methods, such as Principal Components Analysis (PCA).

In this paper, we show how Correspondence Analysis (CA) is superior to PCA and MDA and can be used for the purpose of fault detection and have also defined statistics based on CA that can be used for online process monitoring. It has also been found that the performance of statistics based on CA is better as compared to conventional PCA. The dimensionality reduction achieved using CA is more effective as it takes joint row-column association into account. Also, due the special kind of scaling it employs, CA is also shown to be able to cluster and aggregate the data more effectively (Ding *et al.*, 2002).

The objective of this paper is (i) to demonstrate the usefulness of CA for fault detection, (ii) to define new statistics which are equivalent to Q and T^2 statistics for PCA and (iii) to evaluate and compare performance of PCA, DPCA and CA for detecting faults in a realistic chemical process simulation. We show here that the proposed statistic performs better than the existing PCA and DPCA statistics when applied to the Tennessee Eastman process. The paper is organized as follows. First, PCA and DPCA are briefly presented. Then, CA is described followed by the proposed approach to fault detection using CA based statistics. Finally, PCA, DPCA and CA are applied to the data collected from the Tennessee Eastman process simulator. We conclude with comparative study of results.

2. Principal Components Analysis (PCA)

Any matrix $\mathbf{X}_{m\times n}$ consisting of *m*-observations and *n*-variables, collected from an operating plant has a wealth of information regarding the health of the plant. PCA decomposes the variance in the data, based on dependencies along the columns, to achieve dimensionality reduction. PCA computes a set of new orthogonal principal directions, called loading vectors. Loading vectors are obtained by solving an optimization problem involving maximization of variance explained in the data matrix by each direction. For example, the first direction is obtained as a solution of the optimization problem in the space of the first linear combination $\mathbf{t}_1 = \mathbf{X}\mathbf{p}_1$ as,

$$\max_{\mathbf{p}_1} \left(\mathbf{t}_1^T \mathbf{t}_1 \right) = \mathbf{p}_1^T \mathbf{X}^T \mathbf{X} \mathbf{p}_1 \qquad (1)$$

Such that $\mathbf{p}_1^T \mathbf{p}_1 = 1$.

It has been shown that the singular vector corresponding to the largest singular value provided by the SVD of $\mathbf{X}$, is the solution to the above optimization problem. Because of correlation amongst variables, only first k (substantially smaller than n) loading vectors may explain most of the variance in the data. Thus, PCA decomposes the matrix $\mathbf{X}$ as,

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \qquad (2)$$

where, $\mathbf{P}$ contains only first k ($k << n$) loading vectors. The matrix $\mathbf{T}$ is called the scores matrix. The matrix $\mathbf{E}$ contains the component of variance of matrix $\mathbf{X}$, such as noise, which can not be explained by $\mathbf{T}\mathbf{P}^T$, and is also known as residual matrix.

2.1. Fault detection using PCA

The statistical model developed using PCA, from the normal operating data, can be used for the purpose of online monitoring and fault detection. When employed online, new scores are obtained by projecting the new measurements onto the loading vectors. Normal operation of the plant can be characterized by Hotelling's T^2 statistic (Equation (3)), based on the first *k* loading vectors (principal components) retained. The status of the plant is considered normal if the value of T^2 static stays within its control limit.

$$T^2 = \mathbf{x}^T \mathbf{P} \mathbf{\Lambda}^{-1} \mathbf{P}^T \mathbf{x} \quad (3)$$

where, $\mathbf{x}$ is the new measurement vector and $\mathbf{\Lambda}$ is a diagonal matrix containing first k eigen values of the covariance matrix of $\mathbf{X}$.

The control limit (threshold) for the T^2 statistic T_α^2 can be calculated from Equation (4) (Ku et al., 1995). A value of T^2 statistic greater than the control limit (T_α^2) indicates occurrence of a fault.

$$T_\alpha^2 = \frac{(m-1)k}{(m-k)} F_\alpha (k, m-k) \quad (4)$$

where, $F_\alpha (k, m-k)$ is the upper $100\alpha\%$ critical point of F-distribution with k and $m-k$ degrees of freedom.

However, monitoring only T^2 statistic is not sufficient, as it only detects variation in the direction of the first k PCs. Variation in the space corresponding to $(n-k)$ PCs (having smallest associated singular values) can also be monitored using Q statistic (Jackson and Mudholkar, 1979). The value of Q statistic and its control limit can be calculated as follows:

$$Q = \left[\left(\mathbf{I} - \mathbf{P}\mathbf{P}^T \right) \mathbf{x} \right]^T \left[\left(\mathbf{I} - \mathbf{P}\mathbf{P}^T \right) \mathbf{x} \right] \quad (5)$$

$$Q_\alpha = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{(1/h_0)} \quad (6)$$

where, $\theta_i = \sum_{j=k+1}^{n} (\mu_j)^{2i}$, $h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2}$, c_α is the normal deviate corresponding to $(1-\alpha)$ percentile and μ_j is j^{th} singular value. When a fault occurs that results in change in covariance structure of the normal operating data, it gets reflected by a high Q value.

2.2. Dynamic PCA

Monitoring using PCA statistics implicitly assumes that the measurements at one time instant are statistically independent to the measurements at the past time instances. The assumption is generally not valid for most processes due to dynamics of the plant. The PCA method can be extended to take into account the serial correlations, by augmenting each observation vector with a few past observations and stacking the data in a bigger matrix.

$$\mathbf{X}_A = \left[\mathbf{X}(t) \mathbf{X}(t-1) \cdots \mathbf{X}(t-l) \right] \quad (7)$$

By performing PCA on the augmented data matrix ($\mathbf{X}_A$), a multivariate auto regressive (AR) model is extracted directly from the data (Ku *et al.*, 1995). This however, requires working with considerably larger data matrices than the conventional PCA. The T^2 and Q statistics and their control limits can be generalized directly to DPCA.

3. CORRESPONDENCE ANALYSIS

The aim of correspondence analysis is to develop simple indices to highlight associations between the rows and the columns. Unlike PCA, which canonically decomposes the total variance in the matrix $\mathbf{X}$, CA decomposes a measure of row-column association, typically formulated as the total χ^2 value, to capture the dependencies. CA can be presented in terms of weighted Euclidean space as follows. In general, through an optimization procedure, we seek a lower dimension (say k) approximation of the matrix $\mathbf{X}$ in an appropriate space $\boldsymbol{S}$. In terms of the row and column points, each row of $\mathbf{X}$ can be represented as a point $\mathbf{x}_i$ $(i = 1, 2..m)$ in an n-dimensional space. When one seeks to estimate the lower dimensional space (approximation) $\boldsymbol{S}$ that is closest to this cloud of row points, one could solve optimization problems that are formulated in several possible ways. One such optimization problem to determine the space $\boldsymbol{S}$ could then be minimize a weighted Euclidean distance defined as,

$$d^2 = (\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{D} (\mathbf{x} - \bar{\mathbf{x}}) \quad (8)$$

It can be shown (Greenacre, 1984) that the solution to the problem of minimizing the weighted distances in Equation (8) can be given by decomposition of inertia of row (or column) cloud, i.e. generalized SVD of the matrix $\left[(1/g)\mathbf{X} - \mathbf{r}\mathbf{c}^T \right]$. The vectors $\mathbf{r}$ and $\mathbf{c}$ are the vectors of row sums and column sums of $\left[(1/g)\mathbf{X} \right]$, respectively (Equation (9) & (10)).

$$\mathbf{r} = \left[(1/g)\mathbf{X} \right] \mathbf{1} \quad (9)$$

$$\mathbf{c} = \left[(1/g)\mathbf{X} \right]^T \mathbf{1} \quad (10)$$

where, $\mathbf{1}$ is a vector of all 1's of appropriate dimension. The matrix $\mathbf{D}_r$ is then defined as $\mathbf{D}_r = diag(\mathbf{r})$ and similarly, $\mathbf{D}_c = diag(\mathbf{c})$.

The inertia of the row cloud and the column cloud can be shown to be the same (Greenacre, 1984) and is given by the χ^2 value divided by g. The weight matrix $\mathbf{D}$ is chosen as diagonal matrix of the row sums ($\mathbf{D}_r$) or the column sums ($\mathbf{D}_c$). The generalized SVD of this matrix is defined as

$$\left[(1/g)\mathbf{X} - \mathbf{r}\mathbf{c}^T \right] = \mathbf{A}\mathbf{D}_\mu \mathbf{B}^T \quad (11)$$

such that, $\mathbf{A}^T\mathbf{D}_r^{-1}\mathbf{A} = \mathbf{I}_{m\times m}$ and $\mathbf{B}^T\mathbf{D}_c^{-1}\mathbf{B} = \mathbf{I}_{n\times n}$.

The generalized SVD results of Equation (11) can also be realized via the SVD of an appropriately scaled matrix $\mathbf{X}$, as explained below. We define the matrix $\mathbf{P}$ as,

$$\mathbf{P} = \mathbf{D}_r^{-1/2}\left[(1/g)\mathbf{X} - \mathbf{r}\mathbf{c}^T\right]\mathbf{D}_c^{-1/2} \tag{12}$$

Then, the regular SVD of the matrix $\mathbf{P}$ gives the required singular vectors. The problem of finding principal axis for the row cloud and the column cloud are dual to each other and $\mathbf{A}$ and $\mathbf{B}$ define the principal axes for the column cloud and the row cloud respectively. In general, major part of the χ^2 value can be explained by retaining only first k ($k << m,n$) principal axes corresponding to the largest singular values. The co-ordinates (scores) of the row profile points and column profile points for the new principal axis can be computed by projection on $\mathbf{A}$ and $\mathbf{B}$ (only first k columns are retained), respectively.

$$\mathbf{F} = \mathbf{D}_r^{-1}\mathbf{A}\mathbf{D}_\mu \tag{13}$$

$$\mathbf{G} = \mathbf{D}_c^{-1}\mathbf{B}\mathbf{D}_\mu \tag{14}$$

3.1. Singular values and inertia

The sum of the squared singular values gives the total inertia of the cloud. The inertia explained by each principal axis can then be computed by

$$IN\left(i^{th}\ axis\right) = \frac{\mu_i^2}{\sum_{j=1}^{n}\mu_j^2} \tag{15}$$

where, μ_i is i^{th} singular value.

Similarly, cumulative inertia explained up to the i^{th} principal axis is the sum of inertias explained up to that principal axis. This gives a measure of accuracy (or quality of representation) of the lower dimensional approximation. Although several mathematical criteria do exist for selecting the number of principal axis, there is no generally fixed criterion proposed to determine how many principal axes should be retained.

4. PROCESS MONITORING USING CA

Correspondence analysis has been used to build statistical models for ecological problems, study of vegetation habit of species, social networks, etc. Here we propose to build the statistical model for the plant data pertaining to normal operation using CA. As discussed earlier, CA takes joint row-column association into account while decomposing the χ^2 value. CA has also been shown to give better aggregation and clustering (Detroja *et al.*, 2005). CA also scores over PCA, which assumes statistical independence of samples (rows), as well as DPCA, which requires augmentation of the data matrix.

Once the statistical model is built from the normal operation data, the next task is to define control limits which can be used for the purpose of online statistical process monitoring of the plant. Motivated by Q and T^2 statistics used in PCA and DPCA, we defined here similar statistics for CA.

For online process monitoring, when a new measurement arrives, it is projected onto the PCs to obtain the new row scores (co-ordinates). The new measurement vector $\mathbf{x}$ is given by

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}^T \tag{16}$$

The row sum of this measurement vector, r is given by

$$r = \sum_{i=1}^{m} x_i \tag{17}$$

and the new row scores can be obtained as

$$\mathbf{f} = \left[\frac{1}{r}\mathbf{x}^T\mathbf{G}\mathbf{D}_\mu^{-1}\right]^T \tag{18}$$

4.1. T^2 statistic for CA

Hotelling's T^2 statistic effectively captures normal operating region for the multivariate data in PCA. For the statistical models that are built using CA, a similar statistic can be used to characterize the normal plant behavior. The T^2 value for CA model is defined as in Equation (19).

$$\mathrm{T}^2 = \mathbf{f}^T\mathbf{D}_\mu^{-2}\mathbf{f} \tag{19}$$

where, $\mathbf{D}_\mu$ contains first k-largest singular values, which were retained.

Control limit for the T^2 statistic based on CA, follows from the Equation (4).

4.2. Q statistic for CA

As explained earlier, monitoring the plant using only T^2 statistics is not adequate for fault detection, as it only monitors the variation along the principal axes which were retained in the statistical CA model. Any significant deviation in the direction of $n-k$ PCs (corresponding to smallest singular values), is also indicative of a fault.

The value of Q statistic for CA is defined as in Equation (20).

$$Q = \left[\mathbf{Bf} - \left(\frac{1}{r}\mathbf{x} - \mathbf{c}\right)\right]^T \left[\mathbf{Bf} - \left(\frac{1}{r}\mathbf{x} - \mathbf{c}\right)\right] \quad (20)$$

The control limit for the Q statistic is chosen as 95% confidence limit from the normal operating residual values.

Correspondence analysis, along with the statistics defined here, can be very useful in fault detection. In the next section, we demonstrate the usefulness of CA for fault detection and compare the performance of statistics based on CA, PCA and DPCA.

5. APPLICATION TO TENNESSEE EASTMAN CHALLENGE PROBLEM

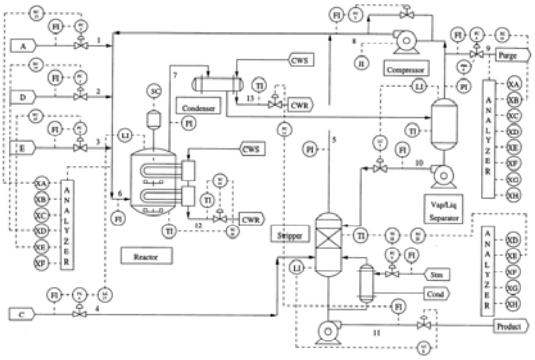


Figure 1: A diagram of the Tennessee Eastman process

The Tennessee Eastman process proposed by Downs and Vogel (1993) has been a benchmark problem for plant-wide control strategy and fault detection (Russell *et al.*, 2000). The test problem is based on an actual chemical process where only the components, kinetics and operating conditions were modified for proprietary reasons. Figure 1 shows a diagram of the process. The simulation code allows 21 pre-programmed major process disturbances as shown in Table 1. The plant-wide control structure recommended in Lyman and Georgakis (1995) and was used by Russell *et al.* (2000) for their study of fault detection using PCA and DPCA was used to generate the closed loop simulated process data for each fault.

The statistical models were built from the normal operation data consisting of 500 samples. All manipulated and measurement variables except the agitation speed of the reactor's stirrer for a total of 52 variables were used. The data was sampled every 3 minutes. Twenty-one testing sets were generated using the pre-programmed faults (IDV 1-21).

The normal operation data was used to build statistical model from PCA, DPCA and CA. Data compression is an important aspect of multivariate statistical tools. The number of PCs to be retained in PCA can be determined via several criteria such as cross validation or scree test. Earlier work (Russell et al., 2000) retained 11 PCs which explained approximately 55% of the variance in the data. When the analysis was done using CA, it was found that when 12 PCs were retained 96.65% of the total inertia was effectively captured. It should be noted here that these values of variance explained by PCA and inertia explained by CA can not be directly compared. Nevertheless, the representation given by CA would appear to be better as through modeling of the row-column associations it is better able to capture inter-relationships between variables and samples.

The objective of the fault detection technique is that it should be independent of the training set, sensitive to all the possible faults of the process, and prompt towards the detection of the fault. Since the fault alarms are inevitable, an out-of-control value of a statistic can be the result of a fault or of a false alarm. In order to decrease the rate of false alarms, a fault can be indicated only when several consecutive values of a statistic have exceeded the threshold. In this study, the fault is indicated only when six consecutive statistic values have exceeded the control limit, and the detection delay is recorded as the first time instance in which the threshold was exceeded. This was done exactly in accordance with what has been reported by Russell *et al.* (2000) so that results can be compared. The missed detection rates for faults 3, 9, and 15 were found to be fairly high, because no observable change in the mean or the variance could be detected by visually comparing the plots of each associated observation variable (Russell *et al.*, 2000). Therefore, these faults are not considered when comparing the methods.

Table 1: Process faults for the Tennessee Eastman process simulator

Fault	Description	Type
IDV(1)	A/C Feed ratio	Step
IDV(2)	B component	Step
IDV(3)	D feed temperature	Step
IDV(4)	Reactor cooling water (RCW) inlet temperature	Step
IDV(5)	Condenser cooling water (CCW) inlet temperature	Step
IDV(6)	A feed loss	Step
IDV(7)	C header pressure loss	Step
IDV(8)	A, B, C feed component	Random
IDV(9)	D feed temperature	Random
IDV(10)	C feed temperature	Random
IDV(11)	RCW inlet temperature	Random
IDV(12)	CCW inlet temperature	Random
IDV(13)	Reactor kinetics	Slow drift
IDV(14)	RCW valve	Sticking
IDV(15)	CCW valve	Sticking
IDV(16)	Unknown	
IDV(17)	Unknown	
IDV(18)	Unknown	
IDV(19)	Unknown	
IDV(20)	Unknown	
IDV(21)	The valve for Stream 4 was fixed (steady state position)	Constant position

The detection delays (in minutes) for all 18 faults (excluding fault 3, 9 and 15), are tabulated in Table 2. Statistic having minimum detection delay is shown

bold faced. All faults could be detected by the statistics defined based on CA. It can also be seen that the Q and T^2 statistics based on CA performed better as compared to statistics based on PCA. In can also be seen that the CA statistics has also performed better than DPCA statistics, which is expected to perform (and have performed) better than PCA statistics. CA based Q statistic is relatively faster in detecting faults when compared with statistics generated via PCA and DPCA. An important observation needs to be made in relation to Fault 19. As seen in Table 2, this fault was not detected by any other statistic except the Q-DPCA and Q-CA. however, even here, CA is seen to detect the fault much more rapidly than the DPCA (30 v/s 246 minutes respectively). Detection delays are also seen to be reduced considerably for other fault cases as well. The false alarms were also fewer in CA when compared to PCA (results are not included due to brevity).

Table 2: Detection delays (in minutes)

Fault	PCA Q	PCA T^2	DPCA Q	DPCA T^2	CA Q	CA T^2
IDV(1)	9	21	15	18	**6**	21
IDV(2)	36	51	39	48	**24**	36
IDV(4)	9	--	**3**	453	**3**	--
IDV(5)	**3**	48	6	6	21	45
IDV(6)	**3**	30	**3**	33	**3**	**3**
IDV(7)	**3**	**3**	**3**	**3**	**3**	**3**
IDV(8)	60	69	63	69	**24**	63
IDV(10)	147	288	150	303	**75**	171
IDV(11)	33	912	21	585	**15**	567
IDV(12)	24	66	24	9	**6**	69
IDV(13)	111	147	120	135	**108**	135
IDV(14)	**3**	12	**3**	18	**3**	--
IDV(16)	591	936	588	597	**27**	84
IDV(17)	75	87	**72**	84	87	711
IDV(18)	**252**	279	**252**	279	261	303
IDV(19)	--	--	246	--	**30**	--
IDV(20)	261	261	252	267	**210**	252
IDV(21)	855	1689	858	1566	**717**	1548

6. CONCLUSION

A new approach to fault detection based on Correspondence Analysis was proposed in this paper. New statistics based on CA, which are similar to Q and T^2 statistics of PCA, were also defined. The Tennessee Eastman process simulation was used to compare the proposed approach to fault detection using CA against conventional PCA and Dynamic PCA.

The process model representation in CA is better as it takes joint row-column association into account without increasing the number of columns in the data. The simulation study also revealed that all the faults in Tennessee Eastman process could be detected. Detection delays for fault detection are significantly reduced for most of the faults when compared with PCA and DPCA statistics.

REFERENCES

Chiang L. H., Russell E. L. & Braatz R. D. (2000), Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis, *Chemometrics & Intelligent Laboratory Systems*, **50**, 243-252.

Detroja K. P., Gudi R. D. & Patwardhan S. C. (2005), Fault detection and isolation using correspondence analysis, *Proceedings of 16th IFAC World congress, Prague*, 2005.

Ding C., He X., Zha H. & Simon H. (2002), Unsupervised Learning: Self-aggregation in Scaled Principal Component Space, *6th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2002)*, **2431**, 112-124.

Downs J. J. & Vogel E. F. (1993), A plant-wide industrial process control problem, *Computers and Chemical Engineering*, **17** (3), 245-255.

Greenacre M. J. (1984), *Theory and Application of Correspondence Analysis*, Academic Press Inc., London.

Greenacre M. J. (1993), *Correspondence Analysis in Practice*, Academic Press, London.

Hardle W. & Simar L. (2003), *Applied Multivariate Statistical Analysis*, Chapter 13, MD Tech, Berlin.

Jackson J. E. and Mudholkar G. S. (1979), Control procedures for residuals associated with Principal Component Analysis, *Technometrics*, **21**, 341--349.

Ku W., Storer R. H. & Georgakis C. (1995), Disturbance Detection & Isolation by dynamic principal components analysis, *Chemometrics & Intelligent Laboratory Systems*, **30**, 179-196.

Lyman P. R. and Georgakis C. (1995), Plant-wide control of the Tennessee Eastman problem, *Computers and Chemical Engineering*, **19**(3), 321-331.

Miller, P., Swanson, R. and Heckler, C. (1998), Contribution Plots: A Missing Link in Multivariate Quality Control, *Appl. Math. and Comp. Sci.*, **8**(4), 775–792.

Nomikos, P & MacGregor J.F. (1995), Multivariate SPC charts for monitoring batch processes, *Technometrics*, **37**(1), 44-59.

Russell E. L., Chiang L. H. & Braatz R. D. (2000), Fault detection in industrial processes using canonical variate analysis and dynamic principal components analysis, *Chemometrics and Intelligent Laboratory Systems*, **51**, 81-93.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

USING SUB-MODELS FOR DYNAMIC DATA RECONCILIATION

Luc Lachance[*], André Desbiens[*], Daniel Hodouin[**]

LOOP (Laboratoire d'observation et d'optimisation des procédés – Process observation and optimization laboratory)
[*] Department of Electrical and Computer Engineering
[**] Department of Mining, Metallurgical and Materials Engineering
Université Laval, Pavillon Pouliot
Québec City (Québec), Canada G1K 7P4
E-mail: desbiens@gel.ulaval.ca

Abstract: The optimal approach for dynamic data reconciliation consists in using a complete and exact process model. Unfortunately, such a model is difficult to obtain in industrial practice. Through an example, several observers based on static, stationary and dynamic sub-models are designed and compared to the optimal approach. The comparisons illustrate that, for the given conditions, static observers generally lead to estimates that are less precise than the measurements. Stationary observers are slightly more precise than static observers but they obviously lack the power of temporal redundancy offered by dynamic models. Deterministic dynamic sub-models, that do not include all physical parameters (thus relatively easy to obtain), which stochastic models are added to, are shown to give good performances.

Keywords: Observers, Singular systems, Measurement noise, Kalman filters, Stochastic modelling.

1. INTRODUCTION

All actions taken to optimize or control a process should be based on the best estimates available for present and past states of the process. Sensors provide measurements of only some of the states and measurement errors are inevitably present. Measurement errors can be gross errors or random measurement noises. Only the later will be addressed in this paper. The objective of data reconciliation is to provide estimates of unmeasured states and to reduce the effects of measurement noise on the measured states. The estimates calculated by data reconciliation must at least satisfy reliable physical constraints such as the equations of mass or energy conservation.

In industry, by far the most popular technique is static data reconciliation which is usually based on static mass conservation as first proposed by Kuehn and Davidson (1961). The main reason for its popularity is certainly that the process modeling is simplified by using only static mass balancing. Several papers have been written on this topic and many references can be found in review papers and books (Crowe, 1996; Romagnoli and Sanchez, 1999; Narasimhan and Jordache, 2000). The fact that applying static data reconciliation on a real-time basis to a dynamic process could be worse than using directly the measurements, as illustrated by Almasy (1990), does not seem to limit the extensive use of this technique in industry. Good results may indeed be expected with static reconciliation for time-averaged data applications such as real-time optimization or metallurgical accounting. However, control applications require dynamic data reconciliation.

On the other hand, although the modeling part is the most difficult task in practice for dynamic data reconciliation, the assumption made by most of the authors is that the exact process model is known. Indeed, the papers are often focused on introducing (Liebman, et al., 1992) or comparing dynamic reconciliation methods (Ramamurthi, et al., 1993). However model mismatch could lead to severely biased estimates (Dochain, 2003). Because of this lack of observer robustness, it is probably better in many practical cases to use a reliable sub-model instead of a complete model with uncertain parameters. The static mass balance is the most popular sub-model but not appropriate for true dynamic applications. Fortunately, other sub-models such as stationary and dynamic sub-models remain very simple while being frequently adequate for dynamic reconciliation.

In their paper, Darouach and Zasadzinski (1991) proposed the use of a generalized Kalman filter for performing dynamic mass balancing. This generalized state space representation and the associated filter and smoother algorithms allow using sub-models for dynamic data reconciliation. They will be extensively used in this paper.

The above process observer is based on the basic lumped dynamic mass conservation equation, without any attempt to model either mixing or material transportation mechanisms or chemical reactions kinetics. Unfortunately, dynamic data reconciliation relying only on this equation usually result in poor filtering ability, and requires the measurement of the process inventories for warranting process observability. Almasy (1990) proposed to use the same dynamic mass conservation equation, while adding the dynamic empirical constraint that species flows behave as random walks. As in the paper by Darouach and Zasadzinski (1991), the mixing and kinetic mechanisms were not addressed in this paper, and the inventory was supposed measured. Furthermore, the problem of tuning the random walk variances is not discussed in the paper. To our knowledge, Stanley and Mah (1977) were the first to introduce the idea of random walks in data reconciliation by coupling them to steady-state conservation equations. A qualitative discussion on how to tune the random walks can be found in their paper.

In this paper, the simulated plant is a continuously stirred tank reactor (CSTR) where only mass balance phenomena are considered. In contrast to most papers, the feed concentration is not supposed to be a deterministic manipulated variable but is defined as a disturbance modeled by a stationary stochastic process. Using the simulated plant, the study objective is to compare thirteen observers based on unbiased sub-models. The benchmark observer is designed from a model identical to the simulator. Four steady-state, four stationary and four dynamic observers are compared to the benchmark. The plant models are of varying complexity including or not including mixing or kinetics information; and adding or not adding empirical stochastic models for stream flow rates modelling.

2. A CONTINUOUSLY STIRRED TANK REACTOR (CSTR)

A CSTR process is simulated to compare the various observers. Heat balance phenomena are not considered, however this would not impair the validity of the presented qualitative discussions and conclusions. From this simple CSTR model, several sub-models are extracted, leading to a variety of observers.

The chemical reaction taking place in the CSTR is a first order irreversible reaction $A \rightarrow B$. The Euler discretization of the differential mass balance equations leads to:

$$\begin{aligned} D_A(k) &= c_{Ai}(k) - c_{Ai}(k-1) \\ &= -K_0 c_{Ai}(k-1) + \frac{Q}{V}\left(c_{Af}(k-1) - c_{Ao}(k-1)\right) \end{aligned} \quad (1)$$

$$\begin{aligned} D_B(k) &= c_{Bi}(k) - c_{Bi}(k-1) \\ &= K_0 c_{Ai}(k-1) - \frac{Q}{V} c_{Bo}(k-1) \end{aligned} \quad (2)$$

where Q is assumed to be constant and perfectly known input and output flowrate; V is the known volume of the tank; c_{Af}, c_{Ai} and c_{Ao} are the variations of concentration of species A around their nominal value respectively in the feed flow, the tank inventory and the output flow; the same notation is used for species B which is not present in the feed (c_{Bi} and c_{Bo}); D_A and D_B are the accumulation rates for species A and B; K_0 is the rate constant of the reaction. Table 1 gives the numerical values of this CSTR simulator.

In addition to mass conservation Equations (1) and (2), perfect mixing is assumed:

$$c_{Ai}(k) = c_{Ao}(k) \quad (3)$$

$$c_{Bi}(k) = c_{Bo}(k) \quad (4)$$

It is straightforward to verify that inserting (3) and (4) into (1) and (2) leads to usual equations for a CSTR. The feed concentration is defined by the following stationary stochastic process:

Table 1. CSTR numerical values

	V	10 L	α	0.9
	Q	2 L/s	σ_ξ^2	0.1
	K_0	1 s^{-1}	σ_{Af}	13.06%
Nominal value	$\overline{C}_{Af}$	5 mol/L	σ_{Ao}	13.32%
	$\overline{C}_{Ao}$	0.833 mol/L	σ_{Bo}	10.81%
	$\overline{C}_{Bo}$	4.167 mol/L		

$$c_{Af}(k) = \frac{\alpha}{1-\alpha Z^{-1}}\xi(k) \qquad (5)$$

where $\xi(k)$ is a zero mean Gaussian white noise of variance σ_ξ^2 that generates variations on the five concentrations. Table 1 lists the standard deviations of these variations (σ_{Af} , σ_{Ao} and σ_{Bo}) expressed in percentage of the corresponding nominal values.

The five measurements are obtained by adding a Gaussian white noise to each concentration. The relative standard deviation of each independent measurement noise is 5% of the corresponding nominal value (thus defining the matrix Σ introduced in Section 3).

Having σ_{Af} , σ_{Ao} and σ_{Bo} larger than the measurement noise standard deviation contributes to obtaining static reconciliation estimates less precise than measurements, as discussed by Almasy (1990). This is a strong incentive for using dynamic data reconciliation.

The objective of the paper is to compare the performances of thirteen observers, using always the same measurement information (the measured values of the five concentrations), while varying the information content in the observer model. The minimum information used is either the steady-state or the dynamic mass conservation constraints. It can be enriched by the information on the reaction kinetics and/or by the information on the mixing properties, and/or by empirical information on the stochastic behaviour of the species flows on the two streams. The observer performances are always compared to the optimal filter based on the complete exact model used for simulation. All the observers are derived using the generalized Kalman filter which is first succinctly described in the next section.

3. THE GENERALIZED KALMAN FILTER

Since all the state variables are measured in this work, the following generalized state space representation is used to design the observers:

$$E\,x(k) = A\,x(k-1) + w(k) \qquad (6)$$

$$y(k) = x(k) + v(k) \qquad (7)$$

where E is usually a singular matrix and

$$x(k) = \left[c_{Af}(k) \;\; c_{Ai}(k) \;\; c_{Bi}(k) \;\; c_{Ao}(k) \;\; c_{Bo}(k)\right]^{\mathrm{T}} \qquad (8)$$

The covariance matrices W and Σ respectively define the properties of the process noise w and the measurement noise v. The corresponding filtering and smoothing algorithms (Darouach and Zasadzinski, 1991) are:

$$\hat{x}(k+1/k+1) = \Sigma E^T \Omega(k) A\hat{x}(k/k) + \left(I - \Sigma E^T \Omega(k)\, E\right) y(k+1) \qquad (9)$$

$$\hat{x}(j/k+1) = \hat{x}(j/k) + P(j/k)\, A^T \Omega(k)\left(E\, y(k+1) - A\hat{x}(k/k)\right) \qquad (10)$$

$$P(k+1/k+1) = \Sigma - \Sigma E^T \Omega(k) E \Sigma \qquad (11)$$

$$P(j/k+1) = P(j/k)\, A^T \Omega(k) E \Sigma \qquad (12)$$

where $\hat{x}(j/k+1)$ is the estimate of the vector x at time j based on the knowledge of measurements up to time $k+1$ ($j<k+1$).

The matrix $\Omega(k)$ is not defined as in Darouach and Zasadzinski (1991) since a process noise is added to the state Equation (6). It is defined by :

$$\Omega(k) = \left(W + E\Sigma E^T + AP(k/k)A^T\right)^{-1} \qquad (13)$$

4. THE VARIOUS OBSERVERS

Thirteen different observers are designed. Only the observer described in Section 4.1 uses the complete and exact information about the process. In this sense, it defines a benchmark for all other observers that rely on sub-models providing incomplete information about the process. The upper part of Table 2 summarizes the model equations used for each observer.

The mass conservation constraint, which is the minimal information used in the observers is :

$$D_A(k) + D_B(k) = \frac{Q}{V}\left(c_{Af}(k-1) - c_{Ao}(k-1) - c_{Bo}(k-1)\right) \qquad (14)$$

It is obtained by adding (1) and (2) and states that the total number of moles must be conserved. An important practical advantage of this equation is that it does not assume any mechanism for mixing and reaction kinetics.

4.1 Observer based on the complete exact process model

The smoothing obtained by this observer (observer 1 in Table 2) corresponds to the best possible results since it uses the same model as the one being used for process simulation, i.e. (1) to (5). The parameter K_0 is supposed to be exactly known (which is the case for all observers requiring Equations (1) and (2)).Measurement redundancy is provided by the fact that the process is perfectly mixed and that the inventory and output concentrations are independently measured.

The generalized state space representation (6) and (7) for the observer model (1) to (5) is obtained with:

$$E = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (15)$$

$$A = \begin{bmatrix} \frac{Q}{V} & 1-K_0 & 0 & -\frac{Q}{V} & 0 \\ 0 & K_0 & 1 & 0 & -\frac{Q}{V} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \alpha & 0 & 0 & 0 & 0 \end{bmatrix} \quad (16)$$

$$W = diag([0 \quad 0 \quad 0 \quad 0 \quad 0.081]) \quad (17)$$

One can observe that, in this particular case, the matrix E is non singular. Defining matrices *E, A* and *W* to put the model equations in a form corresponding to the generalized state space representation will be omitted for other observers since the procedure is very similar.

4.2 Steady-state observers

The model equations used by the four static observers are detailed in Table 2 (observers 2.1 to 2.4). The fundamental characteristic of all static observers is that accumulation rates D_A and D_B are set to zero in (1), (2) or (14). The resulting model for each observer can be described by:

$$E\,x(k) = 0 \quad (18)$$

i.e. $A = 0$ and $W = 0$ in (6). Since $A = 0$, the generalised Kalman filter Equations (9) to (13) does not provide smoothing but only filtering, due to the absence of temporal redundancy.

Instantaneous static observers have been considered in this paper and thus horizon-based static observers are not studied.

4.3 Stationary observers based on the dynamic conservation equation

For these four observers (observers 3.1 to 3.4 in Table 2), the accumulation rates in Equations (1), (2) or (14) are considered as strongly stationary stochastic process instead of being set to zero as for static observers. The resulting models all have the following structure

$$E\,x(k) = w(k) \quad (19)$$

where the covariance matrix W appears as a tuning parameter. The best choice for W is

$$W = E\,X\,E^T \quad (20)$$

where X is the covariance matrix of the generalized state vector x. In practice, calculating (20) is impossible since only the measurements are available (not x). Work is underway to propose practical solutions to this issue. Nevertheless, for fair comparison purposes, W will be calculated using (20). As with static observers, smoothing is not possible, when using only instantaneous observations.

4.4 Observers based only on the dynamic conservation constraint

For setting comparisons to dynamic observers defined in 4.5, a dynamic data reconciliation similar to the one proposed by Darouach and Zasadzinski (1991), i.e. based only on the dynamic conservation Equation (14) (thus $W = 0$), is tested (observer 4 in table 2).

4.5 Dynamic conservation equation with stochastic models for molar flows

These observers are build using the deterministic Equation (14) combined with the following empirical stochastic descriptions of the molar flows:

$$C_{Af}(k) = \beta\, C_{Af}(k-1) + w_2(k) \quad (21)$$

$$C_{Ao}(k)+C_{Bo}(k) = \gamma\left(C_{Ao}(k-1)+C_{Bo}(k-1)\right)+w_3(k) \quad (22)$$

The resulting model again corresponds to (6) with:

$$W = diag([0 \quad W_2 \quad W_3]) \quad (23)$$

The parameters β, γ, W_2 and W_3 have to be determined. Three different cases have been considered (observers 5.1 to 5.3 in Table 2). The first case, inspired by the work of Almasy (1990), assumes random walk behaviours, i.e. $\beta = \gamma = 1$. The second case assumes white noise behaviours, i.e $\beta = \gamma = 0$. For theses two cases, since the parameters β and γ are set, Equations (21) and (22) allow to compute w_2 and w_3 from a set of simulated data, without measurement noise, and therefore to tune W_2 and W_3. The third case uses an identification procedure to estimate β (0.90) and γ (0.94) from the true state signals. The identification residuals provide W_2 (0.081) and W_3 (0.0076). Again, the procedure to obtain the parameter values is based on unavailable true states, but this is an appropriate procedure for the objective to obtain the best possible tuning for each observer.

5. RESULTS AND ANALYSIS

Since the observers are not biased, they can be compared by evaluating the standard deviation of the

relative estimation error (second part of Table 2). For each observer, 2500 samples were simulated.

As already mentioned, it is not surprising that, for this combination of process and instrumentation, steady-state observers are in general worse than measurements. The exception is observer 2.2 which provides good estimates for four variables. This performance is explained by the redundancy provided by the inventory measurements and the perfect mixing constraints. When comparing observers 2.3 and 2.4 to observer 2.2, it can be concluded that in this case no gain is made by adding steady-state kinetic modeling even with the exact knowledge of parameter K_0. Furthermore, all the static observers are unable to increase the precision of the feed concentration probably because its dynamics are too fast to be reconciled in real time with steady state modeling.

Stationary observers produce estimates with better accuracy than measurements, but some variables are significantly less filtered than others. Stationary observer 3.4 which uses the most complete information among the stationary observers gives the best results. This is different from the steady-state observer 2.4 behaviour, because of the adequate tuning of the stationary accumulation rate variances. Nevertheless, the best performance of any stationary observer still remains far from the benchmark, mainly because only instantaneous filtering is possible with stationary observers. To benefit from temporal redundancy and to make smoothing possible, dynamic observers need to be introduced.

The dynamic observer 4 exhibits performances equivalent to those of stationary observers 3.1 and 3.3, because the information contained in the dynamic mass conservation constraint is equivalent to the statistical information on the accumulation rates in the stationary observers. However the dynamic observer 4 is worse than the best stationary observer 3.4 and also worse than 3.2, even if the later ones do not use temporal redundancy. The reason is related to the redundancy created by the perfect mixing constraints.

This gives motivation for using the additional modeling Equations (21) and (22) which are empirical equations that remain to be tuned. Observer 5.1 is indeed doing significantly better than observer 4 because the random walk assumption is a reasonable approximation of the autocorrelated behaviours of the output and input signals. Observer 5.2 is not significantly better than observer 4, because the white noise assumption ignores the temporal correlation of the flowrates variations. Observer 5.3 is a little more precise than observer 5.1. This small improvement is explained by a better selection of β and γ, which are slightly smaller than one, the value used for random walks.

6. CONCLUSION

Several sub-models of the same process were proposed to design steady-state, stationary and dynamic data reconciliation algorithms.

Steady-state observers may produce estimation errors which are larger than measurement errors. This is the consequence of neglecting process dynamics, when the process variable variance due to dynamics is larger than the variance induced by measurement errors, a case that was simulated in the present study. Unfortunately, this is the usual industrial situation, thus precluding the use of such algorithms for real time data reconciliation, although it seems to be an increasingly popular approach in industry.

Stationary observers, which allow mass accumulation rates to statistically deviate from zero, are an efficient alternative to steady-state observers for real time data reconciliation. They produce estimates that are more reliable than measured values, while requiring only a rough estimate of the accumulation rate variances. Although they are noticeably less efficient than filters based on full process models, they are simple to build and tune.

Dynamic observers, based on the minimal dynamic information consisting of the mass conservation constraints (observer 4), are not significantly better than stationary observers, because of the low level of information redundancy. However they are better than steady-state observers forcing static mass conservation (observer 2.1).

The dynamic filter can be improved by adding empirical information to the dynamic mass balance constraint, such as stochastic models of time evolution of flow characteristics evolution. The empirical models can be identified from the experimental data, or simply assumed to be random walks.

The stationary or dynamic proposed observers are all significantly less precise than the optimal observer using the full phenomenological model of the process (observer 1). However, great care must be taken if considering the design of an observer based on a complete dynamic model, since biases may result from badly identified model parameters. The main advantage of the proposed observers is a modeling effort that is considerably less important. This advantage becomes even more important in real applications where several units, such as the one presented in this paper, are present in the plants. Indeed, the observers are mainly based on conservation equations defined by few or no parameters, combined with stochastic modeling that may be tuned from experiments.

The present work was limited to linear systems in the Kalman filtering framework. Similar conclusions could probably be drawn for nonlinear dynamic data reconciliation based on nonlinear programming.

Table 2. The observers and their performances

		Benchmark	Steady-state				Stationary				Dynamic			
		1	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	4	5.1	5.2	5.3
Observation model	A, B species balance - Equations (1) and (2)	X			X*	X*			X**	X**				
	Mixing – Equations (3) and (4)	X		X		X		X		X				
	Mole conservation - Equation (14)		X*	X*			X**	X**			X	X	X	X
	Feed generator – Equation (5)	X										X	X	X
	Stochastic flows – Equations (21) and (22)											X	X	X
Relative estimation error	A feed	2.23	5.82	6.60	5.15	5.66	4.46	4.31	3.97	4.07	4.72	3.30	4.53	3.30
	A inventory	2.63	4.99	3.54	5.19	4.56	4.96	3.52	4.17	3.14	4.91	4.90	4.85	4.90
	B inventory	0.89	5.03	3.95	5.01	5.53	5.06	3.35	4.91	3.25	2.12	1.96	2.10	1.95
	A output	2.63	5.18	3.54	4.92	4.57	4.99	3.52	5.08	3.14	5.05	4.92	4.95	4.92
	B output	0.88	5.52	3.95	6.10	5.53	4.53	3.35	4.33	3.25	4.87	2.38	4.50	2.35
	Sum of relative variances***	20.37	141.38	99.83	139.98	134.88	115.52	65.80	101.81	57.41	100.10	68.61	93.21	68.43

* Accumulation rates D_A and/or D_B are set to zero.
** Accumulation rates D_A and/or D_B are described by a zero mean stochastic processes.
***The measurement relative estimation error is 5.00 % for each variable, and leads to a sum of relative variances of 125.

REFERENCES

Almasy, G.A. (1990), Principles of dynamic balancing, *AIChE J.* **36**, 1321-1330.

Crowe, C.M. (1996), Data reconciliation—progress and challenges, *Journal of Process Control* **6**, 89-98.

Darouach, M. and M. Zasadzinski (1991). Data reconciliation in generalized linear dynamic systems, *AIChE J.* **37**(2), 193-201.

Dochain, D. (2003), State and parameter estimation in chemical and biochemical processes: a tutorial, Journal of Process Control **13**, 801-818.

Kuehn, D.R. and H. Davidson (1961). Computer control II: Mathematics of control. *Chem. Eng. Prog.* **57** (6), 44-47.

Liebman, M.J., T.F. Edgar and L.S. Lasdon (1992), Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques, *Comp. Chem. Eng* **16**, 963-986.

Narasimhan, S. and C. Jordache (2000), *Data reconciliation & gross error detection: an intelligent use of process data,* Gulf Pub. Co., Houston.

Ramamurthi, Y., P.B. Sistu and B.W. Bequette (1993), Control-relevant dynamic data reconciliation and parameter estimation, *Comp. Chem. Eng* **17** (1), 41-59.

Romagnoli, J.A. and M.C. Sanchez (1999), *Data Processing and Reconciliation for Chemical Process Operations*, Academic Press.

Stanley, G.M. and R.S.H. Mah (1977), Estimation of flows and temperatures in process networks, *AIChE J.* **23**(5), 642-650.

Session 6.1

Modeling and Identification

Short and Long Timescales in Recycles

H. A Preisig
Norwegian University of Science and Technology

Finite Automata from First-Principle Models: Computation of Min and Max Transition Times

H. A Preisig
Norwegian University of Science and Technology

Neural Modeling as a Tool to Support Blast Furnace Ironmaking

F. Tadeu, P. de Medeiros, A. Pitasse da Cunha and A. M. F. Fileti
Companhia Siderúrgica Nacional
University of Campinas
MetalFlexi

An Inverse Artificial Neural Network Based Modelling Approach for Controlling HFCS Isomerization Process

M. Yuceer and R. Berber
Ankara University

An Algorithm for Automatic Selection and Estimation of Model Parameters

A. R. Secchi, N. S. M. Cardozo, E. Almeida Neto and T. F. Finkler
Universidade Federal do Rio Grande do Sul

Rigorous and Reduced Dynamic Models of the Fixed Bed Catalytic Reactor for Advanced Control Strategies

E. C. Vasco de Toledo, J. M. F. da Silva, J. F. da C. A. Meyer, and R. M. Filho,
State University of Campinas

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

CONTROL ORIENTATED B-SPLINE MODELLING OF A DYNAMIC MWD SYSTEM

Hong Yue [*,1] **Hong Wang** [**] **Liulin Cao** [***]

[*] *School of Chemistry, The University of Manchester*
[**] *Control Systems Centre, The University of Manchester*
[***] *Department of Chemical Automation,*
Beijing University of Chemical Technology

Abstract: A detailed dynamic model has been developed for the molecular weight distribution (MWD) of styrene bulk polymerization in a continuous stirred tank reactor (CSTR). The moment techniques are applied to formulate the MWD parameters based on the Schultz-Zimm distribution. In order to provide a general model for MWD control, the B-spline approximation has been introduced into the dynamic MWD modelling and the scanning least-square algorithm has been used for parameter estimation of the B-spline weights model. Under simulation environment, this model has been proved to be efficient for feedback MWD control.

Keywords: Molecular weight distribution (MWD), Dynamic model, B-spline approximation, Probability density function (PDF) control

1. INTRODUCTION

The B-spline neural network has been considered as an efficient tool for modelling the output probability density function (PDF) because it provides a general form in describing arbitrary continuous functions. Using the B-spline approximation, the output PDF will be described by the weights of the pre-specified basis functions. Dynamic characteristics of the weights vector can be developed from the data pairs of control input and output PDF so as to formulate the B-spline model for PDF control. In most of the previous works on output PDF modelling and control, it is normally assumed that the weights dynamics are known or the weights vector is available for the PDF approximation. This is partly because some of those works are concentrated on PDF controller design rather than B-spline modelling. It is also because that the B-spline modelling process itself is quite challenging considering the complexity of a dynamic PDF system. Many technical details have to be addressed carefully in order to guarantee the modelling efficiency. A scanning identification algorithm has been developed for the B-spline PDF modelling (Wang, 2000), however, it has been used mainly for static PDF systems or linear dynamic weights systems (Wang and Wang, 1998; Zhang and Yue, 2004). No work has been reported on B-spline modelling using the input and output PDF data from a nonlinear dynamic process so far. This motivates the endeavor of the work in this paper.

A molecular weight distribution (MWD) system has been taken as the case for study. The PDF data used for B-spline modelling are produced from the first-principle MWD model. Although the theory of B-splines is well-developed in approximation theory and linear control (Zhang *et al.*, 1997; Sun *et al.*, 2000; Kano *et al.*, 2003), to our knowledge, no applications to MWD systems have been reported except for a few works by

[1] Partially supported by the Outstanding Overseas Chinese Scholars Fund of Chinese Academy of Sciences (2004-1-4)

the authors (Yue *et al.*, 2004; Wang *et al.*, 2005). The MWD calculation in the previous works is not based on real dynamic models, only the static solution at different situations are considered. In this paper, the MWD model is developed from the polymerization reaction mechanisms with dynamic behaviors. Although the first-principle MWD model can be described by the well-known Schultz-Zimm distribution for this example, the model is further developed by B-splines. This is simply because a general-form MWD model is expected for the purpose of MWD control using PDF control strategies.

Notations

I initiator or its concentration $(mol \cdot L^{-1})$
I^{00} initial initiator concentration $(mol \cdot L^{-1})$
I^0 controlled initial initiator concentration $(mol \cdot L^{-1})$
K_d initiator decomposition rate constant (min^{-1})
K_i initiation rate constant $(L \cdot mol^{-1} \cdot min^{-1})$
K_p propagation rate constant $(L \cdot mol^{-1} \cdot min^{-1})$
K_{trm} chain transfer rate constant $(L \cdot mol^{-1} \cdot min^{-1})$
K_t termination rate constant $(L \cdot mol^{-1} \cdot min^{-1})$
M monomer or its concentration $(mol \cdot L^{-1})$
M^{00} initial monomer concentration $(mol \cdot L^{-1})$
M^0 controlled initial monomer concentration $(mol \cdot L^{-1})$
R_j live polymer of chain length j or its concentration $(mol \cdot L^{-1})$
R total concentration of live polymer radicals $(mol \cdot L^{-1})$
P_j dead polymer of chain length j or its concentration $(mol \cdot L^{-1})$
P total concentration of dead polymer $(mol \cdot L^{-1})$
T reaction temperature (K)
F total feed flow rate $(L \cdot min^{-1})$
V volume of reaction mixture (L)
θ average residential time (min)

2. POLYMERIZATION PROCESS

The process of interest is a styrene bulk polymerization reaction in a continuous stirred tank reactor (CSTR), in which styrene is the monomer for polymerization and azobisisobutyronitrile is used as the initiator. These two flows are injected into the CSTR with the input ratio defined as

$$c = \frac{F_M}{F_I + F_M} \tag{1}$$

where F_M is the flow of monomer and F_I is the flow of initiator. By changing c, the initial concentrations of the two main reaction species will be changed, which will change the output molecular weight distribution. To simplify the process, the reaction temperature is assumed to be kept constant during the control process.

The following free radical polymerization mechanisms are considered for the system.

- Initiation
  $I \xrightarrow{K_d} 2R^*$
  $R^* + M \xrightarrow{K_i} R_1$
- Chain propagation
  $R_j + M \xrightarrow{K_p} R_{j+1}$
- Chain transfer to monomer
  $R_j + M \xrightarrow{K_{trm}} P_j + R_1$
- Termination by combination
  $R_j + R_i \xrightarrow{K_t} P_{j+i}$

Accordingly, the mass balance equations are derived to be

$$\frac{dI}{dt} = (I^0 - I)/\theta - K_d I \tag{2}$$

$$\frac{dM}{dt} = (M^0 - M)/\theta - 2K_i I - (K_p + K_{trm})MR \tag{3}$$

$$\frac{dR_1}{dt} = -R_1/\theta + 2K_i I - K_p M R_1 + K_{trm} M(R - R_1) - K_t R_1 R \tag{4}$$

$$\frac{dR_j}{dt} = -R_j/\theta - K_p M(R_j - R_{j-1}) - K_{trm} M R_j - K_t R_j R \ \ (j \geq 2) \tag{5}$$

$$\frac{dP_2}{dt} = K_{trm} R_2 M + K_t R_1^2 - P_2/\theta \tag{6}$$

$$\frac{dP_j}{dt} = K_{trm} R_j M + \frac{K_t}{2} \sum_{l=1}^{j-1} R_l R_{j-l} - P_j/\theta \quad (j \geq 3) \tag{7}$$

where $\theta = V/F$ is the average residential time of the reactants in the CSTR. Denote

$$R = \sum_{j=1}^{\infty} R_j \tag{8}$$

$$P = \sum_{j=2}^{\infty} P_j \tag{9}$$

as the total concentrations of radicals and polymers, respectively, the following formulations can be established from (4) to (7)

$$\frac{dR}{dt} = -R/\theta + 2K_i I - K_t R^2 \tag{10}$$

$$\frac{dP}{dt} = -P/\theta + K_{trm} M(R - R_1) + \frac{K_t}{2} R^2 \tag{11}$$

R_1 in (11) can be ignored compared with R due to its low concentration, i.e.,

$$\frac{dP}{dt} = -P/\theta + K_{trm} M R + \frac{K_t}{2} R^2 \tag{12}$$

3. FIRST-PRINCIPLE MWD MODEL

3.1 Static MWD Model

The static solution to the concentrations of the reaction species can be derived from their dynamic equations. Denote

$$\alpha = 1 + \frac{K_{trm}}{K_p} + \frac{K_t R}{K_p M} + \frac{1}{K_p M \theta} \tag{13}$$

By taking the differential equations(2),(3), (10) and (12) to be zero, there are

$$I = \frac{I^0}{1 + K_d \theta} \tag{14}$$

$$R = \frac{-1/\theta + \sqrt{1/\theta^2 + 8 K_t K_i I}}{2 K_t} \tag{15}$$

$$M = \frac{M^0}{1 + (K_p + K_{trm}) R \theta} \tag{16}$$

$$P = \theta (K_{trm} M R + \frac{K_t}{2} R^2) \tag{17}$$

Similarly, from equations (4)-(7), the static concentrations of radicals and polymers are

$$R_1 = \frac{2 K_i I + K_{trm} M R}{K_p M \alpha} \tag{18}$$

$$R_j = \alpha^{-1} R_{j-1} = \alpha^{-(j-1)} R_1, \ (j \geq 2) \tag{19}$$

$$P_2 = \theta \left(K_{trm} M R_2 + K_t R_1^2 \right) \tag{20}$$

$$P_j = \theta \left(K_{trm} M R_j + \frac{K_t}{2} \sum_{l=1}^{j-1} R_l R_{j-l} \right), \quad (j \geq 3) \tag{21}$$

Substituting (19) into (20) - (21), and dividing (20) and (21) by the total concentration P, the normalized MWD at static state can be obtained to be

$$P_2 = \frac{\theta}{P} (\alpha^{-1} K_{trm} M R_1 + K_t R_1^2) \tag{22}$$

$$P_j = \frac{\theta}{P} \left(\alpha^{-(j-1)} K_{trm} M R_1 + \frac{j-1}{2} \alpha^{-(j-2)} K_t R_1^2 \right), (j \geq 3) \tag{23}$$

It can be seen that $\sum_{j=2}^{\infty} P_j = 1$. Therefore, the static MWD can be taken as a discrete probability density function of the chain length.

3.2 Dynamic MWD Model

For the dynamic MWD model, the distribution of P_j is not only a function of the chain length, but also a function of time. In this work, the moment method is introduced to setup the dynamic MWD description.

The moments of the number chain-length distributions of radicals and polymers are defined as

$$U_k = \sum_{j=1}^{+\infty} j^k R_j, \ \ k = 0, 1, 2, \cdots \tag{24}$$

$$Z_k = \sum_{j=2}^{+\infty} j^k P_j, \ \ k = 0, 1, 2, \cdots \tag{25}$$

It can be seen from (8) and (9) that $U_0 = R$ and $Z_0 = P$. Using the generation function technique, the differential equations of the leading moments for radicals are derived to be

$$\frac{dU_0}{dt} = -U_0/\theta + 2 K_i I - K_t U_0^2 \tag{26}$$

$$\frac{dU_1}{dt} = -U_1/\theta + 2 K_i I + K_p U_0 M - K_t U_0 U_1 + K_{trm} M (U_0 - U_1) \tag{27}$$

$$\frac{dU_2}{dt} = -U_2/\theta + 2 K_i I + K_p M (2 U_1 + U_0) - K_t U_0 U_2 + K_{trm} M (U_0 - U_2) \tag{28}$$

Similarly, the three leading moments of polymers are derived to be

$$\frac{dZ_0}{dt} = -Z_0/\theta + K_{trm} M U_0 + \frac{K_t}{2} U_0^2 \tag{29}$$

$$\frac{dZ_1}{dt} = -Z_1/\theta + K_{trm} M U_1 + K_t U_0 U_1 \tag{30}$$

$$\frac{dZ_2}{dt} = -Z_2/\theta + K_{trm} M U_2 + K_t U_0 U_2 + K_t U_1^2 \tag{31}$$

The mean and variance of the MWD are linked to the moments by

$$\mu = \frac{\sum_{j=2}^{+\infty} j P_j}{\sum_{j=2}^{+\infty} P_j} = \frac{Z_1}{Z_0} \tag{32}$$

$$\sigma^2 = \frac{\sum_{j=2}^{+\infty} (j - \mu)^2 P_j}{\sum_{j=2}^{+\infty} P_j} = \frac{Z_2}{Z_0} - \frac{Z_1^2}{Z_0^2} \tag{33}$$

Theoretically, an exact formulation of a molecular weight distribution requires countless number of moments, which is infeasible because of the computational load. An alternative method is to choose an appropriate distribution function to approximate the real MWD. For the polymer discussed in this work, the well-known Schultz-Zimm distribution is selected to describe the molecular weight distribution. It makes a simple analytical expression available for the scattering from the distribution. The normalized Schultz-Zimm distribution is defined by (Angerman, 1998)

$$f(n) = \frac{h^h n^{h-1} exp(-hn/M_n)}{M_n^h \Gamma(h)}, \ (n \geq 0) \tag{34}$$

where n is the chain length, h is the parameter indicating the distribution breadth, Γ is the gamma

function, M_n is the number average chain length which is defined as $M_n = Z_1/Z_0$. When $h = 1$, the Schultz-Zimm distribution reduces to the exponential Flory distribution, which is another commonly used distribution for MWD. The mean and variance of the Schultz-Zimm distribution are

$$\mu = \int_0^\infty nf(n)dn = M_n \tag{35}$$

$$\sigma^2 = \int_0^\infty (n-\mu)^2 f(n)dn = \frac{h+1}{h}M_n^2 - \mu^2 \tag{36}$$

By comparing (32), (33) with (35) and (36), the two parameters of the Schultz-Zimm distribution can be obtained to be

$$h = \frac{Z_1^2}{Z_0 Z_2 - Z_1^2} \tag{37}$$

$$M_n = Z_1/Z_0 \tag{38}$$

The calculation of the dynamic MWD can be divided into three steps:
(1) Get Z_0, Z_1, Z_2 from (2), (3), (26)-(31);
(2) Get h and M_n from (37) and (38);
(3) Formulate the MWD by (34).

4. DYNAMIC B-SPLINE APPROXIMATION

Although the dynamic MWD in this case can be described by the analytical Schultz-Zimm distribution function, it is not of a general form for the feedback PDF control scheme. Therefore, the B-spline approximation is introduced for the further model development. Consider a continuous PDF $\gamma(y, u_k)$ defined on $[a, b]$ interval, the linear B-spline neural networks can be used to approximate $\gamma(y, u_k)$ as:

$$\gamma(y, u_k) = \sum_{i=1}^{n} \omega_i(u_k) B_i(y) + e_0 \tag{39}$$

where u_k is the control input at sample time k; $B_i(y)(i = 1, \cdots, n)$ are the pre-specified basis functions defined on the interval $y \in [a, b]$; n is the number of the basis functions; $\omega_i(u_k)(i = 1, \cdots, n)$ are the expansion weighs; e_0 represents the approximation error which satisfies $|e| < \delta_1$ (δ_1 is a known small positive number). To simplify the expression, e_0 is neglected in the following. Denote

$$L(y) = \frac{B_n(y)}{\int_a^b B_n(y)dy} \tag{40}$$

$$c_i(y) = B_i(y) - L(y)\int_a^b B_i(y)dy, \quad i = 1, \cdots, n-1 \tag{41}$$

$$C(y) = [\, c_1(y), \quad c_2(y), \quad \cdots, \quad c_{n-1}(y)\,] \tag{42}$$

$$V_k = [\, \omega_1(u_k), \quad \omega_2(u_k), \quad \cdots, \omega_{n-1}(u_k)\,]^T \tag{43}$$

the static B-spline PDF model (39) can be represented in a compact form as

$$\gamma(y, u_k) = C(y)V_k + L(y) \tag{44}$$

Equation (44) is the static PDF model approximated by the B-spline neural networks, in which $C(y)$ and $L(y)$ are known when the basis functions are chosen. Denote

$$f_k(y) = \gamma(y, u_k) - L(y) \tag{45}$$

and consider the linear dynamics of the weights vector, the output PDF can be described by the following state-space B-spline model:

$$f_k(y) = C(y)V_k \tag{46}$$

$$V_{k+1} = EV_k + Fu_k \tag{47}$$

Here E and F are model parameter matrices. $f_k(y)$ can be further represented as

$$f_k(y) = C(y)(I - z^{-1}E)^{-1}(Fu_{k-1}) \tag{48}$$

and expanded to the following form according to matrix theory (Wang, 2000)

$$f_k(y) = \sum_{i=1}^{n-1} a_i f_{k-i}(y) + \sum_{j=0}^{n-2} C(y)D_j u_{k-1-j} \tag{49}$$

where

$$D_j = (d_{j,1}, \cdots, d_{j,n-1}) \tag{50}$$

By writing

$$\theta = (a_1, \cdots, a_{n-1}, D_0, \cdots, D_{n-2}) \tag{51}$$

$$\phi = (f_{k-1}(y), \cdots, f_{k-n+1}(y), C(y)u_{k-1}, \cdots, C(y)u_{k-n+1}) \tag{52}$$

Equation (49) can be written in the parameterized form of

$$f_k(y) = \theta\phi^T \tag{53}$$

Assume that the definition interval [a,b] can be discretized by a set of sampling points, the parameters a_i and $d_{j,i}$ can be estimated by the so-called scanning identification algorithm with the standard least-square update towards (53) (Wang, 2000). Figure 1 is provided to clarify the scanning process, in which y stands for the chain length varying from 2 to N, m is the total number of sampling points in terms of time.

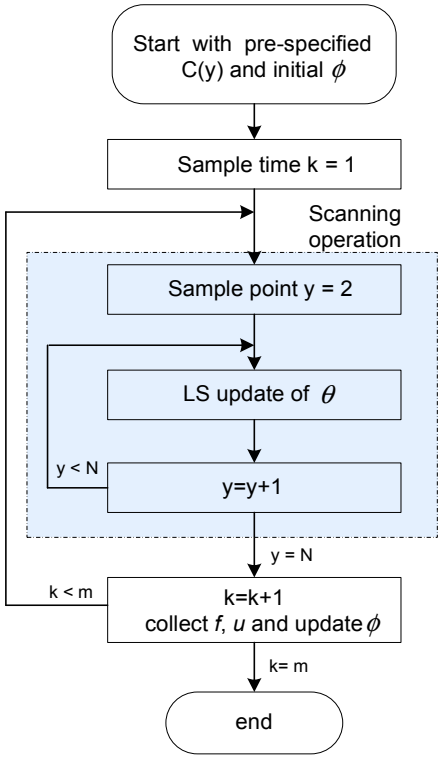


Fig. 1. scanning LS identification algorithm

5. MODEL VALIDATION

For the polymerization system in Section 2, firstly, the dynamic MWD data was produced from the first-principle model (fig. 2). The control input c was created randomly for the training purpose of the B-spline neural networks. Reaction system parameters are given in Table 1. In order to verify the formulation of the dynamic MWD, the steady-state solution of the dynamic model is compared with the results from the static MWD model. When $c = 0.5$, the MWD from the static model is given in fig. 3 and the steady-state MWD from the dynamic model is shown in fig. 4. The two curves are highly closed to each other, although the MWD in static model is an exact solution while the dynamic MWD is produced from the moments method, the generation function technique and finally represented by the Schultz-Zimm distribution. It shows that the moments method and the Schultz-Zimm distribution are appropriate for formulating this dynamic MWD model.

Secondly, the B-spline model was developed with the MWD data produced from the first-principle model. The scanning LS identification algorithm is used to obtain the parameter vector θ in (51). In this simulation, 10 fixed, 3rd-order B-splines are chosen for the MWD approximation. The training data length is 1500. The approximated MWDs with the trained B-spline weights are shown in fig. 5. It can be seen from fig. 5 and fig. 2 that the dynamic approximation is satisfactory.

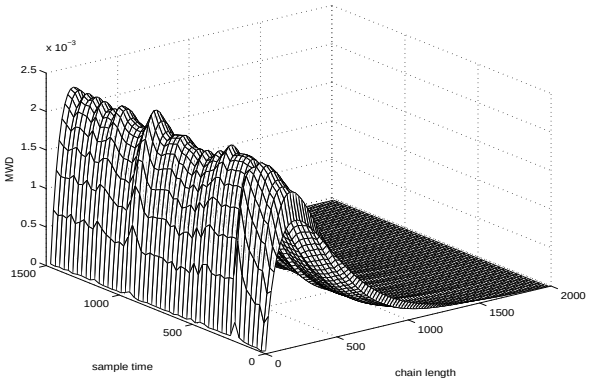


Fig. 2. Original MWDs from first-principle model

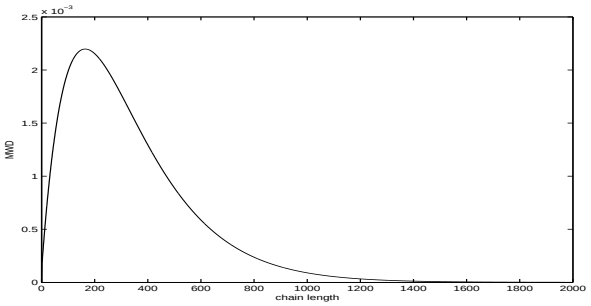


Fig. 3. Static MWD with c=0.5

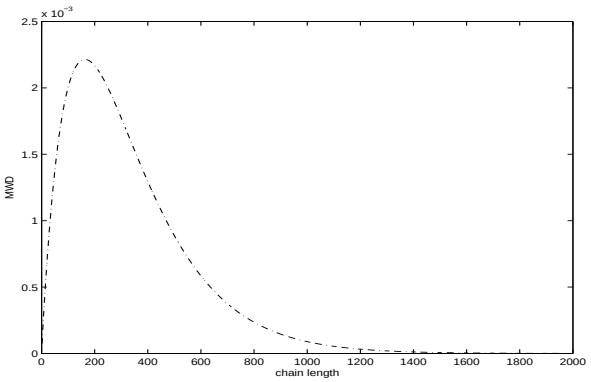


Fig. 4. Steady-state MWD with c=0.5

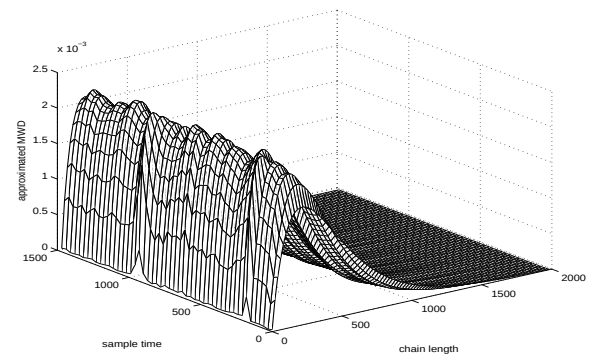


Fig. 5. Approximated MWDs from B-spline model

Table 1. Model parameters

K_d	$9.48 \times 10^{16} exp(-30798.5/rT)$
K_i	$0.6K_d$
K_p	$6.306 \times 10^{8} exp(-7067.8/rT)$
K_{trm}	$1.386 \times 10^{8} exp(-12671.1/rT)$
K_t	$3.765 \times 10^{10} exp(-1680/rT)$
V	3.927
F	0.0286
T	353
I^{00}	0.0106
M^{00}	4.81
r	1.987
c	[0.2,0.8]

Finally, the B-spline model was used for dynamic MWD control to see if the proper feedback control can be achieved with this model. The standard output PDF control is adopted with the following quadratic performance function (Wang, 2000).

$$J = \int_a^b (\gamma(y, u_k) - g(y))^2 dy + \frac{1}{2}\lambda u_k^2 \qquad (54)$$

where $g(y)$ is the target distribution and $\lambda > 0$ is a weighting factor for control energy. Fig. 6 shows the initial, final and target MWDs. Fig. 7 shows the development of output MWDs during the control process. Although there exists a small steady-state MWD tracking error, the controller successfully moves the output MWD from its initial shape towards the target shape. This means that the B-spline model can provide reliable MWD estimation for the feedback MWD control.

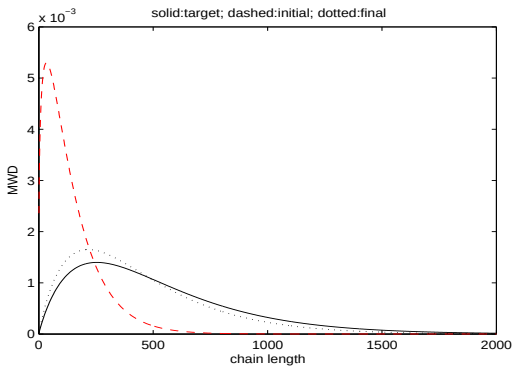


Fig. 6. Initial, final and target MWDs

6. CONCLUSIONS

In this paper, a dynamic first-principle MWD model has been developed and then approximated by the general B-spline functions. It makes the feedback MWD control feasible with the recently developed output PDF control strategies. Based on this model, further progresses on MWD control with different control strategies have been achieved and results will be distributed in the future.

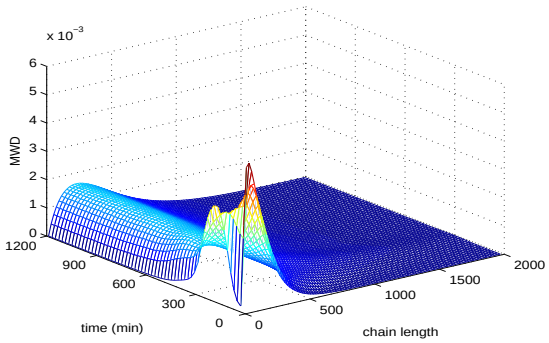


Fig. 7. Output MWDs during control

ACKNOWLEDGMENTS

The authors would like to thank J. Zhang and H. Wu for the help in developing BASIC and MATLAB programs.

REFERENCES

Angerman, H. J. (1998). *The Phase Behavior of Polydisperse Multiblock Copolymer Melts: a Theoretical Study*. http.

Kano, H., M. Egerstedt, H. Nakata and C. F. Martin (2003). B-splines and control theory. *Applied Mathematics and Computation* **145**, 263–288.

Sun, S., M. Egerstedt and C. F. Martin (2000). Control theoretic smoothing splines. *IEEE Trans. Automatic Control* **45**, 2271–2279.

Wang, H. (2000). *Bounded Dynamic Stochastic Systems: Modelling and Control*. Springer Verlag Ltd. London.

Wang, H. and A. Wang (1998). Stable adaptive control of stochastic distributions and its application. In: *Proc. of the UKACC International Conference on Control'98*. pp. 33–38.

Wang, H., J. Zhang and H. Yue (2005). Periodic larning of b-spline models for output pdf control: application to mwd control. In: *Proc. 2005 American Control Conference*. pp. 955–960.

Yue, H., J. Zhang and H. Wang (2004). Shaping of molecular weight distribution using b-spline based predictive probability density function control. In: *Proc. 2004 American Control Conference*. pp. 3587–3592.

Zhang, J. and H. Yue (2004). Improved b-spline neural network based modeling and control of output probability density functions. In: *Proc. IEEE International Symposium on Intelligent Control*. pp. 143–148.

Zhang, J., J. Tomlinson and C. F. Martin (1997). Splines and linear control theory. *Acta Applicand. Math.* **49**, 1–34.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PREDICTION OF GLYCOSYLATION SITE-OCCUPANCY USING ARTIFICIAL NEURAL NETWORKS

Ryan S. Senger and M. Nazmul Karim

Department of Chemical Engineering
Texas Tech University
Lubbock, TX 79409 USA

Abstract: A novel neural network-based model was developed to predict *N*-linked glycosylation site-occupancy characteristics. The model classified potential glycosylation sites as displaying *variable* site-occupancy or *robust* glycosylation when produced by CHO cell cultures under normal growth conditions. The term *variable* site-occupancy describes heterogeneous glycan attachment to a specified protein site. This phenomenon results in a heterogeneous mixture of glycosylated and unglycosylated proteins when produced in mammalian cell culture. The model input consists of amino acid residues around the site of glycosylation. Simulation of the model strongly correlated with previously published experimental results by Kasturi *et al*. (1997) and Mellquist *et al*. (1998).

Keywords: Artificial intelligence, Biotechnology, Computer-aided design, Mathematical models, Neural networks

1. INTRODUCTION

Glycosylated Pharmaceutical Proteins. Protein glycosylation is a vital post-translational modification of many proteins with therapeutic properties. The glycosylation pathway begins in the endoplasmic reticulum of a cell with the attachment of an oligosaccharide to a N-*X*-S/T (where *X* is not praline) polypeptide sequence. The attachment of a glycan structure is then followed by enzymatic trimming and processing of the attached oligosaccharide (glycan) structure (Kornfeld and Kornfeld, 1985; Roth, 1987; Silberstein and Gilmore, 1996). Glycan attachment and remodeling processes occur in the endoplasmic reticulum (ER) and Golgi apparatus of many cell types; however, only a select number of cell types produce glycosylation variants compatible in humans. Glycosylation is of great importance in bioprocessing because of its large influence on *in vivo* properties of therapeutic proteins, including specific activity. In addition, intramolecular influences of glycosylation on protein structure include: proper folding, intracellular location, biological activity, solubility, antigenicity, biological half-life and protease sensitivity. Similarly, intermolecular characteristics affected by protein glycosylation include: targeting to lysosomes, tissue targeting, cell-cell adhesion and binding of pathogens (Stanley, 1992). Cell types commonly selected by the pharmaceutical industry, for expression of glycosylated proteins, include human melanoma cells, baby hamster kidney (BHK) cells, and Chinese hamster ovary (CHO) cells.

Optimization with Respect to Glycosylation. The heterogeneity observed with glycosylation during

bioprocessing and its relevance on the biological activity of therapeutic proteins has led to a new area of optimization in the bioprocessing industry. This new area of glycosylation optimization is currently divided into two parts: (1) the initial attachment of the oligosaccharide to the protein and (2) the processing of glycan branches. For some proteins, the initial glycan attachment process has been found to be *robust*, resulting in a homogenously glycosylated or unglycosylated polypeptide sequence. However, for others, such as the recombinant tissue-type plasminogen activator (r-tPA) protein, this process is *variable*, resulting in a mixture of heterogeneous isoforms (or glycoforms) of fully, partially and unglycosylated species (Kornfeld and Kornfeld, 1985; Grossbard, 1987; Wittwer and Howard, 1990; Andersen *et al*., 2000; Senger and Karim, 2003a). Manipulations of process variables and culture medium conditions have been found to largely impact the degree to which r-tPA is glycosylated at site N184. However, culture conditions resulting in homogenously glycosylated r-tPA have not been found (Andersen *et al*., 2000; Senger and Karim, 2003a,b). Given that other glycosylation sites of r-tPA experience homogenous glycosylation (Grossbard, 1987; Wittwer and Howard, 1990), a better understanding of the mechanisms that cause a particular glycosylation site to display *variable* site-occupancy is desired. Glycosylation optimization is of great benefit to pharmaceutical manufacturing in terms of production costs and would result in tighter control of product specific activity.

Neural Networks in Structural Bioinformatics. Neural network-based models have been developed for the prediction of many structural characteristics of proteins, based on the protein amino acid sequence. In particular, this area of structural bioinformatics has expanded to predict secondary and some three-dimensional structures (Rost and Sander, 1993; Jones, 1999; Kelley *et al*., 2000; Pollastri *et al*., 2002a,b; Baldi and Pollastri, 2003). In general, neural networks have been an intricate part of these model-developments in that their capability for structure prediction has far exceeded that of first-principle (deterministic) models. This is due in large part to the expanding data bank of protein structure and genomic research (Rost, 2001).

Using Neural Networks to Predict Glycosylation Site-Occupancy Characteristics. A novel neural-network model has been developed in this research for predictions of glycosylation site-occupancy as homogeneous (*robust*) or heterogeneous (*variable*). The development of this model has allowed insight into many questions concerning protein glycosylation. The phenomena of variable site-occupancy, in the absence of substrate limitation, was found related to primary sequence characteristics. The number of amino acid residues around the site of glycosylation with influence on glycosylation characteristics was found much larger than what has been cited by previous research. Thus, the goal of this research is to develop a model of glycosylation site-occupancy so the optimization problem of glycosylation site-occupancy may be addressed through site-directed mutations of the protein sequence rather than manipulation of cell culture variables.

2. SYSTEMS AND METHODS

Acquired Data and Amino Acid Residue Quantification. Data for the construction of a glycosylation site-occupancy prediction model consisted of glycosylation sites, the amino acid sequence around this site and whether a particular site promoted homogeneous (*robust*) or heterogeneous (*variable*) glycosylation site-occupancy when produced by mammalian cell cultures. All data was acquired from a literature search interfaced with protein sequence databases. The entire data set consisted of 48 glycosylation sites. Five sequences (~10%) were reserved as a neural network testing data set. Greater than 40% of sequences of the data set were classified as displaying *variable* site-occupancy in the literature. For the input of particular amino acid residues into a neural network model, the identities of all amino acids were first converted to numerical values. Individual amino acid residues were grouped into eleven classes based on similar characteristics, such as charge, size, and hydrophobicity and assigned numerical values based on research by Kasturi *et al*. (1997) and Mellquist *et al*. (1998). Quantification of the target (site-occupancy classification) was also required. Glycosylation sites displaying *variable* site-occupancy were assigned the value *1*, and *robust* sites were assigned *0*. It is noted that a *robust* site may be homogeneously glycosylated or homogenously unglycosylated. Statistical models have been developed to discern between these two types of *robust* site-occupancy (Petrescu *et al*., 2004).

Table 1 Primary Sequence Quantification

Amino Acid Classes	Amino Acids	Assigned Value
Hydroxy	T	1
Hydroxy	S	2
Basic	K R H	3
Thioether	M	4
Alkyl	A V L I	5
Carboxamide	N Q	6
Unsubstituted	G	7
Acidic	D E	8
Mercapto	C	9
Aromatic	F Y W	10
Cyclic	P	11

Neural Network Architecture. Elman recurrent neural networks were used for the construction of the neural network-based model. Recurrent neural networks are renowned for their ability to learn non-causal data sets. The neural network inputs consisted of quantified amino acid residues around the site of glycosylation. Targets consisted of glycosylation site-occupancy assigned values. All neural networks consisted of a single hidden layer with hyperbolic sigmoid transfer functions. A single output neuron was used with a log sigmoid transfer function. A single perceptron neuron was used following the output neuron for two-dimensional classification. Thus, this neuron acted as a rounding function of the recurrent neural network output value. The number of hidden layer neurons was adjusted so that the number of adjustable network parameters (weight and bias values) always remained less than the number of data points used in the training procedure. Initial (prior to training) weight and bias values were assigned random values. Network training was performed using gradient decent with momentum and adaptive learning rate back-propagation. All neural networks were trained for 2000 epochs. Each neural network was independently initiated and trained 100 times. Results were averaged. The entire data set was cross-correlated using a testing set size of approximately 10% of the training set size.

Optimization of the Amino Acid Input Sequence. The number of amino acids on the *N*-terminus and *C*-terminus sites of the glycosylation site was varied, and neural network training and testing set analysis was performed in each case. This goal of this study was to identify relevant amino acids in prediction of glycosylation site-occupancy classification. In this study, the objective function of optimization problem was the mean-square error between the target values of glycosylation site-occupancy and the neural network-predicted values.

Comprehensive Model Construction. Once successful neural networks were identified, with an optimum input sequence length, a comprehensive model was constructed and further tested on published experimental data. The comprehensive model consisted of 20 neural networks that were found to correctly classify all elements of the neural network testing data sets. Networks from all cross-correlation iterations were used to construct the model. This composition of the comprehensive predictive model enabled the model to return an overall prediction value as well as a confidence interval. In particular, all neural networks were simulated, and results were averaged before perceptron classification. This method returned an overall model prediction. The confidence interval represents the fraction of neural networks returning the dominant classification. Thus, the overall model prediction consisted of a value of either *1* or *0*, and the confidence level of prediction was a value between 0.5 (low confidence) and 1 (high confidence).

Further Simulations. Published experimental data by Kasturi *et al*., (1997) and Mellquist *et al*., (1998) was used to further test the comprehensive neural network-based model. The published data focused on the effects of site-directed mutations around *variable* site-occupancy glycosylation site N39 of the rabies virus glycoprotein (rgp). Results of this work found that specific site-directed mutations resulted in the transformation of this glycosylation site from *variable* to *robust* site-occupancy. All of these sequences were simulated using the predictive model. In addition, further simulations were performed on simple theoretical sequences to examine the effects of charged amino acid residues around the site of glycosylation. Alanine (uncharged), lysine (negatively-charged) and aspartate (positively-charged) residues were used in this study.

3. RESULTS AND DISCUSSION

Optimization of the Glycosylation Window Length. The number of amino acids surrounding the site of glycosylation was termed the *glycosylation window*. Optimization of the glycosylation window length was performed by full neural network analysis with various input sequence lengths. Previous research has identified amino acid residues of the N-*X*-S/T-*Y*

glycosylation sequence as having significant influence over glycosylation site-occupancy characteristics (Shakin-Eshleman, 1996; Kasturi *et al*., 1997; Mellquist *et al*., 1998; Petrescu *et al*., 2004). The data-based method of analysis of this research allowed for the impact of 20 amino acid residue sites to be analyzed for influences on glycosylation characteristics. For each input sequence, the mean-square error was calculated between averaged neural network predictions and target values (glycosylation classification). Results are displayed as Figure 1. The starting residue of the input sequence is displayed on the abscissa as (*n*-x). The ending residue of a glycosylation window is displayed on the ordinate axis as (*n*+y), where *n* is the site of glycosylation. For example, a glycosylation window originating at (*n*-5) and extending to (*n*+4) contains a total of 10 amino acids: 5 residues on the *N*-terminus side of the glycosylation site, the glycosylation site itself (*n*) and 4 residues on the *C*-terminus side of the glycosylation site. The average standard deviation of all data points of Figure 1 was calculated as approximately 5% of the given data point value. Results showed a minimum mean-square error value of 0.0767 for the glycosylation window originating at (*n*-5) and extending to (*n*+4). The size of this glycosylation window is larger than others determined by experimental methods, and these are the first results of our knowledge to suggest influence of residues on the *N*-terminus side of a glycosylation site on glycosylation site-occupancy. For comparison, the data set was predicted by random values (in the absence of neural network training). These results were classified by the perceptron to yield and average mean-square error value of 0.7.

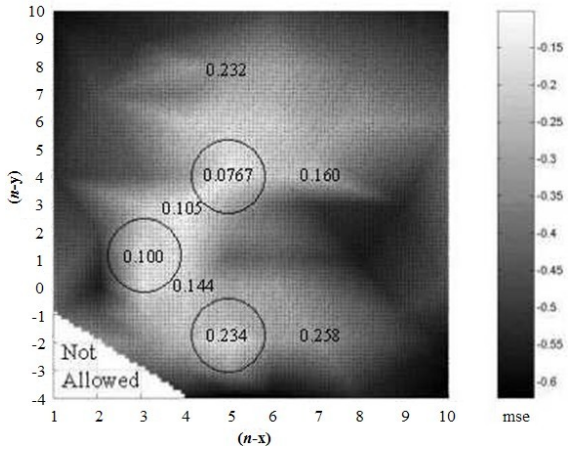


Fig. 1. Glycosylation window length optimization. Starting residue (abscissa). Ending (ordinate).

Simulations of rgp Wild-Type and Mutants Using Comprehensive Predictive Model. The comprehensive neural network-based predictive model was constructed using the optimized glycosylation window input length. The model consisted of 20 independent neural networks, and represented all iterations of the cross-correlation analysis. All neural networks of the comprehensive model classified corresponding testing data sets with 100% accuracy following perceptron classification of recurrent network output values. To further verify the predictive model, wild-type and site-directed mutations of the rgp protein glycosylation site N37 were simulated and compared to published experimental observations by Kasturi *et al*. (1997) and Mellquist *et al*. (1998). These experimental studies examined the influence of amino acid residues at positions *X* and *Y* of the N-*X*-S/T-*Y* glycosylation sequence. These sites correspond to (*n*+1) and (*n*+4) using the terminology developed for the glycosylation window length optimization. Results reported by Kasturi *et al*. and Mellquist *et al*. (1998) reported *glycosylation efficiency*. In short, the glycosylation efficiency is defined as the fraction of fully glycosylated rgp (N37). Thus, reported glycosylation efficiency between 0 and 1 corresponds to *variable* site-occupancy glycosylation. A glycosylation efficiency value of 1 corresponds to homogeneous (*robust*) glycosylation, and a glycosylation efficiency of 0 corresponds to a glycosylation site that is homogeneously (*robust*) unglycosylated. Values of glycosylation efficiency were interpolated from the published experimental studies, taking into account experimental error, and these are listed in Table 2. A total of 19 rgp mutants, in addition to the wild-type protein, were evaluated by the comprehensive predictive model. The sequence identity (details of site-directed mutations), the overall predictive model classification and the confidence level are also reported in Table 3.

Discussion of Prediction Results. Overall, the predictive model showed 95% accuracy in predicting glycosylation site-occupancy characteristics for this set of published experimental data. One rgp mutant (simulation 15a; S39T G40W) was incorrectly classified by the predictive model. However, in this case, the confidence level of the prediction was low (0.65). Although, the unsuccessful model prediction contained an aromatic residue (tryptophan), other predictions involving aromatic residues were classified correctly (simulation 8a). Sequences displaying *variable* site-occupancy, to a glycosylation efficiency of 0.9 (simulations 2a-4a, 8a, 10a-13a, 14a-15a) were correctly classified by the model as promoting *variable* site-occupancy. For

sequences that were constructed that resulted in glycosylation efficiency of or exceeding 0.95, most model predictions (simulations 6a, 7a, 20a) classified these sequences as having *robust* glycosylation. The exception in this case was for the L38N S39T mutant (simulation 6a), which displayed a glycosylation efficiency of approximately 0.95. *Variable* site-occupancy was predicted in this case, but a lower confidence level in this case (0.7) suggested that many neural networks of the predictive model recognized this sequence as promoting *robust* glycosylation. In addition, given the experimental error in the case of simulation 6a of roughly 5%, *variable* site-occupancy may accurately describe this system. The other sequences of simulations 7a and 20a displayed glycosylation efficiencies that exceeded 0.95. Of further importance is that the sensitivity of the predictive model was evaluated by this set of simulations. In short, this set of model predictions correctly classified a polypeptide sequence as having *variable* site-occupancy glycosylation characteristics for glycosylation efficiencies ranging between 0.1 to >0.95.

Simulations of Theoretical Sequences. Due to the success of the comprehensive predictive model in classification of rgp wild-type and variant sequences, the same simulation technique was applied to theoretical sequences. Glycosylation site-occupancy classification in the presence of charged amino acid residues was studied in the following simulations. For consistency, only alanine (A) was used as the uncharged residue outside of the required N-*X*-S/T glycosylation sequence. In addition, aspartate (D) and lysine (K) were used as the negatively and positively-charged residues, respectively. Simulations suggested that a sequence consisting of alanine or aspartate residues throughout the glycosylation window (except for the glycosylation sequence at (*n*) and (*n*+2) would result in robust glycosylation with a high confidence level. With positively-charged lysine residues occupying the glycosylation window and serine or threonine at position (*n*+2), variable site-occupancy glycosylation was predicted with a confidence level of 0.95. Further simulations examined the influence of particular locations within the glycosylation window. For example, a glycosylation window consisting of lysine residues was substituted with aspartate and alanine residues until a robust glycosylation site-occupancy prediction was achieved. A summary of these simulation results is presented in Table 3, and the actual simulation results are given in Table 4. These types of simulation experiments were performed for both serine and threonine in the (*n*+2) position. It was found through simulation that replacement of serine in the glycosylation sequence with threonine increases the robustness of glycan attachment. This evidence further supports this idea, as it was suggested by previous research (Kasturi *et al*., 1997; Mellquist *et al*., 1998; Petrescu *et al*., 2004).

Table 2 rgp variant and wild-type predictions with confidence level and published experimental results

Sim.	Sequence	Overall Model Classification and Confidence Level	Published Glycosylation Efficiency
1a	Wild-type	1 (1.00)	0.35
2a	S39T	1 (0.95)	0.80
3a	L38N	1 (1.00)	0.70
4a	L38S	1 (1.00)	0.90
5a	L38W	0 (0.60)	0.10
6a	L38N S39T	1 (0.70)	0.95
7a	L38G S39T	0 (0.60)	>0.95
8a	G40F	1 (0.75)	0.40
9a	G40P	0 (0.75)	0.05
10a	G40H	1 (1.00)	0.55
11a	G40M	1 (1.00)	0.70
12a	G40N	1 (1.00)	0.80
13a	G40S	1 (1.00)	0.80
14a	G40C	1 (0.90)	0.80
15a	S39T G40W	0 (0.65)	0.80
16a	S39T G40H	1 (1.00)	0.85
17a	S39T G40M	1 (1.00)	0.90
18a	S39T G40N	1 (1.00)	0.90
19a	S39T G40T	1 (1.00)	0.90
20a	S39T G40C	0 (0.55)	>0.95

Sim. is an abbreviation for "Corresponding Simulation."
The Confidence Level is listed in parentheses.
Published glycosylation efficiency values were interpolated from Kasturi *et al*. (1997) and Mellquist *et al*. (1998).

Table 4 Generalizations from simulations

Residues:	Type of influence on glycosylation site-occupancy:
Lysine (positive charge)	Promotes *variable*
Aspartate (negative charge)	Promotes *robust*
Alanine (no charge)	Promotes *robust* with less influence than aspartate
Location in glycosylation window:	**Level of influence on glycosylation site-occupancy:**
(*n*+1)	Highest influence
(*n*+3), (*n*+4), (*n*-1)	Moderate influence
(*n*-5)…(*n*-2)	Lowest influence, but significance was observed
(*n*+2)	Threonine results in more *robust* glycosylation

Table 5 Amino acid residues and simulation results of theoretical sequences

-5	-4	-3	-2	-1	*n*	+1	+2	+3	+4	O.C. (C.L.)
A	A	A	A	A	N	A	S	A	A	0 (0.75)
A	A	A	A	A	N	A	T	A	A	0 (1.00)
A	A	A	A	A	N	P	S	A	A	0 (1.00)
A	A	A	A	A	N	P	T	A	A	0 (1.00)
K	K	K	K	K	N	K	S	K	K	1 (1.00)
K	K	K	K	K	N	K	T	K	K	1 (0.95)
D	D	D	D	D	N	D	S	D	D	0 (1.00)
D	D	D	D	D	N	D	T	D	D	0 (1.00)
A	A	A	A	A	N	K	S	A	A	1 (0.95)
A	A	A	A	A	N	A	S	K	A	1 (0.70)
A	A	A	A	K	N	A	S	A	A	0 (0.60)
A	A	A	A	A	N	K	S	K	A	1 (1.00)
A	A	A	A	A	N	D	S	D	A	0 (1.00)
A	A	A	K	K	N	A	S	A	A	0 (0.60)
A	A	K	K	K	N	A	S	A	A	1 (0.65)
K	K	K	K	K	N	A	S	A	A	1 (0.80)
A	A	K	K	K	N	A	T	A	A	0 (0.60)
A	K	K	K	K	N	A	T	A	A	1 (0.70)
D	D	K	K	K	N	A	S	A	A	0 (0.80)
D	D	D	D	D	N	K	S	D	D	1 (0.90)
D	D	D	D	D	N	A	S	D	D	0 (0.90)
D	D	D	D	D	N	K	T	D	D	1 (0.50)
K	K	K	K	K	N	D	S	K	K	0 (0.85)

O.C. is an abbreviation for "Overall Model Classification."
C.L. is an abbreviation for "Confidence Level."
The Confidence Level is listed in parentheses.

4. CONCLUSIONS

A novel neural network-based predictive model has been developed for the classification of *N*-linked glycosylation as heterogeneous (*variable*) or homogeneous (*robust*) for proteins produced by mammalian cell culture. Amino acid residues around the site of glycosylation were found to impact site-occupancy characteristics. In particular, an optimization study found that 5 residues on the *N*-terminus side and 4 residues on the *C*-terminus side of the glycosylation site directly influence these characteristics. The neural network-based predictive model classified published experimental findings regarding the impact of amino acid residues on site-occupancy characteristics with 95% accuracy. Further simulations with theoretical amino acid sequences revealed negatively-charged promote *robust* glycosylation and that the (*n*+1) position of the glycosylation window had the most influence over glycosylation site-occupancy characteristics. Elimination of *variable* site-occupancy will have significant impact in the pharmaceutical industry. *Robust* glycosylation results in homogenous product production. This is of utmost importance to quality control and optimization of biological activity of glycosylated recombinant proteins with therapeutic properties.

REFERENCES

Andersen, D.C., T. Bridges, M. Gawleitzek and C. Hoy (2000). Multiple cell culture factors can affect the glycosylation of Asn-184 in CHO-produced tissue-type plasminogen activator, *Biotechnol. Bioeng.*, **70**, 25-31.

Baldi P. and G. Pollastri (2003). The principled design of large-scale recursive neural network architectures-DAG-RNNs and the protein structure prediction problem, *Machine Learning*, **4**, 575-602.

Grossbard, E.B. (1987). Recombinant tissue plasminogen activator: A brief review, *Pharm. Res.*, **4**, 375-378.

Jones D.T. (1999). Protein secondary structure predictions based on position-specific scoring matrices, *J. Mol. Biol.*, 292, 195-202.

Kasturi, L., C. Hegang and S.H. Shakin-Eshleman (1997). Regulation of *N*-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem. J.*, **323**, 415-419.

Kelley, L.A., R.M. MacCallum and M. J. E. Sternberg (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM, *J. Mol. Biol.*, **299**, 499-520.

Kornfeld, R. and S. Kornfeld (1985). Assembly of asparagine-linked oligosaccharides, *Ann. Rev. Biochem.*, **54**, 631-664.

Mellquist, J.L., L. Kasturi, S.L. Spitalnik and S.H. Shakin-Eshleman (1998). The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of *N*-linked core glycosylation efficiency. *Biochemistry*, **37**, 6833-6837.

Petrescu, A.J., A.L. Milac, S.M. Petrescu, R.A. Dwek and M.R. Wormald (2004). Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding, *Glycobiology*, **14**, 103-114.

Pollastri G., D. Przybylski, B. Rost and P. Baldi (2002a). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles, *Proteins*, **47**, 228-235.

Pollastri, G., P. Baldi, P. Fariselli and R. Casadio (2002b). Prediction of coordination number and relative solvent accessibility in proteins, *Proteins*, **47**, 142-153.

Rost B. (2001). Review: Protein secondary structure prediction continues to rise, *J. Struct. Biol.*, **134**, 204-218.

Rost, B. and C. Sander (1993). Prediction of protein secondary structure at better than 70% accuracy, *J. Mol. Biol*, **232**, 584-599.

Roth, J. (1987). Subcellular organization of glycosylation in mammalian cells, *Biochim. Biophys. Acta*, **906**, 405-436.

Senger, R.S. and M.N. Karim (2003a). Effect of shear stress on intrinsic CHO culture state and glycosylation of recombinant tissue-type plasminogen activator protein, *Biotechnol. Prog.*, **19**, 1199-1209.

Senger, R.S. and M. N. Karim (2003b). Neural-network-identification of tissue-type plasminogen activator protein production and glycosylation in CHO cell culture under shear environment, *Biotechnol. Prog.*, **19**, 1828-1836.

Shakin-Eshleman, S.H. (1996). Regulation of N-linked core-glycosylation, *Trends Glycosci. Glyc.*, **8**, 115-130.

Silberstein, S. and R. Gilmore (1996). Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase, *FASEB J.*, **10**, 849-858.

Stanley, P. (1992). Glycosylation engineering, *Glycobiology*, **2**, 99-107.

Wittwer, J. and S.C. Howard (1990). Glycosylation at Asn184 inhibits the conversion of single-chain to two-chain tissue-type plasminogen activator by plasmin, *Biochemistry*, **29**, 4175-4180.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

Real time thermal tracking of ladle furnaces: an analytical approach

Zabadal, J.(1), Garcia, R. (2), Salgueiro, M. (2)
Universidade Federal do Rio Grande do Sul
(1) Departamento de Engenharia Nuclear
(2) Programa de pós-graduação em Engenharia Mecânica

Abstract

This work presents a new analytical approach for solving unsteady diffusion problems. In the proposed method, a formal solution is converted in closed form ones, which are obtaining by performing a straightforward procedure that starts with a classical split. The low time processing required to obtain the exact solutions allows performing online control of ladle furnaces. Numerical results are reported.

1 - Introduction

It is widely felt that the significant reduction of time processing due to the recent advances in numerical and analytical methods can make possible to proceed the online control based on direct simulation for some important applications in Engineering, such as in environmental problems (Zabadal, 2005), neutron scattering in nuclear reactors (Bogado, 2004) and casting of steel alloys (Zabadal, 2004). Specifically, for steel casting processes, the simulation of ladle furnaces is particularly advantageous from the operational point of view, because the calculated steel temperature agrees with experimental data even when nonlinear effects are ignored. The major aim of the online control in casting of steel alloys is to ensure that the temperature of the liquid steel which flows from the furnace does not fall out of an interval of roughly 10°C around a mean value about 1600°C (which varies from one specific steel alloy to another) . This control prevents failures in the lattice, which occurs when the temperature is low, and unsuitable flow conditions, when the temperature is higher than a certain value.

The main limitation of the use of numerical schemes to proceed the simulation of casting processes occurs for some scenarios where the ladle remains out of operation for long time intervals (about 24h after the last batch). In these cases the ladle must suffers a slow heating process (during about 10h), and the simulation becomes a very difficult task. The online control based on numerical simulation results unfeasible, due to the large time processing required.

In this work a new analytical method for simulating the heating process is presented. This method, based on the formal solutions of partial differential equations, is employed to overcome the limitations of the methods which were originally conceived to carry out the online control of ladle furnaces. The most important features of the proposed method are the high processing speed and the analytical character of the solutions obtained.

2 – General formulation

The partial differential equation given by:

$$Lf = 0 \quad , \tag{1}$$

where L is a linear operator can be decomposed as

$$Af = Bf \quad , \tag{2}$$

in which A and B are also linear operators. Applying A^{-1} on both sides of (2) we obtain:

$$f = A^{-1}Bf + h_A \ . \tag{3}$$

In this expression h_A stands for the null space of A. Rearranging terms it yields

$$[I - A^{-1}B]f = h_A \ . \tag{4}$$

Solving equation (4) for f it results:

$$f = [I - A^{-1}B]^{-1} h_A \ . \tag{5}$$

The inverse operator appearing in equation (5) can be written as a geometrical series:

$$[I - A^{-1}B]^{-1} = \sum_{k=0}^{k=\infty} \left(A^{-1}B \right)^k \quad . \tag{6}$$

Disregarding the restrictions about the norm of the operator $A^{-1}B$, the solution is then readily obtained in the form:

$$f = \sum_{k=0}^{k=\infty} \left(A^{-1}B \right)^k h_A \quad . \tag{7}$$

In order to obtain a particular solution for equation (1), it becomes necessary to choose a function $f_0 \in N(A)$. In practice, f_0 can be chosen as a function belonging to the intersection of the null spaces of $(A^{-1}B)^n$ and A, in order to convert the series solution into a finite sum. In what follows it will be showed that the closed-form solutions achieved by means of the described method generate high performance algorithms for online control.

3 – Application in online control of ladle furnaces

The online control of ladle furnaces can be carried out by solving the heat equation in the form

$$\frac{\partial f}{\partial t} = \alpha \left(\frac{\partial^2 f}{\partial r^2} + \frac{1}{r} . \frac{\partial f}{\partial r} \right) \quad , \tag{8}$$

in a hollow cylinder, subjected to the following boundary conditions:

$$\left. \frac{\partial f}{\partial r} \right|_{r=o} = 0 \tag{9}$$

and

$$f(l,t) = f_l(\mathrm{t}) \quad . \tag{10}$$

In these equations f represents the temperature, r is the radial coordinate, α is the thermal diffusivity, f_l is the temperature at r=l, and t stands for the time. This model describes a process in which the ladle is heated inside by a flame before receiving liquid steel from the main furnace. The first boundary condition, given by equation (9), states that the external wall does not exchanges energy with the air at room temperature during the process. Although the external wall being not really insulated, equation (9) is a reasonable approximation for thick walls composed of materials whose thermal diffusivity is low, and hence possesses a high thermal inertia. The second boundary condition informs that inner wall is at the flame temperature, which is time dependent.

After the heating process, the ladle receives liquid steel, whose temperature is also time depending. The evolution of the inner wall temperature along the time is fitted using a standard least square procedure available in MapleV. Finally, the initial condition imposes the final profile of the former batch to the next one at t=0.

In equation (8), the operators A, B and A^{-1} are promptly identified as

$$A = \frac{\partial}{\partial t} \quad , \tag{11}$$

$$B = \alpha \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} . \frac{\partial}{\partial r} \right) \tag{12}$$

and

$$A^{-1} = \int (.) dt \quad , \tag{13}$$

whereas in the boundary condition (10) the subscript l refers to the coordinate r = l, where l denotes the thickness of the ladle wall (a typical value for l is about 0,35m). Once T_l is time dependent, the desired solution shall follow the time evolution of the boundary condition given by (10).

The simpler choice of $f_0 \in N(A^{-1}B)^k \cap N(A)$ is given by:

$$f_0 = r^2 \tag{14}$$

Applying operator $A^{-1}B$ it came out

$$f_1 = 4 \alpha t \tag{15}$$

Applying the same operator over f_l it yields:

$$A^{-1}Bf_1 = 0 \tag{16}$$

Therefore $f_0 = r^2$, which belongs simultaneously to the nullspaces of A and $(A^{-1}B)^2$, produces a closed form solution which contains only two terms:

$$f(r,t) = f_0 + f_1 = r^2 + 4\alpha t \quad . \tag{17}$$

Analogously, another particular solution can be easily obtained by setting $f_0 = r^4$. In this case, the closed form solution is expressed as

$$f(r,t) = f_0 + f_1 + f_2 = r^4 + 16\alpha r^2 t + 32\alpha^2 t^2 \quad , \tag{18}$$

and $f_0 = r^4$ belongs to the nullspaces of A and $(A^{-1}B)^3$. Each even power of r, namely r^{2n}, generates a closed form solution belonging to the nullspace of $(A^{-1}B)^{n+1}$. Other examples of solutions obtained from even powers of r are given below:

$$f_0 = r^6 \rightarrow f(r,t) = r^6 + 36\alpha r^4 t + 288\alpha^2 r^2 t^2 + +384\alpha^3 t^3 \quad , \tag{19}$$

$$f_0 = r^8 \rightarrow f(r,t) = r^8 + 64\alpha r^6 t + 1152\alpha^2 r^4 t^2 + +6144\alpha^3 r^2 t^3 + 6144\alpha^4 t^4 \quad , \tag{20}$$

and

$$f_0 = r^{10} \rightarrow f(r,t) = r^{10} + 100\alpha r^8 t + 3200\alpha^2 r^6 t^2 + +38400\alpha^3 r^4 t^3 + 153600\alpha^4 r^2 t^4 + 122880\alpha^5 t^5 \quad . \tag{21}$$

Reminding that the desired solution shall contain some arbitrary parameters in order to fulfill the boundary condition at r = 0, as well as to follow the time evolution of the boundary condition at r = l, a linear combination of the above solutions must be employed in order to simulate the physical scenario. The boundary condition at r = 0 is automatically satisfied because the solution is an even function of r. Hence, all the numerical coefficients in the linear combination are specified in order to fit the boundary condition at r=l. This task is accomplished by means of a conventional curve fitting procedure.

At this point, one may ask why use such a scheme to obtain closed-form solutions, once the analytical one is yet available in literature. The foremost reason is the need to expand the analytical solution in a basis set containing the Bessel functions J_v and Y_v . The oscillations associated with the J_v functions requires a large number of terms in the expansion in order to smooth out the "wigglyness" appearing due to the contributions of the eigenfunctions related to the lowest eigenvalues. Since the definition of both Bessel functions involves the evaluation of the gamma function, which is expressed as a product, a summation or a high degree polynomial, a large number of floating point operations is demanded in order to produce numerical results.

4 – Results and conclusion

The exact solution employed to simulate the heating process is a linear combination given by

$$f = \sum_{k=0}^{5} c_k p_{2k}(r,t) \tag{22}$$

where $p_{2k}(r,t)$ are the polynomials defined by equations (17) to (21), and the coefficients c_0 to c_5 are given in table 1. These coefficients were obtained by fitting the data corresponding to the boundary condition at r=l, as mentioned earlier. The fitting generates a time evolution which reproduces the experimental data at r=l with a mean square deviation about 1°C (notice that c_0 was included in the linear combination, because a constant function is also an exact solution of the heat equation).

Table 1 - Numerical values of the coefficients

Coefficients	Values
∝	1e-7
C_0	102,5
C_1	198,6
C_2	2830
C_3	8116
C_4	235,4
C_5	28,74

Figure 1 shows the corresponding time evolution of the temperature profile along the heating process.

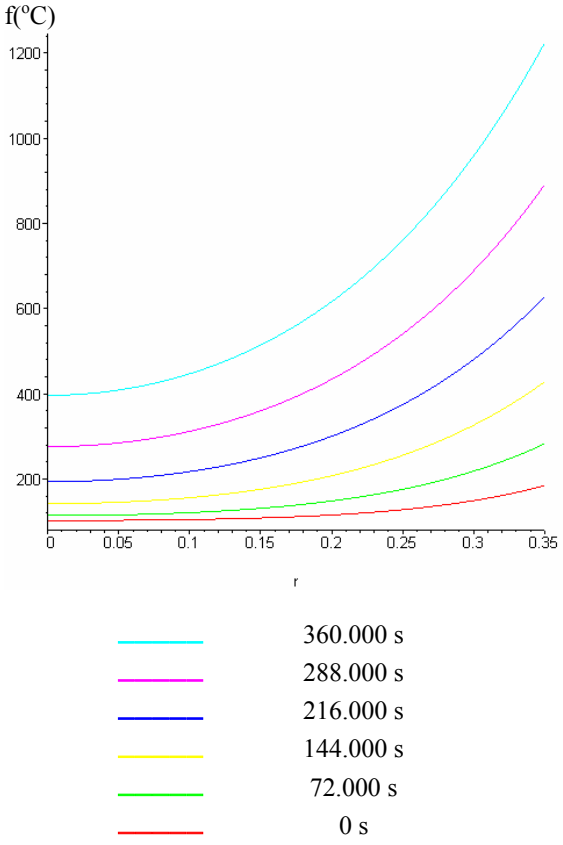


Figure 1 – Time evolution of the temperature profile (°C) during the heating process.

The mean square deviation between the predictions and the experimental data available at the "Aços Finos Piratini" steel casting facilities is about 0,16%, and satisfies the requirement that the temperature of the steel flowing from the furnace does not fall out of an interval of 10°C around the mean value in 79% of the cases (237 batches).

It is important to emphasize that the time required to obtain the temperature profiles are virtually negligible (less than 1s - Sempron 2.4 GHz, 512 Mb RAM, using

Maple V). Nevertheless, the curve fitting procedure must be carried out in advance. It means that, in practice, it becomes necessary to include a fitting routine in the computational code in order to proceed the online control.

Another important question about the method must be answered. Once the proposed method was developed to simulate the thermal behavior of the wall, how to predict the time evolution of the temperature profile at the bottom of the ladle? In this case there exists an exact solution for the corresponding Cartesian problem in the z coordinate, given by

$$f = \mathbf{b_0} + \mathbf{b_1} \mathbf{e}^{\mathbf{b_2} z + \mathbf{b_2^2} t} \tag{23}$$

Notwithstanding the solution in cartesian coordinates were obtained directly by inspection, it is also suitable for real time thermal tracking.

It is also important to remark that the solutions can be employed separately, which means that the thermal coupling between the wall and the bottom of the ladle furnace can be neglected without appreciable loss in accuracy.

Finally, it is convenient to emphasize that all the functions obtained through the iterative scheme, namely, equations (17) to (21), are exact solutions. Hence, equation (22) is not a truncated series which constitutes an approximation to the exact solution, but it is itself an exact solution to the heat equation. Once the functions obtained are conceived to belong to the nullspace of a finite power n of the operator $A^{-1}B$, the "truncated" series defined by

$$f = \sum_{k=0}^{k=n} \left(A^{-1} B \right)^k h_A \quad . \tag{24}$$

is always an exact solution, provided that all the terms beyond n are automatically dropped out. Therefore, no questions about convergence arises along the development of the proposed formulation.

References

Bogado Leite, S., Zabadal, J., Vilhena, M. (2004). Improved Dancoff factors for cluster fuel bundles by the WIMSD code. In: *Kerntechnik*, v 69, pp. 79-83. Munich.

Zabadal, J., Vilhena, M., Bogado Leite, S. (2004). Heat transfer process simulation by finite differences for online control of ladle furnaces. In: *Ironmaking and Steelmaking*, v 31, n 3, pp 227-234. Maney Publisinhg, London.

Zabadal, J., Vilhena, M., Bogado Leite, S., Poffal, C. (2005). Solving unsteady problems in water pollution using Lie symmetries. In: *Ecological Modelling*, v 186, pp 271-279. Elsevier Science Publishers, Copenhagen.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

Solving water pollution problems using auto-Bäcklund transformations

Zabadal, J.[1], Garcia, R. [2], Salgueiro, M. [2]
Universidade Federal do Rio Grande do Sul
(1) Departamento de Engenharia Nuclear
(2) Programa de pós-graduação em Engenharia Mecânica

Abstract

In this work an analytical formulation for solving partial differential equations is proposed. The method dispenses the use of Lie groups to produce maps between exact solutions, and generates nonlocal symmetries admitted by differential equations. The method is applied in water pollution problems, furnishing exact solutions for the advection-diffusion equation which describes the propagation of bacteria and chemicals in water bodies with arbitrary contours.

1 - Introduction

Solving water pollution problems are the first step to prevent environmental damages caused by industrial activity and domestic sewers along rivers and lakes. The simulation of water pollution scenarios provides crucial informations for lowering the costs demanded to treat the emissions and to improve the projects related to the implantation of new sewer systems. The most usual methods employed to simulate dispersion scenarios are finite differences (often implicit and time-marching schemes), and finite elements (usually based on Galerkin and least squares formulations). These numerical methods present some inconvenient features when applied to multidimensional problems in complex-shaped domains. The first is related to the time processing required to obtain the numerical solutions, which seldom can be accomplished using coarse meshes. The second is due to the difficulties associated with changes in the source terms. For many practical problems in Environmental Engineering one must simulate repositioning of the sewer loads or treatments for reducing the corresponding concentration, in order to evaluate the effect of the resulting concentration profile along the water body, an then decide whether a given change in the configuration of the sewer system can reduce the pollutant emissions in certain regions of interest. Since this application requires the simulation of several combinations of loading positions and concentrations, and each change in the source terms requires a new numerical simulation, the resulting time processing becomes prohibitive for planning sewer system configurations.

In order to surmount this difficulty, a new analytical formulation based on auto-Bäcklund transformations (Zwillinger, 1992) is proposed. These transformations are nonlocal Lie symmetries admitted by a given differential equation (Bluman and Kummei, 1989), i.e., changes of variables that converts exact solutions of the equation into new exact ones through a discontinuous map, while the local Lie symmetries are continuous transformations. In this work these nonlocal symmetries are obtained by means of a generalized split formulation, which will be described in what follows, instead of employing Lie group analysis (Ibragimov, 1995), which is the procedure often adopted for solving partial differential equations. The main advantages of the proposed formulation relies on the resulting time processing, which is about 0.1% of the one required by solving two-dimensional advection-diffusion equations via finite differences (time marching schemes, for instance), the use of streamfunction and velocity potential as curvilinear coordinates in the solution obtained, which extends the application of the proposed method for arbitrary geometries, and the simplicity of the auxiliary equations to be solved, when compared with the so-called determining equations (Olver, 2000) for the coefficients of the generators which constitute the Lie group. Besides, there are many cases when some of the determining equations are more difficult to solve than the original one. Moreover, the Lie group approach requires the solution of an additional set of auxiliary equations which comes from the application of exponentials whose argument contains a linear combination of the generators, obtained after solving the determining equations. These advantages will become clear after presenting the proposed formulation.

2. The auto-Bäckund transformation

In this section a new analytical method to construct auto-Bäcklund transformations is described. The method is based on a sequence of non-homogeneuos splits for which the source terms appearing in the corresponding systems of differential equations can be readily obtained from any particular solution of the original equation (even the trivial one). The

novelty of the proposed method relies in the presence of the source term. The methods based on split generates only homogeneous systems of auxiliary equations (Polyanin, 2004). These methods produces solutions which satisfy a very particular set of boundary conditions. In the proposed formulation, an iterative scheme is obtained, in such a way that each iteration produces a new exact solution satisfying a wider set of boundary conditions. In order to start the iterative scheme, let us consider the equation

$$Lf = 0 \quad , \tag{1}$$

where L is a linear differential operator, which can be written in the following form:

$$L = A - B \quad , \tag{2}$$

in which the inverse of A is known. Hence, equation (1) can be expressed as

$$Af = Bf \quad , \tag{3}$$

or, equivalently, as a non-homogeneous system of differential equations:

$$Af = Q \tag{4}$$

and

$$Bf = Q \quad , \tag{5}$$

where the source term must be determined. It will be showed that when the comutator [A,B] is null the source term can be replaced by any exact solution of equation (1). Indeed, applying operator B over equation (4), it yields:

$$BAf = BQ \ . \tag{6}$$

Applying operator A over equation (6) it results

$$ABf = AQ \ . \tag{7}$$

Subtracting equation (7) by (6) and taking into account the linearity of both operators, the following result is obtained:

$$\left[A, B\right] f = AQ - BQ \ . \tag{8}$$

Therefore, when [A,B] = 0 , the source term Q obeys the same differential equation satisfyied by the unknown function f. The former result allows to carry out and iterative scheme which can be recasted in the following fashion:

$$Af_{k+1} = f_k \tag{9}$$

and

$$Bf_{k+1} = f_k \ . \tag{10}$$

The system can be solved in a straightforward way. Starting with any particular solution f_0 of the original equation, which can be even the trivial one, equation (9) is solved, furnishing:

$$f_{k+1} = A^{-1} f_k + h_A \quad , \tag{11}$$

where h_A denotes a function belonging to the nullspace of A. Substituing the solution obtained into equation (10) it results

$$Bf_{k+1} = BA^{-1} f_k + Bh_A \ . \tag{12}$$

Equation (12) is often readily solved for the arbitrary elements contained in h_A . Eventually, this equation must be splitted, producing another system of non-homogeneous differential equations, whose solution is obtained by applying the same procedure already described.

The new solution obtained is then replaced on the right hand side of equations (9) and (10) and the process is repeated. Notice that at each iteration a new exact solution arises. In other words, the procedure above described does not constitute a iterative scheme which converges to a given exact solution. It is important to bear in mind that the iterations stops when the solution obtained becomes flexible enough to satisfy the initial and boundary conditions imposed in a given subdomain. It means that the number and nature of the arbitrary elements (constants or functions) contained in the solution will define the extension of the subdomain in which this solution remains valid. Roughly speaking, the number of iterations determine whether the solution can be used in a “chunk” or in the whole domain.

In the case when A and B do not commute, the proposed method must suffers a slight modification, by defining the g-commutator, denoted by:

$$\left[A, B\right]_g = AgB - BgA \quad , \tag{13}$$

where g is an unknown function. As in the former case, it is easy to show that when the g-commutator is null and analogous iterative scheme can be performed:

$$Af_{k+1} = \frac{f_k}{g} \tag{14}$$

and

$$Bf_{k+1} = \frac{f_k}{g} \quad . \tag{15}$$

In fact, multiplying equations (14) and (15) by g, applying operator B over equation (14) and operator A over equation (15), and finally subtracting the resulting equations, it yields

$$[A,B]_g f_{k+1} = Af_k - Bf_k \quad , \tag{16}$$

In order to ensure the "g-commutativity" between A and B, the function g must satisfy some auxiliary differential equations which are often simpler than the original equation to be solved. Moreover, for most practical purposes, it becomes possible to map the original equation in such a way that [A,B] = 0 , and even when the operators do not commute there are infinite solutions for the auxiliary equations which comes from the condition $[A,B]_g = 0$.

3. Application in water pollution problems

The propagation of conservative pollutants in rivers and lakes for complex-shaped domains is given by

$$\frac{1}{\mathrm{D}}\frac{\partial C}{\partial \Phi} = \frac{\partial^2 C}{\partial \Phi^2} + \frac{\partial^2 C}{\partial \Psi^2} \tag{17}$$

Where D is the mass diffusivity, Ψ is the stream function and Φ is the potential function. In this equation the hydrodinamic boundary layer effects over the concentration profile are not considered, because the boundary layer thickness are negligible when compared to the geographic scale of the water body. In this case the operators A and B are given by

$$A = \frac{\partial^2}{\partial \Psi^2} \tag{18}$$

and

$$B = \frac{1}{\mathrm{D}}\frac{\partial}{\partial \Phi} + \frac{\partial^2}{\partial \Phi^2} \quad . \tag{19}$$

The corresponding system generated by split is written as

$$\frac{\partial^2 C}{\partial \Psi^2} = Q \tag{20}$$

and

$$\frac{1}{\mathrm{D}}\frac{\partial C}{\partial \Phi} + \frac{\partial^2 C}{\partial \Phi^2} = Q \quad . \tag{21}$$

Starting with Q=0 and solving equation (20) it results

$$C(\Phi,\Psi) = f_1(\Phi).\Psi + f_2(\Phi) \quad . \tag{22}$$

Replacing the former result in the equation (22), an auxiliary equation arises:

$$\frac{1}{\mathrm{D}}\frac{df_1}{d\Phi}.\Psi + \frac{1}{\mathrm{D}}\frac{df_2}{d\Phi} + \frac{d^2 f_1}{d\Phi^2}.\Psi + \frac{d^2 f_2}{d\Phi^2} = 0 \quad . \tag{23}$$

The equation above produces two new auxiliary equations:

$$\frac{1}{D}\frac{df_1}{d\Phi} + \frac{d^2 f_1}{d\Phi^2} = 0 \tag{24}$$

and

$$\frac{1}{D}\frac{df_2}{d\Phi} + \frac{d^2 f_2}{d\Phi^2} = 0 \quad , \tag{25}$$

whose solutions are obtained by direct integration:

$$f_1(\Phi) = c_1 + c_2.e^{-D\Phi} \tag{26}$$

and

$$f_2(\Phi) = c_3 + c_4.e^{-D\Phi} \quad . \tag{27}$$

Substituing (26) and (27) in (22) the first exact solution is obtained

$$C(\Phi,\Psi) = (c_1 + c_2.e^{-D\Phi}).\Psi + c_3 + c_4.e^{-D\Phi} \quad . \tag{28}$$

Since [A,B] = 0 , the process can be restarted with

$$Q = (c_1 + c_2.e^{-D\Phi}).\Psi + c_3 + c_4.e^{-D\Phi} \quad , \tag{29}$$

which is replaced on the right hand side of equations (20) and (21) . Following the same steps above mentioned, a new exact solution arises:

$$\begin{aligned} C(\Phi,\Psi) = {} & \frac{1}{6}(c_1 + c_2.e^{-D\Phi}).\Psi^3 + \frac{1}{2}(c_3 + \\ & + c_4.e^{-D\Phi}).\Psi^2 + (c_1.\Phi + c_2(-e^{-D\Phi}.\Phi - e^{-D\Phi}) - \\ & - c_5.e^{-D\Phi} + c_6).\Psi + c_3.\Phi + c_4(-e^{-D\Phi}.\Phi - \\ & - e^{-D\Phi}) - c_7.e^{-D\Phi} + c_8 \quad . \end{aligned} \tag{30}$$

Although the process can be easily continued, the former solution is suitable to simulate a wide class of water pollution problems since the shape of the domain is depends only upon the expressions for Ψ and Φ. In order to write the solution in the original variables, it becomes necessary to define the streamfunction and the velocity potential. The stream function near an arbitrary contour is given by

$$\Psi(x,y) = U_\infty \cdot y + a \arctan\left[b\left(y - m(x)\right)\right] \quad , \tag{31}$$

where a and b are numerical parameters which accounts for the mean declivity of the margins, and m(x) is the function describing (locally) the countors. The velocity potential, which is obtained by means of the Cauchy-Riemann conditions (Churchill, 1975),

$$\frac{\partial \Phi}{\partial x} = \frac{\partial \Psi}{\partial y} \tag{32}$$

and

$$\frac{\partial \Phi}{\partial y} = -\frac{\partial \Psi}{\partial x} \quad , \tag{33}$$

results

$$\Phi(x,y) = U_\infty \cdot x - \int \frac{\partial f(x,y)}{\partial x} dy + k \quad . \tag{34}$$

4. Results and conclusions

The proposed method was applied to obtain two-dimensional concentration distributions along the Guaiba lake (Figure 1), by solving equation (17), with boundary conditions of second kind imposed along the margins ($\Psi = \Psi_0$) and a boundary condition of first kind upstream, which specifies the concentration profile at $\Phi = \Phi_0$. In this water body, a~800 and b~0,01 in equation (31).

The concentration distribution for coliforms, showed in Figure 1, presents reasonably agreement with the experimental data. The mean square deviation between numerical and experimental values is about 20%, the same magnitude of the dispersion between the own measurements.

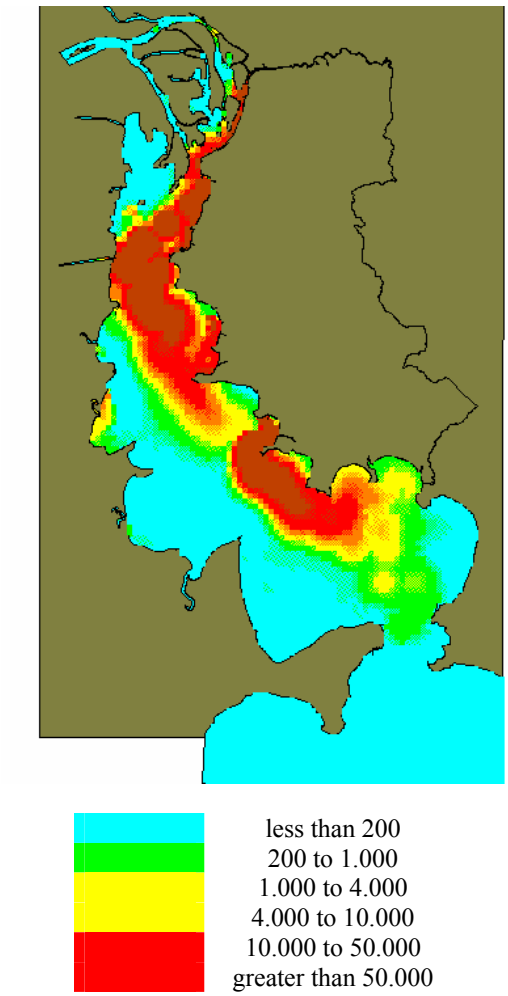


Figure 1 – Concentration distribution for coliforms (org/100ml)

Figure 2 shows the concentration distribution for phosphorus (PO_3 and PO_4 forms). In this case, the mean square deviation are roughly about 10%, which is also the same uncertainty verified between the experimental data.

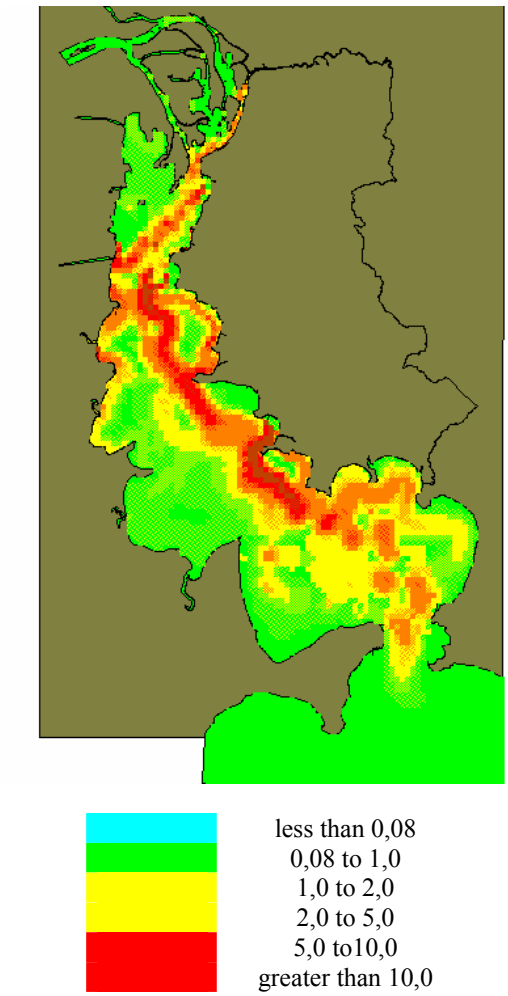


Figure 2 – Concentration distribution for phosphorus (mg/L)

In both cases, the time processing required to perform the simulations is about 5 minutes (Sempron 2.8 GHz, 512 Mb RAM, using Visual Basic 6.0).

References

Bluman, G., Kummei, S. – Symmetries and differential equations – Springer-Verlag, N. York (1989).

Chuchill, R. – Variáveis complexas e suas aplicações – Mc Graw-Hill, N. York (1975).

Ibragimov, N. – CRC handbook of Lie group analysis of differential equations – CRC press, Boca Raton (1995).

Olver, P. – Applications of Lie groups to differential equations – Springer-Verlag, N. York (2000).

Polyanin, A., Zaitsev, V. – Handbook of nonlinear partial differential equations – Chapman-Enskogg, N. York (2004).

Zwillinger, D. – Handbook of differential equations – Academic press, N. York (1992).

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

IDENTIFICATION OF UNCERTAIN WIENER SYSTEMS

Jose Figueroa [*,1] **Silvina Biagiola** [*]
Osvaldo Agamennoni [*]

[*] *Departamento de Ingeniería Eléctrica y de Computadoras, Universidad Nacional del Sur, Av. Alem 1253; (8000) Bahía Blanca, Argentina*

Abstract: A significant research work has been carried out on modeling, identification and control of processes represented by Wiener models. These models include a cascade connection of a linear time invariant system and a static nonlinearity. Several approaches can be found in the literature to perform the identification process. In this article, we describe a parametric description for the system, that allows to describe the uncertainty as a set of parameters. The proposed algorithm is illustrated through a pH neutralization process.

Keywords: Wiener Models, Process Control, Uncertainty

1. INTRODUCTION

Nonlinear model-based control has been widely diffused among the chemical engineering community. The use of models based entirely on fundamental process understanding has the advantage of possessing a clear physical interpretation. However, these models tend to be highly complex making impossible their application in popular model-based control strategies (Pottmann and Pearson, 1998).

On the other hand, purely empirical models (black-box), based entirely on input/output data, lack of physical interpretation. However, they are known to be "successful" and to have good flexibility.

A third approach is used when some physical insight is available, but several parameters remain to be determined from observed data. In this category, Pearson and Pottmann (2000), include three model structures: the Wiener model, the Hammerstein model and the feedback block-oriented model. These models are built from the combination of two components: a static (memoryless) nonlinearity $N(.)$ and a linear time invariant (LTI) system $H(z)$.

In this paper we are interested in Wiener models: a cascade connection of $H(z)$ followed by the static nonlinearity $N(.)$. The use of these models has been treated in literature in different contexts (Pearson and Pottmann, 2000; Lussón *et al.*, 2003; Biagiola *et al.*, 2004). Some representation and identification algorithms for uncertain Wiener Models will be presented. The goal is to obtain a nominal model of the process plus a parametric description of the uncertainty, which is the main contribution of this work. For this purpose, Laguerre polynomials are used to model the linear dynamic block, and a piecewise linear (PWL) representation of the nonlinear static block is provided. This modeling approach shows to be advantageous due to its simplicity, easy use and good application results. Moreover, the model

[1] Corresponding author. Email: figueroa@uns.edu.ar. Phone: +54 291 4595101 ext. 3325. FAX: +54 291 4595154. This work was financially supported by the CONICET, CIC and the Universidad Nacional del Sur.

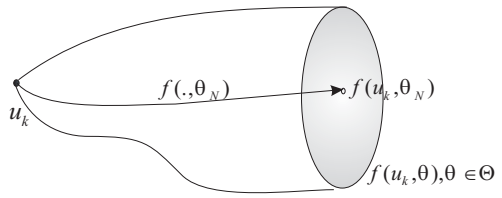


Fig. 1. Model under uncertainties

uncertainty can be easily mapped on to the model parameters.

The paper is organized as follows. In Section 2, general concepts about models and uncertainties are introduced. In Section 3 some usual descriptions and identification techniques of Wiener systems are reviewed. The proposed uncertainty model is presented in Section 4 and an algorithm for parameter uncertainty characterization is introduced. In Section 5, the results are evaluated on the basis of a simulation of a pH neutralization process. Final remarks are addressed in Section 6.

2. PROCESS INFORMATION, MODELS AND UNCERTAINTIES

Let us consider that process data are available in the form of two sets of process inputs ($\mathbf{u} = \{u_0, u_1, \cdots, u_N\}$) and outputs ($\mathbf{y} = \{y_0, y_1, \cdots, y_N\}$). Then, we aim at finding a mathematical model which approximates these data. This is performed in a two steps procedure.

In the first step, a "type model" is selected. We use the previous knowledge about the process:

$$\hat{y}_{k+1} = F(\hat{y}_k, \cdots, \hat{y}_{k-N_y}, u_k, \cdots, u_{k-N_u}, \theta) \quad (1)$$

where the predicted output at time $k+1$ depends of the previous inputs and predicted outputs and of the set of parameters (θ) to be determined.

In the second step, the parameters (θ) are computed to minimize the difference between the process and model outputs ($y_k - \hat{y}_k$) to any time. This is usually performed by minimizing the least squared error. In what follows we denote this set of parameters as *nominal parameters* θ_N.

When the interest aims at obtaining an uncertainty related with this nominal model, a typical approach is to define a set of possible models to represent all the process behaviours. This is performed by considering a set of model parameters Θ such that when these parameters $\theta \in \Theta$ are used, the whole set of exciting inputs $\mathbf{u}$ is "mapped" onto an output set which contains the set of the output data (see Fig. 1). In this way, we assume the same format for all the possible models in the uncertain set. This models family is defined in terms of a set of parameters.

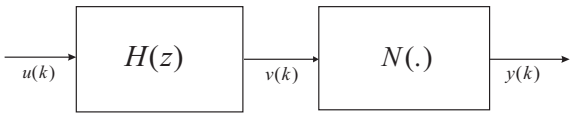


Fig. 2. The Wiener model structure.

3. WIENER MODEL IDENTIFICATION

3.1 Model Description

Figure 2 depicts a Wiener model. It consists of a LTI system $H(z)$ followed by a static nonlinearity $N(.)$. That is, the linear model $H(z)$ maps the input sequence $\{u(k)\}$ into the intermediate sequence $\{v(k)\}$, and the overall model output is $y(k) = N(v(k))$. In the following, there is no loss of generality in assuming $H(1) = 1$, since that any other value of this gain can be included in the nonlinear block (Pearson and Pottmann, 2000).

One of the most common choices for the representation of the linear block are the *Rational Transfer Functions* (Pearson and Pottmann, 2000; Figueroa *et al.*, 2004). Another usual option are the *Linear State Space Models* (Lussón *et al.*, 2003). A drawback of these models is that we need a large number of parameters to describe a system with a slow impulse response or a damped system. Alternative representations, where prior knowledge about the dominant poles can be used, are the *Laguerre and Kautz Models*. For example, the Laguerre model describes the transfer function $H(z)$ with the following basis function expansion,

$$H(z) = \sum_{i=0}^{N_L} h_i L_i(z, a) \quad (2)$$

$$L_i(z, a) = \frac{\sqrt{1-a^2}}{z-a} \left(\frac{1-az}{z-a} \right)^{i-1} \quad (3)$$

where the parameters of the model are the coefficients h_i and $a \in \Re$ is a filter coefficient chosen a priori. The nonlinear block $N(.)$ is, in general, a real-value function of one variable, i.e. $y = N(v)$. We describe the nonlinear function as

$$y = \sum_{i=0}^{N_n} \tilde{f}_i \tilde{B}_i(v) \quad (4)$$

where the basis functions $\tilde{B}_i(v)$ have been predetermined, the values $\tilde{f}_i$ are the parameters that should be computed and N_n will be referred to as "order" of the nonlinearity. Once the basis functions $\tilde{B}_i$ are fixed, the output is a linear function of the parameters. This allows us to use a linear regression to estimate the parameters. The two basic advantages of this approach are the low complexity and the uniqueness of the solution. Some possible choices for the basis functions are *Power*

Series, *Chebyshev Polynomials*, *Sigmoid Neural Networks* or *Piecewise Linear Function* (PWL). In particular, the PWL functions have proved to be a very powerful tool in the modeling and analysis of nonlinear systems. The general formulation of PWL functions allows us to represent a non-linear system through a set of linear expressions, each of them valid in a certain operation region. To make this approximation, the domain of variables $\aleph$ is partitioned into a set of σ non-empty regions $\aleph^i$, such that $\aleph = \bigcup_{i=1}^{\sigma} \aleph^i$. In each of these regions the non-linear function is approximated using a linear (affine) representation. These functions allow a systematic and accurate treatment of the approximating functions. It can be proved (Julián *et al.*, 1999) that any nonlinear continuous function $N(v) : \Re^m \rightarrow \Re^1$ can be uniquely represented using PWL functions in the form of Eq. (4) as:

$$\tilde{B}_i(v) = \Lambda\left(v, \beta_i\right) \tag{5}$$

where β_i are given parameters that define the partition of the domain of v, and Λ are functions that involve nested absolute values. In this paper we use an orthonormal description of the basis due to its local properties.

3.2 Nominal Model Identification

Different methods for Wiener models identification have been reported, and they can be grouped in three main approaches. The first one is an iterative algorithm for Hammerstein models identification (Narendra and Gallman, 1966). If the system is adequately parameterized, then the prediction error can be linearly separated into each set of parameters (the those of the linear and the nonlinear blocks). The estimation is then performed by minimizing alternatively, with respect to each set of parameters.

A second approach, based on correlation techniques (Billings and Fakhouri, 1978), relies on a separation principle, but with the rather restrictive requirement on the input to be white noise.

A recent approach for the identification of block-oriented models is based on least squares estimation and singular value decomposition (Bai, 1998). Due to the particular parameterization used, this method applies only for single input/single output systems. Gómez and Baeyens (2004) performed a more general parameterization to deal with multiple input/ multiple output (MIMO) systems. This approach will be herein followed for nominal model identification.

Let us assume that an input-output data set is available, noted as u_k and y_k, respectively. To obtain these data sets, several aspects should be taken into account. For example, the process should be persistently excited in the whole domain of the nonlinear block, such that all the relevant dynamics is captured.

From Fig. 2, the signal v_k can be written as

$$v_k = H(z) \bullet u_k, \text{ as well as } v_k = N^{-1}(y_k) \tag{6}$$

Equating both sides of these equations (with the inclusion of an error function $\epsilon(k)$ to allow for modeling error) the following equation is obtained

$$\sum_{i=0}^{N_n} f_i B_i(y_k) = h_0 l_0(u_k) + \sum_{i=1}^{N_l} h_i l_i(u_k) + \epsilon(k) \tag{7}$$

or, equivalently,

$$\epsilon(k) = \sum_{i=0}^{N_n} f_i B_i(y_k) - h_0 l_0(u_k) - \sum_{i=1}^{N_l} h_i l_i(u_k) \tag{8}$$

which is a linear regression. Defining

$$\theta = [f_0, f_1, \cdots, f_{N_n}, h_1, h_2, \cdots, h_{N_l}]^T \tag{9}$$

$$\phi = [B_0(y_k), B_1(y_k), \cdots, B_{N_n}(y_k), -l_1(u_k), -l_2(u_k), \cdots, -l_{N_l}(u_k)]^T, \tag{10}$$

Then, Eq. (8) can be written as

$$\epsilon(k) = \theta^T \phi - l_0(u_k) \tag{11}$$

Now, an estimate $\hat{\theta}$ of θ can be computed by minimizing a quadratic criterion on the prediction errors $\epsilon(k)$ (i.e. the least squares estimate). It is well known that this estimate is given by:

$$\hat{\theta} = \left(\Phi_N \Phi_N^T\right)^{-1} \Phi_N \Gamma \tag{12}$$

where $\Gamma = [-l_0(u_1), \cdots, -l_0(u_N)]^T$ and $\Phi = [\phi(1), \cdots, \phi(N)]$ are formed using the set of the N data available from the process.

Now, estimates of the parameters $\hat{f}_i$ $(i = 0, \cdots, N_n)$, $\hat{h}_0 = 1$ and $\hat{h}_i$ $(i = 1, \cdots, N_l)$ can be computed by partitioning the estimate $\hat{\theta}$, according to the definition of θ in (9). It is important to remark that we are identifying the inverse of the nonlinearity, which is frequently used in many control applications.

4. UNCERTAINTY CHARACTERIZATION

In this section we develop an algorithm, based on the ideas of Section 2, to characterize the uncertainties of the model obtained in Section 3. We introduce a set of parameters $\mathcal{H}$ for the linear dynamic block and a set $\mathcal{F}$ for the parameters of the inverse of the nonlinear block:

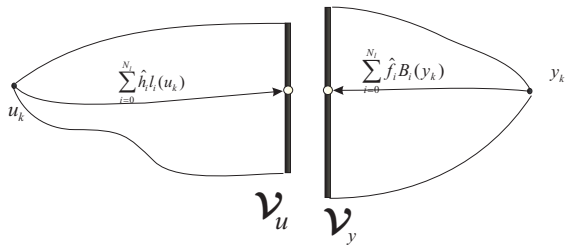


Fig. 3. Uncertainty sets in Wiener Model

$$\mathcal{H} = \left\{ h : h = \hat{h} + \delta^h, h_i^l \leq \delta_i^h \leq h_i^u \, 1 \leq i \leq N_l \right\} \quad (13)$$

$$\mathcal{F} = \left\{ f : f = \hat{f} + \delta^f, f_i^l \leq \delta_i^f \leq f_i^u \, 1 \leq i \leq N_n \right\} \quad (14)$$

To define these bounds, let us define some sets. Given the input data u_k, the linear uncertain system defined by $\mathcal{H}$ maps at some specific time k over a set

$$\mathcal{V}_u = \left\{ v : v = \sum_{i=0}^{N_l} h_i l_i(u_k), h \in \mathcal{H} \right\} \quad (15)$$

Given an input u_k, the Laguerre term of order i, $l_i(u_k)$ is a real number and the set $\mathcal{V}_u$ takes the form of $\mathcal{V}_u = \{v : v_l \leq v \leq v_u\}$.

On the other hand, if we consider the uncertain description of the parameters in $\mathcal{F}$, a given output y_k maps at some specific time k over a set

$$\mathcal{V}_y = \left\{ v : v = \sum_{i=0}^{N_n} f_i B_i(y_k), f \in \mathcal{F} \right\} \quad (16)$$

This situation is showed in Fig. 3. From this picture it is clear that the parameters set will describe the uncertainties description of Section 2 if $\mathcal{V}_y \cap \mathcal{V}_u \neq \emptyset$. In this way, the point u_k is mapped onto $\mathcal{V}_u$ through $\mathcal{H}$. Then, since $\mathcal{V}_y \cap \mathcal{V}_u \neq \emptyset$, this point will be mapped in y_k through the inverse of $\mathcal{F}$. Then, it is only necessary to compute the parameters bounds to satisfy this condition. The nominal linear model parameters $\hat{h}_i$ can be written as a vector, by considering that the Laguerre basis $l_i(u_k)$ are a set of real numbers for each input u_k. Let $l(u_k)$ be the vector which i^{th} entry is the Laguerre basis $l_i(u_k)$. Then, the expression of the linear model is

$$\hat{v}(k) = \hat{h}^T l(u_k). \quad (17)$$

In a similar way, the PWL basis $B_i(y_k)$ are a set of positive real numbers for each output y_k. $B(y_k)$ is the vector whose i^{th} entry is the PWL basis $B_i(y_k)$. Then, the linear model expression is:

$$v(k) = \hat{f}^T B(y_k). \quad (18)$$

In the following, let us analyze the bounds on the parameters.

4.1 Uncertainty concentrated in the linear block

In this case, let us assume that the uncertainty is concentrate in the linear block. Then, we are looking for the uncertain model that maps the set of data $\mathbf{u}$ to the set $\mathbf{v} = \hat{f}^T B(\mathbf{y})$. To define an uncertain model that allows to describe the complete set of data, we should compute the set $\left\{ h : h = \hat{h} + \delta^h, h_i^l \leq \delta_i^h \leq h_i^u \right\}$. Now, since that the entries of $l(u_k)$ could be positive or negative, it is possible to split the vector $l(u_k)$ by defining $l^+(u_k) = max(l(u_k), 0)$ and $l^-(u_k) = min(l(u_k), 0)$. Then, forming the vector $\gamma = \left[-(l^-(u_k))^T, (l^+(u_k))^T \right]^T$, we can compute the uncertainties bounds as

$$\min_{h^l, h^u} \sum_{i=1}^{N_l} \left(h_i^l + h_i^u \right) \quad (19)$$

s.t.

$$\left[(h^l)^T, (h^u)^T \right] \gamma \geq e(k), \text{ if } e(k) \geq 0; k = 1, \cdots, N$$

$$- \left[(h^l)^T, (h^u)^T \right] \gamma \leq e(k), \text{ if } e(k) \leq 0; k = 1, \cdots, N$$

$$h_i^l, h_i^u \geq 0$$

$$\text{where} \qquad e(k) = \hat{c}^T B(y_k) - \hat{h}^T l(u_k) \quad (20)$$

4.2 Uncertainty concentrated in the nonlinear block

In this case, let us assume that the uncertainty is concentrated in the nonlinear stationary block. Then, we are looking for the uncertain model that maps the set of data $\mathbf{y}$ to the set $\mathbf{v} = \hat{h}^T l(\mathbf{u})$. Then, to define an uncertain model that allows to describe the complete set of data, we should compute the set $\left\{ f : f = \hat{f} + \delta^f, f_i^l \leq \delta_i^f \leq f_i^u \right\}$. Now, since that the entries of $B(y_k)$ are positive, we can compute the upper bound uncertainties as

$$\min_{f^u} \sum_{i=1}^{N_n} f_i^u \quad (21)$$

$$\text{s.t.} \quad (f^u)^T B(y_k) \geq e(k), k = 1, \cdots, N$$

$$f_i^u \geq 0$$

and the lower bound as

$$\min_{f^l} \sum_i f_i^l \quad (22)$$

$$\text{s.t.} \quad -(f^l)^T B(y_k) \leq e(k), k = 1, \cdots, N$$

$$f_i^l \geq 0$$

4.3 Uncertainty in both the linear and nonlinear blocks

In this case, we consider the most general case, where uncertainty is present in both models. Note

that the intersection of the uncertainties in the linear and nonlinear models should be non empty. This can be solved as:

$$\min_{h^l,h^u,f^l,f^u} \sum_i \left(h_i^l + h_i^u + f_i^l + f_i^u\right)$$
$$\text{s.t. } \left[-(h^l)^T, -(h^u)^T, (f^u)^T\right] \begin{bmatrix} \gamma \\ B(y_k) \end{bmatrix} \geq e(k),$$
$$\text{if } e(k) \geq 0; k = 1, \cdots, N$$
$$\left[-(h^l)^T, -(h^u)^T, (f^l)^T\right] \begin{bmatrix} \gamma \\ B(y_k) \end{bmatrix} \leq e(k),$$
$$\text{if } e(k) \leq 0; k = 1, \cdots, N$$

5. PROCESS DESCRIPTION

To illustrate the identification procedure, simulation results were obtained. The example consists of the neutralization reaction between a strong acid (HA) and a strong base (BOH) in the presence of a buffer agent (BX) (Galán, 2000). The neutralization takes place in a CSTR with a constant volume V. An acidic solution with a time-varying flow $q_A(t)$ of composition $x_{1i}(t)$ is neutralized using an alkaline solution with flow $q_B(t)$ of known composition made up of base x_{2i} and buffer agent x_{3i}. For this specific case, under some assumptions, the dynamic behavior of the process can be described considering the state variables: $x_1 = [A^-]$, $x_2 = [B^+]$ and $x_3 = [X^-]$. Then, the mathematical model of the process is:

$$\dot{x}_1 = q_A/V\ x_{1i} - (q_A + q_B)/V\ x_1 \quad (23)$$
$$\dot{x}_2 = q_B/V\ x_{2i} - (q_A + q_B)/V\ x_2 \quad (24)$$
$$\dot{x}_3 = q_B/V\ x_{3i} - (q_A + q_B)/V\ x_3 \quad (25)$$
$$F(x,\xi) \equiv \xi + x_2 + x_3 - x_1 - K_w/\xi - x_3/[1 + (K_x\,\xi/K_w)] = 0 \quad (26)$$

where $\xi = 10^{-pH}$. The parameters of the system are addressed in Table 1. Using this model a set of data is generated by simulating 2000 samples with a sample time $T_s = 0.5$. A random signal uniformly distributed in $[0, 1]$ is applied to the manipulated variable q_B, this input changes each five samples. A random gaussian noise with zero media and variance 0.5 is added to the measured pH. Before proceeding with the identification, the steady values are removed from input ($q_B = 0.5$) and output ($pH = 7.7182$), respectively.

Table 1. Neutralization Parameters

Parameter	Value
x_{1i}	0.0012 $mol\ HCL/l$
x_{2i}	0.0020 $mol\ NaOH/l$
x_{3i}	0.0025 $mol\ NaHCO_3/l$
K_x	10^{-7} mol/l
K_w	10^{-14} mol^2/l^2
q_A	1 l/m
V	2.5 l

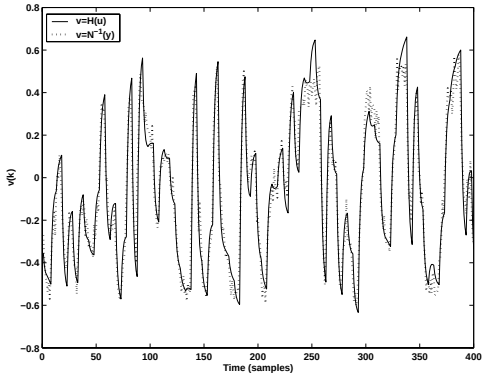


Fig. 4. Simulation for the nominal Wiener model

In a first step, we compute a nominal Wiener Model as described in Section 3. We consider three Laguerre polynomials (i.e. $N_l = 3$) with $a = 0.7$ to represent the linear model and a PWL with 8 sections partition to describe the nonlinear static gain. The identification is performed using a set of 1000 data, and the remaining data are used for validation. Figure 4 shows a set of these results, restricted to 400 samples (half for identification and half for validation). Two curves are shown: the signal $v(k)$ as the output of the linear block and as the output of the inverse of the nonlinear block $N^{-1}(y(k))$. The parameters are:

$$h^T = \begin{bmatrix} 1 & -0.2022 & 0.1386 \end{bmatrix}$$
$$f^T = [-0.660\ -0.445\ -0.416\ -0.389\ -0.374\ -0.303\ -0.042\ 0.132\ 0.204\ 0.219\ 0.557]$$

for the linear and the nonlinear blocks, respectively.

In a second step, we assume the uncertainty is concentrated in the linear block. By solving the problem described in Section 4.1, the uncertainty (see Fig. 5) in the parameters is described by:

$$h^u = \begin{bmatrix} 0.5320 & 0.120 & 0.315 \end{bmatrix}$$
$$h^l = \begin{bmatrix} 0.427 & 0.174 & 0.319 \end{bmatrix}$$

The case with uncertain nonlinear parameters is now considered. Solving the problem of Section 4.2, the parameter bounds (see Fig. 6) are:

$$f^u = [0.000\ \ 0.083\ \ 0.060\ \ 0.074\ \ 0.056\ \ 0.135\ \ 0.293\ \ 0.355\ \ 0.216\ \ 0.478\ \ 0.053\ ]^T$$
$$f^l = [0.000\ \ 0.137\ \ 0.260\ \ 0.000\ \ 0.273\ \ 0.304\ \ 0.404\ \ 0.054\ \ 0.295\ \ 0.206\ \ 0.079]^T$$

Finally, let us consider the case with uncertainty in both blocks. Solving the problem of Section 4.3, the parameter bounds (see Fig. 7) are:

$$f^u = [0.029\ \ 0.156\ \ 0.082\ \ 0.131\ \ 0.124\ \ 0.147$$

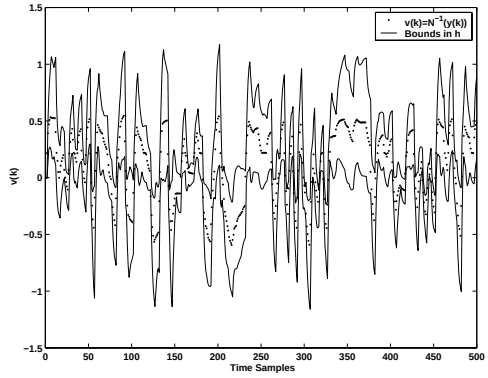


Fig. 5. Uncertainty in linear parameters

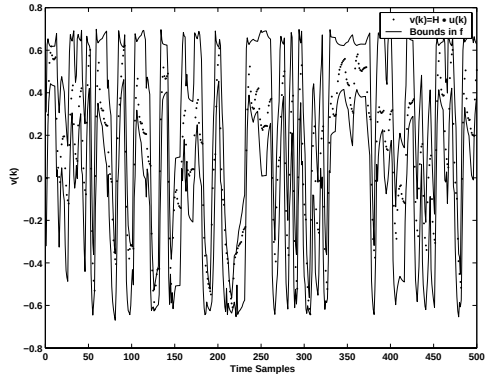


Fig. 6. Uncertainty in nonlinear parameters

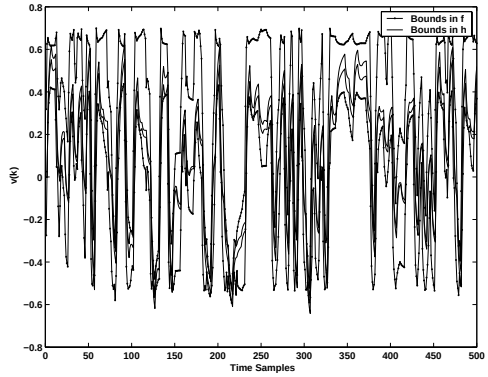


Fig. 7. Uncertainty in linear and nonlinear parameters

$$
\begin{aligned}
& 0.312 \;\; 0.341 \;\; 0.215 \;\; 0.479 \;\; 0.053]^T \\
f^l = & [0.000 \;\; 0.000 \;\; 0.174 \;\; 0.000 \;\; 0.133 \;\; 0.231 \\
& 0.342 \;\; 0.055 \;\; 0.267 \;\; 0.163 \;\; 0.106]^T \\
h^u = & \left[\, 0.000 \;\; 0.000 \;\; 0.046 \,\right] \\
h^l = & \left[\, 0.0833 \;\; 0.000 \;\; 0.000 \,\right]
\end{aligned}
$$

6. CONCLUSIONS

In this article, identification and robustness analysis of Wiener systems are considered. Different representations had been compared in terms of robust modeling capabilities. PWL functions were used to represent the nonlinear gain, with benefits due to its good approximation level. The simultaneous identification approach herein used showed a slight advantage in terms of approximation errors. These errors exhibit a linear dependence on the model parameters, which reduces the complexity of the identification formulation.

REFERENCES

Bai, E. (1998). An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. *Automatica* **34, 3**, 333–338.

Biagiola, S., O. Agamennoni and J.L. Figueroa (2004). H_∞ control of a Wiener type system. *International Journal of Control* **77, 6**, 572–583.

Billings, S. and S. Fakhouri (1978). Identification of nonlinear systems using the correlation techniques. *Proceedings IEE* **125**, 691–697.

Figueroa, J.L., J. Cousseau and R. de Figueiredo (2004). A simplicial canonical piecewise linear adaptive filter. *Circuits, System and Signal Processing* **23**, 365–386.

Galán, O. (2000). *Robust multi-linear model-based control for nonlinear plants*. PhD thesis. University of Sydney, Australia.

Gómez, J.C. and E. Baeyens (2004). Identification of block-oriented nonlinear systems using orthonormal bases. *Journal of Process Control* **14**, 685–697.

Julián, P., A.C. Desages and O.E. Agamennoni (1999). High level canonical piecewise linear representation using a simplicial partition. *IEEE Transactions on Circuits and Systems* **CAS-46**, 463–480.

Lussón, A., O. Agamennoni and J.L. Figueroa (2003). A nonlinear model predictive control scheme based on Wiener piecewise linear models. *Journal of Process Control* **13**, 655–666.

Narendra, K.S. and P.G. Gallman (1966). An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions on Automatic Control* **AC-11**, 546–550.

Pearson, R.K. and M. Pottmann (2000). Gray-box identification of block-oriented nonlinear models. *Journal of Process Control* **10**, 301–315.

Pottmann, M. and R.K. Pearson (1998). Block-oriented NARMAX models with output multiplicities. *AIChE Journal* **44, 1**, 131–140.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

A COMPARATIVE STUDY OF PREDICTION OF ELEMENTAL COMPOSITION OF COAL USING EMPIRICAL MODELLING

A. Saptoro, H.B. Vuthaluru and M.O. Tadé*

Department of Chemical Engineering, Curtin University of Technology
GPO BOX U1987, Perth, Western Australia 6845, Australia

Abstract: This paper presents empirical modelling approach in predicting elemental composition of coal. The model is developed to estimate carbon, hydrogen and oxygen content of coal. In the present work, several methods are applied to formulate the model including multiple regression (MR), principal component regression (PCR), partial least squares (PLS) and back propagation neural networks (BP-ANN). The use of BP-ANN shows the best result among the tested methods and appears to be a promising tool for predicting elemental composition of coal because it gave the least root mean square of error (RMSE) and the highest correlation coefficient (R^2).

Keywords: BP-ANN, elemental composition of coal, empirical modelling, multiple regression, PCR, PLS, ultimate and proximate analysis of coal

1. INTRODUCTION

Coal properties have many significant impacts on boiler operation and performance during coal combustion. Burning an unfamiliar fuel can reduce the efficiency of a power plant, increase pollutant emissions and, in some cases, actually damage the boiler or other system components. This can seriously affect the profitability and safety of a power plant. Power plant operators need to be confident that they can adequately know the coal properties and predict the consequences of using off-design or unfamiliar coals before they are fed into the boiler. Consequently, it is very important to provide coal properties data for power plant operators.

Unfortunately, there are some limitations of existing assessment of coal properties especially by using both conventional laboratory procedures and current on-line analysers. In laboratory procedure, chemical analysis was done on samples taken to the laboratory. Proximate analysis is relatively easy and quick to perform because it can be obtained using common laboratory equipment and is useful in practical application, however, it does not present detail information about the actual composition of coal. On the other hand, elemental analysis of coal requires highly trained analyst compared with proximate analysis, which only requires standard laboratory equipment and can be run by any competent scientist or even a skilled operator. Additionally, elemental analysis can take a day to obtain the results and in process control point of view, this traditional laboratory analysis of coal samples does not allow real time control if some adjustments need to be made to the system, for example controlling air to fuel ratio. Meanwhile, current on-line analysers of coal such as prompt gamma neutron activation analysis (PGNAA) are expensive. On-line analyser can cost from about £ 30,000 to £ 150,000 for a single parameter unit, with prices rising to as much as £ 400,000 for a prompt gamma neutron activation analysis (PGNAA) unit (Carpenter, 2002). Moreover, PGNAA still cannot be placed anywhere at the interface between the mine and the power plant (Yao et al., 2005). For these reasons, efforts must be directed to develop fast, reliable and inexpensive coal analyser which has capability for real time measurement / prediction.

One of the approaches is to establish suitable correlations to enable prediction of coal properties as a function of other available and easily obtainable coal properties. The mathematical models are developed using lab- and full-scale plant data as inputs to assess and predict other variables as on-line system outputs. Several variables exist in full scale and experimental data documented in power plant database and these data are used in the present work to develop process models. For determining elemental composition, proximate analysis can be used mainly due to its availability in power plant data base. However, due to great variability of coal properties, it is difficult to propose a suitable model, which can

* Corresponding author. Tel.: +61 8 92667581; fax: +61 8 92662681.
Email address: M.Tade@exchange.curtin.edu.au (M.O. Tadé)

represent the correlation between properties of coal. Coals are complex materials and can vary in qualities even from one mine to another or from one seam to another one. Therefore, extreme care must be taken in formulating a suitable model for representing the relationship between elemental composition and proximate analysis data.

To date, there is very limited work in the literature relevant to the elemental prediction of coal using proximate analysis data. To the best of our knowledge, only Yao et al. (2005) has developed a model for predicting hydrogen content and demonstrated the potential use of BP-ANN to tackle the difficulties in predicting elemental composition. This paper will present a comparative study of empirical modelling to predict carbon (C), hydrogen (H) and oxygen (O) content in coal using proximate analysis. Several methods are applied to formulate the model including multiple regression (MR), principal component regression (PCR), partial least square (PLS) and back propagation neural networks (BP-ANN).

2. THEORETICAL BACKGROUND

Several mathematical tools are considered to arrive at an empirical model which can predict elemental composition from proximate analysis. The description of several mathematical techniques and associated algorithms is presented below.

2.1 Multiple Regression (MR)

In simple linear regression, a dependent variable (y) is predicted from a single independent variable (x). In multiple regressions, a dependent variable is predicted from several independent variables. For predicting C, H and O content using proximate analysis (ash content (ash), volatile matter (VM), moisture content (MC) and fixed carbon (FC)), the models are formulated as follows:

$$C = \alpha_{0C} + \alpha_{1C} ash + \alpha_{2C} VM + \alpha_{3C} MC + \alpha_{4C} FC + \varepsilon \quad (1)$$

$$H = \alpha_{0C} + \alpha_{1C} ash + \alpha_{2C} VM + \alpha_{3C} MC + \alpha_{4C} FC + \varepsilon \quad (2)$$

$$O = \alpha_{0O} + \alpha_{1O} ash + \alpha_{2O} VM + \alpha_{3O} MC + \alpha_{4O} FC + \varepsilon \quad (3)$$

where $\alpha_o, \alpha_1, \alpha_2, \alpha_3 \; and \; \alpha_4$ are the model parameters and ε is the error term. As a note, sometimes, multiple regression models are developed involving interaction terms among independent variables to improve its prediction performance. Based on linear regression, one can estimate $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ and α_4 with reasonable accuracy. The estimates of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ and α_4, will be denoted as $a_o, a_1, a_2, a_3, and \; a_4$. The predicted values of C, H and O using these estimates, will be further denoted as $\hat{C}$, $\hat{H}$ and $\hat{O}$ so that

$$\hat{C} = a_0 + a_{1C} ash + a_{2C} VM + a_{3C} MC + a_{4C} FC \quad (4)$$

$$\hat{H} = a_{0H} + a_{1H} ash + a_{2H} VM + a_{3H} MC + a_{4H} FC \quad (5)$$

$$\hat{O} = a_{0O} + a_{1O} ash + a_{2O} VM + a_{3O} MC + a_{4O} FC \quad (6)$$

To get estimates for $a_o, a_1, a_2, a_3, and \; a_4$, use the values of $\alpha_o, \alpha_1, \alpha_2, \alpha_3 \; and \; \alpha_4$ that result in minimum values of the sum of squared errors (SSE). In other words, if Ci, Hi and Oi is an observed value of C, H, and O, the values of $a_o, a_1, a_2, a_3, and \; a_4$ are obtained so that the following parameters are as small as possible (Berk and Carey, 2004; Draper and Smith, 1998; Brereton, 2003)

$$SSE_C = \sum_{i=1}^{N} (C_i - \hat{C}_i)^2 \quad (7)$$

$$SSE_H = \sum_{i=1}^{N} (H_i - \hat{H}_i)^2 \quad (8)$$

$$SSE_O = \sum_{i=1}^{N} (O_i - \hat{O})^2 \quad (9)$$

2.2 Principal Component Regression (PCR)

If **X** is the matrix of predictor / independent variables (proximate analysis: ash, VM, MC and FC) and **Y** is the matrix of response / dependent variables (elemental composition: C, H or O), principal components of **X** are constructed through principal component analysis (PCA) which can be expressed as follow

$$\mathbf{X} = \mathbf{U}\mathbf{V}^{\mathbf{T}} + \mathbf{E} \quad (10)$$

where **U** is **X**-scores ,**V** is **X**-loadings and **E** is **X**-residuals. Principal component regression (PCR) takes the scores from the decomposed **X** matrix and regresses them on the dependent data set, **Y** (Beebe et al., 1998; Martin et al, 1995; Brereton, 2003).

2.3 Partial Least Square (PLS)

If **X** is the matrix of predictor / independent variables (proximate analysis: ash, VM, MC and FC) and **Y** is the matrix of response / dependent variables (elemental composition: C, H or O), the correlation

between **Y** as function of **X** usually can be described as follows

$$\mathbf{Y} = \mathbf{X}\,\mathbf{b} \quad (11)$$

where vector b contains the model coefficient. The PLS model has the form

$$\mathbf{X} = \mathbf{U}\mathbf{V}^{\mathbf{T}} + \mathbf{E} \quad (12)$$

$$\mathbf{Y} = \mathbf{W}\mathbf{Z}^{\mathbf{T}} + \mathbf{F} \quad (13)$$

The matrices on the right-hand side of these models are defined by **U**=**X**-scores, **W**=**Y**-scores, **V**=**X**-loadings **Z**=**Y**-loadings, **E**=**X**-residuals, and **F**=**Y**-residuals. The final PLS prediction model can be re-expressed as

$$\hat{\mathbf{Y}} = \mathbf{X}\boldsymbol{\beta}^{\mathbf{T}} \quad (14)$$

$$\mathbf{B}^{\mathbf{T}} = \mathbf{P}(\mathbf{P}^{\mathbf{T}}\mathbf{P})^{-1}\mathbf{Q}^{\mathbf{T}} \quad (15)$$

where $\hat{\mathbf{Y}}$ are the predictions of **Y** and **β** are the regression coefficient vectors (Ramadhan, 2005; Martin et al., 1995; Brereton, 2003).

2.4 Back Propagation Artificial Neural Networks (BP-ANN)

Let **X** be a set of n input neurons, **Y** a set of m output neurons, ξ_i real input potential and y_i real output of neuron i. The neuron's output for this type of network is defined by the equation:

$$y_i = \sigma(\xi_i) \quad (16)$$

where the activation function, σ, maybe linear, threshold, sigmoid, hyperbolic tangent or radial basis function.

The network error E(w) related to a training set is defined as a sum of square errors $E_k(w)$ of network concerning each training example and depends on the configuration of network w.

$$E(w) = \sum_{k=1}^{p} E_k(w) \quad (17)$$

$$\text{where } E_k(w) = 0.5\sum_{j\in Y}\left(y_j(w, x_k) - d_{kj}\right)^2 \quad (18)$$

Partial network error $E_k(w)$ related to the k-th Training example is directly proportional to a sum of square of difference between the real network value and the desired output where $y_j(w, x_k) - d_{kj}$ is the j-th output error for the k-th training sample. An error of zero would indicate that all the outputs pattern computed by the neural network perfectly match the expected values and the network is well trained (Haykin, 1999, Bulnová and Kostúr, 2003, Pham, 1995).

3. RESULTS AND DISCUSSION

Four methods are used to develop elemental composition predictor of C, H and O content in coal using proximate analysis. The total data set sizes are 167 x 4 for **X** and 167 x 1 for **Y**. The data from different countries and mines (Pisupati et a.l, 1992; Furimsky et al., 1990; Artos and Scaroni, 1993; Peralta et al., 2001; Coimbra et al., 1993; Fan et al., 1999; Visona and Stanmore, 1997; Armesto et al., 2003; Bailey et al., 1990; Peralta et al, 2002; Lockwood et al., 1998; Brewster et al., 1995; Su, 1999, Carlson, 1996; McLennan et al., 2000; Guo et al., 1997; Charland et al., 2003) were considered for developing and testing the models. 75 % of the data is used to develop the models and a quarter of the total data is used as independent data for testing the models. The range of data used for training and testing is presented in Table 1.

Table 1 Proximate analysis and elemental composition of coals (%)

	Average	Range
Proximate analysis		
Ash	11.33	0.5 - 40.4
VM	30.34	2.6 - 54.9
MC	6.76	0.1 - 36.8
FC	52.72	23.6 - 87.6
Elemental Composition		
C	74.07	40.6 - 94.6
H	4.45	0.4 - 6.7
O	10.61	0.2 - 38

3.1 Multiple Regression (MR)

The multiple regression model predicts C, H and O as a linear function of ash, VM, MC and FC. All the results, root mean square of error (RMSE) and coefficient correlation of the linear fit of measured and predicted (R^2) data are summarized in Tables 2, 3 and 4.

3.2 Principal Component Regression (PCR)

PCR is done for each possibility where we generate matrices **U** and **V** into one, two, three and four components. The result for each number of components (RMSE and R^2) is listed in Tables 2, 3 and 4.

3.3 Partial Least Square (PLS)

PLS is done for each possibility where we generate matrices **U**, **V**, **W** and **Z** into one, two, three and four components. The result for each number of components (RMSE and R^2) is shown in Tables 2, 3 and 4.

3.4 Back Propagation Artificial Neural Networks (BP – ANN)

Training process is done to develop suitable models for predicting C, H and O content. In this case,

Levenberg-Marquardt algorithm is used for the learning process and constructing the topology of BP-ANN containing input layer with four nodes (ash, VM, MC, and FC), one hidden layer and one output layer with one node (C, H, or O). To find the optimum number of neurons in the hidden layer, the number of neurons was changed from 2 up to 50 and arrived at optimum results which give minimum value of RMSE and maximum value of R^2. The 4-22-1 network for predicting C, 4-10-1 network for predicting H and 4-9-1 for predicting O give the most accurate prediction for this present study. The prediction performances are summarized in Tables 2, 3 and 4.

Table 2 Summary of Prediction Performance of C

Model		Training		Test
		RMSE	R^2	R^2
MR		6.259187	0.6915	0.6762
PCR	1 comp	6.71869	0.6568	0.7866
	2 comp	6.56509	0.6672	0.776
	3 comp	6.51491	0.6741	0.7678
	4 comp	6.40771	0.6874	0.6737
PLS	1 comp	6.69081	0.6546	0.7853
	2 comp	6.48453	0.6757	0.7494
	3 comp	6.3821	0.6875	0.6776
	4 comp	6.40771	0.6874	0.6737
BP-ANN		4.99097	0.879	0.91

Table 3 Summary of Prediction Performance of H

Model		Training		Test
		RMSE	R^2	R^2
MR		0.56654	0.705	0.3969
PCR	1 comp	1.24138	0.0063	0.0055
	2 comp	0.6664	0.6628	0.4262
	3 comp	0.61319	0.6628	0.4563
	4 comp	0.58019	0.7006	0.3942
PLS	1 comp	1.18171	5E-05	0.0142
	2 comp	0.63392	0.6436	0.4288
	3 comp	0.57805	0.7005	0.3998
	4 comp	0.58019	0.7005	0.3942
BP-ANN		0.44272	0.899	0.888

3.5 Discussion

From the results of each methods used in the present work, it is obvious that BP – ANN based predictor of C, H, and O content of coal show the best result compared with other methods (the prediction performance of BP-ANN is the best). Overall, it appears that BP-ANN based model is a valuable tool to assess the coal properties for any coal-fired power plant. The residual plots for prediction of C for each method are as shown in Figures 1, 2, 3 and 4. It can be seen that the Figure 4 for BP-ANN show the most random residual plot indicating the good fit of this model compared with the other three methods.

Table 4 Summary of Prediction Performance of O

Model		Training		Test
		RMSE	R^2	R^2
MR		4.273568	0.5902	0.5118
PCR	1 comp	7.59932	0.3684	0.5101
	2 comp	4.69999	0.5137	0.4032
	3 comp	4.47801	0.5608	0.5516
	4 comp	4.34347	0.5902	0.5115
PLS	1 comp	7.28019	0.5902	0.4428
	2 comp	4.55755	0.542	0.4474
	3 comp	4.32669	0.59	0.5186
	4 comp	4.34347	0.5902	0.5115
BP-ANN		2.33773	0.939	0.894

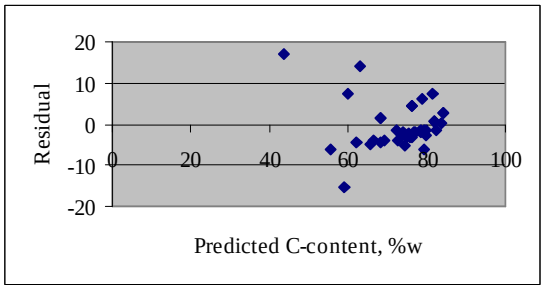


Fig. 1. Residual Plot for Prediction of C using MR

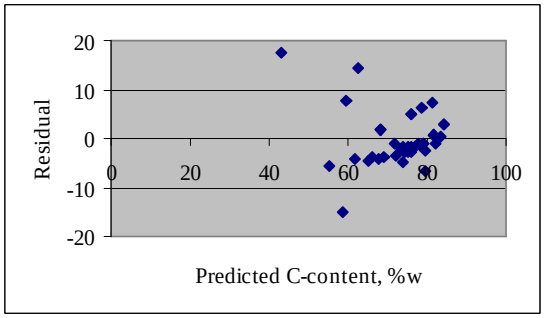


Fig. 2. Residual Plot for Prediction of C using PCR

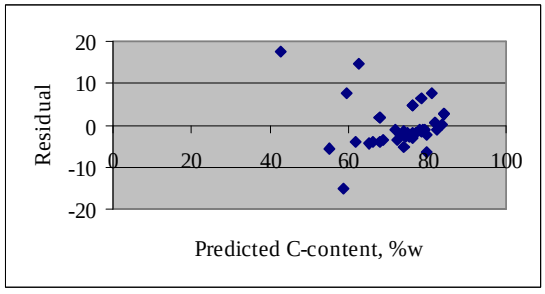


Fig. 3. Residual Plot for Prediction of C using PLS

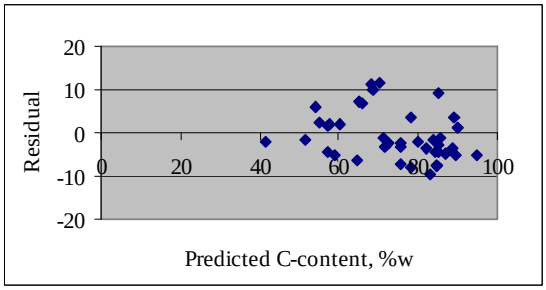
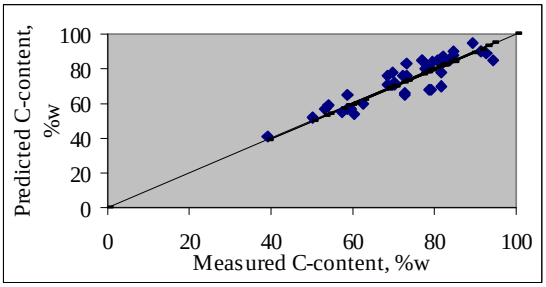


Fig. 4. Residual Plot for Prediction of C using BP-ANN

The linear fit of measured values of C, H, and O and their predicted values using BP-ANN based model are shown in Figures 5, 6 and 7. Predicted values are within ± 10 % from the measured elemental compositions.



.

Fig.5. Measured and Predicted Values of Carbon

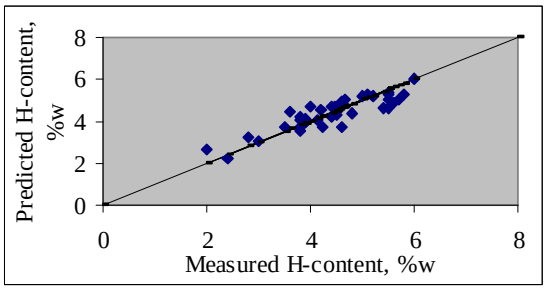


Fig.6. Measured and Predicted Values of Hydrogen

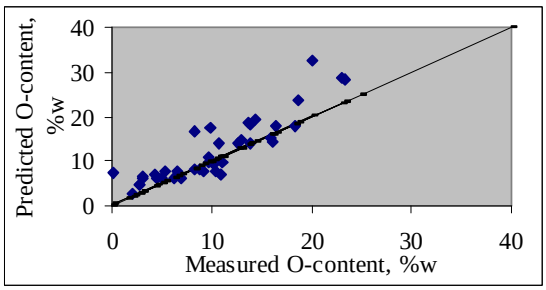


Fig. 7. Measured and Predicted Values of Oxygen

5. PRACTICAL IMPLICATIONS

Most of the coal-fired power stations receive coal from different mining operations. The coal, generally vary in quality even from the same coal mine. Moreover, due to the changing nature of the coal, off-line ultimate analysis may not be so accurate when the coal reaches the mill inlet. The challenge is to monitor the coal through the process and the quality of coal at the mill inlet is known so that the combustion can be appropriately controlled. Empirical model such as BP-ANN model would be a useful tool in this regard to provide the on-line information of elemental composition of coal, which can be used to determine the stoichiometric air requirement for various coal samples. Thus, the empirical model can provide fast and reliable prediction of elemental composition of coal to enhance performance of the combustion control system for power utilities.

6. CONCLUSION

In this paper, four empirical modelling approaches were applied to predict C, H, and O content in coal based on the proximate analysis data. The methods included multiple regression (MR), principal component regression (PCR), partial least square (PLS) and back propagation neural networks (BP-ANN). The use of BP-ANN gave the best result among the tested methods and appears to be a promising tool as elemental composition predictor. However, further improvements are needed for BP-ANN by utilizing additional data for training the model and using other learning algorithms. Also it is important to find the optimum value of number of epoch and learning rate of the networks. Furthermore, development of a good estimator for predicting complete elemental compositions of coal (C, H, N, S, and O) is also challenging which could potentially provide useful and valuable data for power plant operators.

ACKNOWLEDGEMENTS

Authors wish to acknowledge Prof Richard Brereton and Centre for Chemometrics, School of Chemistry, University of Bristol, UK, for the permission to use multivariate analysis add-in software.

REFERENCES

Armesto, L., H. Boerrigter, A. Bahillo and J. Otero (2003). N_2O emissions from fluidised bed combustion: The effect of fuel characteristics and operating conditions. *Fuel*, **82**, 1845-1850.

Artos, V. and A.W. Scaroni (1993). T.g.a. and drop-tube reactor studies of the combustion of coal blends. *Fuel*, **72**, 927-933.

Bailey, J.G., A. Tate, C.F.K. Diessel and T.F. Wall (1990). A char morphology system with applications to coal combustion. *Fuel*, **69**, 225-239.

Beebe, K.R., R.J. Pell and M.B. Seasholtz (1998). *Chemometrics A Practical Guide*, pp. 250-335. John Wiley & Sons, Inc., New York.

Berk, K.N. and P. Carey (2004). *Data Analysis with Microsoft Excel*, pp. 294-367. Thomson Brooks/Cole, Toronto.

Brereton, R.G. (2003). *Chemometrics data analysis for the laboratory and chemical plant*, pp. 271-338. John Wiley & Sons, Chicester.

Brewster, B.S., L.D. Smoot and S.H. Barthelson (1995). Model comparison with drop tube combustion data for various devolatilization submodels. *Energy & Fuels*, **9**, 870-879.

Bulnová, A. and K. Kostúr (2003). Development of control system by Studgard neural network simulator. *Acta Montanistica Slovaca*, **8**, 217-219.

Carlson, K.E. (1996). Fossil Fuels. In: *Power Plant Engineering* (L.F. Drbal, P.G. Boston and K.L. Westra, Ed.), Chap. 4, pp. 71-123. Chapman & Hall, New York.

Carpenter, A.M. (2002). *Coal Quality Assessment – The Validity of Empirical Test*, pp. 5-21. IEA Clean Coal Centre, London.

Charland, J.P., J.A. MacPhee, L. Giroux, J.T. Price and M.A. Khan (2003). Application of TG-FTIR to the determination of oxygen content of coals. *Fuel Processing Technology*, **81**, 211-221.

Coimbra, C.F.M., J.L.T. Azevedo and M.G. Carvalho (1994). 3 – D numerical model for predicting NO_x emissions from an industrial pulverized coal combustor. *Fuel*, **73**, 1128-1134.

Draper, N.R. and H. Smith (1998). *Applied Regression Analysis*, pp. 217-231. John Wiley & Sons, Inc., New York.

Fan, J.R., P. Sun, Y.Q. Zheng, Y.L. Ma and K.F. Cen (1999). Numerical and experimental investigation on the reduction of NO_x emission in a 600 MW utility furnace by using OFA. *Fuel*, **78**, 1387-1394.

Furimsky E., A.D. Palmer, W.D. Kalkreuth, A.R. Cameron, G. Kovacik (1990). Prediction of coal reactivity during combustion and gasification by using petrographic data. *Fuel Processing Technology*, **25**, 135-151.

Guo, B., Y. Shen, D. Li and F. Zhu (1997). Modeling coal gasification with a hybrid neural network. *Fuel*, **76**, 1159-1164.

Haykin, S. (1999). *Neural Network A Comprehensive Foundation*, pp. 156-175. Prentice Hall, New Jersey.

Lockwood, F.C., T. Mahmud and M.A. Yehia (1998). Simulation of pulverised coal test furnace performance. *Fuel*, **77**, 1329-1337.

Martin, E.B., A.J. Morris and J. Zhang (1995). Artificial neural networks and multivariate statistics. In: *Neural Networks for Chemical Engineers* (A.B. Bulsari, Ed.), Vol. 6, Chap. 26, pp. 627-658. Elsevier, Amsterdam.

McLennan, A.R., G.W. Bryant, B.R. Stanmore and T.F. Wall (2000). Ash formation mechanism during pf combustion in reducing conditions. *Energy & Fuels*, **14**, 150-159.

Peralta, D., N.P. Paterson, D.R. Dugwell and R. Kandiyoti (2001). Coal blend performance during pulverised-fuel combustion: estimation of relative reactivities by a bomb-calorimeter test. *Fuel*, **80**, 1623-1634.

Peralta, D., N.P. Paterson, D.R. Dugwell and R. Kandiyoti (2002). Development of a reactivity test for coal-blend combustion: The laboratory-scale suspension firing reactor. *Energy & Fuels*, **16**, 404-411.

Pham, D.T. (1995). An introduction to artificial neural networks. In: *Neural Networks for Chemical Engineers* (A.B. Bulsari, Ed.), Vol. 6, Chap. 1, pp. 1-19. Elsevier, Amsterdam.

Pisupati, S.V., B.G. Miller and A. Scaroni (1992). Effect of blending low-grade anthracite products with bituminous coals on combustion characteristics in a bench-scale stoker simulator. *Fuel Processing Technology*, **32**, 159-179.

Ramadan, Z., P.K. Hopke, M.J. Johnson and K.M. Scow (2005). Application of PLS and Back-Propagation Neural Networks for the estimation of soil properties. *Chemometrics and intelligent laboratory systems*, **75**, 23-30.

Su, S (1999). Combustion behaviour and ash deposition of blended coals, *PhD Thesis*, p. 10. Department of Chemical Engineering, The University of Queensland.

Visona, S.P. and B.R. Stanmore (1997). Modelling NO formation in a swirling pulverized coal flame. *Chemical Engineering Science*, **53**, 2013-2027.

Wu, H., G. Bryant and T. Wall (2000). The effect of pressure on ash formation during pulverized coal combustion. *Energy & Fuels*, **14**, 745-750.

Yao, H.M., H.B. Vuthaluru, M.O. Tadé and D. Djukanovic (2005). Artificial neural network-based prediction of hydrogen content of coal in power station boilers. *Fuel*, **84**, 1535-1542.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ENERGY BASED DISCRETIZATION OF AN ADSORPTION COLUMN

A. Baaiu * **F. Couenne** * **L. Lefevre** * **Y. Le Gorrec** *
M. Tayakout *

* *Laboratoire d'Automatique et de Génie des Procédés,*
C.N.R.S. UMR 5007, Université Lyon 1
ESCPE, Bat 308 G, 43 Bvd du 11 Nov 1918
69622 Villeurbanne cedex, France
Email:name@lagep.cpe.fr

Abstract: A new method for the spatial discretization of complex multi-scale systems described by partial differential equations is presented. This method allows to preserve the global power balance equation and the geometric structure of the system. The modelling of the adsorption column is based on a network approach. The key notions are the energy function and the description of the power transfers within the system and through its boundaries with the help of a power-conserving geometric structure. The proposed discretization method preserves this geometric structure and is thermodynamically consistent.

Keywords: Energy based Modelling, Distributed parameter systems, Discretization

1. INTRODUCTION

An adsorption column is a complex system which may be mathematically described by multi-scale partial differential balance equations. It may also be modelled using a network approach. This approach, which is an extension of the infinite dimensional port based modelling approach, consists in splitting each phenomena into atomic elements with particular energetic behaviors. Then, these atomic elements are connected via an *interconnection* structure which characterizes energy exchanges within the system and through its boundaries. This kind of modelling presents many advantages compared to the classical PDEs approach. First of all, each *atomic* element is well characterized from an energetic point of view. The interconnection between these *atomic* elements is done using power conjugated variables, named the port variables. The use of these port variables makes the interconnection between elements from different physical domains consistent. As a consequence, the second advantage of such modelling is the modularity that it offers. Submodels and laws can be changed without taking into account problems linked to the interconnection : causality, consistency of the port variables etc ... Finally the third advantage that particularly interests us and that is the center of interest of this paper is the discretization method that derives directly from this formalism. This spatial discretization method preserves the energetic behavior of each subsystems, the geometry of the energy flows and the global power balance.

The main phenomena that occurs within the column (diffusion, mixing and convection) can be represented using a conservative part, a dissipative part and an interconnection structure named "Dirac structure" which is also used in the Port Hamiltonian Systems definition (Golo *et al.*, 2004; Maschke and Schaft, 2001; Schaft

and Maschke, 2002). This representation is related respectively to energy conservation, passivity and instantaneous power conservation. These underlying properties are fundamental in control theory since they may be used for stability analysis or control purposes.

Finite dimensional approximation is a fundamental concern for the control of distributed parameters complex systems. One of the difficulties is to develop a reduction method which preserves some interesting qualitative features of the original model (such as stability, passivity, etc.). In this paper, we present a method of spatial discretization for an adsorption column model. One of the interests of this method is that the Dirac structure is preserved as well as the associated global power balance equation.

In section 2, we briefly recall some basic features of port-based modelling. In section 3 we introduce the adsorption model and give its associated multi-scale geometric model. The section 4 is devoted to the presentation of the discretization method. The section 5 presents some simulation results issued from this discretization scheme.

2. PORT BASED MODELLING FOR DISTRIBUTED SYSTEMS

In this section we recall the main differences between the classical and the port based modelling approaches. The port based modelling is using a network type language. This kind of approach takes place within an unified approach for the energetic modelling of complex multidomain systems. Let us restrict the presentation to matter conservation. In this case, the general mass balance equations for species i in a 3-dimensional spatial domain V issued from the conservation laws take the familiar form:

$$\int_V \frac{\partial q_i}{\partial t} = -\int_V \mathrm{div}(N_i) \ \ ,\forall i \in \{1, \cdots n_c\} \qquad (1)$$

where n_c denotes the number of components, q_i the molar density or concentration, N_i the molar flux of the specie i going through the boundary ∂V of the domain V. A distributed source term f_i may appear in the balance equation but is omitted here as it is not useful to outline the instantaneous power preserving interconnection structure. Let denote d the exterior derivative of a differential form and note that the previous mass balance equation can be written in the local form :

$$\frac{\partial q_i}{\partial t} = -dN_i \ \ ,\forall i \in \{1, \cdots n_c\} \qquad (2)$$

The port based modelling defines the network variables according to the Gibbs equation. Let g denotes the Gibbs free energy density. We assume that this specific energy depends on the molar concentration vector,*i.e* $g = g(q)$. The time variation of the total Gibbs free energy can be written ($\delta_c g$ denotes the variational derivative of g) :

$$\frac{\partial}{\partial t}\int_V g = \int_V (\delta_c g)^T \wedge \dot{q} \qquad (3)$$

Using the mass balance equation (2) the global Gibbs equation becomes :

$$\frac{\partial}{\partial t}\int_V g = -\int_V (\delta_c g)^T \wedge dN \qquad (4)$$

and after integration by parts of the right hand term :

$$\int_V (\delta_c g)^T \wedge dN + \int_V d\left((\delta_c g)^T\right) \wedge N = \int_{\partial V} (\delta_c g)^T \wedge N \qquad (5)$$

The network variables are then defined as the pairs $(\delta_c g, dN)$ and $(d(\delta_c g), N)$. The variables $\delta_c g = \mu_i = e_1$ and $d(\delta_c g) = d\mu = e_2$ are called the *effort variables* and the variables $dN = \Phi_1$ and $N = \Phi_2$ are called the *flow variables*. The pair of effort and flow variables is called power conjugated variables as their product has the unit of a power. Consequently, the equation (5) links the power within the spatial domain and the power flux at the boundary. With these notations the interconnection structure can be written

$$\begin{pmatrix} \Phi_1 \\ \mathrm{e}_2 \end{pmatrix} = \begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix} \begin{pmatrix} \Phi_2 \\ e_1 \end{pmatrix} \qquad (6)$$

In addition to the equation (6) we define the boundary variables by :

$$\Phi_\partial = -\Phi_2|_{\partial V} \quad , \quad e_\partial = e_1|_{\partial V} \qquad (7)$$

Using these notations, we can state that $((e_1, e_2), (\Phi 1, \Phi_2), e_\partial, \Phi_\partial)$ defines a *Dirac structure* (See (Schaft and Maschke, 2002), (Maschke and Schaft, 2001) for details).

3. THE PORT BASED MODEL OF AN ADSORPTION COLUMN

Adsorption processes are multi-scale processes. If a zeolite is used as adsorbent medium, the mass transfer phenomena description may be decomposed at three different scales namely the extra-granular, macroporous and microporous scales (see Fig. 1).

In this section we present the port-based model of the microporous scale and the coupling with the macroporous scale. The two other scales models are similar and not detailed in this paper. Furthermore all the considered models are isothermal and we assume that only one of the components is diffusing at the microporous scale.

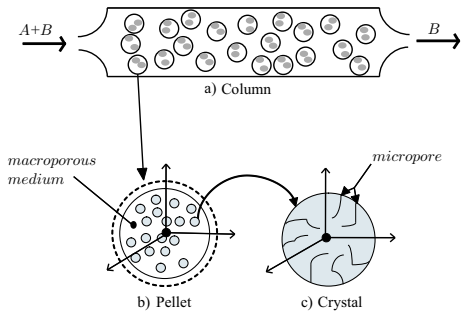


Fig. 1. Adsorption column

In the context of a three-dimensional spatial domain, we distinguish between zero-forms (functions), one-forms, two-forms and three-forms. Basically functions such as μ_i can be evaluated at any point of the spatial domain, one-forms can be integrated over every 1-dimensional curve, two-forms such as molar flux N_i (of species i) can be integrated over every 2-dimensional surface and three-forms such as concentrations q_i can be integrated on every sub-volume of the spatial domain.

In the following we consider a spherical symmetry in spherical coordinates (r, θ, ϕ). Consequently :

- the molar flux may be reduced to the 0-form (on a 1D domain) ϕ_{i2}^{mic} such that $\phi_{i2}^{mic} = 4\pi r^2 N_i^{mic}$,
- the chemical potential of species i, the 0-form μ_i^{mic} becomes the 0-form denoted by e_{i1}^{mic},
- the concentration of species i, the 3-form q_i becomes the 1-form denoted by $q_i^L = 4\pi r^2 q_i$.

Let now consider the figure Fig. 2 and first the left hand part of the picture relative to the conservation phenomenon.

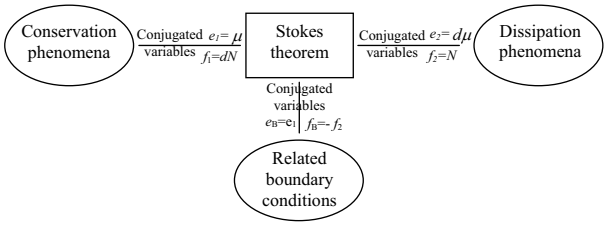


Fig. 2. Energy based model of the microporous scale

The conserved variable is the linear concentration of species i, the 1-form q_i^L, which obeys the conservation equation:

$$\frac{\partial q_i^L}{\partial t} = -div\phi_{i2}^{mic} = -\phi_{i1}^{mic} \tag{8}$$

The closure equation representing the thermodynamical equilibrium in the adsorbed scale is derived from a Langmuir's model such as ($*$ denotes the Hodge star product which, in the one-dimensional case, transforms 0-forms into 1-forms and conversely) :

$$e_{i1}^{mic} = \mu_i^0(T, P_0) + RT \ln\left(\frac{1}{P_0 k}\frac{*q_i^L}{(*q_s^L - *q_i^L)}\right) \tag{9}$$

Let now consider the right hand part of the picture of Fig. 2 that represents the diffusive phenomenon. Maxwell-Stefan law is used for representing the diffusion in the microporous scale. The only considered friction is the one exerted by the solid on species i. In this case, the Maxwell-Stefan law becomes:

$$\phi_{i2}^{mic} = -\frac{D^{mic} * q_i^L}{RT}\ \frac{\partial e_{i1}^{mic}}{\partial r} = -\frac{D^{mic} * q_i^L}{RT}\ * e_{i2}^{mic} \tag{10}$$

Equations (10) and (8) make appear the interconnection structure depicted in the center of the figure Fig. 2 and defined by :

$$\begin{cases} \phi_{i1}^{mic} = \dfrac{\partial \phi_{i2}^{mic}}{\partial r} \\ e_{i2}^{mic} = \dfrac{\partial e_{i1}^{mic}}{\partial r}, \end{cases} \tag{11}$$

We also define some *port variables* as :

$$\phi_B = -\phi_{i2|\partial}^{mic}, \quad e_B = e_{i1|\partial}^{mic} \tag{12}$$

The flow variables $(\phi_{i1}^{mic}, \phi_{i2}^{mic})$ and the effort variables $(e_{i1}^{mic}, e_{i2}^{mic})$ are respectively the extensive and intensive variables. $(\phi_{i1}^{mic}, e_{i1}^{mic})$ and $(\phi_{i2}^{mic}, e_{i2}^{mic})$ are two couple of power conjugated variables. This interconnection structure is power preserving and makes the link between the energy within the spatial domain and the boundary power flows. In the case of crystal, the power flux at the boundaries is composed with power flux in the center of the crystal and power exchanges with the macroporous medium.

The coupling between microporous and macroporous scales is done using two kinds of variables, the intensive and extensive variables. The coupling relation between the intensive variables is derived from the assumption of local equilibrium at the interphase between the two scales. This leads to:

$$\mu_i^{mic}(x, z)_{|z \in \partial V^{mic}} = \mu_i^{mac}(x) 1_{\partial V^{mic}(x)}(z) \tag{13}$$

where $1_{\partial V^{mic}(x)}(z)$ denotes the function taking the value 1 if $z \in \partial V^{mic}(x)$, 0 else. Another coupling relation is defined on the conjugated extensive variables, the volumetric density flux variable at the macroporous scale $f_i^{mac}(x)$ and the flux variable of microporous scale $N_i^{mic}(x, z)_{|z \in \partial V^{mic}}$ restricted to the boundary of its domain. This coupling relation between the extensive variables

is the volumetric mass balance equation which expresses the continuity of molar flux at the boundary of the two scales at the point $x \in V^{mac}$:

$$f_i^{mac}(x)+\left(\int_{\partial V^{mic}(x)} N_i^{mic}(x,z)\mathrm{d}S(z)\right).\rho_p(x)=0 \tag{14}$$

with $\rho_p(x)$ the *volumetric density* of crystals in the pellet.

4. DISCRETIZATION

We shall follow the discretization procedure based on a mixed finite element method and adapted to Port Hamiltonian systems in (Golo *et al.*, 2004). The purpose is to preserve the energetic behavior of each basic element of the figure Fig. 2. For that purpose, we propose appropriated interpolation functions for both effort and flow variables.

4.1 Approximation of flows and efforts

In the sequel we shall derive a discretized power conserving structure for a finite element defined on some radial interval $\mathcal{R}=[a,b]\subset Z=[0,R_{mic}]$. Hence the port variables of such a finite element are:

$$\begin{array}{ll} e_\partial^a(t) = e_1(t,a) & e_\partial^b(t) = e_1(t,b) \\ \phi_\partial^a(t) = -\phi_2(t,a) & \phi_\partial^b(t) = -\phi_2(t,b) \end{array} \tag{15}$$

The exchange of power between the element and its environment takes place in the port at the spatial boundary of the element.

The variables defined around the power conserving structure are the 1-forms ϕ_1 and e_2, and the 0-forms ϕ_2 and e_1. Let us define the following approximation of the one-forms ϕ_1 and e_2 :

$$\begin{cases} \overline{\phi_1}(t,r)=\phi_1^{ab}(t)\omega_1^{ab}(r) \\ \overline{e_2}(t,r)=e_2^{ab}(t)\omega_2^{ab}(r) \end{cases} \tag{16}$$

where $\omega_1^{ab}(r)$ and $\omega_2^{ab}(r)$ are one-forms satisfying:

$$\int_a^b \omega_i^{ab}=1 \ \text{ for } i=1,2 \tag{17}$$

The 0-forms ϕ_2 and e_1 are approximated by:

$$\begin{cases} \overline{e_1}(t,r)=e_1^a(t)\omega_1^a(r)+e_1^b(t)\omega_1^b(r) \\ \overline{\phi_2}(t,r)=\phi_2^a(t)\omega_2^a(r)+\phi_2^b(t)\omega_2^b(r) \end{cases} \tag{18}$$

where the 0-forms satisfy :

$$\begin{array}{l} \omega_i^a(a)=1, \ \ \omega_i^a(b)=0, \\ \omega_i^b(a)=0, \ \ \omega_i^b(b)=1, \ \ \text{for } i=1,2 \end{array} \tag{19}$$

in order to satisfy to the boundary conditions (15).

4.2 The discretization of the power conserving structure

In the following, we propose to discretize the power conserving structure depicted at the center of the figure Fig.2. Let us first recall the constitutive relation of the conservative structure defined in (11) and associated with the exterior derivative :

$$\begin{cases} \phi_1 = d\phi_2 \\ e_2 = de_1 \end{cases} \quad \begin{cases} e|_\partial = e_1|_{\partial\mathcal{R}} \\ \phi|_\partial = -\phi_2|_{\partial\mathcal{R}} \end{cases} \tag{20}$$

The approximations of equalities in (20) gives:

$$\overline{\phi_1}=d\overline{\phi_2} \quad \overline{e_2}=d\overline{e_1} \tag{21}$$

$$\overline{e}|_\partial=\overline{e_1}|_{\partial\mathcal{R}} \quad \overline{\phi}|_\partial=-\overline{\phi_2}|_{\partial\mathcal{R}} \tag{22}$$

Now let us introduce the approximation formulas (16) and (18) in (21) and integrate along the interval $[a,b]$ the resulting equations. Thank to (17) and (19) the following equations summarizing the discretized interconnection are obtained:

$$\begin{bmatrix} e_\partial^a \\ e_\partial^b \\ \phi_\partial^a \\ \phi_\partial^b \\ \phi_1^{ab} \\ e_2^{ab} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 \\ -1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} e_1^a \\ e_1^b \\ \phi_2^a \\ \phi_2^b \end{bmatrix} \tag{23}$$

Moreover it implies some choice for the forms of the approximations:

$$\begin{array}{ll} d\omega_2^a(r)=-\omega_1^{ab}(r) & d\omega_2^b(r)=\omega_1^{ab}(r) \\ d\omega_1^a(r)=-\omega_2^{ab}(r) & d\omega_1^b(r)=\omega_2^{ab}(r) \end{array} \tag{24}$$

In order to insure the power conservation of the structure we have to define the internal approximation variables $\phi_2^{ab}(t)$ and $e_1^{ab}(t)$ such that the approximated power relation be expressed in the following way:

$$\begin{array}{l} P_{ab}(t)=\int_{ab}\phi_1^{ab}(t)e_1^{ab}(t)+ \\ \int_{ab}\phi_2^{ab}(t)e_2^{ab}(t)+[e_\partial(b)\phi_\partial(b)-e_\partial(a)\phi_\partial(a)] \end{array} \tag{25}$$

After computation, the approximated power is given by:

$$\begin{array}{l} \overline{P}_{ab}(t) = \left(\overline{e}_1(b)\overline{\phi}_2(b)-\overline{e}_1(a)\overline{\phi}_2(a)\right)+ \\ \quad \left[\overline{e}_\partial(b)\overline{\phi}_\partial(b)-\overline{e}_\partial(a)\overline{\phi}_\partial(a)\right] \end{array} \tag{26}$$

With relations given in (25) and (26), it appears that a general choice for the internal variables can be proposed :

$$e_1^{ab}=\alpha_{ab}e_1^a+\beta_{ab}e_1^b \quad \phi_2^{ab}=\gamma_{ab}\phi_2^a+\delta_{ab}\phi_2^b \tag{27}$$

Taking into account the two last relations of (23), it appears that $\alpha_{ab}+\beta_{ab}=1$, $\gamma_{ab}=\beta_{ab}$ and $\delta_{ab}=$

α_{ab} . In order to obtain a balanced discretization, we will choose for simulation results $\alpha_{ab} = \frac{1}{2}$.

The elimination of e_1^a, e_1^b, ϕ_2^a and ϕ_2^b thank to (23) and (27), the use of (22) and the fact that $\beta_{ab} = 1 - \alpha_{ab}$ permits to write :

$$\begin{bmatrix} e_1^a \\ e_1^b \\ \phi_2^a \\ \phi_2^b \end{bmatrix} = \begin{bmatrix} 0 & \alpha_{ab}-1 & 1 & 0 \\ 0 & \alpha_{ab} & 1 & 0 \\ -\alpha_{ab} & 0 & 0 & 1 \\ 1-\alpha_{ab} & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \phi_1^{ab} \\ e_2^{ab} \\ e_1^{ab} \\ \phi_2^{ab} \end{bmatrix} , \begin{bmatrix} e_\partial^a \\ e_\partial^b \\ -\phi_\partial^a \\ -\phi_\partial^b \end{bmatrix} = \begin{bmatrix} e_1^a \\ e_1^b \\ \phi_2^a \\ \phi_2^b \end{bmatrix} \tag{28}$$

One can show that the previous approximation of the initial power conserving structure remains power conserving. This discretized power conserving structure gives the expression of the boundary variables in function of the discretized internal port variables.

4.3 The discretization of thermodynamical properties

We focus our attention on the right hand side of the figure Fig. 2 relative to the dissipation. Since the linear concentration q^L belongs to the same space as the flow ϕ_1 (a 1-form on the spatial domain), its approximation has to be chosen as :

$$\overline{q}^L(t,r) = -\Psi_1^{ab}(t)\omega_1^{ab}(r) = n^{ab}(t)\omega_1^{ab}(r) \tag{29}$$

with

$$\frac{d\Psi_1^{ab}(t)}{dt} = \phi_1^{ab}(t) = -\dot{n}^{ab}(t) \tag{30}$$

The energy on the element [a,b] is defined as : $H_C = \int_0^t \left(\int_{ab} \dot{q}^L(t,r)\mu(t,r) \right) dt$. Equation (9) gives the thermodynamical law in the adsorbed scale.

Using the approximation and integrating along the interval $[a,b]$, the approximated energy on the considered volume of micropore is given by :

$$\overline{H}_C = \int_0^t \dot{n}^{ab}(t) \left(\mu^0(T,P_0) + RT \ln \left(\frac{n^{ab}(t)}{P_0\ k\ (n_s^{ab} - n^{ab}(t))} \right) \right) dt$$

Taking into account the expression of the approximated effort $\overline{e}_C(t,r)$, which is a 0-form, one obtains the following constitutive relation for the chemical potential internal variable :

$$e_1^{ab} = \mu^0(T,P_0) + RT \ln \left(\frac{n^{ab}(t)}{P_0\ k\ (n_s^{ab} - n^{ab}(t))} \right)$$

Finally we obtain a relation for $\overline{H}_C$ which is consistent with the fundamental relations of the thermodynamics :

$$e_1^{ab} = \frac{\partial \overline{H}_C}{\partial n^{ab}}. \tag{31}$$

4.4 The discretization of the diffusion equations

The discretized constitutive relation defining the flux due to diffusion may be obtained in an analogous way. Consider the power associated with the diffusion in the microporous medium :

$$P_R = \int_{ab} e_R \phi_R$$

and compute its expression in the discretized variables :

$$\overline{P}_R = \int_{ab} \overline{e}_R \overline{\phi}_R = -\frac{K_{ab} D}{RT} n^{ab}(t) \left(e_2^{ab}(t) \right)^2$$

with $K_{ab} = \int_{ab} *\omega_1^{ab}(r) * \omega_2^{ab}(r)\omega_2^{ab}(r)$. $q^L(t,r)$ having been approximated by $\overline{q}^L(t,r) = n^{ab}(t)\omega_1^{ab}(r)$ one can deduce, by identification of the power variables, the discretized flux ϕ_2^{ab} :

$$\phi_2^{ab} = \frac{\partial \overline{P_R}}{\partial e_2^{ab}} = -R_{ab} e_2^{ab}(t) \tag{32}$$

with $R_{ab}(n^{ab}) = \frac{2K_{ab}D}{RT} n^{ab}$.

5. SIMULATION RESULTS

In order to satisfy the condition (17), the one-forms ω_i^{ab} are defined by $\omega_i^{ab} = \frac{dz}{b-a}$. The zero-forms ω_i^a and ω_i^b are defined such that Eq. (15) is satisfied. So we have $\omega_i^a = \frac{b-r}{b-a}$, $\omega_i^b = \frac{r-a}{b-a}$. The spatial domains (all the scales are considered for simulation) are discretized in equal meshes in each scale (10 meshes for each scale). The discretized model is simulated with the physical parameters presented in (Jolimaitre, 1999).

This simulation is performed for a separation of mixture of two constituents, O_2 and N_2. The simulated experiment is the response of an adsorption column, initially saturated in N_2 to a steam of air. The process is initially at equilibrium which corresponds to $\mu = 0$ all along the profile (in each point in the simulated column).

In Fig. 3-a, the output of the column is initially saturated with N_2 is shown. In Fig. 3-b, the concentration of the first constituent at the first, sixth and last mesh of the extragranular phase is represented. In Fig.3-c, the concentrations in the macroporous medium attached to the last discretized mesh of the column is given. The curves correspond to the first, the sixth and the last mesh. In Fig.3-d, the concentration in the microporous medium attached at boundaries of the pellet (the last mesh) which is itself attached in the last mesh of the column.

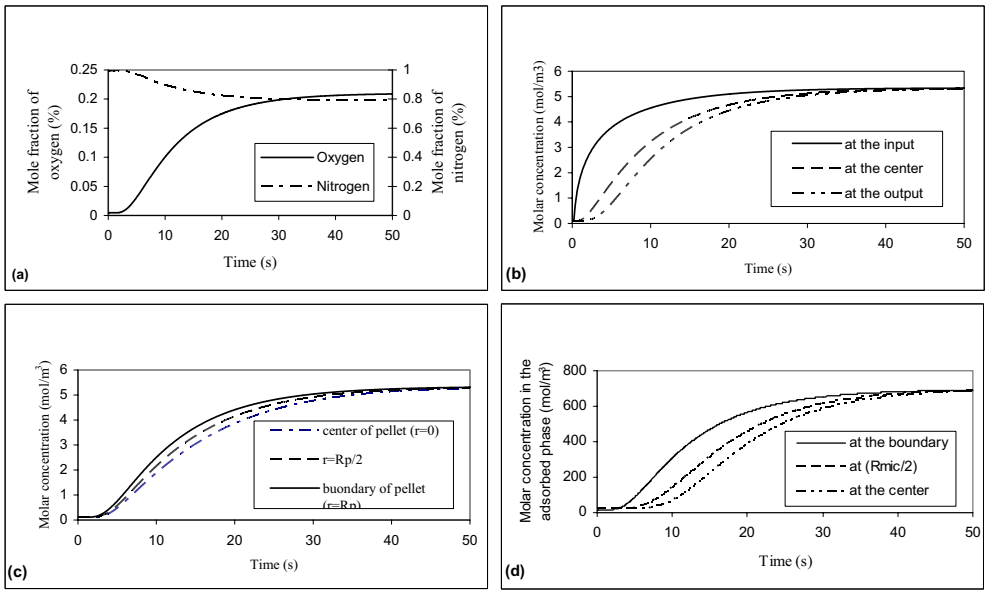


Fig. 3. Simulation results

6. CONCLUSION

In this paper we discussed the port-based modelling and spatial discretization of distributed parameter systems. In order to illustrate this approach, a model of adsorption column has been derived directly from its thermodynamical description. The modelling methodology presented exhibits some interesting features :

- The modelling is coordinate free.
- The model is a network model where each element represents a specific phenomenon which may be identified from a thermodynamics point of view.
- The instantaneous power conservation and the description of the power transfers within the system and through its boundaries are explicitely represented.

These properties of the model have several important consequences :

- The derived model requires parameters that have a clear physical meaning. This considerably simplifies the parameters estimation task.
- The model is acausal, hence postpones the choice of boundary conditions (for instance depending here on the model of the gaseous phase in the adsorption column) and is thus clearly reusable.
- The central geometric Dirac structure is a direct generalization of Poisson structure in Hamiltonian systems. It suggests and allows the use of passivity-based or energy-shaping techniques for control purposes.

These considerations strongly encourage the development of a discretization method which preserves both the nature of the interconnection structures and the physical properties of the connected elements. Such a method has been presented in this paper. Its numerical effectiveness has been established. But the key point is that we now possess a reduced model which allows a direct use of the geometric and thermodynamics properties of the PDEs model to develop estimation or control algorithms. Both the model and the discretization method apply for a large class of distributed parameters thermodynamics systems.

Acknowledgements : This work has been done in the context of the European sponsored project GeoPlex with reference code IST-2001-34166. Further information is available at http://www.geoplex.cc.

REFERENCES

Golo, G., V. Talasila, A. Van Der Schaft and B. Maschke (2004). Hamiltonian disctrization of boundary control systems. *automatica* **40**, 757–771.

Jolimaitre, E. (1999). Mass transfer and adsorption equilibrium study in MFI zeolites:aplication to the separation and debranched hydrocarbons in silicalite. PhD thesis. University of Lyon1.

Maschke, B. and A. Van Der Schaft (2001). Canonical interdomain coupling in distributed parameter systems: an extension of the symplectic gyrator. In: *Int. Mechanical Engineering Congress and Exposition , New-York, USA*.

Schaft, A. Van Der and B. Maschke (2002). Hamiltonian formulation of distributed-parameter systems with boundary energy flow. *Journal of Geometry and Physics* **42**, 166–194.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

INFERENCE OF OIL CONTENT IN PETROLEUM WAXES BY ARTIFICIAL NEURAL NETWORKS

Lima, Anie D. M.[1,2]; Silva, Danilo do C.S. [1]; Silva[1], Valéria S., De Souza Jr., Maurício B.[2]

[1]*Research And Development Center (CENPES), Lubricants and Special Products, PETROBRAS, Rio de Janeiro, Brazil*
[2]*School of Chemistry, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*

Abstract: The reduction of the time required to determine oil content is important in the production of petroleum waxes. Here, it is aimed to generate a model whose output (the inferred oil content) is obtained from inputs given by other characterization parameters (needle penetration, viscosity, density and refractive index) that are obtained from simpler experiments. Laboratory experiment data together with industrial data were employed in the modeling. These 'real' data were compared with predictions made by the linear models and artificial neural networks. The networks outperform the linear models, as they generate smaller residuals in the whole operational range considered.

Keywords: Nonlinear analysis system, Artificial intelligence, Process parameter estimation, Product quality, Neural-network models, Backpropagation algorithms

1. INTRODUCTION

Oil content in petroleum waxes is presently measured by standard experiments recommended by ASTM: ASTM D 721 (to oil content lower than 15% m/m) and D 3235 (to oil content bigger than 15% m/m). The experiments are done with complex glass apparatus and demand a lot of time. It is possible to develop correlations between physic properties and oil content to help works that need the result of this property in a short time. To do this, one of the methods is the use of artificial neural networks.

Artificial neural networks are computational technics that present a mathematic model inspired in neuron structures of intelligent organisms and that acquire knowledge from experience (Haykin, 1999). The use of neural networks depends on the ability to adapt it to the problem under consideration, by changing the synaptic weights (in the 'learning' phase) to increase efficiency.

Neural networks have been extensively used to represent non linear input-output dependencies, as it has been proved that they can approximate arbitrary well any continuous function (Funahashi; Hecht-Nielsen, 1989; Hornik, 1989).

This work comprises two kinds of investigation: experimental and modeling. In the first approach, values of needle penetration, viscosity, density and refractive index from samples of one kind of wax with different oil contents were acquired. These properties depend on composition and crystallization of wax. In the second one, those data were processed to develop linear models and neural networks in order to predict this characteristic. This technique allows the development of a calculation program to be used works in a refinery environment, so that, based on it, the operator can decide about variables of the process. It is also possible to design a control procedure that acts on the process based on inference of the model

2. SCIENTIFIC METHODOLOGY

2.1 Production and analysis of petroleum waxes

One of the processes of production of waxes is deoiling, that is, extraction of oil in waxes. The process consists in cooling slack wax until a temperature in which only waxes get solid, allowing their separation by filtration. The kind of crystallization determines if the wax will get more oil content during this process, determining the solvent consumption. Hydrocarbon waxes constituted mostly by n-alkanes (macrocrystallines), with crystals like 'dishes', have a structure easier to remove oil. Branches Waxes (microcrystallines), with crystals like 'needles', present more difficulty to remove oil in deoiling (Speight, 2001).

The excess of oil in a wax reduces hardness, and this is inconvenient to store the final product. Oil is also responsible for the appearance of spots, which is a bad characteristic when the end use product is candle. Hardness is measured by needle penetration (ASTM D1321, 2004).

Oil in a wax means that the product has some structures that have more affinity with oil than wax. These can be identified by the following experiments: viscosity (ASTM D445, 2004), density (ASTM D4052, 1996) and refractive index (ASTM, D1218, 2002) in waxes and that is why these parameters and needle penetration are important to predict oil content (Lima et. al, 2005).

2.2 Experimental methodology

Oil Content (ASTM D 721, 1985)

The sample is dissolved in methyl ethyl ketone, afterwards the solution is cooled to –32°C (–25°F) to precipitate the wax, and filtered. Evaporating the methyl ethyl ketone and weighing the residue determine the oil content of the filtrate.

Viscosity (ASTM D445)

This method specifies a procedure for the determination of the kinematic viscosity by measuring the time for a volume of liquid to flow under gravity through a calibrated glass capillary viscometer. The dynamic viscosity can be obtained by multiplying the kinematic viscosity by the density of the liquid.

Density (ASTM D4052)

This experiment covers the determination of the density or relative density of petroleum distillates and viscous oils. A small volume (approximately 0.7 mL) of liquid sample is introduced into an oscillating sample tube and the change in oscillating frequency caused by the change in the mass of the tube is used in conjunction with calibration data to determine the density of the sample.

Needle Penetration (ASTM D1321)

The depth penetrated (0.1 mm) in a cylinder of wax by a standard needle, with a load of 100g, in a specific temperature, during 5 seconds, corresponds to the 'needle penetration' measurement.

Refractive Index (ASTM D1218)

The refractive index is measured using a high-resolution refractometer of an optical-mechanical or automatic digital type with the prism temperature accurately controlled. The instrument principle is based on the critical angle concept.

2.3 Basic concepts about neural networks

A common network has multilayer configuration with parallel processing. The most used is MLP (multilayer perceptron), with an input layer, a hidden layer and an output layer (Fig. 3.).

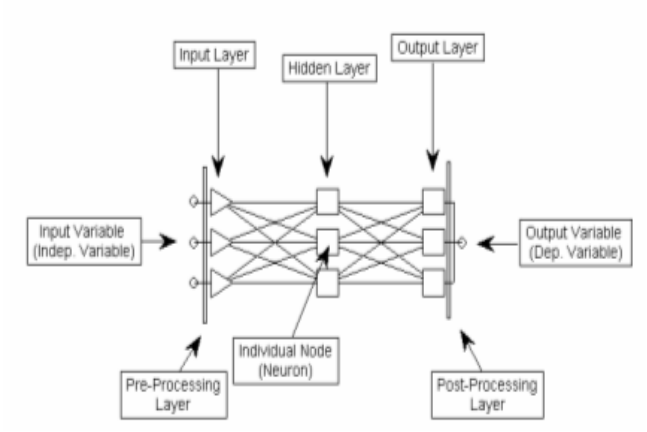


Fig. 3. Example of MLP network (De Souza Jr.,1993).

Data are fed in the input layer, which has a neuron per each input variable. Each one of the neurons in input layer is connected to each neuron of the hidden layer. Seemingly, each hidden neuron is connected to each unit of the output layer. The number of neurons in the output layer is the same number of output variables. Signals arriving on a neuron go to cell body, where they are added to others that come from other neurons of the previous layer. The 'j' neuron (Fig. 4.) from the layer (k+1) receives a set of inputs $s_{pi,k}$ (i = 1, ..., nk) corresponding to the outputs of nk neurons from previous layer. These outputs were influenced by wjik weights that correspond to each connection. The neuron sums inputs and the resultant value is added to a bias (an inner limit of activation) represented by $\theta_{j,k+1}$. The response $s_{pj,k+1}$ is produced by 'j' neuron to this signal, according to an activation function f( ) called transfer function (De Souza Jr., 1993).

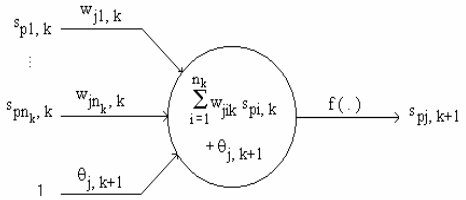


Fig. 4. The jth Neuron (De Souza Jr., 1993).

Common transfer functions are:

- Linear function:

$$f\left(\lambda_{pj,k+1}\right) = \lambda_{pj,k+1} \qquad (1)$$

- Sigmoidal function:

$$f\left(\lambda_{pj,k+1}\right) = \left[1 + \exp\left(-\lambda_{pj,k+1}\right)\right]^{-1} \qquad (2)$$

- Hyperbolic function:

$$f\left(\lambda_{pj,k+1}\right) = \tanh\left(\lambda_{pj,k+1}\right) \qquad (3)$$

The training phase of a neural network consists in giving a set of data, with inputs and outputs known, so weights and biases for each neuron of network are adjusted with training algorithms using prediction errors, until network results in correct predictions of

outputs. The procedure is iterative and continues until minimization of the global error function is reached. A second subset of data is randomly chosen for selection section or validation. These data are not used in the adjustment of weights and biases during training section, but performance of network is checked during the training with them. If error of selection data is not decreasing or begins to increase for a specified number of iterations, training is stopped. If training does not have restrictions, neural networks can describe training data very well, but usually describe new data poorly. Because of this, a third subset of data is randomly chosen as an additional check of the capacity of generalization of the neural network (De Souza Jr., 1993).

2.4 Experimental Procedure

In this section a sample of macrocrystalline wax 150/155 produced from heavy oil was utilized. The experiment of oil content was performed according to ASTM D 721 and resulted 0.69% m/m. Different fractions of heavy base oil were added to 15 portions of original sample to obtain new samples with oil content ranging from 1 to 15% m/m. The experiments of needle penetration 25ºC, refractive index 70ºC, viscosity 80ºC and density 70ºC were performed three times for each sample.
In addiction to the laboratory data, results of experiments performed on final products of a refinery like 120/125, 130/135, 140/145 and 150/155 waxes, produced by the same petroleum, shown in Table 1 were considered in the study. The complete set of results (industrial plus laboratory) is presented on Figures 5,6,7 and 8.

Table 1. Results from experiments with different kinds of final waxes from the same petroleum.

Oil Content (%m/m)	Needle Penetration (1/10 mm)	Viscosity 80ºC (cSt)	Density a 70ºC	Refractive Index
0.97	21.0	9.228	0.7895	1.4410
0.98	21.2	9.285	0.7892	1.4410
0.99	21.8	9.069	0.789	1.4405
1.47	22.5	8.149	0.7871	1.4401
1.06	25.2	8.417	0.7877	1.4402
1.06	22.0	8.369	0.7876	1.4402
1.01	22.0	8.671	0.7885	1.4404
1.07	24.0	8.348	0.7876	1.4402
0.99	20.0	9.018	0.7889	1.4408
1.09	16.0	5.254	0.7757	1.4355
0.99	16.0	5.183	0.8097	1.4352
0.91	25.0	5.322	0.7752	1.4353
0.44	15.2	5.19	0.7748	1.4339
0.54	25.0	5.449	0.7752	1.4369
0.94	43.0	5.343	0.7759	1.4344
0.54	15.0	5.141	0.7747	1.4352
0.82	20.0	5.235	0.7751	1.4341
3.01	31.6	4.655	0.7733	1.4335
2.96	43.6	4.217	0.7712	1.4320

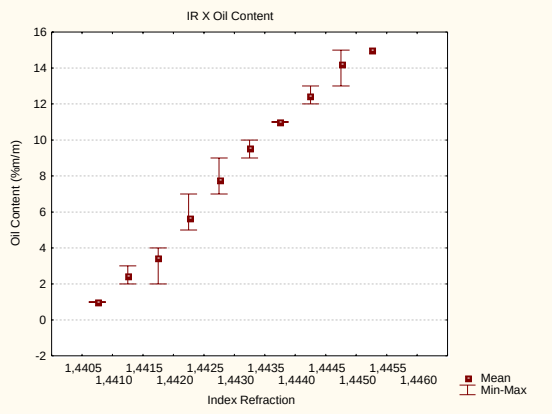


Fig. 5. Results of Index Refraction vs Oil Content

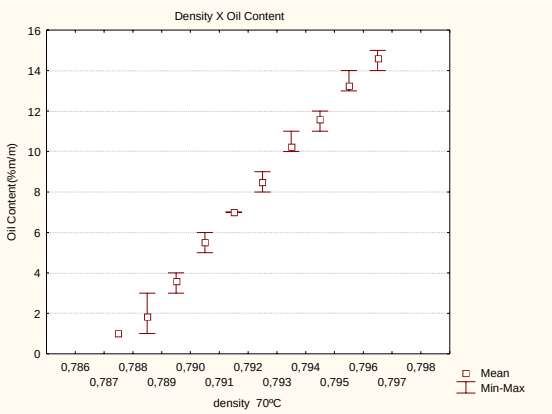


Fig. 6. Results of Density vs Oil Content

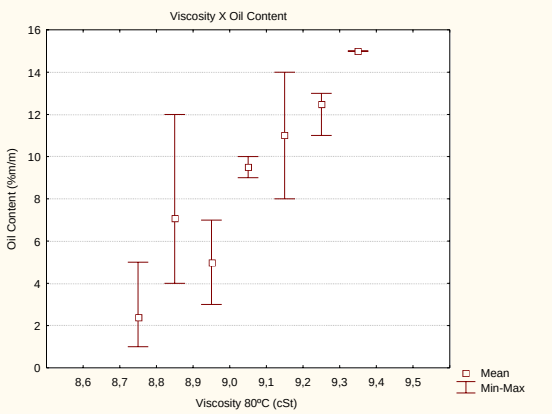


Fig. 7. Results of Viscosity vs Oil Content

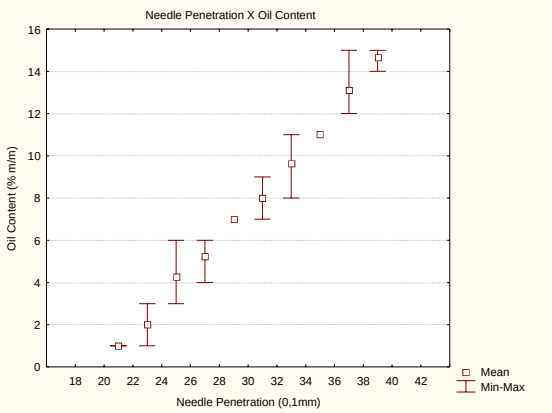


Fig. 8. Results of Needle Penetration vs Oil Content

2.5 Simple Linear Regression

Prior to linear regression, the data were analysed for autlier detection (value outside the range between + or – $2{,}5\sigma/n^{1/2}$) and no outliers where found. The observation of Figures 5 to 8 shows that the viscosity measurement has a large variability. Additionally, it is noticed that for medium and high oil content values an approximate linear dependence is observed between this characteristic and the other ones studied. So, simple linear regression models were tested first.

The simple linear regression between each variable and the oil content is presented in Figures 9 to 12.

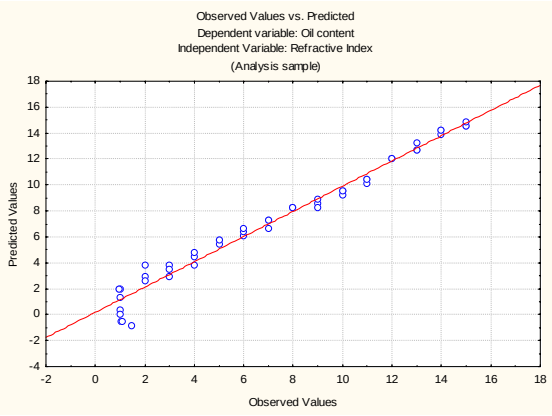


Fig. 9. Linear Regression – Refractive Index X Oil Content

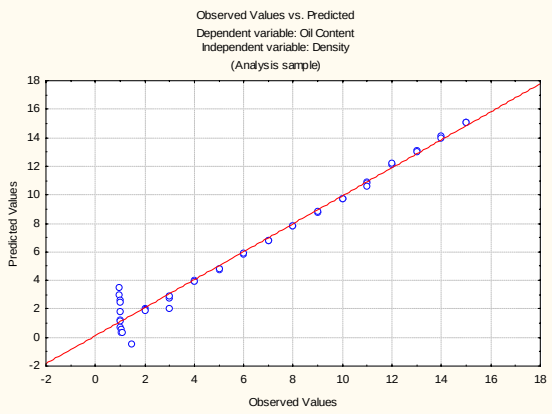


Fig. 10. Linear Regression – Density vs Oil Content

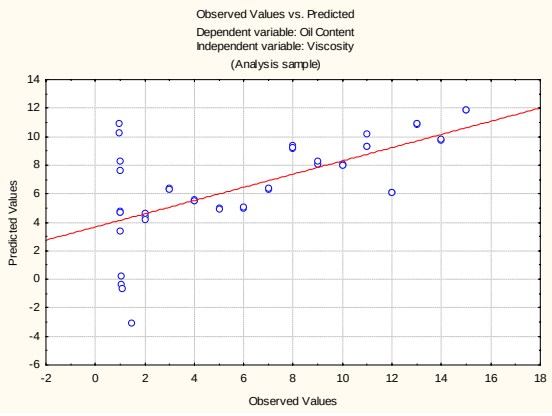


Fig. 11. Linear Regression – Viscosity vs Oil Content

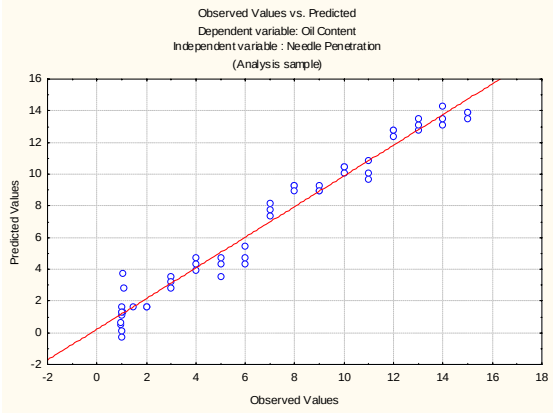


Fig. 12. Linear Regression – Needle Penetration vs Oil Content

The R-square (also known as determination coefficient) values of the linear regressions are in Table 2. This parameter indicates the percentage of the data variation that is explained by the linear model.

Table 2. R-square of linear regressions

Variables X Oil Content	
	R-square
Penetration	0.9564
Viscosity	0.6908
Density	0.9823
Refractive Index	0.9725

Viscosity has the worst result for linear analysis, due to its higher variation. For values between 0 and 6% m/m of oil content, linear predictions show the higher deviations. However, this range is the most interesting one, taking into consideration the final products. This motivated the use of non-linear models in this work. Neural networks were assumed, as they do not demand that the specific non-linear dependence is explicitly described.

2.6 Building neural network

The complete set of data was randomly separated in three subsets (in proportion 2:1:1) for training, selection (validation) and test (a second independent validation).

The software S TATISTICA© 6.0 was used to select the best network, based on data training and validation, through statistical analysis of the residuals of the predictions.

3. DISCUSSIONS AND RESULTS

A descriptive statistic analysis can be observed on Table 3. SD ratio is the ratio between standard deviation of difference between predict values and outputs, and Standard deviation between outputs and its average, that is, ratio between prediction deviation and deviation of real data from its average.

The best regression model is the one with lower SD ratio and networks with the best performances has this value closer to 0. An MLP with performance ranked as 'Excellent' was obtained, that uses backpropagation method of calculation, with an input layer operating with 4 neurons and linear functions, 9 neurons in hidden layer with hyperbolic functions and 1 linear neuron in output layer. The performance of non-linear network was compared with the best results obtained from the best linear network, with 4 input neurons and one output, as shown in Table 3.

An observation of SD ratio indicates that the networks can be used to predict the property proposed and MLP network is better than linear, as the MLP network has a SD ration five times lower than the linear model.

Table 3. Results from Statistica© of statistics analysis of networks

Descriptive Statistics	Linear	MLP
Average Error	0.2472	-0.04306
SD Error	1.465	0.3129
Absolute Average Error	0.9363	0.2544
SD Ratio	0.3053	0.06517
Correlation	0.9525	0.9979

In Table 4, it is possible to see the performance parameters for training, selection, and test that represent SD ratios for those sections. The MLP network has lower values to SD ratios and errors, proving itself as more adequate to be used for prediction.

Table 4. Results from Statistica© for training , selection and test of networks.

Training Summary- Performances						
Parameter	**Train**	**Selection**	**Test**	**Error Train**	**Error Selection**	**Error Test**
MLP	0.06162	0.05458	0.09213	0.02151	0.02449	0.02541
Linear	0.1840	0.4667	0.2866	0.06054	0.1610	0.09181

In Figure 12 it is possible to see the results of predicted versus observed values for MLP and Linear networks. The straight line to MLP shows that predictions are satisfactory. The linear model presents very high deviations at low oil content values, which makes it unadvisable to use in that range.

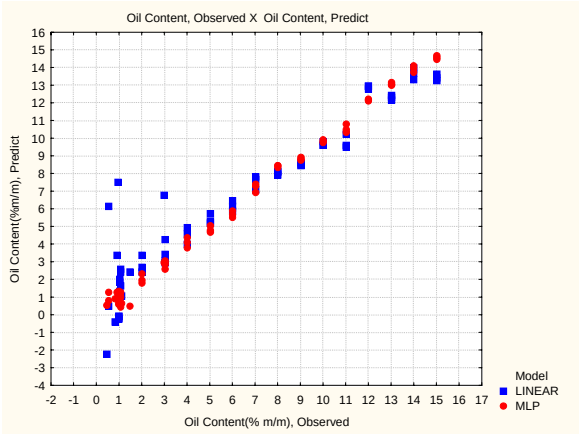


Fig.12. Predict oil contents X observed oil contents

The sensitivity analysis of inputs shows the relative contribution of each variable. Each variable is treated as if it were unavailable for analysis, being substituted by its average value. Global error of network when the variable is not available is divided by global error when it is available, resulting a ratio that should be bigger than 1.0, if variable contributes to the solution of the problem. The results are in Table 5. All variables have influence on oil content, receiving rank 1st the most influent variable.

Table 5. Results of sensitivity analysis from Statistica© for networks

Sensibility Analysis				
	Needle Penetration	**Viscosity**	**Density**	**Refractive Index**
MLP Ratio	4.287	11.51	7.855	9.258
MLP Rank	4 th	1st	3rd	2nd
Linear Ratio	1.779	3.070	1.034	4.895
Linear Rank	3rd	2nd	4th	1st

4. CONCLUSIONS

It is possible to see that the trained neural network makes good predictions to the set of data obtained from the proposed experiments and to the set of results obtained from final products of a refinery. The model proposed of a multilayer network as a MLP (4-9-1), with hyperbolic functions in the hided layer, presented correlation 0.9979 against 0.9525 of the linear model, beyond of best train, selection and test performances.

Additionally, as neural networks may outfit the data, special care was taken here on order to avoid this risk by using two validation (selection and test) data sets

The linear models can be considered as satisfactory if the range of oil content is above 6 % m/m. However for values between 0 to 6 %, linear models produce very large errors (even negative values may be obtained).
As this low range is important for the final product, the ANNs reveal itself as the best option to infer oil content from characteristics that may be obtained through more rapid experiments.

The neural network can be converted in a program written in C++, and interfaced, using a man machine interface, to the operator, so that he or she can use to get the result of oil content from the inputs considered. Additionally, a controller can be implemented to act on process by changing variables based on that inference.

The characteristics of petroleum processed to production of waxes have influence on the physical properties used to the prediction context proposed in this paper. This study was performed with a product of one kind of petroleum. To another kind of petroleum, it would be necessary to retrain the neural network including the new data.

REFERENCES

Baughman, D. R. and Y. A. Liu (1995); *Neural Networks In Bioprocessing and Chemical Engineering,* Academic Press, San Diego.

De Cerqueira, E. O, J. C. De Andrade and R. J. Poppi (2001); *Redes Neurais e suas Aplicações em Calibração Multivariada*, Quim. Nova, **Vol. 24**, **No. 6**, 864-873, 2001.

De Souza JR., M.B (1993); *Redes Multicamadas Aplicadas a Modelagem e Controle de Processos,* D.Sc. Dissertaion, PEQ/COPPE/UFRJ (in portuguese).

Funahashi, K.; *On the Approximate Realization of Continuous Mappings by Neural Networks*, Neural Networks, **2**, 183-192.

Haykin, S. (1999). *Neural networks – A comprehensive foundation*, Prentice Hall, Upper Saddle River, New Jersey.

Hecht-Nielsen, R.; *Theory of the Backpropagation Neural Network*, IEEE Int. Conf. on Neural Networks, **Vol. I**, 593-605, San Diego, 1989.

Hornik, K.; M. Stinchcombe & White; *Multilayer Feedforward Networks are Universal Approximators*, Neural Networks, **2**, 359-366, 1989.

Lima, A. D. M. & N. F. Pereira (2005); *Manual de Parafinas - Qualidade e Toxicologia*; Workshop de Avaliação dos Objetivos da Produção, Petrobras, CENPES/PDAB/LPE, 2005.

Speight, James G. (2001); Handbook of Petroleum Analysis, John Wiley & Sons.

Stipanovic, A. J.; G. P. Firmstone and M. P. Smith (1997); *Having Fun With Base Oils: Predicting Properties Using Neural Networks; Symposium on Worldwide Perspectives on the Manufacture*, Characterization and Application of Lubricant Base Oils, American Chemical Society, San Francisco, CA, April-1997

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

SHORT AND LONG TIMESCALES IN RECYCLES

Heinz A Preisig*
Department of Chemical Engineering
Norwegian University of Science and Engineering
N-7491 Trondheim, Norway

Abstract

A new awareness on modelling is growing in the control-oriented community recognising the fact that control is dominantly model based. Since control is about manipulating certain characteristics of the plant, it is no surprise that modelling for control focuses on extracting exactly those characteristics of the plant that are to be controlled. This invariably induces the use of time scale assumptions and consequently model reduction methods. These assumptions lead to a time-scale separation, which results in a layered control structure, with the control loops getting slower as one moves upwards in the hierarchy of time-scales.
Recycle structures are very common. The components may be fast, but the overall structure including the loop is usually much slower because of the recycle. These structures thus lend themselves to the application of time-scale assumptions. We demonstrate that any of these structures can be analysed. Starting with a first-principle based representation that makes no particular assumption on the nature of the process except that of a large recycle and fast internal dynamics, we derive a first-order approximation of a system. The result is generic and not dependent on the particular nature of the individual processes other structural properties of the process.

Keywords: Computer-aided, modelling, model reduction, process systems engineering, control

Background

One observes that currently larger and larger systems are being controlled and since the control methods are increasingly model based, the dimensionality of the model becomes a serious issue [2]. In many cases, whilst the control algorithms are available, computing is not up to solve the thus-formulated problems in real-time. It is also observed that one often gets quite satisfactory results from low-order models, which brings about the thought that one should be able to extract the control-relevant dynamics from the complex models and use the reduced-order model instead for control [1]. Feedback makes control rather robust to a certain class of modelling errors. Order of the approximation is one of them, if one does not insist on very fast control. The consequence of this thinking leads in recent years to a revitalisation of model reduction based on time-scale assumptions.

Time-scale assumptions are done all the time when modelling processes. Mostly assumptions seem to just "occur" – they are mostly done intrinsically, for example one makes the assumption of an ideally-stirred tank reactor, which in terms of time scales implies that the internal flows are much faster than the flows in and out of the tank. For the illustrative example, we shall use an abstraction introduced by this group over the past years as part of the Modeller project [4, 5, 7] to demonstrate that any recycle process can be captured in this framework. Thus models, such as they are published in for example [3], which also motivated this derivation, fold into the discussion below and in terms of assuming event dynamics for reactions into [4] thus covering the two important domains of time scale assumptions made in process engineering.

*E-mail: Heinz.Preisig@chemeng.ntnu.no
web: http://www.chemeng.ntnu.no/~preisig/

Proposition

Processes with (multiple) internal recycles exhibit in the large time scale uniform intensities in the recycle and in the fast time scale, a change in the input will result in an internal profile of the intensities. The time scales introduced are related to the dynamics of the capacitive elements in the recycles[1].

More precisely, in the **large time scale**, slow changes in the continuous (positive) input will result in a uniform profile of the intensive quantities if the internal time constants are relatively small. Also, pulses of extensive quantity spread instantaneously in the recycle's capacitive elements. On the **short time scale** the effects invert: the slow variations in the positive input streams have no visible effect on the intensities of the elements in the recycles and fast injections cause a distribution in the intensities.

Process Definition

For the analysis we choose a general recycle process, which consists of a number of capacitive elements[2] in the recycles and in general any number of inflows and outflows. For the purpose of simplicity but without any limitation to the applicability of the result, though, we limit the process to one inflow and one outflow. Furthermore, we firstly limit the discussion to a single recycle process, which readily extends to a multiple, interlocking recycle process later. The choice of the model is motivated by such models as they were used in [6] but also models that are constructed in computational fluid mechanic packages. We assume that no transformation of extensive quantity (i.e. mass transformation in the form of reaction) or a very fast transformation is taking place in the plant. In the figures, the circles represent capacitive elements, here modelled as single lumped systems. The arrows mark mass flow, here for simplicity unidirectional, that is, the flow does not change direction during the viewed time period. The plant has two special elements, namely the one where the input stream enters the recycles, here labelled with 1 and the one where the outflow is attached, here shown as e. The generic element in the cycle is labelled with an i.

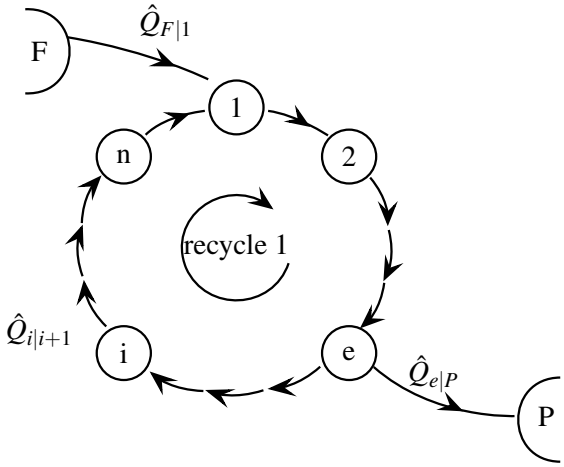


Figure 1: *A one recycle, one-input, one-output process. The recycle represents the plant being modelled. The plants environment has two elements both reservoirs, that is infinitely large capacities. The F indicates the feed reservoir, the P the product reservoir. The lump i is an arbitrary system in the recycle without an inflow or an outflow to the environment. The system 1 is where the inflow is connected and the outflow is connected at the element e.*

The basic dynamic equations are then the conservation of fundamental extensive quantities, Q for each lump, which balance the change in the system with the in and outflows of fundamental extensive quantity, $\hat{Q}$:

$$\begin{aligned} \frac{dQ_1}{dt} &= \hat{Q}_{n|1} - \hat{Q}_{1|2} + \hat{Q}_{F|1}\,, \\ \frac{dQ_e}{dt} &= \hat{Q}_{e-1|e} - \hat{Q}_{e|e+1} - \hat{Q}_{e|P}\,, \\ \frac{dQ_i}{dt} &= \hat{Q}_{i-1|i} - \hat{Q}_{i|i+1}\,. \end{aligned}$$

With the appropriate definitions these equations are cast into a matrix equation:

$$\frac{d\underline{\mathbf{Q}}_I}{dt} = \underline{\underline{\mathbf{A}}}_1 \underline{\hat{\mathbf{Q}}}_I + \underline{\underline{\mathbf{B}}}_E \underline{\hat{\mathbf{Q}}}_E\,. \qquad (1)$$

In what follows, we shall refer to the matrix $\underline{\underline{\mathbf{A}}}_1$ as the *recycle matrix*, here the one stands for recycle one. The two matrices $\underline{\underline{\mathbf{A}}}_1$ and $\underline{\underline{\mathbf{B}}}_E$ are direction coefficient matrixes for the internal flows and the external flows, respectively and represent the graph of vertices (capacities) and arcs (flows):

$$\underline{\underline{\mathbf{A}}}_1 := \left(\; -\underline{\mathbf{s}}_1 + \underline{\mathbf{s}}_2, -\underline{\mathbf{s}}_2 + \underline{\mathbf{s}}_3, \ldots, -\underline{\mathbf{s}}_n + \underline{\mathbf{s}}_1 \; \right) \quad ,$$
$$\underline{\underline{\mathbf{B}}}_E := \left(\; \underline{\mathbf{s}}_1, \underline{\mathbf{s}}_e \; \right) \quad .$$

[1] It will be necessary to make assumptions about the distribution of the relative dynamics of the elements in the recycles asking for relative uniform distribution.

[2] What is here called *capacitive element* is in other parts of the literature often called *compartment*

$\underline{\mathbf{s}}_i$ :: vector with zeros with a 1 at the ith position.

The basic model balances an arbitrary fundamental extensive quantity, which we denoted with Q. The Q is thus placed into the role of the state. The proposition suggests that we expect the intensities in the recycle to converge to the same level. This makes it necessary to introduce a state variable transformation changing the representation from the fundamental extensive quantity, being the state, to an arbitrary intensive variable being the state. For this purpose, we have to introduce a second extensive quantity q, which is used to norm the fundamental extensive quantity Q, thereby defining the arbitrary intensive quanitity:$\xi := \frac{Q}{q}$. The second extensive quantity is usually chosen such that it changes only insignificantly as a consequence of the process. Often it is a quantity such as volume, which implies assumptions on the changes of the volume with flow conditions and concentration changes. These are order-of-magnitude assumption. If indeed a second fundamental extensive quantity is chosen, we need to label the fundamental extensive quantities: $\xi := \frac{Q^a}{Q^b}$. Both will satisfy the balance equations (1). Rewriting the balance equations for the fundamental extensive quantity Q^a one finds:

$$\frac{d\underline{\underline{\mathbf{Q}}}^b\,\underline{\xi}_I}{dt} = \underline{\underline{\mathbf{A}}}_1\,\underline{\underline{\hat{\mathbf{Q}}}}_I^b\,\underline{\xi}_I + \underline{\underline{\mathbf{B}}}_P\,\hat{Q}^b_{e|P}\,\underline{\xi}_I + \underline{\mathbf{b}}_F\,Q^b_{F|1}\,\xi_F\,,$$

with the matrices $\underline{\underline{\mathbf{Q}}}^b$ and $\underline{\underline{\hat{\mathbf{Q}}}}_I^b$ being diagonal matrices. The index I is used to mark internal quantities, such as internal flows. The recycle matrix $\underline{\underline{\mathbf{A}}}_1$ is not changed, whilst the factor with the direction matrix $\underline{\underline{\mathbf{B}}}_E$ is split into two. The second part describes the inflow part (term with $\underline{\mathbf{b}}_F$) and the first part describes the outflow part (term with $\underline{\underline{\mathbf{B}}}_P$). This bi-sectioning is a reflection of the fact that the streams inherit the property of the source system. Since we assumed unidirectional flow, the inflow inherits the properties of the feed system and the outflow the one of the system where it is connected to the recycle, namely the system labelled e. The outflow part, is the vector $\underline{\mathbf{s}}_e$ padded with the appropriately sized zero matrix to form the $\underline{\underline{\mathbf{B}}}_P$ matrix such that this matrix operates on the full internal intensive property vector $\underline{\xi}_I$.

This representation is readily extended to multiple-recycle systems. For each recycle a term with a recycle matrix is added. Because the dimension of the model changes, the other matrices are padded with zero blocks accordingly.

$$\frac{d\underline{\underline{\mathbf{Q}}}^b\,\underline{\xi}_I}{dt} = \left(\sum_{\forall r}\underline{\underline{\mathbf{A}}}_r\,\underline{\underline{\hat{\mathbf{Q}}}}_I^b + \underline{\underline{\mathbf{B}}}_P\,\hat{Q}^b_{e|P}\right)\underline{\xi}_I + \\ + \underline{\mathbf{b}}_F\,Q^b_{F|1}\,\xi_F \qquad (2)$$

The vector $\underline{\xi}_I$ is a collection of the intensities of all plant-internal subsystems. The running index r indicates the recycle loops. The further extension to the case of multiple inflows and outflows are also readily accommodated by modifying the $\underline{\underline{\mathbf{B}}}_P$-matrix and $\underline{\mathbf{b}}_f$-vector as well as the inflow intensity vector accordingly.

Time Scale Analysis

In the large time scale two extreme cases are of interest. Firstly, it is of interest to analyse the behaviour of the fast part of the plant, here the internal recycles, as the external flows are changing on the large time scale, namely slowly. Secondly it is the response to very fast changes, approximated by impulses, of the fast part of the plant, though on the large time scale.

Slowly Changing Inputs (Approx. Const.)

For the **first case**, the internal system will approach the equilibrium when making the order-of-magnitude dynamics assumption of a constant input. Thus for the fundamental extensive quantity Q^a we can write using the intensities:

$$\underline{\mathbf{0}} = \underline{\underline{\mathbf{A}}}_1\,\underline{\underline{\hat{\mathbf{Q}}}}_I^b\,\underline{\xi}_I + \underline{\underline{\mathbf{B}}}_P\,\hat{Q}^b_{e|P}\,\underline{\xi}_I + \underline{\mathbf{b}}_F\,\hat{Q}^b_{F|1}\,\xi_F\,, \quad (3)$$

which has the solution: $\underline{\xi}_I = \underline{\mathbf{e}}.\xi_F$

Proof. The latter is proven easily by noticing the fact that both fundamental extensive quantities satisfy the balance equation. Thus

$$\underline{\mathbf{0}} = \underline{\underline{\mathbf{A}}}_1\,\underline{\underline{\hat{\mathbf{Q}}}}_I^b + \underline{\mathbf{s}}_e\,\hat{Q}^b_{e|P} + \underline{\mathbf{b}}_F\,\hat{Q}^b_{F|1}\,,$$

and by rewriting the vector of extensive quantities as the product of a diagonal matrix, the said vector as diagonal, with a vector of ones $\underline{\mathbf{e}} := [1,1,\ldots,1]^T$, and noticing that $\underline{\mathbf{s}}_e := \underline{\underline{\mathbf{B}}}_P\,\underline{\mathbf{e}}$, the desired form is obtained:

$$\underline{\mathbf{0}} = \underline{\underline{\mathbf{A}}}_1\,\underline{\underline{\hat{\mathbf{Q}}}}_I^b\,\underline{\mathbf{e}} + \underline{\underline{\mathbf{B}}}_P\,\hat{Q}^b_{e|P}\,\underline{\mathbf{e}} + \underline{\mathbf{b}}_F\,\hat{Q}^b_{F|1}\,.$$

It is now apparent that

$$\underline{\mathbf{b}}_F \hat{Q}^b_{F|1} \quad := \quad -\left(\underline{\underline{\mathbf{A}}}_1 \underline{\underline{\hat{\mathbf{Q}}}}^b_I + \underline{\underline{\mathbf{B}}}_P \hat{Q}^b_{e|P}\right) \underline{\mathbf{e}},$$

which when substituted into equation (3) yields

$$\underline{\mathbf{0}} \quad = \quad \left(\underline{\underline{\mathbf{A}}}_1 \underline{\underline{\hat{\mathbf{Q}}}}^b_I + \underline{\underline{\mathbf{B}}}_P \hat{Q}^b_{e|P}\right) \left(\underline{\xi}_I - \underline{\mathbf{e}}\,\xi_F\right)$$

proving the fact of the solution to equation (3). □

Keys

Assumption *The dynamics of the input is assumed slow, so slow that it does not change significantly in the time scale of the fast process.*

Assumption *Internal process dynamics are fast compared to external dynamics.*

Result *Intensities approaches equilibrium quickly.*

Result *Internal intensities are uniform and identical to the input, in the single input case, otherwise the weighted average.*

Result *At steady state, all extensive quantities do not change, thus norming may be done with any extensive quantity.*

The last statement is worth elaborating: Often the volume is chosen as the norming extensive quantity. As the above analysis shows, the requirement of being conserved is implied. If the result is applied to slowly changing inputs, the volume, the density and the internal volumes are not to change significantly.

A Slightly More Restricted Model

For the further development, we first generalize our model and use the above-given definition for the intensity. The norming of the fundamental extensive quantity thus forming an intensive quantity is often based on the norming extensive quantity to not change significantly in the attainable region in which the process operates. Probably the most common example is the volume. Constant volumes and constant densities are frequently applicable assumptions as the neglected nonlinearity is often very mild.

Let $\underline{\mathbf{q}}^b$ be the vector of norming extensive quantities with each element referring to the respective loop. The assumption is then constant norming quantities in each loop. Thus:

$$\frac{d\underline{\mathbf{q}}^b}{dt} \quad = \quad 0\,. \tag{4}$$

and use it to slightly generalise the model (2):

$$\frac{d\underline{\underline{\mathbf{q}}}^b\,\underline{\xi}_I}{dt} \quad = \quad \left(\sum_{\forall r} \underline{\underline{\mathbf{A}}}_r \underline{\underline{\hat{\mathbf{q}}}}^b_I + \underline{\mathbf{s}}_e \hat{q}^b_{e|P}\right) \underline{\xi}_I + \underline{\mathbf{b}}_F\, q^b_{F|1}\, \xi_F\,.$$

Here the norming, constant extensive quantities have been wrapped into a diagonal matrices $\underline{\underline{\mathbf{q}}}^b$ and $\underline{\underline{\mathbf{q}}}^b_I$. With the assumption (4) this yields:

$$\underline{\underline{\mathbf{q}}}^b \frac{d\underline{\xi}_I}{dt} \quad = \quad \left(\sum_{\forall r} \underline{\underline{\mathbf{A}}}_r \underline{\underline{\hat{\mathbf{q}}}}^b_I + \underline{\mathbf{s}}_e \hat{q}^b_{e|P}\right) \underline{\xi}_I + \underline{\mathbf{b}}_F\, q^b_{F|1}\, \xi_F\,,$$

The assumption (4) has further the consequence that the flows of the extensive quantity $\hat{q}^b$ in the individual recycle loops are the same and that the inflow is identical to the outflow: $\hat{q}^b_{i-1|i} = \hat{q}^b_{i|i+1} =: \hat{q}^b_r$ for all i in loop r and , $\hat{q}^b_{F|1} = \hat{q}^b_{e|P} =: \hat{q}^b_E$. For simplicity reasons, we further assume that all capacitive elements in the plant are of equal size when the capacity is measured in the quantity q. Thus $q^b_m := q^b_i$ for all i. These assumptions and the substitution of the consequently defined quantities simplifies the model to

$$\frac{d\underline{\xi}_I}{dt} \quad = \quad \left(\sum_{\forall r} \frac{\hat{q}^b_r}{q^b_m} \underline{\underline{\mathbf{A}}}_r + \frac{\hat{q}^b_E}{q^b_m} \underline{\mathbf{s}}_e\right) \underline{\xi}_I + \frac{\hat{q}^b_E}{q^b_m} \underline{\mathbf{b}}_F\, \xi_F\,,$$

The fractions of flows of extensive quantity and capacity of elements measured in the same extensive quantity are the inverse of the time constants associated with mixing in each element and the effect of the in- and outflow of to and from the plant. The inverse of these time constants can be interpreted as frequencies, in simple cases corner frequencies: $\nu_r := \frac{\hat{q}^b_r}{q^b_m}$, $\nu_E := \frac{\hat{q}^b_E}{q^b_m}$. The model is now cast in its final form:

$$\begin{aligned} \frac{d\underline{\xi}_I}{dt} \quad &= \quad \left(\sum_{\forall r} \nu_r \underline{\underline{\mathbf{A}}}_r + \nu_E \underline{\mathbf{s}}_e\right) \underline{\xi}_I + \nu_E\, \underline{\mathbf{b}}_F\, \xi_F\,, && (5) \\ &= \quad \underline{\underline{\mathbf{A}}}\, \underline{\xi}_I + \underline{\mathbf{b}}\, \xi_F\,. && (6) \end{aligned}$$

Lumping : a First-Order Model

The idea of reducing the order is to lump all recycles into one big lump. The result of this lumping is a first-order differential equation, which under the same mild conditions as assumed before, is linear an can be readily integrated for simple inflow profiles, which change only the intensive properties of

the inflow. Similarly the detailed model can be integrated. The difference is the approximation error made when using the single-lump model instead of the model with the recycles for a given inflow profile in the intensive property of the feed stream.

Simply summing up all the small elements does the lumping, which mathematically is achieved by multiplying the extended version of equation (1) with the left null matrix of the matrix $\underline{\underline{\mathbf{A}}}$, which is the transposed of a vector of ones $\underline{\mathbf{e}}$ of corresponding length :

$$\underline{\mathbf{e}}^T \frac{d\underline{\mathbf{Q}}_I}{dt} = \underline{\mathbf{e}}^T \left(\sum_{\forall r} \underline{\underline{\mathbf{A}}}_r \underline{\hat{\mathbf{Q}}}_I + \underline{\underline{\mathbf{B}}}_E \underline{\hat{\mathbf{Q}}}_E \right).$$

This operation eliminates all internal flows. This result can also be derived by recognising that column sum of the recycle matrices $\underline{\underline{\mathbf{A}}}_r$ is null. Defining the lumped quantity: $\bar{Q}_I^a := \underline{\mathbf{e}}^T \underline{\mathbf{Q}}_I$. Thus $\frac{d\bar{Q}_I^a}{dt} = \underline{\mathbf{e}}^T \underline{\underline{\mathbf{B}}}_E \underline{\hat{\mathbf{Q}}}_E$. Again, the intensive quantity is of interest. Thus we define the intensity for the lumped system in the same way as before:$\bar{\xi}_I := \frac{\bar{Q}_I^a}{q_I^b}$. Again assuming that the extensive quantity q_I^b does not change appreciably, and splitting the term with the matrix $\underline{\underline{\mathbf{B}}}_E$ as before, the model is cast into a new form:

$$\frac{d\bar{\xi}_I}{dt} = \frac{\hat{q}_E^b}{q_I^b} \left(\underline{\mathbf{e}}^T \underline{\underline{\mathbf{B}}}_P \underline{\xi}_I + \underline{\mathbf{e}}^T \underline{\mathbf{b}}_F \xi_F \right).$$

In the case of the single inflow, single outflow process, and defining the corner frequency $\bar{\nu} := \frac{\hat{q}_E^b}{q_I^b}$, which reduces to

$$\frac{d\bar{\xi}_I}{dt} = \bar{\nu} \left(-\xi_e + \xi_F \right). \qquad (7)$$

This model describes the process still based on the recycle model because it uses the intensive state ξ_e, which can be obtained by integrating the recycle model (6). The observation, though, that at steady state all the internal intensities approach the same value, stimulates the idea of simply replacing intensity of the exit element with the averaged intensity $\bar{\xi}_I$. Introducing a new intensive variable, indicating with a ˜ the approximation of ξ_I, results the final approximate first-order model:

$$\frac{d\tilde{\xi}_I}{dt} = \bar{\nu} \left(-\tilde{\xi}_I + \xi_F \right). \qquad (8)$$

Impulse Responses-A Comparison

A comparison of the impulse response of models provides, when plotted, an excellent visual measure for the fidelity of the models. Both, the impulse response of the recycle model and the lumped model are readily computed. Starting with the definition of the inflow change: $\xi_F(t) := \xi_F^o \, \delta(t-0)$. The recycle model is to be integrated for the solution, assuming zero initial conditions:

$$\begin{aligned} \underline{\xi}_I(t) &:= \nu_E \int_0^t \underline{\underline{\mathbf{e}}}^{\mathbf{A}(t-\tau)} \underline{\mathbf{b}}_F \, \xi_F(\tau) \, d\tau, \\ &:= \nu_E \, \underline{\underline{\mathbf{e}}}^{\mathbf{A}t} \underline{\mathbf{b}}_F \, \xi_F^o. \end{aligned}$$

The intensity of the element where the outflow is

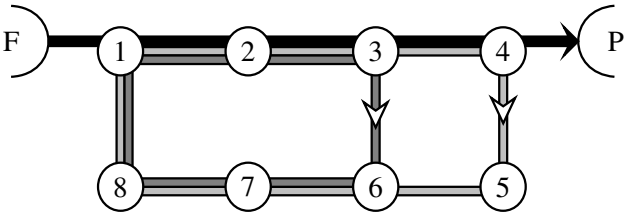


Figure 2: *A sample plant with 8 lumps and three streams*

connect is the eth element in the solution vector. It is selected by multiplying the solution with the transposed of the e-selection vector, being zero except the eth element, which is one. The such found intensity $\xi_e(t)$ is substituted into model (7):

$$\frac{d\bar{\xi}_I}{dt} = \bar{\nu} \left(-\nu_E \underline{\mathbf{s}}_e^T \underline{\underline{\mathbf{e}}}^{\mathbf{A}t} \underline{\mathbf{b}}_F \, \xi_F^o + \xi_F(t) \right),$$

which needs to be integrated again:

$$\begin{aligned} \bar{\xi}_I(t) &= \bar{\nu} \left(-\nu_E \underline{\mathbf{c}}_e^T \int_o^t \underline{\underline{\mathbf{e}}}^{\mathbf{A}\tau} d\tau \, \underline{\mathbf{b}}_F \, \xi_F^o + \int_0^t \xi_F(t) \, d\tau \right), \\ &= \bar{\nu} \left(-\nu_E \underline{\mathbf{c}}_e^T \underline{\underline{\mathbf{A}}}^{-1} \left(\underline{\underline{\mathbf{e}}}^{\mathbf{A}t} - \underline{\underline{\mathbf{I}}} \right) \underline{\mathbf{b}}_F + 1 \right) \xi_F^o. \end{aligned}$$

The solution for the approximate model (8) is simple: $\tilde{\xi}_I(t) = \bar{\nu} \, \underline{\underline{\mathbf{e}}}^{-\nu_I t} . \xi_F^o$ Finally we can compute various errors for ex.: $e(t) := \bar{\xi}_I - \tilde{\xi}_I$. The attached plots, show the solutions for the two recycle process shown in Figure 2. The volumetric flows in the two recycles are 1, and the inflow, respective the outflow, is 0.1. The total volume of the plant is set to 1 and the internal volumes are all the same, thus the individual volume is 1/8 in this case, as there are 8 lumps all together.

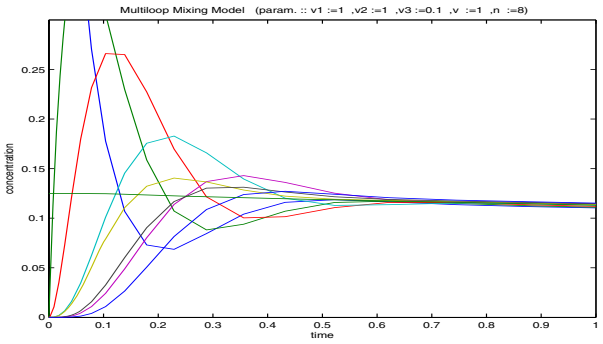


Figure 3: *Simulated impulse response of 8 equally sized tanks and a single-lumped approx.*

Plot 3 shows the impulse response of the 8 lumps, model 6, and the smooth middle one is the impulse response of the single lump approximation of model 7. Figure 4 finally shows the difference between the model (7) and the model and (8).

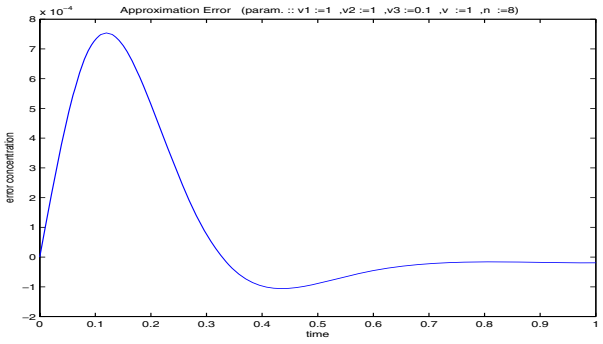


Figure 4: *Error between model 7 and 8*

Conclusions

The concise representation, as it was developed as part of our Modeller project ([7]) of process systems enables a generic analysis of such processes. The here-discussed recycle process is generic and not dependent of the nature of the sub-processes. The analysis also makes the different steps and assumptions nicely visible: 1) Split plant into fast and slow section. 2) Assume slow interaction between the fast and the slow section, thus limiting the spectrum of interaction and implicitly defining fast and slow. 3) In the long time scale one assumes fast internal dynamics, which results in uniform intensive quantities inside the system. 4) Using the above result, all the internal dynamics can be lumped into one. 5) Notice: Time constants of the plant become explicitly visible once one transforms from the space of the conserved extensive quantities into the space of the conjugate intensive quantities.

Results published in the literature, such as [3] on special processes can be nicely generalised using the generic, physics-based representation as it is presented here and in [4] if event dynamic reactions are involved.

References

[1] P Astrid. *Reduction of process models: a proper orthogonal decomposition approach*. PhD thesis, Eindhoven University of Technology, 2004.

[2] Jogchem Berg van den. *Model reduction for dynamic real-time optimization of industrial processes*. PhD thesis, Delft University of Technology, 2005.

[3] Aditya Kumar and Prodromos Daoutidis. Nonlinear dynamics and control of process systems with recycles. *Journal of Process Control*, 12:476–484, 2002.

[4] H A Preisig. On concentration control of fast reactions in slowly-mixed plants with slow inputs. pages FP 06–6, Anchorage, Alaska, 2002. American Control Conference.

[5] H A Preisig and M R Westerweele. Effect of timescale assumptions on process models and their reconciliation. ESCAPE 13, 2003.

[6] J G A Vusse Van de. New model for the stirred tank reactor. *Chemical Engineering Science*, 17:507–521, 1962.

[7] M R Westerweele. *Five Steps for Building Consistent Dynamic Process Models and Their Implementation in the Computer Tool MODELLER*. PhD thesis, TU Eindhoven, Eindhoven, The Netherlands, ISBN 90-386-2964-8 2003.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

FINITE AUTOMATA FROM FIRST-PRINCIPLE MODELS: COMPUTATION OF MIN AND MAX TRANSITION TIMES

Heinz A. Preisig *

* *Dept of Chemical Engineering, NTNU, 7491 Trondheim, Norway*

Abstract: Supervisory control schemes of (complex) plants utilize different forms of automata or related structures such as Petri-nets. Empirical, knowledge-based mapping of the plant's operation into such a structure cannot be complete or correct. These automata can be computed by a model-based approach, which guarantees completeness and correctness within the limits of the given model. The result is a non-deterministic automaton (Philips 2001), which however contains no information about the range of transition time that may be expected. This information would be extremely useful for the design of the derived operational procedures such as supervisory controllers on all levels and fault detection and fault isolation schemes. The problem has been formulated several times in the past, for example (Kowalewsky 1999, Engell 1997). Here a solution to the problem is described, which applies to plants generating a monotone flow field for constant inputs.

Keywords: Discrete-event dynamic systems, timed automaton, fault detection, supervisory control, modelling, hybrid systems

1. CURRENT STATE OF AFFAIRS

The increasing complexity of plants and the request for closer interaction between plants asks for more and increasingly sophisticated automation. Traditionally, process units were controlled separately, but increased interaction and required co-ordination make it necessary that the process is viewed and analysed in its full entity, giving rise to the subject of plant-wide control. On the supervisory level, which also links to the management levels such as planning and sequencing of operations and capacity allocation, the plant is event-driven. Currently used empirical modelling techniques cannot guarantee the completeness or correctness of the description, thus one branch of research focused on the computation of one-step automaton representations for continuous plants that are observed by an event detection mechanism. These problems can now be seen as solved. Algorithms exist for linear plants Preisig 1993, (monotone: Preisig 1996, general: Philips et al 1997, Pijpers 1996) and nonlinear plants (Preisig et al 1997, Bruinsma 1997), which can also handle all important exceptions. Also the state explosion problem, which was seen as one of the major drawbacks of these automaton computations, has been completely removed (Philips 2001, Foerstner 2001).

The computation of the automata models is based on the representation depicted in Figure 1, the first box representing the continuous (or fast sampled time-discrete) plant, the second the event detection mechanism, which assumes knowledge of the state and noise-free data. We term this mechanism *domain observer*[1], thereby indicating that the extended event detection mechanism reconstructs the continuous state from the output, if it is not directly accessible, and generates a

[1] –in deviation to Lunze, who uses the term quantizer. By choosing the term *domain observer*, we want to place emphasis on the required knowledge of the state, as it is the state that is discretised and not the output.

signal as the continuous state changes from one subdomain into another defined through boundaries placed into the state space of the continuous system. The resulting non-deterministic automaton models have been used in a first study of DEDS control *synthesis* methods (Philips 1998b, Philips 1999) and fault detection (Philips 1998a, Ramkumar 1998, Ramkumar 1999b, Ramkumar 1999a, Lunze 2000, Lunze 1999).

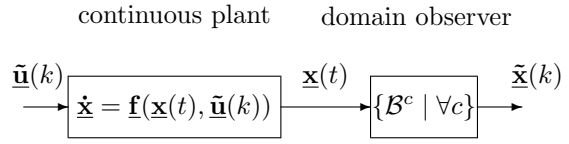


Fig. 1. *Discrete modelling of a discretely observed plant. The tilde quantities represent discrete-event signals*

In both applications it is apparent that knowledge of minimum and maximum transition times would be a very useful piece of information. Thus the problem is formulated, if such information can be obtained from the equations. Here we shall focus on linear plants, though it should be noted that linearity is not limiting, rather limitations on the flow field are imposed, as we shall see below.

2. PROBLEM FORMULATION

Given a linear system with a continuous state, $\underline{\mathbf{x}}$, and an input, $\underline{\tilde{\mathbf{u}}}$ that, whilst continuous, is changing only at event times and stays constant in between. The derivation may start from a model that is as general as a linear-in-state, time-varying model of the form:

$$\frac{d\underline{\mathbf{x}}(t)}{dt} = \underline{\underline{\mathbf{M}}}(t)\,\underline{\mathbf{x}}(t) + \underline{\mathbf{h}}(t;\underline{\tilde{\mathbf{u}}})\,, \qquad (1)$$

with $\underline{\mathbf{x}} \in \mathbb{R}^n, \underline{\tilde{\mathbf{u}}} \in \mathbb{R}^m$, which for simplicity of algebra we shall reduce to the standard linear, time-constant plant:

$$\frac{d\underline{\mathbf{x}}(t)}{dt} = \underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}}(t) + \underline{\underline{\mathbf{B}}}\,\underline{\tilde{\mathbf{u}}}(k)\,. \qquad (2)$$

We shall also assume direct knowledge of the state. If the state is not directly accessible, an observer must be added to the plant with the dynamics being fast enough so as to be negligible on the time scale the discrete-event dynamic system operates.

For the automaton representation, we split the continuous state domain into a set of hypercubes by defining a set of ordered boundary values $\beta^c_{d_c}$ with c identifying the state co-ordinate and d_c the membership of the value in the ordered set of boundary values, $\beta^c_1 < \beta^c_2 < \cdots < \beta^c_{n_c}$ and $[\beta^c_1, \beta^c_{n_c}]$ the validity range of x_c, defined on the co-ordinate c. For the arbitrary co-ordinate c the boundary set is then:

$$\mathcal{B}^c := \{\beta^c_{d_c} | d_c := 1, ..., n^c\}\,,$$

with. In practice, these sets are part of the definition of the domain observer. The domain observer assigns membership of the state to an interval dynamically, that is, the boundary point belongs to the interval from where the trajectory enters the boundary (Philips 2001). The hypercubes are conveniently defined in the form of a matrix

$$\underline{\underline{\mathbf{H}}} := \left[\left[\beta^c_s, \beta^c_{s+1}\right]\right] := \left[\underline{\mathbf{b}}^c_{-1}, \underline{\mathbf{b}}^c_{+1}\right]\,,$$

with the $\underline{\mathbf{b}}$ vectors being introduced for the elegance of notation later (Equation (3) ff). Each hypercube has $n!$ faces, each of which is a hyperplane. An event E^S is defined as a crossing of the boundary between two hypercubes, thus a crossing of the actual continuous trajectory through a face S of a hypercube. At this time, the domain observer will emit a signal indicating this event. This definition of an event excludes simultaneous crossing of boundaries; thus, passing through corner points of the hypercubes, defined by the intervals, is not possible. The latter is justified assuming a sequential output line from the domain observer. The computation of the discrete behaviour of the plant as shown in Figure 1 has been reported elsewhere (Preisig 1993, Philips et al 1997, Preisig 1996). Here we wish to compute the minimum and maximum time it takes for the system to move from one transition to the next.

3. WHAT'S THE NEXT POSSIBLE TRANSITION

Having defined the task of computing the minimal and maximal time it takes for event E^B to occur after event E^A, we need first to find what event E^B is possible after E^A has occurred. For this purpose a number of objects are required. Having defined the hypercube representing a discrete state in the continuous state space, and having defined an event as a crossing of the surface of the hypercube, we define a trajectory as

$$\mathcal{X}(\underline{\mathbf{x}}_i) := \{\underline{\mathbf{x}}(t) \mid \forall t, \underline{\mathbf{x}}(t_i) = \underline{\mathbf{x}}_i\}\,,$$

and a bundle of trajectories being

$$\mathcal{T}^A := \{\mathcal{X}(\underline{\mathbf{x}}_i) \mid \mathcal{X}(\underline{\mathbf{x}}_i) \cap \in \mathcal{A} \neq 0\}\,,$$

whereby $\mathcal{A}$ is a bounded piece of a hyperplane.

With these definitions we can define the surface elements of the hypercube connected by a bundle of trajectories, and thus the *connected events*, by identifying the connecting bundle:

$$\mathcal{T}^{A\ B} := \mathcal{T}^A \cap \mathcal{T}^B\,;$$

yielding the respective surface pieces:

$$\Omega^{A|B} := \mathcal{T}^{A\ B} \cap A\,,$$
$$\Omega^{B|A} := \mathcal{T}^{A\ B} \cap B\,.$$

The task is thus to find the connecting trajectory bundle. For this purpose, we split the surface of the hypercube into two sets, namely one set where the flow enters $\mathcal{F}^{in}$ and a set where the flow exits $\mathcal{F}^{out}$.

At this point, the main assumption is introduced, namely that the flow field is monotone within the extent of the hypercube. At first, this assumption appears rather restrictive. However, one must keep in mind that the flow field is here for a process for which all the inputs are being kept constant. Most natural processes show under these conditions a monotone behaviour. We also exclude the trivial case in which the flow is parallel with a hypercube's surface.

With these conditions, the direction of the flow is:

$$\underline{\mathbf{s}} := \text{sign}\left(\dot{\underline{\mathbf{x}}}(t)\right), t < \infty\,, \tag{3}$$

and the centre point of the entry surface and the exit surface of the hypercube can be determined:

$$\underline{\mathbf{x}}^{in} := \left[b_i^j\right]_{\forall j}, i := -s_j\,,$$
$$\underline{\mathbf{x}}^{out} := \left[b_i^j\right]_{\forall j}, i := s_j\,.$$

These points are the intersection of a set of hyperplanes:

$$\mathcal{P}^{in} := \left\{\mathcal{P}(x_i^{in}), \forall i\right\}.$$
$$\mathcal{P}^{out} := \left\{\mathcal{P}(x_i^{out}), \forall i\right\}.$$

with the individual hyperplanes:

$$\mathcal{P}(x_j) := \left\{\underline{\mathbf{x}} \mid x_j := b_i^j, i \in \{-s_j\}\right\}.$$

Now the different connected pieces of the surfaces can be computed:

$$\mathcal{R}^{A,B} := \mathcal{T}^A \cap \mathcal{P}(x_j^{out})\,,$$

and the exit surface piece

$$\Omega^{B|A} := \mathcal{R}^{A,B} \cap B\,. \tag{4}$$

where $A \in \mathcal{F}^{in}$ and $B \in \mathcal{F}^{out}$. If the forward intersection $\Omega^{B|A}$ exists, thus the intersection is non-empty, the corresponding next event does exist and the opposite piece of surface on the entry face is the intersection of the trajectory bundle defined by the exit piece $\Omega^{A|B}$ [2]:

$$\Omega^{A|B} := \mathcal{T}^{\Omega^{B|A}} \cap A\,.$$

4. TRANSITION TIME

For either of the two models (1, 2) and knowing what next transitions may occur, the transition times can be calculated for any entry point by solving [3] the transcendental equation for T:

$$\begin{aligned} x_k^b &:= \underline{\mathbf{e}}_k^T\, \underline{\mathbf{x}}^b(T)\,, \\ &:= \underline{\mathbf{e}}_k^T \left(\underline{\underline{\mathbf{e}}}^{\int_0^T \underline{\underline{\mathbf{M}}}(t)\,dt} \left(\underline{\mathbf{x}}^a(0) + \right.\right. \\ &\qquad \left.\left. + \int_0^T \underline{\underline{\mathbf{e}}}^{-\int_0^t \underline{\underline{\mathbf{M}}}(\tau)\,d\tau}\, \underline{\mathbf{h}}(t; \underline{\tilde{\mathbf{u}}})\,dt\right)\right), \end{aligned}$$

$$\begin{aligned} x_k^b &:= \underline{\mathbf{e}}_k^T \left(\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,T} \left(\underline{\mathbf{x}}^a(0) + \int_0^T \underline{\underline{\mathbf{e}}}^{-\underline{\underline{\mathbf{A}}}\,t}\, \underline{\underline{\mathbf{B}}}\, \underline{\tilde{\mathbf{u}}}\,dt\right)\right), \\ &:= \underline{\mathbf{e}}_k^T \left(\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,T} \underline{\mathbf{x}}^a(0) + \underline{\underline{\mathbf{A}}}^{-1} \left(\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,T} - \underline{\underline{\mathbf{I}}}\right) \underline{\underline{\mathbf{B}}}\, \underline{\tilde{\mathbf{u}}}\right), \\ &:= \underline{\mathbf{e}}_k^T \left(\underline{\mathbf{g}}(T, \underline{\mathbf{x}}^a)\right), \end{aligned}$$

where $\underline{\mathbf{x}}^a(T) \in \Omega^a$ and $\underline{\mathbf{x}}^b(T) \in \Omega^b$ and $\underline{\mathbf{e}}_k^T$ the unity vector $[0, 0, ..., x_k, 0, ..., 0], x_k := 1$ selecting the co-ordinate that defines the exit face.

5. THE 3-D SAMPLE SYSTEM

The sample system, being linear and time constant, $\Sigma := \{\underline{\underline{\mathbf{A}}}, \underline{\underline{\mathbf{B}}}\}$ being used as an illustration in the continuation is given by the matrices

$$\underline{\underline{\mathbf{A}}} := \begin{pmatrix} 0.8642 & -0.6340 & -0.0672 \\ 15.4736 & -5.3626 & -0.6678 \\ 10.2891 & -2.4301 & -1.5016 \end{pmatrix}, \tag{5}$$

$$\underline{\underline{\mathbf{B}}} := \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \tag{6}$$

with the input being kept constant at a given value. With the eigenvalues $\underline{\lambda} := [-1, -2, -3]$ the system is asymptotically stable.

The Figures 2, 3, 4, 5, 6, 7 show the different pairs of surface elements for the sample system with a zero input. The left-lower front corner being the centre of the entering surface and the right-upper back corner being the centre of the exit surface of the cube.

5.1 *An Alternative View*

An interesting insight is obtained by looking at the problem from a slightly different angle: One

[2] We use here a more detailed notation by indicating the sequence with which the elements of the respective faces are obtained. One may read $B|A$ as (face element B given face element A)

[3] For a reference of solving linear, time-variant ODE's see for example Walter 1960, 1993

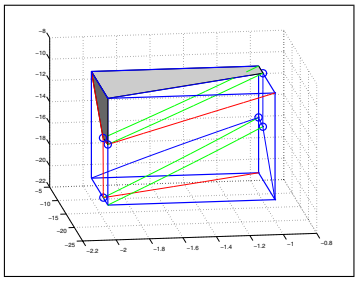


Fig. 2. Front (dark) to attached top (light).

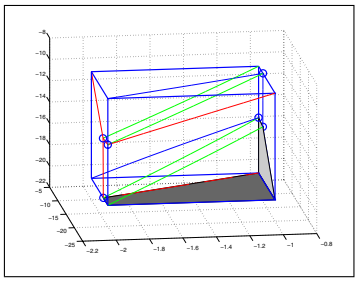


Fig. 3. Bottom (dark) to opposite back (light).

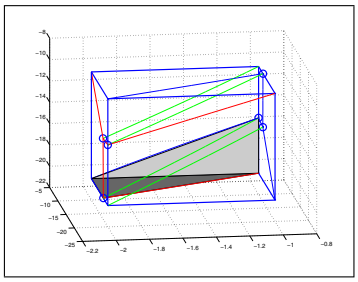


Fig. 4. Bottom (dark) to attached back (light).

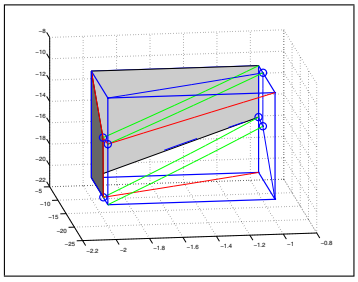


Fig. 5. Front (dark) to attached back (light).

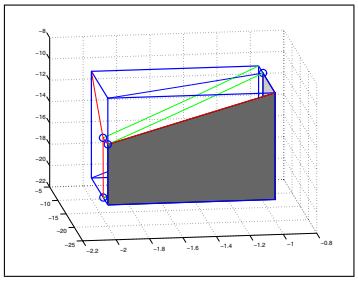


Fig. 6. Front side (dark) to attached back (light).

can view the sectioning of the exit (entry) faces as a projection of the entry (exit) edges onto the opposite side with the dynamic system being the mechanism of projection. Figure 8 shows the projection of the entry edges on the exit surface, which is done forward in time, and Figure 9 the

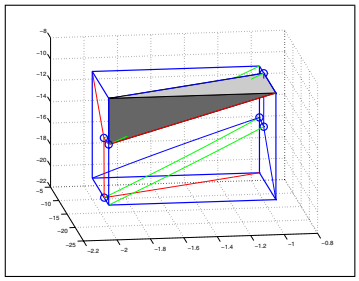


Fig. 7. Front side (dark) to attached top (light).

projection of the exit edges on the entry surface, done backward in time. In the Figure 8 the entry edge is shown in thick lines and the projections in medium lines. In the Figure 9, it is the exit edges in thick lines and the backward projections in medium lines.

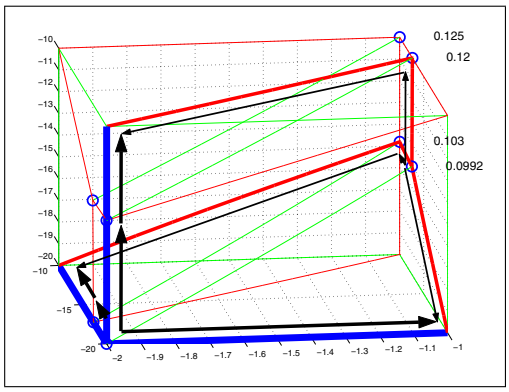


Fig. 8. The view of projecting the entry edges onto the flow-opposite faces.

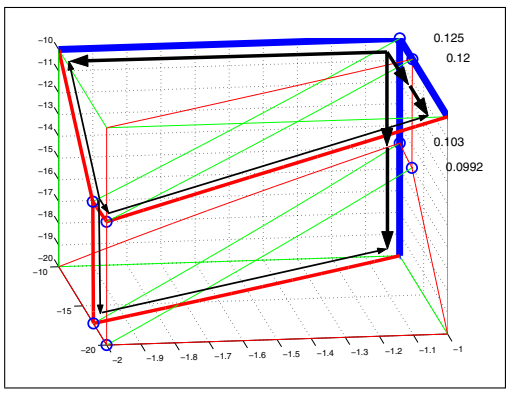


Fig. 9. The backwards projection of the exit edges onto the flow-opposite faces. The arrows indicate the progress of the direction of the begin points as related to the locus of the projected points. The numbers to the left of the marked points indicate the respective transition times.

6. FINDING THE LONGEST AND THE SHORTEST TRAJECTORY IN A MONOTONE FIELD

In a monotone flow field, the computation of the longest and the shortest time is an optimisation

problem where the starting point, being element of the entry hypercube surface, is changed such that one finds the minimum and the maximum transition time: In more colloquial terms to find the longest and the shortest trajectory starting on the entry surface of the hypercube.

The optimisation is rather simple if the objective function, namely the transition time changes monotonicly with the adjustable variables, here the position on the entry surface, because in a monotone field the two extremes are associated with opposite corner points of the boundary Gill 1980. It is sufficient to prove monotonic properties of the transition time as a function of the starting point, which is identical of analysing the gradient of the transition time changing with the co-ordinate on the boundary is not changing sign. Let

$$f(T, \underline{\mathbf{x}}^a) := \underline{\mathbf{s}}\,(\underline{\mathbf{x}}^b(T) - (\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,T}\underline{\mathbf{x}}^a + , \\ + \underline{\underline{\mathbf{A}}}^{-1}\,(\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,T} - \underline{\underline{\mathbf{I}}})\,\underline{\underline{\mathbf{B}}}\,\underline{\tilde{\mathbf{u}}}))\,,$$

then, since the transition time T cannot be computed analytically, the implicit function theorem is to be used to compute the desired gradient:

$$\frac{dT}{d\underline{\mathbf{x}}^a} := -\frac{f_{\underline{\mathbf{x}}^a}(T, \underline{\mathbf{x}}^a)}{f_T(T, \underline{\mathbf{x}}^a)} \\ := \frac{-\underline{\mathbf{s}}\,\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,T}}{\underline{\mathbf{s}}\,\left(\underline{\underline{\mathbf{A}}}\,\underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,T}\underline{\mathbf{x}}^a + \underline{\underline{\mathbf{e}}}^{\underline{\underline{\mathbf{A}}}\,T}\,\underline{\underline{\mathbf{B}}}\,\underline{\tilde{\mathbf{u}}}\right)}\,.$$

Monotonic behaviour breaks down as the above gradient passes through a zero in one of its components. At a first glance, the change of sign could be caused by either of the numerator or the denominator. A brief analysis though reveals that it is the denominator that determines the location of the change.

Proof : Consider the boundary Ω^b to initially be close to the starting boundary Ω^a. The transition time can thus be brought arbitrarily close to zero. As the target boundary is moved away, the starting boundary can be moved as well. Again, the difference can be kept arbitrarily small. As long as the gradient does not change, direction, the derivative remains in the same half plain. The sum, or the integral does thus also change in the same direction, which proves the fact that the transition time changes monotonic with the initial location on the starting surface, until the denominator changes sign. The latter is the locus of a derivative in one co-ordinate being zero, which is on a flat plane cutting the space into two monotonic sub-domains. These local equilibrium plains intersect, if we constrain the discussion to asymptotically stable (non-oscillatory) systems, at the global equilibrium point.

□

Alternatively one can prove that the function $T(\underline{\mathbf{x}}_a)$ is monotone as long as the the right-hand-side of the dynamic model equations does not change sign:

Proof : Given that $\underline{\underline{\mathbf{A}}}\,\underline{\mathbf{x}} + \underline{\underline{\mathbf{B}}}\,\underline{\tilde{\mathbf{u}}}$ does not change sign (asymptotic behaviour), the inverse does not change sign either and the integral with time is monotone and so is the integral of the inverse. The monotone behaviour changes as the sign of the integrand changes.

□

With the accumulated information, it is trivial now to provide the minimal and maximal transition times for each transition. In the cases where the entry face is attached to the exit face, the minimal transition is always zero. The maximal transition is given by the longest trajectory forming the tube running across the hypercube, which is attached to the respective piece of the entry face. Thus only four different maximal transition times occur in the whole, independent of the dimension of the problem. The transition times for the example are shown in Figure 8.

7. CONCLUSIONS

The surface of the hypercube splits into two sections, the entry section and the exit section. If the flow is *not running in parallel with the coordinates*, there is only one central entry corner and only one central exit corner. Each of the faces of the hypercube belongs to one of the two surfaces, namely the entry or the exit section. Each face is split into sections whereby each of the entry sections is connected with an exit section, thus defining the reachable pieces of the surface as a function of the entry location.

The computation of the different surface sections is done by finding the forward projection of the centre entry corner onto the exit surface and the backward image of the centre exit point onto the entry surface. The edges of the entry faces project onto the exit surfaces using the dynamics of the process for the projection. The result is the lines subdividing the exit faces. The inverse computation, namely the backward projection of the centre exit point and the exit edges onto the entry surface results the other set of face-sectioning lines.

The minimal and the maximal times for a transition are associated with the centre corner points and the additional two trajectories cutting across the hypercube. Because the objective function, namely the transition time is a monotone function of the location on the entry surface, the maximum and the minimum are associated with transitions from the corner and edge points or to the corner and edge points. Only four trajectories must be computed.

The principle of the computation is not limited to linear systems. Monotonicity is the only condition being used. Note that monotonicity is only requested for the region of the continuous state space being covered by the discrete state space at constant inputs.

8. REFERENCES

Bruinsma U B D M R; (1997); State-Event Discrete Modelling of Nonlinear Plants; *M.Sc Thesis, NR 1984*; Eindhoven University of Technology, Eindhoven, The Netherlands.

Engell S, Kowalewski S, Krogh B; (1997); Discrete events and hybrid systems in process control; *Proc. 5th Intern Conference on Chem Process Control, Tahoe City, (J. E. Kantor, C. E. Garcia, B. Carnahan, Eds.), AIChE Symp Series, No. 316, Vol. 93*; 165-176.

Förstner D; (2001); Aualitative Modellierung für die Prozessdiagnose und deren Anwendung auf Dieseleinspritzsysteme; *PhD Thesis*; Universität Hamburg-Harburg, Hamburg and Bosch AG, Stuttgart; 138.

Gill P E; (1980); Practial Optimisation; *Strange*; USA.

Kowlewski S, Engell S, Preussig J, Stursberg O; (1999); Verification of logic controllers for continuous plants using timed condition/event-system models; *Automatica*; **35**; 505-518.

Lunze J, Schröder J; (1999); Process diagnosis base on a discrete-event description; *Automatisierungstechnik*; **47**, 780; 358-365.

Lunze J, Schröder J; (2000); State observation and diagnosis of discrete-event systems described by stochastic automata; *Discrete Event Dynamic Systems*; in press.

Philips P P H H; (2001); Modelling, Control and Fault Detection of Discretely-Observed Systems; *PhD Thesis*; TU-Eindhoven, Eindhoven, The Netherlands, ISBN 90-386-1729-1; 157p.

Philips P P H H, Bruinsma U B D M R, Weiss M, Preisig H A; (1997); A Mathematical Approach to Discrete-event Dynamic Modelling of Hybrid Systems; *IFAC Symposium on AI in Real-Time Control*; Kuala Lumpur, Malaysia; 185-190.

Philips P P H H, Ramkumar K, Lim K W, Preisig H A, Weiss M; (1998a); Automation-based fault detection and isolation; *ESCAPE-99, Comp & Chem Eng*; Budapest, Hungary; S215-S218.

Philips P P H H, Weiss M, Preisig H A; (1998b); A design strategie for discrete control of continuous systems; *American Control Conference 98*; San Diego, USA.

Philips P P H H, Weiss M, Preisig H A; (1999); Control based on discrete-event models of continuous systems; *European Control Conference 99*; Karlsruhe, Germany.

Pijpers J H; (1996); Modelling for supervisory control; *M.Sc. Thesis, NR-1956*; Eindhoven University of Technology, Eindhoven, The Netherlands.

Preisig H A; (1993); More on the Synthesis of a Supervisory Controller From First Principles; *12th IFAC World Congress 93*; Sydney, Australia; Vol. V, 275.

Preisig H A; (1996); A Mathematical Approach to Discrete- Event Dynamic Modelling of Hybrid Systems; *Computers & Chemical Engineering*; **20**, Suppl; S1301- S1306.

Preisig H A, Pijpers M J H, Weiss M; (1997); A discrete modelling procedure for continuous processes based on state-discretization; *MATHMOD 2*; Vienna, Austria; 189-194.

Ramkumar K B; (1999a); Fault Detection and Diagnosis in Process Plants Using Finite-State Automaton; *M.Sc. Thesis*; National University of Singapore; 103 p.

Ramkumar K B, Philips P P H H, Ho W K, Preisig H A, Lim K W; (1999b); A real-time realisation of fault-detection and diagnosis using finite-state automaton; *Int Fed of Auto Control '99, 14th World Congress*; Bejing, China, Proceedings; P-211.

Ramkumar K, Philips P P H H, Preisig H A, Ho W K, Lim K W; (1998); Structured fault-detection and diagnosis usig finite-state automation; *24th Annual conference of the IEEE Industrial Electronis Society*; Aachen, Germany.

Walter W; (1960, 1993); Gewöhnliche Differentialgleichungen, eine Einführung; *Springer*; Germany; 325 S.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

NEURAL MODELING AS A TOOL TO SUPPORT BLAST FURNACE IRONMAKING

F. T. P de Medeiros[a], A. Pitasse da Cunha[b], A. M. Frattini Fileti[c]

[a]*Companhia Siderúrgica Nacional (CSN), Brazil*
[b]*MetalFlexi, Brazil*
[c]*Chemical Systems Engineering Department, (UNICAMP), SP, Brazil*

Abstract: This paper describes the development of a hybrid model based on artificial neural network and its industrial application to the ironmaking at *Companhia Siderúrgica Nacional* (CSN -Volta Redonda/Brazil). The Iron Blast Furnace is highly complex process subject to oscillations in raw material characteristics. A precise model is essential to adjust © *2002* charging and blow conditions to match productivity, chemical quality and costs targets. A neural model was developed in order to estimate chemical and thermal parameters to feed a first principles model capable of evaluating alternative operation standards. As a consequence, operation efficiency is enhanced leading to higher productivity and lower costs. *Copyright © 2002 IFAC*

Keywords: modelling, neural network, ironmaking, iron blast furnace.

1. INTRODUCTION

The impulse from the domestic market and the abundance of quality raw materials have favored the development of the Brazilian steel industry, which is viewed as playing a fundamental part in the process of industrialization and development. CSN is one of the largest steelmaking groups in Latin America, with a production capacity of 5.8 million tons of raw steel per year.

The Iron Blast Furnace reduces iron ore, producing liquid iron (hot metal) which is converted to steel by exothermic oxidation of metaloids dissolved in the iron in the basic oxygen steelmaking process.

The Blast Furnace is a very complex processes in terms of chemistry, fluidodynamics and heat exchange. The composition of the burden material to be loaded and the blast to be blown determines productivity, quality and costs. Designing burden and blast requires a fairly accurate process model to define an appropriate operation standard from an almost infinite set. Particular characteristics, associated to both materials and equipment, are to be considered in the model requiring actual data to be analysed before applying first principle models.

Many simple models exist to analyse the blast furnace process based on heat, mass and chemistry balance and some are even ingenuous. However, chemical equilibrium mismatches and kinetics parameters need to be estimated based on materials and equipment characteristics in order to quantify performance indexes. Usually, to close that gap it is necessary to apply a comprehensive statistic model. Chemical composition analysis of every furnace stream need to be taken (raw material, blow, overhead gas and liquid metal), which introduces a dead time to the performance calculations.

One of the alternative and efficient tools, which enable one to obtain a numerical description of this kind of complex process, is the artificial neural

network (ANN). Interactions and the dynamics among variables are readily captured from operating data base presentation to the network. From past operating conditions and calculated mismatch parameters, a network model allows performance indexes computation.

Neural networks are becoming an effective component of the steel manufacture automation system. There are various applications of neural networks in the steel industry. Schlang et al. (1997) describes the use of neural networks in the control of flat steel rolling mills and electric furnaces (Siemens AG). Cox et al. (2002) explore the use of neural networks to predict oxygen and coolant requirements during the end-blow period of the Port Talbot basic oxygen steelmaking - BOS - plant (Corus Group). However, the authors report that the application of the neural model 'in situ' was to be carried out just in future work. Ping et al. (2003) describe the implementation of an intelligent model for controlling BOS end-points at WISCO's No 2 steel shop. This static control model combines neural networks and first principles. Indeed for the iron Blast Furnace process there are few papers on neural networks. Radhakrishnan and Mohamed (2000) describe a successful application of a neural network for the identification and control of blast furnace hot metal quality.

A growing literature within the field of chemical processes describing the use of artificial neural networks has evolved for a diverse range of engineering applications such as fault detection, signal processing, process modelling and control (Himmelblau, 2000). According to the author, because neural networks are nets of basis functions, they can provide good empirical models of complex nonlinear process useful for a wide variety of purposes.

Considering the difficulties outlined above, obtaining accurate mismatch parameters for first principles models in iron and steelmaking has proved to be a very hard task. Usually two kinds of models are employed to blast furnace operation: those very simple using estimated mismatch parameters that are corrected as operation goes on and complex models with to many parameters to be of practical use.

The present work is concerned with developing a hybrid model - neural network and mass and heat balances – and its application to the ironmaking blast furnace at CSN (Brazil). The main goal is to obtain a tool to design burden and blast conditions in order to match the targets of productivity, chemical quality and costs of the liquid metal.

1. METHODS

2.1 Process Description

Companhia Siderúrgica Nacional's steelworks entails three blast furnaces, two of them in operation and one out of service. Blast Furnace 2 produces nearly 4,000 tons of hot metal per day whereas Blast Furnace 3 produces between 9,500 and 11,000 tons per day. Iron ore sinter, pellets and lumpy hematite constitute the ferrous burden. As reducing agents, metallurgical coke and pulverised coal are used, being the latter injected through the tuyeres. Sometimes dolomite or quartzite are used as fluxes to control slag composition. The blast composition (air, oxygen and steam) and the rate of coal injection are the main and most sensitive parameters of control. Operation aims at production rate, hot metal chemical composition and temperature and ultimately, low cost. Because the plant is self-sufficient in coke, a small proportion of it is imported bringing significant characteristic variations to the mixture.

The core of the process is a counter-flow reactor where a series of chemical and thermal exchanges are performed in several ent internal zones (Figure 1).

As the ferrous burden descends it is first dried, then reduced by the up-coming process gas containing CO and H_2. This zone, called upper granulated zone, or preparation zone or even indirect reduction zone, ideally produces wustite (Fe_yO) to be reduced to metallic Fe in the inferior zones. The index y, in this case, approaches 0.95. In real terms, however, the wustite will have an excess of oxygen which is quantified in terms of kg-mol of O / kg-mol of Fe. This parameter is necessary to establish a proper mass and thermal balance of the process and will be designated by ω (Rist and Maysson, 1967). The thermal balance also needs a parameter to represent real conditions. This is the heatloss that will be represented by λ.

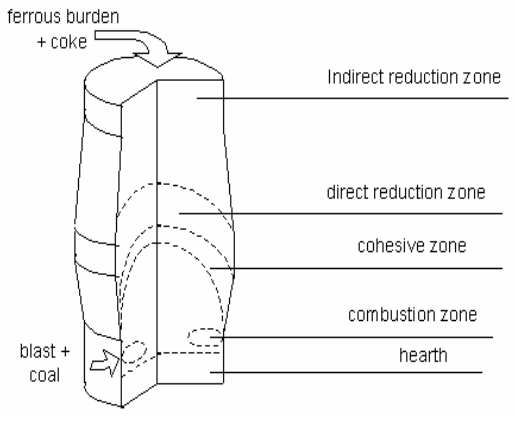


Fig. 1. The Iron Blast Furnace internal zones.

The main heat source is the combustion of coal and coke that produces mixture of CO, CO_2, H_2, H_2O and N_2. CO is regenerated in the direct reduction zone and below by the Boudouard reaction (Eq. 2). H_2 also plays an important role and the C-H-O will be in equilibrium in most sub-processes.

The basic chemical reactions involved are:

$$CO + 1/2\ O_2 \rightarrow CO_2 \quad (1)$$
$$CO_2 + C \rightarrow 2CO \quad (2)$$
$$H_2O + CO \rightarrow H2 + CO_2 \quad (3)$$
$$3Fe_2O_3 + CO \rightarrow 2Fe_3O_4 + CO_2 \quad (4)$$
$$yFe_3O4 + CO \rightarrow 3Fe_yO + (4y - 3)/2\ CO_2 \quad (5)$$
$$Fe_yO + CO \rightarrow yFe + CO_2 \quad (6)$$
$$SiO_2 + C \rightarrow SiO_{(gas)} + CO \quad (7)$$
$$SiO_{(gas)} + C \rightarrow \underline{Si} + CO \quad (8)$$

Silicon is partially reduced from silica into gaseous silica monoxide and incorporated to the liquid by further reduction. This process is rather complex and the metal silicon content is very hard to estimate.

The other impurities in the metal, manganese and phosphorous, do not represent a difficult estimation task depending more on the raw materials composition than on the process conditions.

2.2 Artificial Neural Network (ANN)

Theory. ANN are mathematical models constituted by several neurons, arranged in different layers (input, hidden and output), interconnected though a complex network. The multi-layer feedforward, that is the most popular of the much architectures currently available, was used. According to Equation (9), a neuron is responsible for the summation of all signals from previous layer's neurons, $y_{i,j}$ (amplified or weakened by weight values, $w_{i,j,k}$) and a value called bias, $b_{i,j}$. i represents the order of the layer whereas j and k indicate the order of the neuron in the layer. An activation function, f - such as hyperbolic tangent, sigmoid or linear function – is used for the activation of the neuron output, $y_{i,k}$.

$$y_{i,k} = f(\Sigma\ (w_{i,j,k}\ y_{i-1,j}) + b_{i,k}) \quad (9)$$

The data processing within the ANN structure is executed collectively and simultaneously through the dense network of neurons and their connections. Those characteristics were crucial for the this technique to be chose and not other multivariate regression ones which tend to give too much weigh to extreme values of the input variables.

Training the ANN. Once the network weights and biases have been initialized, the network is ready for training. The training process requires a set of examples of proper process behavior - network inputs and target outputs. During training the weights and biases of the network are iteratively adjusted to minimize the network objective function. The basic training algorithm is the backpropagation algorithm, in which the weights are moved in the direction of the negative gradient (Demuth and Baele, 2002).

The first method for improving generalization is called regularization. This involves modifying the objective function, which is normally chosen to be the sum of squares of the network errors on the training set. It is possible to improve generalization if we modify the objective function by adding a term that consists of the mean of the sum of squares of the network weights and biases:

$$F = \beta.\ SSE + \alpha SSW \quad (10)$$

where *SSE* is the sum of squared errors, *SSW* is the sum of squares of the network weights, and β and α are objective function parameters (Demuth and Baele, 2002).

According to Foresse and Hagan (1997), using this objective function will cause the network to have smaller weights and biases, and this will force the network response to be smoother and less likely to overfit. One feature of this algorithm is that it provides a measure of how many network parameters, (weights and biases) are being effectively used by the network. This effective number of parameters will be called *p*. *P* is the total number of parameters in the network.

Neural network training can be made more efficiently if certain preprocessing steps are performed on the network inputs and targets. Then, before training the network the training data was normalized in the range [0.1, 0.9], as follows:

$$y_{0,j} = 0.8\ ((x_j - xmin_j)\ /\ (xmax_j - xmin_j)) + 0.1 \quad (11)$$

where $y_{0,i}$ is the normalized value for the variable x_j, and $xmin_j$ and $xmax_j$ are the minimum and maximum of each variable x_j.

Modeling and data set. A neural model was developed in the present work to predict: the equilibrium mismatch parameter for the oxygen mass balance (ω), the thermal loss parameter for the heat balance (λ), the gas flow resistance parameter (ρ), the hot metal Silicon content ([*Si*]) and the sulfur partition coefficient between slag and metal (K_S). Feeding those parameters to a simple mass and heat balance, a precise operation pattern is defined to guide operators and technical staff for immediate and strategic decision making.

Table 1 shows the final variables selection and their meaning. Coke drum ($x15$) and reactivity ($x16$) indexes quantify physical strength and chemical activity, respectively, and are important both to gas flow and chemistry in the process.

Three years of Blast Furnace 3 operation were observed. Records were condensed into 23 input variables. Sets corresponding to days with missing or inconsistent data were filtered out . Records include those acquired by the furnace digital automation system, works and mines laboratories. Finally a 28 x 820 data bank was gathered, randomized and fed into a MATLAB® program. The final data-base was then split into two sets, one for training and the other for generalization tests (15% of the data). It was

carefully checked the range of each variable since it should be similar to both sets.

In the search for the architecture that could yield the best possible prediction model accuracy, a study was performed to establish the number of nodes in the network hidden layer.

Table 1. Input (x) and output (y) variables used for the neural modeling.

	Blast variables
x1	kg-mol of N2 in blast / ton of metal
x2	kg-mol de H2O in blast / ton of metal
	Burden variables
x3	kg of slag / ton of metal
x4	Primary slag B4
x5	Hearth slag B4
x6	blast temperature (^oC)
x7	kg of small-coke / ton of metal
x8	kg of injected coal / ton of metal
x9	kg of lumpy hematite / ton of metal
x10	kg of pellets / ton of metal
x11	kg of quartzite / ton of metal
x12	external coke / total coke
x13	pulverized coal ash content
x14	pulverized coal oxygen content
x15	average coke Drum Index
x16	average coke Reactivity Index
x17	coke mean size (mm)
x18	hematite < 6,35 mm fraction
x19	hematite > 38 mm fraction
x20	hematite Decrepitation Index
x21	kg of stock sinter / ton of metal
	Equipment and environmental variables
x22	rain fall index (mm)
x23	tapping hole campaign index (1 or 0)
	Output variables
y1	wustite stoichiometric index (ω)
y2	gas flow resistance (ρ)
y3	metal silicon content ($[Si]$)
y4	heat losses in MJ / ton of metal (λ)
y5	sulfur in slag / sulfur in metal (Ks)

The predicted parameters are combined with other variables in a deterministic model to estimate the overall process pattern. The parameter θ represents the ratio between metal and gas produced. ϕ represents the unity gas flow calculated from the predicted gas resistance, ρ, and the pressure imposed by the equipment, blower and reactor. The overall performance index, π, is the final product of the model, meaning the amount of metal produced in a unity time for each cross section area unit and results form the product ϕ x θ. Figure 2 shows a cause and effect diagram for the hybrid model. The four final variables: $[Si]$, $[S]$, μ and π are efficient process performance indexes. The first two indicate metal silicon and sulfur content, respectively. The parameter μ, as defined by Rist (Rist and Misson, 1967) quantifies the specific consumption of reducing agents ($C + H_2$) and ultimate the metal cost.

The index π quantifies the amount of metal per unity area, therefore, the overall process productivity.

2.3 Experimental Industrial Application

The operation process using the model as supporting toll at Blast Furnace 3 is shown on Figure 3. The application will be extended to Blast Furnace 2 after the experimental application to Blast Furnace 3.

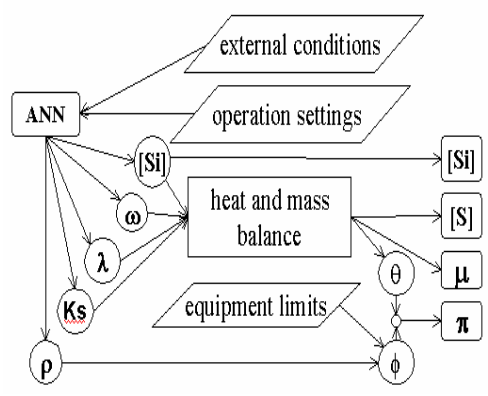


Fig 2. Data flow diagram for the hybrid model.

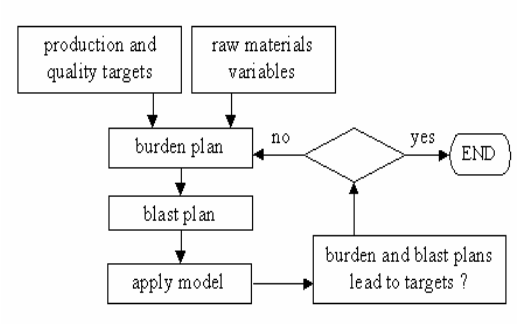


Fig 3. Industrial application procedure flow diagram.

3. RESULTS

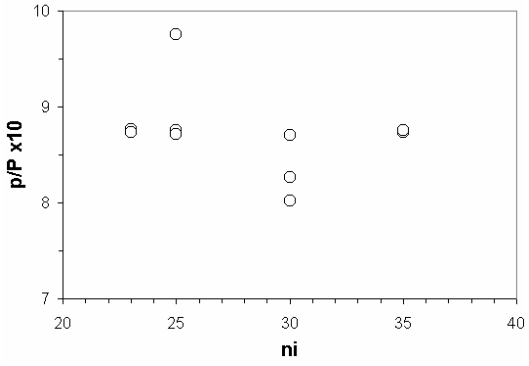


Fig 4. Relationship between effective parameter ratio (p/P) and the number of neurons in the hidden layer (n_1).

Figure 4 shows how the ratio between initial number (P) of network parameters - weights and bias – and the number of effective parameters after training (p)

behaves with the increase in the number of neurons in the hidden layer (n_1).

According to Foresse and Hagan (1997) the decreasing effective parameters ratio (p/P) indicate that the number of neurons is excessive. Another criterion leads to the same conclusion, as illustrated by Figure 5. It is clear that the larger number of hidden layer neurons does not contribute to a smaller mean quadratic error for the generalization set although the mean error for the training set decreased. In conclusion, the best network architecture was found to be 23x23x5.

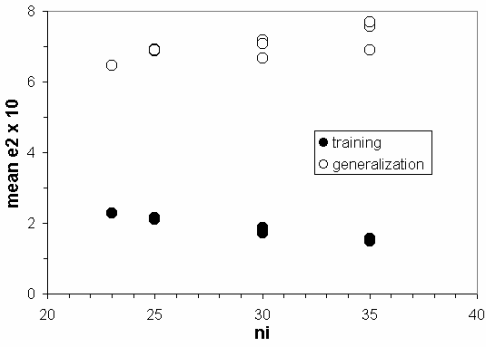


Fig 5. Relationship between mean quadratic errors (*mean* e^2) and the number of neurons in the hidden layer (n_1).

In this study and for the chosen architecture, the neuron activation function used in the hidden layer was a sigmoid one while a linear function was chosen for the output layer neurons.

Table 2 shows the mean square error for each of the 5 output variables expressed in terms of respective standard deviations. As expected, smaller mean quadratic errors are obtained for training sets. Mean errors for generalization sets were considered acceptable.

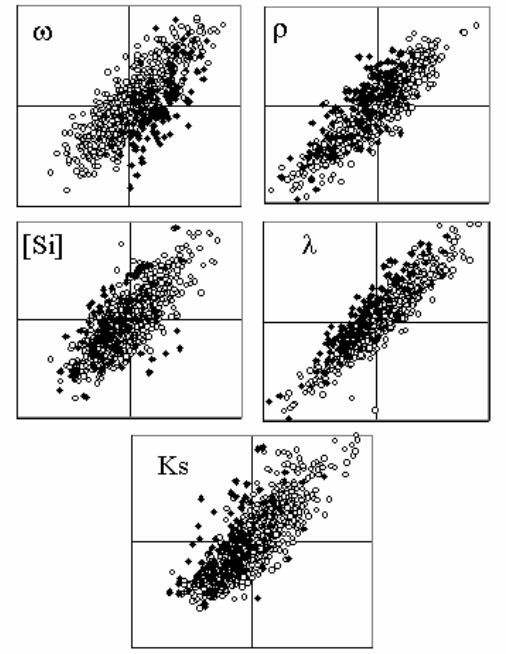
Table 2 – Square mean errors for the output variables in terms of respective standard deviations

	training	generalization
ω	0.503	0.722
ρ	0.400	0.651
[Si]	0.560	0.822
λ	0.354	0.496
Ks	0.531	0.729

Figure 6 shows how estimated standardized values (horizontal axis) match actual ones. It could be also noted from Figure 6 the tendency of experimental seen and unseen points to follow the diagonal line, indicating the accuracy of the network approach. The estimation of low values of ω and high vales of *Ks* was deficient for a few cases.

It should be pointed out that ϕ, θ and, consequently, π will depend not only on the values estimated by the network but also on other process variables. Therefore there is no point in estimating them at this moment.

Fig 6. Dispersion plots of the network output



variables (predicted values x actual values) for both training (O) and generalization (♦) sets. Axes cross at the mean value.

3.2 Experimental Industrial Application

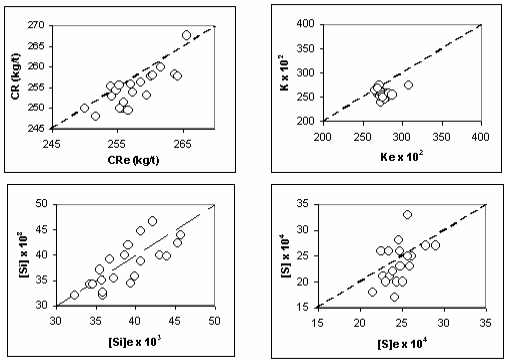
Following the steps previously described (Figure 3), the experimental industrial application was carried out during a twenty-day period. During the first twenty days of September 2005, the Blast Furnace number 3 operation was guided by the model. According to Figure 2, four variables were taken to access the prediction capacity of the model: coke-rate *(CR)*, metal silicon content *[Si]* , sulfur metal content *[S]* and Ergun fuidodynmic resistance index *(K)*. Figures 7 shows the results of the industrial observations while in Table 3 results can be numerically compared.

Table 3 – Statistical analysis of indexes observed during the test period without and with model

	error mean	error sd	set sd	population sd
CR (kg/t)?	2.8	2.7	4,6	36.9
[Si] x10^4	0.8	2.9	4,4	13.5
[S] x 10^5	1.5	3.8	4,4	7.9
K	19.0	11.42	8,9	24.0

It can be observed from table 3 that the error standard deviation was smaller than the relative standard for the experimental set and much smaller than the standard deviation observed in actual operation. For the fluidodynamic resistance it can be pointed out that the test period did not present sufficient variation

for the adequate assessment of the model on this aspect.



.
Fig 7 – Results of experimental industrial tests.

Because operational corrective actions were still too timid, fuel-rate corrections were allowed some hot metal temperature variations which contaminated sulfur control. This can be observed in Figure 8. In future, better heat control, with more confident use of the model, will also improve chemical quality, because chemical equilibrium is strongly connected to temperature.

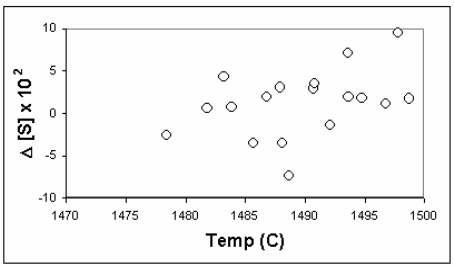


Figure 8 – Influence of hot metal temperature on the prediction error for hot metal sulfur

4. CONCLUSIONS

The main contribution of the present work is the development of a neural model which can increase prediction accuracy and operation performance while reducing costs for the blast furnace process at *Companhia Siderúrgica Nacional* (CSN - Volta Redonda/ RJ/ Brazil). Obtaining liquid iron in stable conditions is a very hard task, because the Blast Furnace is a complex process, conjugating several sub-processes. Some of them are continuous, some transient, occurring in the same reactor and still subject to oscillations in raw material composition.

The developed hybrid model, based on mass and heat balance associated to an artificial neural network, aims at supporting both operational and strategic decision making.
A 23x23x5 feedforward network proved to be an efficient architecture, using sigmoid and linear activation functions for the hidden and output neurons, respectively.

Except for fluidodynamic resistance, in other words, permeability, the period in which the model was used to guide industrial furnace operation proved to be predictable and consistent. For assessment of the permeability prediction a longer period will be necessary to allow for significant variation of that parameter.

The analysis of alternative raw materials or practice standards can be held also with the support of the model as long as the variables are kept inside the operating range studied.

It could be concluded that the neural model is a relevant tool to support an iron Blast Furnace operation since some corrections and retraining are carefully carried out by expert human operators in a systematic way. These procedures are crucial for adopting the neural model as a standard operating practice.

REFERENCES

Demuth, H., Beale, M., 2002. *Neural Network Toolbox User's Guide for Use with MATLAB®* The Mathworks Inc., Version 4, Reading:

Hagan, M., Chapter 5: *Backpropagation.*

Foresee, F.D., Hagan, M.T., 1997. Gauss-Newton approximation to Bayesian Learning. Proc. IJCNN, 1930-1935.

Himmelblau, D.M., 2000. Applications of artificial networks in chemical engineering. *Korean J. Chem. Eng.* **17** (4), 373-392.

Ping, H., Liu, L., Lihong, Y., Zhigang, H., Rong, D., Jingbo, X., Wei, C., Yisheng, T., Chenghuan, Y., Fengxi, L., 2003. Combining intelligent and mathematical models for BOS control at WISCO. *Steel Times International* **27** (8), 31-32.

Radhakrishnan, V.R. , Mohamed A.R.,2000. Neural networks for the identification and control of blast furnace hot metal quality. *Journal of Process Control* **10**, 509-524.

Rist, A.; Meysson, N. 1967. *Journal of Metals*, April, **50**..

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

AN INVERSE ARTIFICIAL NEURAL NETWORK BASED MODELLING APPROACH FOR CONTROLLING HFCS ISOMERIZATION PROCESS

Mehmet Yuceer and Ridvan Berber

Department of Chemical Engineering, Faculty of Engineering, Ankara University
Tandoğan 06100, Ankara - Turkey

Abstract: Isomerization of the glucose content of high fructose corn syrup (HFCS) into fructose needs to be strictly controlled in order to obtain a balanced product for sweetness and solubility, creating a non-trival problem. This work presents an approach to modelling of a real industrial isomerization reactor by artificial neural networks (ANN) pre-processed with principal component analysis (PCA). The initial model considered the exit fructose concentration as the output variable while the substrate flow rate to the reactor as the principal input (manipulated) variable. Then the neural network model was restructured and inversely trained by assuming the exit fructose concentration as the input variable and the feed flow rate as the output variable. Results indicate good performance by application of the developed strategy to an extensive industrial data set.

Keywords: ANN, PCA, HFCS.

1. INTRODUCTION

Glucose syrups and blends are used as an alternative to sugar or sucrose in many applications such as food, chiefly confectionery but also bakery and soft drinks. High Fructose Corn Syrup (HFCS) is a nutritive sweetener with high commercial potential. Most of HFCS is produced by the hydrolysis of starch into glucose. Glucose has only about 70 % of the sweetness of sucrose and is less soluble. At high concentrations, glucose syrup must be kept warm to avoid crystallization. On the other hand, fructose is 30 % sweeter than sucrose and twice as soluble as glucose at low temperatures (Asif and Abaseed, 1998). Using enzyme technology, the conversion of glucose to fructose by at least 50 % overcomes both problems giving a stable high-fructose corn syrup (HFCS) that is as sweet as a sucrose solution. Therefore, large percentage of the glucose derived from starch hydrolysis is converted into its sweeter-tasting isomer fructose, by the use of enzymes. The crystal clear syrup performs many of the same functions as sugar, but sold at a price considerably below sugar. Thus, HFCS is finding an increased use in soft drinks manufactured in the advanced countries. Presently, two normal grades i.e. 42 wt % HFCS and 55 wt % HFCS and an enriched grade 90 wt. % HFCS are commercially available.

In HFCS production it is important that the concentration of the product from isomerisation reactor maintains a constant value for consumer satisfaction. The process is complicated because of the interrelated influences arising from the enzyme activity, inflow concentrations, temperature and dry substrate. In practice, this control problem is generally solved by relying on the past experience of the operators with help from current daily process measurements. Therefore the industry is eager for sophisticated techniques that will allow them to control the process strictly and with ease. This in turn requires that a representative model of the isomerisation reactor be identified.

On the other hand, in cases where abundant data (i.e. process measurements) is available, one of the major developments in model building and control has been in the field of artificial neural networks (ANNs). During the last decade ANNs evolved from only a research tool into a tool that is applied to many real world engineering problems, statistics, even medical and biological fields. The number of European patents obtained in the last decade corroborates the trend of increased applications of ANNs (Kappen, 1996).

The fact that glucose content of HFCS is sweeter but less soluble than its fructose content dictates that the conversion of glucose to fructose is required, but at a certain level. Thus, this izomerization process needs to be strictly controlled in order to obtain a balanced product, creating a non-trival problem due to the complexities in the enzyme technology and interrelation of variables involved. The industry

tends to rely on data based techniques that would be representative of their past experience.

In this paper, we have investigated the modeling and control of the production of high-fructose corn syrup in an industrial isomerisation process by artificial neural networks with pre-processing with principal component analysis. An inverted control-oriented model was developed to regulate the fructose concentration in the reactor by manipulating the substrate flow rate. The back-propagation algorithm was applied to training and testing the network. The compliance of the prediction of the regular model to the real industrial data, and the possible use of the inverted control model is presented and discussed here.

2. PROCESS DESCRIPTION

The first step in the manufacture of HFCS is the production of aqueous starch slurry. For HFCS processing, corn is cleaned, and soaked in hot water containing a preservative such as dissolved SO_2. Determination and consequent removal of oil bearing germs is achieved through partial grading of corn. The oil-bearing germs are separated, dried and expelled to extract the oil, which is a by-product with high market value. Oil bearing genus free corn grains are ground and processed to remove fibrous materials, and proteins. The refined starch slurry is sent to a jet-cooking unit wherein an appropriate dose of enzyme alfa-amylase catalyses its conversion into meltodextrins which is a low dextrose equivalents (DE) oligosaccharide. The next step is saccharification, where the low DE syrup is further converted to dextrose by the action of glucoamylase enzyme. Most modern plants use continuous saccharifications process. It takes 65-75 hrs to obtain a 94-96% dextrose hydrolysate, which is then sent for isomerisation after proper refining. This dextrose syrup (94-96 % DS, dry substrate) is passed over columns (reactors) packed with immobilized isomerase enzyme to obtain 42 weight % HFCS. The degree of isomerisation can be controlled by the flow of the substrate. This part of the process, which is the subject of this study, is crucial in the sense that the concentration of the fructose in the reactor exit determines the final product specification, and thus needs to be strictly regulated. The enriched grade i.e. 90 % HFCS is obtained form the 42 % HFCS by chromatographic separation technique. 55 % HFCS is produced by blending 42 % HFCS with the enriched HFCS. Figure 1 depicts the input and output variables on a block representation of the isomerisation reactor.

3. ARTIFICIAL NEURAL NETWORK MODELLING WITH PCA PREPROCESSING

An Artificial Neural Network (ANN) is an information processing system that roughly replicates the behaviour of a human brain by emulating the operations and connectivity of biological neurons. It performs a human-like reasoning, learns the attitude and stores the relationship of the processes on the basis of a representative data set that already exists. In general, the neural networks do not need much of a detailed description or formulation of the underlying process, and thus appeal to practicing engineers who tend to mostly rely on their own data.

Depending on the structure of the network, usually a series of connecting neuron *weights* are adjusted in order to fit a series of inputs to another series of known outputs. When the weight of a particular neuron is updated it is said that the neuron is *learning*. The training is the process that neural network learns. Once the training is performed the verification is very fast. Since the connecting weights are not related to some physical identities, the approach is considered as a black-box model. The adaptability, reliability and robustness of an ANN depend upon the source, range, quantity and quality of the data set.

The feedforward neural networks consist of three or more layers of nodes: one input layer, one output layer and one or more hidden layers. The input vector passed to the network is directly passed to the node activation output of input layer without any computation. One or more hidden layers of nodes between input and output layers provide additional computations. Then the output layer generates the mapping output vector. Each of the hidden and output layers has a set of connections, with a corresponding strength-weight, between itself and each node of preceding layer. Such structure of a network is called a *Multi-Layer Perceptron* (MLP).

Neural network training can be made more efficient if certain preprocessing steps are performed on the network inputs and targets. Particularly in some situations where the dimension of the input vector is large and the components of the vectors are highly correlated, it is useful to reduce the dimension of the input vectors. An effective procedure for performing this operation is principal component analysis (PCA). This technique has three effects: it orthogonalizes the components of the input vectors (so that they are uncorrelated with each other); it orders the resulting orthogonal components (principal components) so that those with the largest variation come first; and it eliminates those components that contribute the least to the variation in the data set. The input vectors are multiplied by a matrix whose rows consist of the eigenvectors of the input covariance matrix. This produces transformed input vectors whose components are uncorrelated and ordered according to the magnitude of their variance.

A feed-forward back-propagation artificial neural network (BPNN) is chosen in the present study since it is the most prevalent and generalized neural network currently in use, and straightforward to implement. Figure 2 illustrates the basic

configuration of the network, particularly for the case of control-oriented inverted model. Each interconnection in the model has a scalar weight associated with it, which modifies the strength of the signal. The function of the neuron is to sum the weighted inputs to the neuron and pass the summation through a non-linear transfer function. In addition, a bias can also be used, which is another neuron parameter that is summed with the neuron's weighted inputs. Back-propagation refers to the way the training is implemented and involves using a generalized delta rule. A 'learning' rate parameter influences the rate of weight and bias adjustment, and is the basis of the back-propagation algorithm. The set of input data is propagated through the network to give a prediction of the output. The error in the prediction is used to systematically update the weights based upon gradient information. The network is trained by altering the weights until the error between the training data outputs and the network predicted outputs is small enough. There are many back-propagation training algorithms available. The choice of algorithm depends on the type of problem and may require experimentation of different algorithms. The algorithms have different computation and storage requirements, and train data at different speeds. The goal of selection is to efficiently and accurately train the network while keeping the speed of training relatively fast. In this work the Levenberg–Marquardt algorithm was used.

After generating sets of training patterns, appropriate NN architecture and associated parameters must be chosen for the particular application. The main design parameters are the number of hidden layers, number of neurons in each layer, and the neuron processing functions. The choice of these parameters will depend on the complexity of the system being modelled and they will affect the accuracy of the model. There is no exact guide for the choice of the numbers. The architecture of most ANN model is designed by trial and error.

In this work, a three-layer feed-forward network was created. The first layer has six hyperbolic tangent sigmoid neurons, the second layer has twenty logarithmic sigmoid neurons and last layer has one linear neuron. The performance function was calculated by using mean squared error. The network was trained for 2000 epochs.

4. RESULTS AND DISCUSSION

In this study the HFCS isomerisation process is modelled with ANN, and ANN with preprocessing PCA. The first regular model was created by considering the fructose concentration in the isomerization reactor as the output variable, such as illustrated in the process input-output diagram in Figure 1. Then the model was inverted such that the substrate flow rate was regarded as the output, so that a simple control strategy can be created to regulate the fructose concentration at a certain level by changing the input flow rate. The inverted model is depicted in Figure 2 where the network structure is also identified. Comparison of Figures 1 and 2 reveals the interchange of the substrate flow rate and the fructose concentration between the regular model and inverted model.

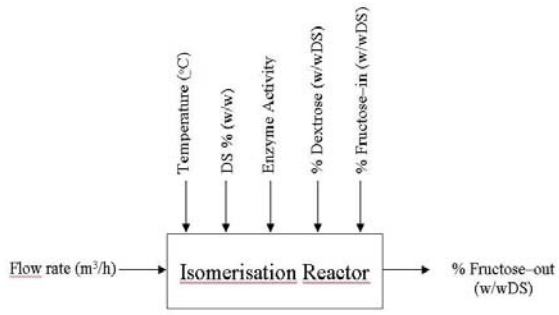


Fig. 1. Process block diagram represented by the regular model

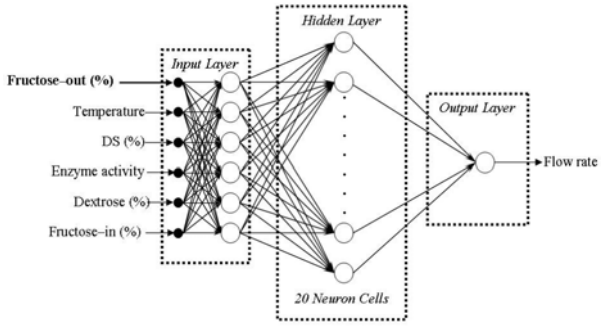


Fig. 2. Selected neural network structure depicting the form of inverted model

For development of neural network models the Neural Network Toolbox 4 and MATLAB 6.5 (The Mathworks Inc.) were used. A MATLAB script was written, which loaded the data file, trained and validated the network and saved the model architecture. The input and output data were normalised and de-normalised before and after the actual application in the network. A personal computer with Pentium-4 1.2 GHz processor and 512 Mb internal memory was used for processing neural network models.

The training and testing data was provided by Cargill Inc. from their Orhangazi–Turkey plant. Each data represents the measurements of one day, and the whole set is a mixture of data recorded from different reactors. Prior to conducting the network training operation using the back propagation algorithm, the industrial data set (1200 data altogether) was divided into two sections; a training set, which consisted of 1000 data, and a testing set that was formed by remaining 200 data. Dividing of training and testing sections was made by random selection.

The ANN model used first for regular representation of the reactor consists of six input nodes corresponding to (a) flow rate (manipulated variable for control) (m3/h), (b) temperature (^{o}C), (c) dry substrate DS (w/w %), (d) cumulative flow rate as

enzyme activity, (e) dextrose concentration in the input stream to the reactor (w/wDS %), (f) fructose concentration in the input stream to the reactor (w/wDS %). The single output was the fructose concentration in the reactor (w/wDS %). Thus a three layer feed-forward neural network was chosen for modeling purposes. In the hidden layer, twenty hidden neurons were used. For training, the classical back-propagation algorithm was used. Activation functions used were logarithmic sigmoid and tangent sigmoid. The selected 'control-oriented' network structure is shown in Figure 2 where the substrate flow rate and the fructose concentrations are interchanged to form the inverted model. MLP network structure and ANN parameters are shown in Table 1.

Table 1. ANN parameters

Parameter	ANN	ANN+PCA	Invers ANN+PCA
Learning rate	0.1	0.1	0.1
Epochs	2000	2000	2000
No. MLP layers	3	3	3
Input nodes	6	6	6
Hidden nodes	20	20	20
Output nodes	1	1	1
Training accuracy	86.6	99.47	92.8
Test accuracy	85	98.1	91.4

The model was first trained for the regular input-output behavior of the isomerisation reactor, whose results are shown in Figure 3 in parity plot form. The behaviour of the network for the test data is reflected in the following Figure 4. As can be detected from Figure 4, the network model captures the general trend in the output but obviously does not give very close prediction.

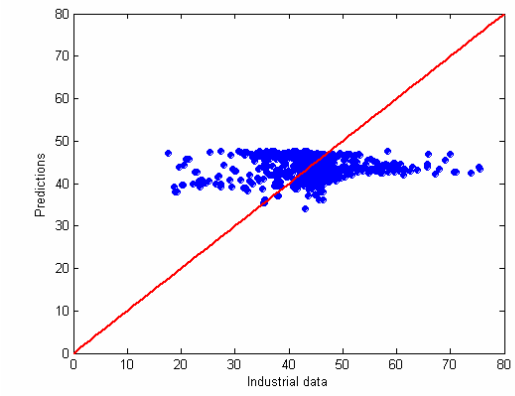


Figure 3. ANN model for learning data (% fructose in reactor)

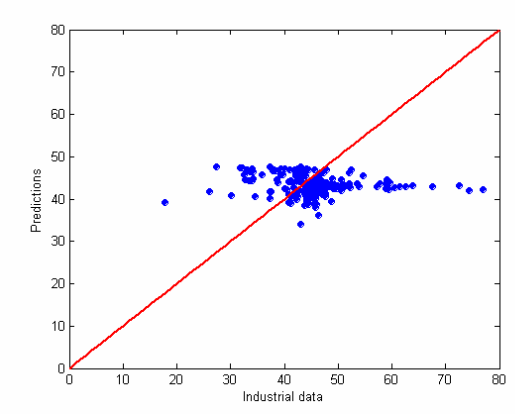


Figure 4. ANN model for test data (% fructose in reactor)

Therefore, preprocesssing with PCA technique was applied, whose results are depicted only for the test data in Figure 5 for predicting the fructose concentration. We have conservatively retained those principal components which account for 99.8 % of the variation in the data set. There was apparently redundancy in the data set, since the principal component analysis has reduced the size of the input vectors from 6 to 5.

When figures 4 and 5 are compared, it becomes evident that the pre-processing improves the prediction capability of the model tremendously.

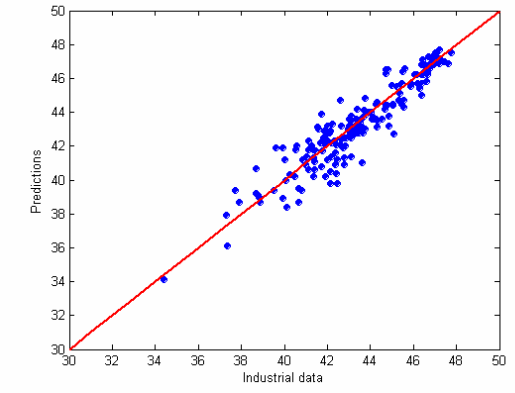


Figure 5. ANN model with PCA preprocessing for test data (% fructose in reactor)

Satisfied with the results obtained from pre-processed ANN model, the structure of model was then inverted to form the ANN model shown in Figure 2 with interchanged substrate input flow rate and fructose concentration in the reactor. Figure 6 shows the learning results of the network for predicting the flow rate of the substrate to the reactor. The inverted model was then validated with the test data revealing the results shown in Figure 7, which demonstrates close agreement of the model predictions with the industrial data.

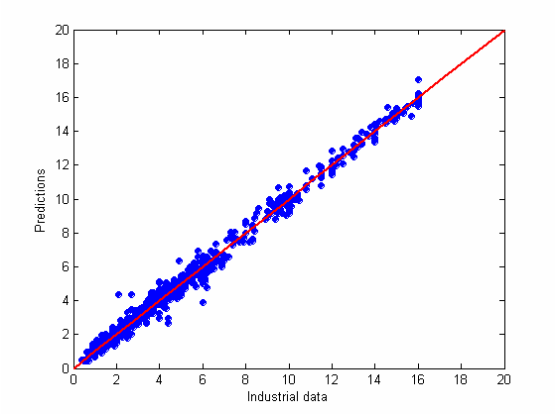


Figure 6. Inverted ANN model with PCA preprocessing for learning data (substrate input flow rate to the reactor, m^3/h)

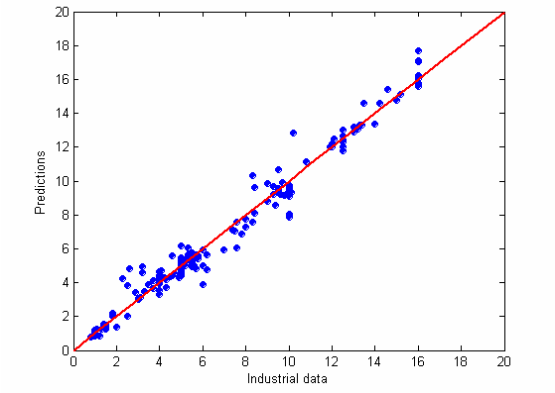


Figure 7. Inverted ANN model with PCA preprocessing for test data (substrate input flow rate to the reactor, m^3/h)

Overall evaluation of these results demonstrate that the ANN model preprocessed with PCA accurately predicts the behaviour of HFCS isomerisation reactor, and therefore holds promise for application in real industrial world.

5. CONCLUSIONS AND FUTURE WORK

An artificial neural network model with pre-processing with PCA was developed in this work to predict the substrate feeding rate to the isomerisation reactor in HFCS processing for control purposes. Since the sampling rate is inherently slow in the process, the results allow the interpretation that by implementation of the suggested model it will be possible to regulate the fructose concentration of the exit stream at the desired level.

With the promising results obtained from this work, our current efforts are directed towards two directions:

(a) Developing a graphical user interface (GUI) to implement the suggested strategy for controlling the fructose concentration in the HFCS isomerisation reactor. The suggested method, with the help of this GUI, will allow operators to decide what flow rate of input substrate is to be used for maintaining the fructose level at the output, based on daily measurements of other process variables.

(b) We are also considering a model-based control strategy with the ANN model. Although the sampling rate in practice is very slow to merit such a technique, it would be interesting to see the performance of a model predictive neural network controller for such real industrial data.

Acknowledgement: We truly acknowledge the industrial data provided by CARGILL Inc. Orhangazi-Turkey through Yontem Beyazkus and Ercan Erdas.

REFERENCES

Asif, M. and A. F. Abaseed (1998). Modelling of glucose isomerization in a fluidized bed immobilized enzyme bioreactor. *Bioresource Technology,* **64,** 229-235.

Kappen, H. J. (1996). An overview of neural network applications. *Proceedings 6th ICCTA*, Wageningen, the Netherlands, pp. 75-79.

The Mathworks Inc. 2003. http://www.mathworks.com

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

AN ALGORITHM FOR AUTOMATIC SELECTION AND ESTIMATION OF MODEL PARAMETERS

Argimiro R. Secchi[1], Nilo Sérgio M. Cardozo[2], Euclides Almeida Neto[3], Tiago F. Finkler[4]

1,2,4 - Grupo de Modelagem, Simulação, Controle e Otimização de Processos (GIMSCOP)
Departamento de Engenharia Química – Universidade Federal do Rio Grande do Sul
Rua Sarmento Leite, 288/24 – CEP: 90050-170 – Porto Alegre –RS – Brazil
Phone: +55-51-3316-3528 – Fax: +55-51-3316-3277
3 - PETROBRAS S/A – Brazil
E-mail: {[1]arge,[2]nilo,[4]tiago}@enq.ufrgs.br, [3]ean@petrobras.com.br

Abstract: An algorithm for automatic selection and estimation of model parameters is presented. The algorithm uses a sensitivity matrix based calculation of the parameters effects on the measured outputs and of a linear-independence metric. A predictability degradation index and a parameter correlation degradation index are used as stop criteria and the method is extended to dynamic models and multiple operating points. The applicability of the developed algorithm is illustrated through a hypothetical nonlinear input-output model and through the analysis of data from an experimental isothermal batch bioreactor. The obtained results show the effectiveness of the algorithm.

Keywords: Parameter Estimation, Sensitivity Matrix, Parameter Selection, Principal Component Analysis.

1. INTRODUCTION

Parameter estimation constitutes a key step in the identification and calibration of models. However, often only a subset of the parameters of the model can be estimated, due to limitations in the experimental window and the amount of data. In such a situation, the quality of the estimation is strongly dependent on the selection of the subset of parameters to be estimated. Consequently, a reasonable amount of effort has been made to automating the selection of parameters through the development of adequate criteria and procedures to the execution of this task (Weijers and Vanrolleghem, 1997; Brun *et al.*, 2002; Calvello and Finno, 2004; Ioslovich *et al.*, 2004; Li *et al.*, 2004).

Analysis of sensitivity has proven to be a valuable tool for identifying relevant and uncorrelated parameters. Different strategies based on the use of the sensitivity matrix have been proposed (Weijers and Vanrolleghem, 1997; Li *et al.*, 2004). A particularly systematic and effective identifiability measure method has been proposed by Li *et al.* (2004). In this method the magnitude of each parameter effect on the measured variables is quantified by applying principal-component analysis to a steady-state parameter-output local sensitivity matrix and the determination of the least uncorrelated parameters is accomplished recursively by computing the minimum distance between the sensitivity vector of a candidate parameter and the vector spaces spanned by sensitivity vectors of the parameters already selected for estimation.

Although the method proposed by Li *et al.* (2004) provides an effective ranking of the parameters of a given model, it does not provide criteria to the determination of the optimum number of parameters to be selected for the parameter estimation.

In this work, an algorithm for automatic selection of model parameters based on an extension of the identifiability measure of Li *et al.* (2004) is presented. In this algorithm a predictability degradation index and a parameter correlation degradation index are proposed to be used as stop

criteria. Additionally, the method is extended to dynamic models and multiple operating points.

2. FUNDAMENTALS

The proposed algorithm generates a ranking of the parameters according to their identifiability, measured through the magnitude of their effects on the output variables and a linear-independence metric. The magnitude of the effects of the parameters and the linear-independence metric are calculated from the sensitivity matrix, as proposed by Li *et al.* (2004).

Additionally, a predictability degradation index and a parameter correlation degradation index are defined to be used as stop criteria for the parameter selection algorithm, addressing the question of the number of parameters that should be estimated. The use of the predictability degradation index accounts for the fact the variability of the prediction is expected to increase when the optimum number of selected parameters is overcome. The use of the parameter correlation degradation index is intended to avoid the selection of an unnecessarily high number of parameters, which would increase the correlation between the parameters.

Another important feature in the proposed algorithm is the usage of global sensitivity matrix, which is composed by the information at each experimental point. In this way, the local calculations proposed by Li *et al.* (2004) for the magnitude of the effects of the parameters and the linear-independence metric was easily extended to deal with multiple operating points and dynamic data.

Li *et al.* (2004) have proposed a different dynamic extension procedure, based on a sensitivity matrix to be obtained as the weighted average of the local sensitivity matrices. Although the authors have not implemented this procedure and, consequently, there are not results to be used as basis of comparison, the usage of the global sensitivity matrix is expected to be a more reliable approach. The reason for this statement is that an average sensitivity matrix could lead to loss of information, mainly in problems where the sign of the gains are expected to change.

The definition of the degradation indexes for predictability and parameter correlation as well as the procedure to the calculation of the sensitivity matrix are presented in the next section.

3. AUTOMATIC PARAMETER SELECTION AND ESTIMATION ALGORITHM

The proposed algorithm for automatic selection of model parameters with simultaneous parameter estimation is based on an extension of the identifiability measure of Li *et al.* (2004) and on the proposed predictability degradation index.

For a given set $\{y \in \Re^{ny}, u \in \Re^{nu}\}$ of N experiments (or available process data in N steady-state operating conditions or N dynamic time-points) with r repetitions and a given nonlinear model of the process, the following algorithm is applied to estimate the best possible parameters within a set of $\theta \in \Re^{np}$.

Algorithm SELEST:
1) Evaluate the mean values of y and u for each experiment:

$$\bar{y} = \frac{1}{r}\sum_{k=1}^{r} y_k \in \Re^{ny \cdot N} \quad \text{and} \quad \bar{u} = \frac{1}{r}\sum_{k=1}^{r} u_k \in \Re^{nu \cdot N} \qquad (1)$$

and, if not given, compute the normalized measurement covariance matrix ($V_y \in \Re^{ny \cdot N \times ny \cdot N}$):

$$V_y = \frac{(y - \bar{y} \cdot 1_{1,r})(y - \bar{y} \cdot 1_{1,r})^T}{r-1} \otimes^{-1} \bar{y}\bar{y}^T \qquad (2)$$

where $\otimes^{-1}$ denotes element-by-element division, $1_{1,r}$ denotes a row vector of ones, and $y \in \Re^{ny \cdot N \times r}$.

2) Compute the normalized parameter-output sensitivity matrix, $S \in \Re^{ny \cdot N \times np}$, using an initial estimate of the model parameters, θ_o:

$$S = [S_1^T \; S_2^T \cdots S_N^T]^T \qquad (3)$$

where $S_j = (\Diamond \hat{y}_j)^{-1} \hat{S}_j \, \Diamond \theta_o \in \Re^{ny \times np}$, $\Diamond(\cdot)$ denotes the diagonal matrix of a vector, $\hat{y}_j \in \Re^{ny}$ is the model prediction for the j-th experimental point, using the input mean value $\bar{u}_j$:

$$\begin{aligned} &F(t_j, x, \dot{x}, \bar{u}_j; \theta_o) = 0 \quad , \quad x(0) = \bar{x}_o \\ &\hat{y}_j = H(x, \bar{u}_j; \theta_o) \end{aligned} \qquad (4)$$

and $\hat{S}_j$ is the parameter-output sensitivity matrix evaluated at the j-th point:

$$\hat{S}_j = \frac{\partial H}{\partial x} W_x + \frac{\partial H}{\partial \theta}. \qquad (5)$$

The parameter-state sensitivity matrix, $W_x = \frac{\partial x}{\partial \theta}$, is obtained by solving the following initial-value problem for dynamic processes:

$$\frac{\partial F}{\partial \dot{x}} \dot{W}_x + \frac{\partial F}{\partial x} W_x + \frac{\partial F}{\partial \theta} = 0 \quad , \quad W_x(0) = \frac{\partial x_o}{\partial \theta} \qquad (6)$$

or the linear system:

$$W_x = -\left(\frac{\partial F}{\partial x}\right)^{-1} \frac{\partial F}{\partial \theta} \qquad (7)$$

for steady-state processes.

3) Set m = min {np, $ny \cdot N$} and carry out the singular values decomposition of S left-weighted by the inverse of the normalized standard deviation of the measurements, $\sigma_i = \sqrt{(V_y)_{i,i}}$:

$$(\Diamond\sigma)^{-1} S = U \Sigma V^T \tag{8}$$

or, similarly, carry out the descending-ordered characteristic values decomposition of the Fisher information matrix:

$$F = S^T (\Diamond V_y)^{-1} S = V \Lambda V^T, \quad \Lambda = \Sigma^T \Sigma \tag{9}$$

where $\Diamond V_y$ denotes the diagonal matrix composed by the elements of the diagonal of V_y. Then, determine the overall effect of each parameter on the outputs by using the first m principal components (first m column vectors of matrix V, denoted by $V_m \in \Re^{np \times m}$) and the magnitude measure E (Li *et al*., 2004):

$$E = \frac{|V_m|\,\lambda}{\sum_{j=1}^{m} \lambda_j} \quad \in \Re^{np} \tag{10}$$

where $|V_m|$ denotes the matrix with absolute value of the elements of V_m, and λ are the first largest m characteristic values in Λ.

4) Select the highest ranked parameter $p_1 = \{\theta_k \mid E_k = \max_j E_j\}$ and set the number of selected parameters to $n = 1$ and the parameter index set to $\Omega_n = \{k\}$, representing the index set of the best possible parameters to be estimated with the given data set (in descending order).

5) Compute the reduced Fisher information matrix, F_n, regarding to the selected parameters p and the corresponding covariance matrices estimates of the parameters, V_p, and output predictions, $V_{\hat{y}}$:

$$F_n = S_\Omega^T (\Diamond V_y)^{-1} S_\Omega \quad \in \Re^{ny \cdot N \times n} \tag{11}$$

$$V_p = F_n^{-1} \tag{12}$$

$$V_{\hat{y}} = S_\Omega \, V_p \, S_\Omega^T \tag{13}$$

where S_Ω denotes the sub-matrix of S containing only the Ω_n columns. Also, compute the correlation coefficients of these covariance matrices, ρ_p and $\rho_{\hat{y}}$, and the condition number, κ, of F_p:

$$\rho_p = V_p \otimes^{-1} \sqrt{\Diamond V_p \; \Diamond V_p^T} \quad , \quad \overline{\rho}_p = \| \rho_p - I_n \|_\infty \tag{14}$$

$$\rho_{\hat{y}} = V_{\hat{y}} \otimes^{-1} \sqrt{\Diamond V_{\hat{y}} \; \Diamond V_{\hat{y}}^T} \quad , \quad \overline{\rho}_{\hat{y}} = \| \rho_{\hat{y}} - I_{ny} \|_\infty \tag{15}$$

$$\kappa = \|F_n\| \cdot \|V_p\| \tag{16}$$

where I_n denotes the identity matrix of size n, and $\| \cdot \|_\infty$ denotes the highest element of a matrix in absolute value. With this norm definition, $\overline{\rho}_p$ gives the highest correlation among the parameters.

6) Keeping the remaining parameters at the initial estimate θ_o, obtain a new estimate vector $\hat{p}_n$ for the parameters p by least square (or maximum likelihood) parameter estimation for the selected parameters. Also compute the normalized residuals ξ, the predictability degradation index ψ_n, and the parameter correlation degradation index η_n

$$\xi = \frac{1}{r} \sum_{k=1}^{r} [y_k - \hat{y}_k(\hat{p}_n)] \otimes^{-1} y_k \quad \in \Re^{ny \cdot N} \tag{17}$$

$$\psi_n = \overline{\rho}_{\hat{y}} + \| \xi \|_\infty \tag{18}$$

$$\eta_n = \overline{\rho}_p + \delta_{1,n} \tag{19}$$

where $\delta_{i,j}$ is the Kronecker delta. The addition of $\delta_{1,n}$ in Eqn. (19) is necessary to avoid an early stop in step 7 when $n = 2$.

7) Apply the following stop criteria, using a maximum allowed parameter correlation, ρ_{max}:

7.a) If $n > 1$ and ((($\psi_{n-1} < 1$ or ($\eta_{n-1} < \rho_{max}$ and $\eta_n > \rho_{max}$)) and $\psi_{n-1} < \psi_n$) or $\kappa^{-1} < \varepsilon$), then Ω_{n-1} is the solution index set and $\hat{p}_{n-1}$ is the corresponding estimated parameter vector, and terminate the algorithm. ε is the floating-point relative accuracy of the machine.

7.b) If $n = np$, then Ω_n is the solution index set and $\hat{p}_n$ is the corresponding estimated parameter vector, and terminate the algorithm.

8) If $n < m$, then compute the linear-independence metric d_j (Li *et al*., 2004) for each remaining parameter with respect to previously selected parameters:

$$d_j = \sin\left[\cos^{-1}\left(\frac{s_j^T V_\Omega \, s_j}{\| s_j \| \cdot \| V_\Omega \, s_j \|} \right) \right], \; \forall j \notin \Omega_n \tag{20}$$

where $V_\Omega = S_\Omega \left(S_\Omega^T S_\Omega \right)^{-1} S_\Omega^T$. Otherwise, i.e. $n \geq m$, compute the linear-independence metric $d_{q,j}$ for each remaining parameter with respect to all possible (m–1)-tuples Ωq of the previously selected parameters, for

$$1 \leq q \leq \frac{n!}{(m-1)!(n-m+1)!}, \tag{21}$$

where $\Omega q \subset \Omega_n$ and $|\Omega q| = m–1$, using Eqn. (22).

$$d_{q,j} = \sin\left[\cos^{-1}\left(\frac{s_j^T V_{\Omega q} \, s_j}{\| s_j \| \cdot \| V_{\Omega q} \, s_j \|} \right) \right], \; \forall j \notin \Omega_n \tag{22}$$

where $V_{\Omega q} = S_{\Omega q}\left(S_{\Omega q}^T S_{\Omega q}\right)^{-1} S_{\Omega q}^T$. And determine the worst-case metric: $d_j = \min_q d_{q,j}$.

9) Calculate the identifiability index I_j (Li *et al.*, 2004) for each remaining parameter θ_j:

$$I_j = E_j\, d_j \ , \ \forall\, j \notin \Omega_n. \qquad (23)$$

Select the next highest ranked parameter $p_{n+1} = \{\theta_k \mid I_k = \max_j I_j\}$, set the number of selected parameters to $n = n + 1$ and the index set to $\Omega_n = \{\Omega_{n-1}, k\}$, and return to step 5.

It is also possible to add the following diagnostic information in the exit conditions of step 7, evaluated at the stage n-1 (7.a) or n (7.b):

If $\overline{\rho}_{\hat{y}} \geq \rho_{\max}$ and $\overline{\rho}_p < \rho_{\max}$ then the outputs are too much correlated due to possibly high inputs correlation;

If $\overline{\rho}_{\hat{y}} \geq \rho_{\max}$ and $\overline{\rho}_p \geq \rho_{\max}$ then the outputs are too much correlated due to high parameter correlation;

If $\overline{\rho}_p \geq \rho_{\max}$ then the parameters are too much correlated.

The design constant $\rho_{\max}$ of the algorithm is an upper bound for the degree of parameter correlations. This limit is much easier to set than a threshold for the identifiability index I_j, whose value depends much more on experiments than statistic meanings.

4. ILLUSTRATIVE EXAMPLES

In order to illustrate the application of the algorithm SELEST, consider the following hypothetical nonlinear input-output model:

$$\begin{aligned} y_1 &= \theta_1 e^{-\theta_2/u_1} u_2\, u_3 + \theta_3 e^{-\theta_4/u_1} u_2\, u_4 \\ y_2 &= 1 - \theta_1 e^{-\theta_2/u_1} u_2\, u_3 + \theta_5 e^{-\theta_6/u_1} u_3 \\ y_3 &= \theta_7\, u_1 + \theta_8 \left(\theta_1 e^{-\theta_2/u_1} u_2\, u_3 + \theta_5 e^{-\theta_6/u_1} u_3\right) \end{aligned} \qquad (24)$$

with $y \in \Re^3$, $u \in \Re^4$, and $\theta \in \Re^{8+}$. The limited experimental data is composed by $N = 3$ operating points (OP) and $r = 3$ repetitions, shown in Table 1 for three cases. The initial estimate of the model parameters and their exact solution are given in Table 2. The repetitions were generated considering no errors in the inputs, using the exact parameters and adding to the outputs a noise with normal distribution, zero mean, and variance of 5% within 98% of significance level.

In the case 1, the most important input variable, u_1, is kept constant, reducing the estimation capability of the measurements. In the case 3, the last two OPs are correlated. The case 2 is the most favourable data set among the three cases.

Table 1. Experimental data sets for example 1.

	var.	OP_1	OP_2	OP_3
case 1	u_1	0.98	0.98	0.98
	u_2	0.73	0.13	0.43
	u_3	0.23	0.45	0.72
	u_4	0.67	0.47	0.13
	y_1	0.676/0.700/0.710	0.178/0.174/1.175	0.758/0.765/0.762
	y_2	0.623/0.621/0.614	0.933/0.919/0.911	0.332/0.338/0.341
	y_3	7.810/7.508/7.383	6.959/6.903/6.871	9.093/8.889/9.348
case 2	u_1	0.98	0.52	0.75
	u_2	0.73	0.13	0.43
	u_3	0.23	0.45	0.72
	u_4	0.67	0.47	0.13
	y_1	0.676/0.700/0.710	0.058/0.058/0.059	0.537/0.517/0.536
	y_2	0.623/0.621/0.614	0.954/0.969/0.975	0.527/0.545/0.539
	y_3	7.810/7.508/7.383	3.417/3.655/3.532	6.573/6.800/6.761
case 3	u_1	0.98	0.52	0.52
	u_2	0.73	0.13	0.13
	u_3	0.23	0.45	0.45
	u_4	0.67	0.47	0.57
	y_1	0.676/0.700/0.710	0.058/0.058/0.059	0.059/0.059/0.058
	y_2	0.623/0.621/0.614	0.954/0.969/0.975	0.988/0.984/1.002
	y_3	7.810/7.508/7.383	3.417/3.655/3.532	3.489/3.507/3.494

Observing the maximum normalized residuals in Table 2, the reduced-space parameter estimation had similar performance than the full space, showing that the additional parameters would have insignificant improvement in the model predictions. In fact, the index ψ shows the degradation of the predictability for the case 1 when adding the next parameter given by the identifiability index, and a very small improvement in the case 3. In both cases, according to η, the next parameter is highly correlated with the previous selected parameters ($\rho_{\max}$ was set to 0.99). Moreover, the full-space estimations were more sensitive to the initial estimates. The high residuals for the exact parameters are due to the random nature of the noise added to the outputs.

Table 2. Model parameters estimates for example 1. Bold results mean the estimates of the selected parameters in the order shown between parentheses.

par.	exact	θ_o	case 1	case 2	case 3
θ_1	7.65	6.50	6.50	**7.429(5)**	**7.551(5)**
θ_2	1.15	2.40	**0.988(1)**	**1.130(2)**	**1.133(3)**
θ_3	3.89	2.70	2.70	**3.161(3)**	**3.558(4)**
θ_4	1.75	1.50	**1.465(2)**	**1.603(1)**	**1.739(2)**
θ_5	0.23	0.01	**0.116(5)**	**0.136(7)**	**0.086(7)**
θ_6	0.79	0.15	0.15	**0.586(8)**	0.15
θ_7	6.32	4.25	**6.434(3)**	**6.400(4)**	**6.461(1)**
θ_8	3.42	5.50	**3.651(4)**	**3.608(6)**	**3.084(6)**
$\|\xi\|_\infty$	final estimate		0.0259	0.0172	0.0136
	exact		0.0447	0.0330	0.0306
	initial estimate		1.4702	0.8066	0.7774
	full estimate*		0.0261	0.0170	0.0130
ψ	n		1.0206	0.6435	1.0136
	n+1		1.0225	-------	1.0130
η	n		0.7027	0.9983	0.9697
	n+1		1.0000	-------	1.0000

*estimating all np parameters using the exact solution as initial guess.

Applying the diagnostic conditions at exit of step 7 of the algorithm SELEST, the results for the cases 1 and 3 say the outputs are too much correlated due to possibly high inputs correlation, and the parameters are too much correlated in the case 2. Indeed, case 3 was designed with high input correlation in OP_1 and OP_2 and in case 1 all operating points are correlated by the input variable u_1. In case 2, the correlation between θ_1 and θ_2 was the responsible for the high degradation index η_n.

The model parameters were estimated by the least square technique using the Levenberg-Marquardt method with the BFGS (Broyden, Fletcher, Goldfarb, and Shanno) updating scheme for the Hessian matrix (Edgar and Himmelblau, 1988) and relative error tolerance of 10^{-6} for variables and objective function.

Consider now a real example of a multi-route, non-structured kinetic model for microbial growth and substrate consumption of an experimental isothermal batch bioreactor to produce β-galactosidase by *Kluyveromyces marxianus* growing on cheese whey (Longhi *et al.*, 2004). The model is described by a set of five ordinary differential equations:

$$F(t,x,\dot{x},u;\theta) = \dot{x} - f(t,x,u;\theta) = 0 \quad , \quad x(0) = x_o$$
$$y = H(x,u;\theta) \qquad (25)$$

where the states $x \in \Re^5$ are biomass, lactose, ethanol, liquid-phase and gas-phase oxygen concentrations, $u \in \Re$ is the reactor temperature, and $y \in \Re^4$ are x_1, x_2, x_3, and oxygen saturation percentage (pO_2), which is a function of x_4 and u (Longhi *et al.*, 2004). Only one operating condition was used to test the algorithm, at u = 38ºC. The experimental data set is shown in Table 3, and the initial conditions for the state variables are {0.16, 48.90, 0, 0.0075, 1.152}. The model has 12 parameters and their initial estimates are given in Table 4.

Table 3. Experimental data for example 2.

time (h)	y_1 (g/L)	y_2 (g/L)	y_3 (g/L)	y_4 (%)
0	0.16	48.90	0.000	102.5
2	0.19	51.14	0.263	95.5
4	0.30	47.35	0.149	81.9
6	1.68	46.43	0.215	25.5
8	6.59	33.00	2.126	0.8
10	13.09	9.86	11.057	0.2
15	20.42	0.30	6.750	0.1
24	22.74	0.00	0.000	91.2
27	23.11	0.00	0.000	100.6
30	22.65	0.00	0.000	102.9

As shown in Table 4, the tuning of the first ten parameters ranked by the identifiability index got the best predictive capacity from the limited available experimental data. This is also proved in Figure 1, when comparing the prediction of the models adjusted by reduced-space and full-space parameter estimations with the experimental data. The most pronounced difference between the models appears in the dissolved oxygen and ethanol concentrations. In this example ρ_{max} was set to 0.98 and the system of ordinary differential equations was integrated by an implicit BDF method of variable order (Brenan *et al.*, 1989) with relative error tolerance of 10^{-6} and absolute error tolerance of 10^{-8}.

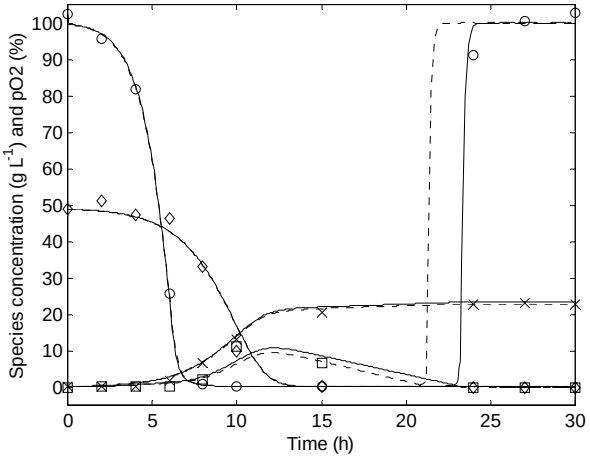


Fig. 1. Experimental data (symbols) for batch culture at 38ºC: (×) biomass concentration (y_1), (◊) substrate concentration (y_2), (□) ethanol concentration (y_3), and (o) dissolved oxygen concentration (y_4); and model predictions with reduced-space (solid line) and full-space (dotted line) parameter estimations.

Table 4. Model parameters estimates for example 2. Bold results mean the estimates of the selected parameters in the order shown between parentheses.

parameters*	θ_o	$\hat{p}$	$\hat{p}$ full
$\theta_1 = \mu_{1max}$	0.60	**0.600(2)**	0.600
$\theta_2 = \mu_{2max}$	0.06	0.06	0.054
$\theta_3 = \mu_{3max}$	0.16	**0.190(10)**	0.206
$\theta_4 = k_1$	20.00	**20.00(4)**	19.953
$\theta_5 = k_{ox1}$	1.00	**1.001(1)**	1.032
$\theta_6 = k_2$	4.26	**4.419(3)**	4.183
$\theta_7 = \phi_{X/S}^{oxid} / Y_{X/S}^{oxid}$	0.63	0.63	0.676
$\theta_8 = \phi_{X/E}^{oxid} / Y_{X/E}^{oxid}$	6.20	**6.257(7)**	6.206
$\theta_9 = 1/Y_{X/S}^{ferm}$	2.44	**2.411(5)**	2.466
$\theta_{10} = 1/Y_{X/S}^{oxid}$	2.63	**2.480(9)**	2.664
$\theta_{11} = \phi_{X/S}^{ferm} / Y_{X/S}^{ferm}$	0.85	**0.935(6)**	0.914
$\theta_{12} = 1/Y_{X/E}^{oxid}$	6.67	**6.257(8)**	6.669
$\|\| \xi \|\|_\infty$	0.1199	0.0787	0.0880
ψ	n	1.0787	1.0880
	n+1	1.0880	-------
η	n	0.9209	0.9860
	n+1	0.9867	-------

* See (Longhi *et al.*, 2004) for parameters definitions.

5. CONCLUSION

An algorithm for automatic selection and estimation of model parameters, based on the identifiability index of Li *et al.* (2004), has been proposed. The predictability and parameter correlation degradation indexes presented good performance as criteria for the determination of the number of parameters that

should be estimated. The usage of the global sensitivity matrix showed to be an adequate strategy to analyze the parameters effects on the outputs when dealing with multiple operating points or dynamic data. The employed examples showed that the algorithm was effective for estimating the best possible subset of parameters within a full set of model parameters, both for steady-state and dynamic models.

REFERENCES

Brenan, K, S. Campbell and L. Petzold (1989). *Numerical solution of initial-value problems in differential-algebraic equations*. Elsevier, New York.

Brun, R., M. Kühni, H. Siegrist, W. Gujer and P. Reichert (2002). Practical identifiability of ASM2d parameters—systematic selection and tuning of parameter subsets. *Water Res.*, **36**, 4113-4127.

Calvello, M. and R. J. Finno (2004). Selecting parameters to optimize in model calibration by inverse analysis. *Comp. & Geotech.*, **31**, 411–425.

Edgar, T.F. and D.M. Himmelblau (1988). *Optimization of Chemical Processes*. McGraw-Hill, New York.

Ioslovich, I., P.-O. Gutman and I. Seginer (2004). Dominant parameter selection in the marginally identifiable case. *Math. Comp. in Simul.*, **65**, 127–136.

Li, R., M.A. Henson and M.J. Kurtz (2004). Selection of Model Parameters for Off-Line Parameter Estimation. *IEEE Transactions on Control Systems Technology*, **12** (3) 402-412.

Longhi, L.G.S., D.J. Luvizetto, L.S. Ferreira, R. Rech, M.A.Z. Ayub and A.R. Secchi (2004). A Growth Kinetic Model of Kluyveromyces marxianus Cultures on Cheese Whey as Substrate. *Journal of Industrial Microbiology*, **31** (1) 35–40.

Weijers, S.R. and P.A. Vanrolleghem (1997). A procedure for selecting best identifiable parameters in calibrating activated sludge model no.1 to full-scale plant data. *Water Sci. Technol.*, **36**, 69-79.

Rigorous and Reduced Dynamic Models of the Fixed Bed Catalytic Reactor for Advanced Control Strategies

Eduardo C. Vasco de Toledo[a], Delba Nisi C. Melo[a], José Marcos F. da Silva[a], Viktor O. C. Concha[a], João Frederico da C. A. Meyer[b], Rubens Maciel Filho[a].

[a]*School of Chemical Engineering*
[b]*Institute of Mathematics, Statistics and Scientific Computing*
State University of Campinas - UNICAMP
PO Box: 6066 – Zip Code: 13081-970 - Campinas, SP - Brazil. FAX +55-1937883965
email: delba@lopca.feq.unicamp.br

Abstract: Rigorous and reduced heterogeneous dynamic models for fixed bed catalytic reactor were developed in this work. The models consist on mass and heat balance equations for the catalyst particles as well as for the bulk phase of gas. They also consider the variations in the physical properties and in the heat and mass transfer coefficients, the continuity equation for the fluid phase as well as the heat exchange through the jacket of the reactors. The models were used to describe the dynamic behaviour of the ethanol oxidation to acetaldehyde over Fe-Mo catalyst. The proposed models were able to predict the main characteristics of the dynamic behaviour of the reactors, and it was possible to compare the results obtained in simulations of models with different degrees of formulation complexity, thus indicating which model is more suitable for a specific application. This information is important for the real time integration implementation procedure.

Keywords: fixed bed, catalytic reactors, dynamic models, reduced models, dynamic behaviour.

1. INTRODUCTION

The design of cooled tubular reactors, involves complex tradeoffs between tube geometry, pressure drop, and heat-transfer area. Thus the behaviour of chemical reactors depends on variations in the inlet conditions, as well as in other physical and chemical parameters of the system. If the solid phase takes part in the process, e.g. heterogeneous catalysis, the task becomes very complex. The models commonly applied in simulation of heterogeneous fixed bed reactors usually gives results with quite good accuracy, but solving the model equation set is rather than difficult, when compared with pseudo-homogeneous models, because of the complex dynamic behaviour resulting from the non-linear distributed features which, among other things, give rise to inverse response resulting in catastrophic instabilities such as temperature runaway. This non-linearity is a consequence of heat generation by chemical reaction, and the inverse response arises from the presence of different heat capacities of the fluid and solid as well as the bulk flow of fluid causing interactions between heat and mass transfer phases. This causes differential rates of propagation of heat and mass transfer, which influence the heat generation through reaction on the solid catalyst. Therefore such models can be used only in a limited range, i.e. for preliminary calculation of reactor operation conditions. Solving pseudo-homogenous models is simple, but often their accuracy is low.

Without a detailed model it is not possible to make reliable predictions because the location of the regions with unstable behaviour, change in space and time are dependent on the input disturbances and control action. Obviously, it is essential to predict this behaviour for application in control. Reliable models depend on the insight of how the dominant physic-chemical mechanisms and external factors affect the overall performance. However, when on-line applications are required, reduced models have to be used, which can keep the essential characteristics of the system (McGreavy and Maciel Filho, 1989).

In this work, rigorous and reduced heterogeneous models for fixed bed catalytic reactors were developed. The proposed heterogeneous dynamic models for fixed bed catalytic reactors consist on mass and heat balance equations for the catalyst particles as well as for the gas phase, include the resistances to mass and heat transfer at the gas-solid interface and consider (or may not consider) the resistances inside the catalyst particle. The rigorous heterogeneous dynamic models are used in applications where computational accuracy may be more emphasized than computational speed, for instance, reactor design, planning of start-ups, shutdowns and emergency procedures (Martinez et al., 1985; Pellegrini et al., 1989; Elnashie and Elshishine, 1993). For real time implementation, as control and on-line optimisation, it is required to overcome the computational burden with a faster and easy numerical solution when compared to rigorous heterogeneous models. Bearing this in mind, reduced heterogeneous models were developed (McGreavy and Maciel, 1989).

The reduced models were obtained through mathematical order reduction, which eliminates the spatial co-ordinate of the catalyst particle and promote radially lumped-differential formulations (Maciel Filho, 1989; Vasco de Toledo, 1999). Therefore, one or more independent variable can be integrated leading to approximated formulations that retain detailed local information in the

remaining variables as well as medium information in the directions eliminated by the integration.
Considering the fixed bed catalytic reactor, the spatial dimension can be eliminated of the catalyst particle model (radial variable) thus generating bidimensional reduced models.
As a case study, the catalytic oxidation of the ethanol to acetaldehyde over Fe-Mo catalyst was considered, (Vasco de Toledo, 1999). It is a strongly exothermic reaction, representative of an important class of industrial processes.

2. REDUCTION TECHNIQUES

The solution of diffusion/reaction multidimensional problems present difficulties associated with a large analytic involvement and also request considerable computational effort. Consideration of such facts, becomes of interest for practical application in engineering, so that it is convenient to propose simpler formulations for the original system of partial differential equations, through the reduction of the number of its independent variables. Therefore, one or more independent variables can be integrated, leading to approximate formulations that retain detailed local information in the remaining variable as well as medium information in the directions eliminated by the integration. This information comes from the boundary conditions related to the eliminated directions. In this work, different reduction approaches(Classic, Hermite, Finlayson,Dixion,Generic) generating differentiates lumped formulations were investigated.
These techniques generate models that describe the reactor axial and radial profiles as a function of the time for the convenient explicit elimination of the dependence in the radial variable of the catalyst particle (bidimensional reduced model).Others approaches based on the use of wave propagation principle to describe the hot sport motion may be used since the could lead to have a smaller number of state what could be advantageous (Gilles and Epple, 1982; Marquardt, 1989). However, such approaches are not suitable to represent the system on all possible operating range without further identification. On the other hand models reduction based on mechanistic models are supposed to be valid in the whole domain.
Another possible way to generate a reduced model, which does not eliminate any dimension of the reactor, is to apply the method of the orthogonal collocation to the rigorous model of the reactors with only one internal collocation point in the radial direction. This technique does not simplify the mathematics/numeric solution of the reactor.
The proposed approaches for model reduction lead to different results since the model parameters are dependent upon the reduction techniques, Table 1.

2.1 Classic Reduction

This technique makes uses of the theorem of the mean value, given by equation (1) to produce the reduced models. In Table 1 is shown the parameters obtained for the Classic Reduction approach, as developed by Maciel Filho, 1989 and Vasco de Toledo, 1999.
For spherical co-ordinates:

$$[\ ]m = 3\int_0^1 r^2[\ ]dr \qquad (01)$$

where [ ]m defines a mean radial value of the amount inside of the left bracket.

2.2 Hermite Reduction

The technique makes use of the approach $H_{0,0}$ or $H_{1,1}$ and simultaneously of the theorem of the mean value for spherical co-ordinates, leading to the generation of the radial medium variables (Vasco de Toledo, 1999).

$$[\ ]m = 3\int_0^1 r^2[\ ]dr \qquad (02)$$

$$H_{0,0} = \int_0^1 y(x)dx \cong \frac{1}{2}\left[y(0) + y(1)\right] \qquad (03)$$

$$H_{1,1} = \int_0^1 y(x)dx \cong \frac{1}{2}\left[y(0) + y(1)\right] + \frac{1}{12}\left[y'(0) - y'(1)\right] \qquad (04)$$

where [ ]m defines a mean radial value of the amount inside of the left bracket.

2.3 Finlayson Reduction

Through the discretization of the radial term of the model for orthogonal collocation, and only using one collocation point, Finlayson manipulated the obtained equations analytically and generated a reduced model similar to apply the Hermite technique (Finlayson, 1971).

2.4 Generic Reduction

This technique is the generic mathematical representation of the reduction techniques mentioned previously. It leads to the generation of other parameters besides of the generated by other techniques (Hermite and Finlayson).

2.5 Dixion Reduction

Following the same methodology of Finlayson, Dixion discretizated the radial term of the model through orthogonal collocation. He manipulated the obtained equations analytically and generated a reduced model similar to apply other reduction techniques (Dixion, 1996).

2.6 Reduction Technique Using One Point Internal Radial Collocation

Another technique used to derive a reduced model (which does not eliminate a dimension of the reactor as the other techniques above described) refers to the application of the orthogonal collocation method to the model with only one internal collocation point in radial direction, for the

solid phase. In this work, in the rigorous models, 5 internal orthogonal collocation points were employed in the catalyst particle model.

In conclusion, the use of these techniques aims to evaluate the following points: the reduced models ability to predict the dynamic behaviour of the reactor; the computational time demanded to obtain dynamic profiles; the easiness/difficulties of the implementation and numerical convergence of these techniques.

One important feature when considering the reduced models is the less quantity of heat and mass transfer parameters necessary for their formulation, which does not imply in loss of prediction capability when compared to rigorous models. It may be important in case that is not possible to obtain trustful correlations to estimate some heat and mass parameters. This may be a restriction to the development of rigorous heterogeneous models.

3. DYNAMIC MODELS

The models developed here are based on the models proposed by Jutan et al., 1977; Martinez et al.,1985; Maciel Filho, 1989; Pellegrini et al., 1989; Elnashie and Elshishine, 1993; and Vasco de Toledo, 1999.

The models of the reactor were generated under the following considerations: variation of physical properties, mass and heat transfer coefficients, along the reactor length; intraparticle gradient negligible (reduced heterogeneous models); axial dispersion was neglected. Axial dispersion is found to be no significant for reactors with relatively high length /diameters, which is typical situation of industrial fixed bed catalytic reactors.

A possible way to have further model reduction is to neglect the time derivatives in respect to concentrations, since it leads to a significant reduction in the number of states. (Maciel Filho, 1989 and Vasco de Toledo, 1999). This was not considered in this paper because for chemical reactions with several by products.

Therefore, the heterogeneous models of the fixed bed catalytic reactor developed are:

3.1 Heterogeneous Model I - Rigorous Model (Tridimensional)

Reactant Fluid Mass Balance:

$$\varepsilon \frac{\partial X_g}{\partial t} = \frac{D_{ef}}{R_t^{\,2}} \frac{1}{r} \frac{\partial}{\partial r}\left[r \frac{\partial X_g}{\partial r} \right] - \frac{G}{\rho_g L} \frac{\partial X_g}{\partial z} + k_{gs}\, a_v\, (X_s^s - X_g) \tag{05}$$

Reactant Fluid Energy Balance:

$$\rho_g C_{pg} \varepsilon \frac{\partial T_g}{\partial t} = \frac{\lambda_{ef}}{R_t^{\,2}} \frac{1}{r} \frac{\partial}{\partial r}\left[r \frac{\partial T_g}{\partial r} \right] - \frac{G\, C_{pg}}{L} \frac{\partial T_g}{\partial z} + h_{gs}\, a_v\, (T_s^s - T_g) \tag{06}$$

Catalyst Particle Mass Balance:

$$\varepsilon_s \frac{\partial X_s}{\partial t} = \frac{D_s}{R_p^{\,2}} \frac{1}{r_p^{\,2}} \frac{\partial}{\partial r_p}\left(r_p^{\,2} \frac{\partial X_s}{\partial r_p} \right) + \frac{PM\, \rho_s\, R_W(X_s, T_s)}{\rho_g} \tag{07}$$

Catalyst Particle Energy Balance:

$$\rho_s\, C_{ps} \frac{\partial T_s}{\partial t} = \frac{\lambda_s}{R_p^{\,2}} \frac{1}{r_p^{\,2}} \frac{\partial}{\partial r_p}\left(r_p^{\,2} \frac{\partial T_s}{\partial r_p} \right) + \frac{\rho_s\, (-\Delta H_R)\, R_W(X_s, T_s)}{(R+1)} \tag{08}$$

Continuity Equation:

$$\frac{\partial}{\partial z}\left(\rho_g u_g \right) = 0 \tag{09}$$

Coolant Fluid Balance:

$$\frac{\partial T_R}{\partial t} = -\frac{u_R}{L} \frac{\partial T_R}{\partial z} + \frac{4\,U}{D_t\, \rho_R\, C_{pR}} (T_g(1, z, t) - T_R) \tag{10}$$

Momentun Balance:

$$\frac{\partial p}{\partial z} = -\frac{G^2\, L}{\rho_g\, D_p\, g_c} f \tag{11}$$

with the following boundary conditions:

$$\begin{aligned}
&r = 0 \quad \frac{\partial X_g}{\partial r} = \frac{\partial T_g}{\partial r} = 0, \\
&r_p = 0 \quad \frac{\partial X_s}{\partial r_p} = \frac{\partial T_s}{\partial r_p} = 0 \\
&r = 1 \quad \frac{\partial X_g}{\partial r} = 0, \frac{\partial T_g}{\partial r} = B_{ih}\, (T_R - T_g(1, z, t)) \\
&r_p = 1 \quad \frac{\partial X_s}{\partial r_p} = \frac{k_{gs}\, R_p}{D_s} \left(X_g - X_s^s \right), \\
&r_p = 1 \quad \frac{\partial T_s}{\partial r_p} = \frac{h_{gs}\, R_p}{\lambda_s} \left(T_g - T_s^s \right), \\
&z = 0 \quad X_g = X_s = 0, \quad T_g = T_{go}, \quad T_s = T_{so}, \\
&T_R = T_{Ro}, \quad p = p_o
\end{aligned} \tag{12}$$

The Equations 09, 10 and 11 are valid also for the remaining models.

The following notations are used, a_v is the external particle surface area per unit of catalyst volume, m^{-1}; B_{ih}, Biot number; C_p, calorific capacity, kcal/kg.K; D_{ef}, radial effective diffusivity, m/h; D_p, particle diameter, m; f, friction factor; G, mass flow velocity, $kg/m^2.h$; gc, conversion factor; h_{gs}, particle to fluid convective heat transfer coefficient, $kcal/m^2.h.K$; h_W, convective heat transfer coefficient in the vicinity of the wall, $kcal/m^2.h.K$; k_{gs}, particle to fluid mass transfer coefficient, m/s; L, length of the reactor, m; p, pressure of the reactor, atm; PM, the mean molecular weight, kg/kmol; r, dimensionless radial distance of the reactor; r_p, dimensionless radial distance of the particle; R, air/ethanol ratio; R_t, reactor radius, m; R_p, particle radius, m; R_W, rate of the oxidation, kmol of reactant mixture/h.kgcat; T, reactor temperature, K; T_{fo}, feed temperature, K; T_{go}, feed temperature, K; T_{so}, catalyst feed temperature, K; T(1,z,t), wall temperature of the reagent fluid, K;

T_R, coolant temperature, K; T_{Ro}, coolant feed temperature, K; t, time, h; u, velocity, m/h; U, global heat transfer coefficient, kcal/m^2.h.K; X, conversion; z, dimensionless axial distance. Greek letter: λ, conductivity, kcal/m.h.K; ΔH_R, enthalpy of reaction molar, kcal/kmol; ρ, density, kg/m^3; ρ_B, catalyst density, kgcat/m^3; ρs, catalyst density, kgcat/m^3; ε, porosity. Subscripts: ef, effective; f, fluid; g, gas; i, interstitial; o, feed; R, refrigerant; s, solid. Superscripts: s, condition at external surface.

3.2 - Heterogeneous Model II - Reduced Models (Bidimensional)

The use of the reduction techniques leads to the generation of the radial mean variables. The generate model describes the axial and radial profiles as a function of the time for the convenient explicit elimination of the dependence in the radial variable of the catalyst particle. The reduced models were generated applying the reduction techniques (Classic, Hermite, Finlayson, Generic and Dixion) to the Heterogeneous Model I (Rigorous Model).

Reactant Fluid Mass Balance:

$$\varepsilon \frac{\partial X_g}{\partial t} = \frac{D_{ef}}{R_t^{\,2}} \frac{1}{r} \frac{\partial}{\partial r}\left[r \frac{\partial X_g}{\partial r} \right] - \frac{G}{\rho_g L} \frac{\partial X_g}{\partial z} + \alpha_m k_{gs} a_v (\lambda X_{sm} - X_g) \tag{13}$$

Reactant Fluid Energy Balance:

$$\rho_g C_{pg} \varepsilon \frac{\partial T_g}{\partial t} = \frac{\lambda_{ef}}{R_t^{\,2}} \frac{1}{r} \frac{\partial}{\partial r}\left[r \frac{\partial T_g}{\partial r} \right] - \frac{G\, C_{pg}}{L} \frac{\partial T_g}{\partial z} + \alpha_t h_{gs} a_v (\lambda T_{sm} - T_g) \tag{14}$$

Catalyst Particle Mass Balance:

$$\varepsilon_s (1-\varepsilon) \frac{\partial X_{sm}}{\partial t} = -\alpha_m k_g a_v (\lambda X_{sm} - X_g) + \frac{PM (1-\varepsilon) \rho_s R_W (X_{sm}, T_{sm})}{\rho_g} \tag{15}$$

Catalyst Particle Energy Balance:

$$(1-\varepsilon) \rho_s C_{ps} \frac{\partial T_{sm}}{\partial t} = -\alpha_t h_{gs} a_v (\lambda T_{sm} - T_g) + \frac{(1-\varepsilon) \rho_s (-\Delta H_R) R_W (X_{sm}, T_{sm})}{(R+1)} \tag{16}$$

with the following boundary conditions:

$$\begin{aligned} &r = 0 \quad \frac{\partial X_{gm}}{\partial r} = \frac{\partial T_{gm}}{\partial r} = 0, \\ &r = 1 \quad \frac{\partial X_{gm}}{\partial r} = 0, \frac{\partial T_{gm}}{\partial r} = B_{ih} (T_R - T_{gm}(1, z, t)) \\ &z = 0 \quad X_{gm} = 0, \quad T_{gm} = T_{go}, T_R = T_{Ro}, p = p_o \end{aligned} \tag{17}$$

X_{sm} and T_{sm} are mean radial conversion and temperature of the reactor solid phase.

In equations 13 to 16 the parameters α_i and λ vary according to the reduced model, as described below in Table 1.

Table 1. Bidimensional Reduced Models

MODELS	PARAMETERS	
	α_i	λ
Classic	$\alpha_m = 1$ $\alpha_t = 1$	1
Hermite $H_{0,0}$	$\alpha_m = 1$ $\alpha_t = 1$	2/3
Hermite $H_{1,1}$ factor = 4 Finlayson factor = 3; Generic factor ≠ 4.	$\alpha_m = \left(\frac{\text{factor}}{\text{factor} + B_{im}} \right)$, $\alpha_t = \left(\frac{\text{factor}}{\text{factor} + B_{ih}} \right)$, $B_{im} = \frac{K_{gs} R_p}{D_s}$, $B_{ih} = \frac{h_{gs} R_p}{\lambda_s}$	1
Dixion	$\alpha_m = \frac{3B_{im}(B_{im}+4)}{B_{im}^{\,2} + 6B_{im} + 12}$ $\alpha_t = \frac{3B_{ih}(B_{ih}+4)}{B_{ih}^{\,2} + 6B_{ih} + 12}$	1

As a case study, the catalytic oxidation of the ethanol to acetaldehyde over Fe-Mo catalyst was considered. It is a strongly exothermic reaction, representative of an important class of industrial processes (Vasco de Toledo, 1999).

$$\begin{aligned} B &= 2K_1K_2P_{O2}P_{ET} \\ A &= K_3K_1P_{ET}P_{AC} + K_1P_{ET} + 2K_2P_{O2} + K_3K_4P_{AC}P_{H2O} \\ R_W &= \frac{B}{A} \end{aligned} \tag{18}$$

P_{O2}, P_{ET}, P_{H2O}, P_{AC} are partial pressure of oxygen, ethanol, water and acetaldehyde, respectively; K_i are the kinetic constants in the Arrhenius form (Vasco de Toledo, 1999).

The numeric solution of the models was obtained using the method of the lines in conjunction with the orthogonal collocation, which showed to be an effective procedure for the space discretization (radial and axial directions), employing the LSODAR algorithm for the integration in time (Villadsen and Michelsen, 1978; Vasco de Toledo, 1999).

4. RESULTS AND DISCUSSIONS

Initially, Figures 1 and 2 show at the axial distance the dynamic profiles (surfaces) of temperature using the Heterogeneous Model I (Rigorous), which is a more realistic representation of the dynamic behaviour since it is considered a more detailed representation of the physic-chemical phenomena taking place in the system.

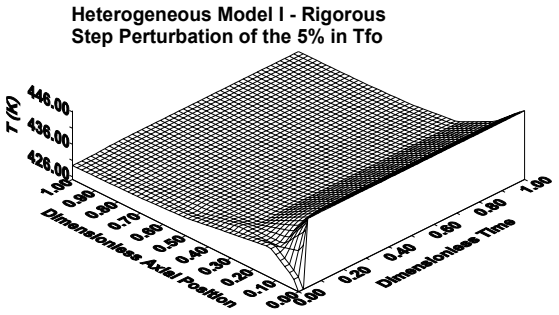


Figure 1. Temperature of the reactant fluid in the reactor: step perturbation in $T_{fo.}$

After a step perturbation in the reactant feed temperature, T_{fo}, the inverse response phenomenon is observed in the temperature profile at the beginning of the reactor (Figure 1). This is a typical characteristic of fixed bed catalytic reactors. It is also clear the great influence of the variation of T_{fo} in the dynamic behaviour. The identification of this phenomenon is very important in the elaboration of safe and efficient control strategies, since it determines the choice of manipulated and controlled variables as well the axial positions where the control of the system must be done.

Another important feature of this reactor is the presence of a hot spot located in the region where the inverse response phenomenon takes place.

In Figure 2 is represented the temperature dynamic behaviour of the reactor after a step perturbation in the coolant fluid feeding temperature, T_{ro}, where is observed an asymptotic dynamic behaviour.

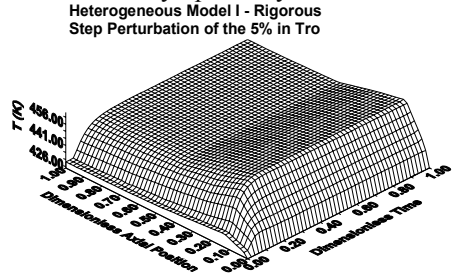


Figure 2. Temperature of the reactant fluid in the reactor: step perturbation in T_{ro}.

Comparing the results obtained in Figure 1 and 2, T_{ro} had a greater influence in the thermal profile at the axial position and this perturbation did not caused the inverse response phenomenon. Therefore, T_{ro} is more adequate to be chosen as a manipulated variable. Nevertheless, this conclusion must not be generalized for all operational and design conditions, since the determination of the manipulated variables depends on this conditions. For this reason, care should be taken in the choice of the manipulated and controlled variables in the elaboration of an efficient and safe control strategy.

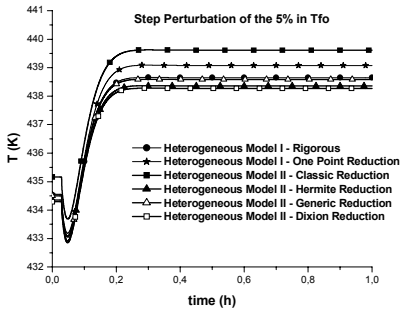


Figure 3. Temperature of the reactant fluid in the reactor ($z = 0.03$ m): step perturbation in T_{fo}

The following figures present the dynamic behaviour of the reactor predicted by the bidimensional reduced models and the rigorous tridimensional model. Figures 3, 4 and 5 show the thermal and conversion dynamic profile after a step perturbation in T_{fo}. At this condition, near the entrance of the reactor ($z = 0.03$ m) and at $z = 0.13$ m, there is an inverse response of the temperature (Figures 3 and 4, respectively). The profiles generated by the reduced models are in good agreement to those obtained with the rigorous model. This is also observed in the reactor conversion (Figure 5). It is important to mention that the computational time during simulations of the reduced models was up to 10 times less than that of the rigorous model. This time gain, without loss of prediction quality, justifies the use of the reduced models for applications in control and optimization cases represented in Figures 6 and 7 (step perturbation of T_{ro} and G, respectively).

Despite of the better reproduction of the rigorous model by the generic reduction technique, it is important to bear in mind that this result may differ according to new design and operation conditions.

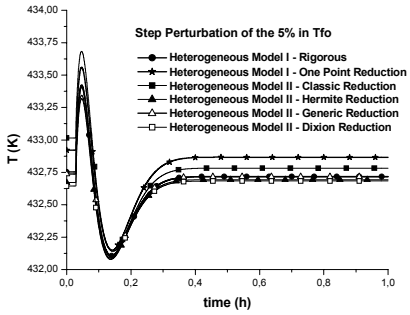


Figure 4. Temperature of the reactant fluid in the reactor ($z = 0.13$ m): step perturbation in T_{fo}

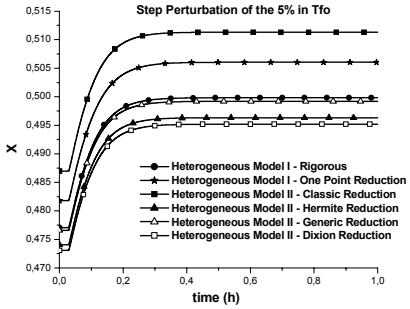


Figure 5. Conversion of the reactant fluid in the reactor (z = 0.03 m): step perturbation in T_{ro}.

In a nutshell, the results obtained with the models proposed to represent the fixed bed catalytic reactor show that the bidimensional reduced models had a good capacity to predict the dynamic behaviour described by the tridimensional model. Since all bidimensional models demanded similar computational time, the choice of a specific model will depend on the ability of the model to describe the real reactor behaviour.

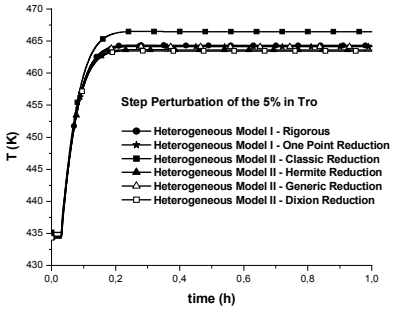


Figure 6. Temperature of the reactant fluid in the reactor (z = 0.03 m): step perturbation in T_{fo}.

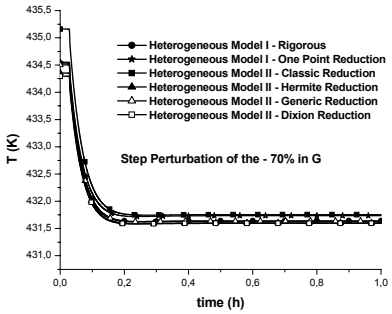


Figure 7. Temperature of the reactant fluid in the reactor (z = 0.03 m): step perturbation in G.

5. CONCLUSIONS

The proposed heterogeneous models were able to predict the main characteristics of the dynamic behaviour of the fixed bed catalytic reactor, including the inverse response phenomena and the hot spot present in the former. This knowledge is essential to design and control these reactors. The computational time demanded for the solution of the Heterogeneous Model I (rigorous) is high in comparison to the Reduced Models, which restricts the use of the rigorous models in cases where time is not a limiting factor. Otherwise, when on-line applications are required, the reduced models showed to be more adequate. The models based on reduction techniques overcame computational burden with a faster and easier numerical solution, as well as other difficulties found in rigorous heterogeneous models, especially related to the large number of parameters and sophisticated numerical procedures required to the solution. It is important to mention that for different design and operational conditions used in this work, the performance of the reduced models must be re-evaluated.

ACKNOWLEDGEMENTS

The authors are grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq for the financial support.

REFERENCES

Dixion, A., G. (1996). "An Improved Equation for the Overall Heat Transfer Coefficient in Packed Beds", Chemical Engineering and Processing, **35**, 323-331.

Elnashaie, S. S. E. H. and Elshishini, S. S. (1993). "Modelling, Simulation and Optimization of Industrial Fixed Bed Catalytic Reactors, Topics in chemical engineering Volume 7", Gordon and Breach Science Publishers .

Finlayson, B. A. (1971). "Packed Bed Reactor Analysis by Orthogonal Collocation", Chem. Eng. Sci., **26**, 1081-1091.

Jutan, A., Tremblay, J. P., McGregor, J. F., and Wright, J. D. (1977). "Multivariable Computer Control of a Butane Hydrogenolysis Reactor: Part I". State Space Reactor Modelling, AIChE Journal, **23**, 5, 732-742.

Maciel Filho, R. (1989). "Modelling and Control of Multitubular Reactors", Ph.D. Thesis, University of Leeds, Leeds,1989.

Marquardt, W. (1989)."Wellenausbreitung in verfahrenstechnischen Prozessen", Chemie Ingenieur Technik, **61**, 5, p. 362-377.

Gilles, E.D.; Epple, U. (1982). "Model reduction of the fixed-bed reactor", in Book: Recent advances in adsorption and ion exchange, AICHE Symposium Series.

Martinez, O. M., Pereira Duarte, S. J., and Lemcoff N., O. (1985). "Modeling of Fixed Bed Catalytic Reactors", Comp. Chem. Eng., **9**, 5, 535-545.

McGreavy, C. and Maciel Filho, R. (1989). "Dynamic Behaviour of Fixed Bed Catalytic Reactors", in "IFAC Dynamics and Control of Chemical Reactors", Maastricht, the Netherlands .

Pellegrini, L., Biardi, G. and Ranzi, E. (1989). "Dynamic Model of Packed-Bed Tubular Reactors", Comp. Chem. Eng., **13**, 4/5, 511-518.

Vasco de Toledo, E. C. (1999). "Modelling, Simulation and Control of the Fixed Bed Catalytic Reactors", Ph.D. Thesis, University of Campinas, São Paulo, Brazil .

Villadsen, J., Michelsen, M. L. (1978)."Solution of Differential Equation Models by Polynomial Approximation". Prentice-Hall, New Jersey.

Session 6.2

Optimization and Scheduling

Dynamic Penalty Formulation for Solving Highly Constrained Mixed-Integer Nonlinear Programming Problems

C. M. Silva and E. C. Biscaia Jr.
Universidade Federal do Rio de Janeiro

Application of Genetic Algorithms to the Optimization of an Industrial Reactor

I. R. de Souza Victorino and R. M. Filho
State University of Campinas

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

MODELING OF NLP PROBLEMS OF CHEMICAL PROCESSES DESCRIBED BY ODEs

Tvrzská de Gouvêa, M. [1] **and Odloak, D.**[2]

[1]*Universidade Presbiteriana Mackenzie, Department of Materials Engineering*
Rua da Consolação 896, 01302-907, São Paulo-SP, Brazil; miriamtg_br@yahoo.com
[2]*Universidade de São Paulo, Department of Chemical Engineering*
odloak@usp.br

Abstract: Both real-time and off-line optimizations are commonly performed in order to enhance productivity. The optimization problem is often posed as a non-linear programming (NLP) problem solved by a SQP algorithm. When processes need to be described by differential equations, difficulties will arise in using SQP algorithms, since Jacobians of constraints described by differential equations will have to be evaluated. In this paper, we show how to derive analytical expressions for both Jacobian and Hessian matrices for the constraints described by ordinary differential equations, without increasing the dimension of the resultant NLP problem to be solved.

Keywords: differential equations, nonlinear programming, process models, optimization problems.

1. INTRODUCTION

Globalization has lead to the necessity of optimally operating the chemical plants. Thus, not only is it necessary to adequately control the chemical processes, but, moreover, optimal operating conditions must be continuously forecast and implemented, which may be achieved by solving a real-time optimization (RTO) problem. As far as continuous processes described by concentrated parameters models are under regard, optimal operating policies can be obtained by solving nonlinear programming (NLP) problems, possessing constraints described solely by algebraic equations. SQP algorithms may effectively solve NLP problems and successful implementations of real time optimization strategies are well known (Zanin et al., 2002, Jakhete et al., 1999, Ellis et al., 1998, Agrawal et al., 1996). When it comes to optimize continuous processes described by distributed models or semi-batch and batch processes, one difficulty arises when the corresponding optimization problem is to be solved by a SQP algorithm. The latter will require that Jacobians or even a Hessian matrix be evaluated for the constraints represented by differential equations. One common approach to override the difficulties arisen from the mathematical description of the process by differential equations is to discretize the latter (Cuthrell and Biegler, 1989), i.e., the continuous state variables are transformed into several discrete variables. This approach leads to an increase in the dimension of the NLP problem to be solved and to the loss of information due to the discretization performed. In this paper, we aim to present a general procedure to analytically model Jacobians and Hessian matrices of the constraints described by ordinary differential equations (ODEs). The resulting model for evaluation of the Jacobians and Hessians is composed by ordinary differential equations that are coupled to the differential equations describing the process model. Thus, evaluation of the Jacobians and Hessians may be obtained with the same numerical precision as the solution of the process model and without any loss of information.

In section 2, we briefly review the available different SQP algorithms emphasizing the need for the evaluation of Jacobians and even Hessians. In section 3 we present guidelines on how to write the

optimization problem. In sections 4 and 5, the models for calculating Jacobians and Hessians are presented. In section 6, we apply the proposed procedure to the optimization problem of a batch reactor. The paper is finally concluded in section 7.

2. GENERAL STEPS OF SQP ALGORITHMS

Given a NLP problem as in (1), the SQP algorithms produce a sequence of values as in (2), where the search direction d_k is the solution of the QP problem (3) and α_k is so as to guarantee $f(x_{k+1}) < f(x_k)$. Hence, the general idea of the SQP algorithms can be summarized in algorithm A_o.

$$f(x^*) = \min_{x \in S \cap B_g(x^*)} f(x) \tag{1}$$

where,

$S = \{x \in R^n : h(x) = 0; g(x) \le 0\}$,

$f: R^n \to R; h: R^n \to R^m; g: R^n \to R^p$, $B_\delta(x^*)$ is an open–ball of radius δ with center in x^*.

$$\{x_k\} \to x^* : x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

$$\min_{d_k \in S_l} \frac{1}{2} d_k^T H_k d_k + \nabla f^T(x_k) d_k \tag{3}$$

where, H_k is either equal to $\nabla^2 L(x_k, \lambda_k, \mu_k)$ or an approximation to it and $\nabla^2 L(x_k, \lambda_k, \mu_k)$ is the Hessian evaluated at the point x_k, λ_k, μ_k of the Lagrangian function associated with the NLP defined as in (4), S_l is the set of constraints, which is either chosen as in (5) or (6).

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x) \tag{4}$$

where, λ and μ are the Lagrange multipliers of the equality and inequality constraints of the NLP.

$$S_I = \begin{Bmatrix} d_k \in R^n : \nabla h^T(x_k) d_k = -h(x_k); \\ \nabla g^T(x_k) d_k \le -g(x_k) \end{Bmatrix} \tag{5}$$

$$S_I = \begin{Bmatrix} d_k \in \Delta_k : \nabla h^T(x_k) d_k = -h(x_k); \\ \nabla g^T(x_k) d_k \le -g(x_k) \end{Bmatrix} \tag{6}$$

where, Δ_k is the trust region where the linear approximation S_I of S is expected to hold well.

Algorithm A_o

1. $k = 0$; $x_k = x_o$
2. Solve the QP problem to obtain d_k
3. Obtain α_k so that $f(x_k + \alpha_k d_k) < f(x_k)$
4. $x_{k+1} = x_k + \alpha_k d_k$ and check the optimality condition on x_{k+1}. If it is satisfied stop otherwise $k = k + 1$ and return to 2.

Differences in the SQP algorithms are related to whether analytical expressions are provided for the Hessian matrix H_k or if it is estimated from the Jacobian matrix of the constraints and enforced to be positive definite, to whether (5) or (6) are chosen as the constraints of the QP problem and in what manner α_k is calculated (Tvrzská de Gouvêa and Odloak, 1998, Ternet and Biegler, 1998, Lucia et al., 1996, Bartholomew-Biggs and Hernandez, 1995, Schmid and Biegler, 1994). Different implementations affect convergence properties and robustness of the SQP algorithm. Analytical expressions for both the Jacobians and the Hessian matrix may make the SQP algorithms achieve a quadratic convergence rate and by properly managing the nonconvex QP problems obtained with analytical expressions of H_k, robustness of the SQP algorithms may be increased (Tvrzská de Gouvêa and Odloak, 1998). So it may be desired to have available not only analytical expressions for $\nabla h^T(x_k)$ and $\nabla g^T(x_k)$, but also for H_k. As far as the constraints in (1) are solely described by algebraic equations, difficulties in establishing analytical expressions for the Jacobians and Hessian matrix are restricted to the difficulties in obtaining derivatives. When there are constraints described by ODEs, analytical expressions are not readily available and so the common practice (Cuthrell and Biegler, 1989) is to simply discretize the differential equations. By doing so, important process information may be lost and the SQP algorithm will have its convergence rate deteriorated. If discretization is not adequately performed, numerical instabilities may also occur. So it may be advantageous to have analytical expressions for the Jacobian and Hessian matrices of the constraints described by ODEs. In sections 4 and 5 we show how to derive differential equations that analytically describe the Jacobians and Hessian matrix for processes described by ODEs, which enables one to use SQP algorithms that explicitly deal with nonconvex QP subproblems.

3. THE GENERAL OPTIMIZATION PROBLEM OF PROCESSES DESCRIBED BY ODEs

Equation (7) generally describes optimization problems of processes described by ODEs. The economical objective function (f) is modeled by an algebraic equation. It may correspond to the batch time or to the heat consumption or to operational costs or any other desired economical specification. So it will typically depend on the initial conditions of the process (x_o), on the operational time or on the length of the equipment, both of these latter variables denoted by t, on state variables x^e, on the degrees of freedom of the process given by the manipulated variables (u) and on any other process variables z. In the formulation presented in equation (7), x^e corresponds to state values calculated by the SQP algorithm, i.e., these values must be equal to the solutions x of the differential equations that describe the process model given in equation (8). The general analytical solution $x(x_o,u,t)$ of (8) is not known, just its numerical one that must equal to x^e. In the formulation of the NLP problem presented in (7), the equality constraints h were written in an algebraic form and were divided into two groups. By means of the first group of equality constraints h_1, the state variables are to be evaluated. Though h_1 is written as an algebraic equation, x is actually the numerical solution of a system of ODEs. Therefore, for the evaluation of the Jacobian matrix of the constraints h_1, derivatives in terms of x_o, u and t must be

evaluated. In constraints h_2, just algebraic equations are considered. Note that these constraints are written in terms of the state variables evaluated by the SQP algorithm (x^e) and not in terms of the state variables evaluated by the solution of the ODEs (x). This is a subtle way to eliminate the dependence on the independent variable t of all variables not explicitly appearing in the differential term of the ODEs. Also note that x^e may contain values for a same physical variable evaluated at different values of t. For example, if bound constraints are to be made on the temperature along a tubular reactor, different values for the state variable temperature must be available for different positions. Inequality constraints are presented in g. The remaining constraints correspond to bound constraints on the decision variables of the NLP problem described in (1). As to the dimension of the NLP problem (1), it is assumed that: $z \in R^{n_z}$; $x^e, x_o \in R^{n_x}$, $u \in R^{n_u}$, $t \in R$, $f : R^{n_z+2n_x+n_u+1} \to R$; $h_1 : R^{2n_x+n_u+1} \to R^{n_x}$; $h_2 : R^{n_z+2n_x+n_u+1} \to R^{m-n_x}$; $g : R^{n_z+2n_x+n_u+1} \to R^{p}$.

$$\begin{aligned}
&\min_{z,x^e,x_o,u,t} f(z,x^e,x_o,u,t)\\
&s.t. \quad h_1(x^e,x_o,u,t) = -x^e + x(x_o,u,t) = 0\\
&\qquad h_2(z,x^e,x_o,u,t) = 0\\
&\qquad g(z,x^e,x_o,u,t) \le 0\\
&\qquad z_{\min} \le z \le z_{\max} \qquad (7)\\
&\qquad x^e_{\min} \le x^e \le x^e_{\max}\\
&\qquad x^o_{\min} \le x^o \le x^o_{\max}\\
&\qquad u_{\min} \le u \le u_{\max}\\
&\qquad 0 \le t \le t_{\max}\\
&\dot{x} = \bar{h}(x_o,u,t) \qquad (8)
\end{aligned}$$

Since f, h_2 and g are algebraic equations, the evaluation of their derivatives with respect to the decision variables is straightforward and well known. So, no further comments will be made. Equation (7) can be discretized by either finite differences or orthogonal collocation methods and put in an algebraic form as in (9), where now x^e are estimates for the state variables taken at nd discretization points. Thus, instead of n_x state variables, one will now have $n_x \times nd$ state variables and the dimension of the NLP problem (7) will be significantly increased, which will affect the performance of the SQP algorithms. Not only there will be the need to discretize the state variables, but if the manipulated variables are dependent on t, they will also have to be discretized, leading to another increase in the dimension of the SQP. At the same time, loss of information is inevitable. So it is expected that the number of iterations for convergence to an optimal solution will augment as well as the computational cost of each SQP iteration.

$$h_1(x^e,x_o,u,t) = 0 \qquad (9)$$

So there is the interest in not performing any discretization and derive analytical expressions for both Jacobain and Hessian matrices of the constraints in the partition h_1. In order to facilitate the derivation of analytical expressions for the Jacobians and Hessians of h_1, it is convenient that some further assumptions be taken on the way the equations are written in (7), which are:

- All manipulated variables are assumed to be independent.
- The initial conditions are independent from the manipulated variables u.

With those assumptions, simple ordinary differential equations may be derived by means of which, the Jacobians and Hessians of the constraints h_1 may be evaluated, as will be shown in the next two sections. The procedure will not increase the size of the NLP, will not result in any loss of process information and allows the manipulated variables to have a continuous or discrete dependence on t. Since analytical expressions are made available by the described procedure, SQP algorithms may be chosen in order to deal with nonconvex QPs and thus better convergence properties may be expected. There are some drawbacks, though. The analytical derivation of the expressions shown in sections 4 and 5 may be complex and the number of differential equations needed to be solved in order to evaluate both the Jacobians and Hessians may be large if either n_x or n_u are too large. Fortunately, chemical processes typically have small number of degrees of freedom and so n_u will be restricted to a small number.

4. EVALUATION OF THE JACOBIAN MATRIX

Since the constraints h_1 as defined in equation (7) do not depend on z and depend linearly on x^e, and taking (8) into account, the Jacobian matrix ∇h_1^T of the constraints h_1 will have the general structure given in equation (10).

$$\left[\begin{array}{c|c|ccc|ccc|c} & & \frac{\partial x_1}{\partial x_{o,1}} & \cdots & \frac{\partial x_1}{\partial x_{o,n_x}} & \frac{\partial x_1}{\partial u_1} & \cdots & \frac{\partial x_1}{\partial u_{n_u}} & \bar{h}_1(x_o,u,t)\\ 0 & -I & \multicolumn{3}{c|}{\vdots} & \multicolumn{3}{c|}{\ddots} & \vdots\\ & & \frac{\partial x_{n_x}}{\partial x_{o,1}} & \cdots & \frac{\partial x_{n_x}}{\partial x_{o,n_x}} & \frac{\partial x_{n_x}}{\partial u_1} & \cdots & \frac{\partial x_{n_x}}{\partial u_{n_u}} & \bar{h}_{n_x}(x_o,u,t) \end{array}\right] \qquad (10)$$

Let $y_{1,i,j} = \frac{\partial x_i}{\partial x_{o,j}}; j = 1 \ldots n_x; i = 1 \ldots n_x$ and $y_{2,i,j} = \frac{\partial x_i}{\partial u_j}$; $j = 1 \ldots n_u; i = 1 \ldots n_x$. So, for evaluating (10) at any point (x_o, u, t), one has to obtain $y_{1,i,j}$ and $y_{2,i,j}$. Since, n_x differential equations in t will be solved in (8), the idea is to augment the number of differential equations in time and by means of the added ODEs, each $y_{1,i,j}$ and $y_{2,i,j}$ are to be evaluated. This may be done by adequately applying the chain rule as is shown in equation (11) for the evaluation of $y_{1,i,j}$.

$$\frac{\partial y_{1,i,j}}{\partial t} = \frac{\partial}{\partial t}\left(\frac{\partial x_i}{\partial x_{o,j}}\right) = \frac{\partial}{\partial x_{o,j}}\left(\frac{\partial x_i}{\partial t}\right) \qquad (11)$$

Since, from (8) $\frac{dx_i}{dt} = \bar{h}_i(x_o, u, t)$, equation (11) becomes equation (12).

$$\frac{\partial y_{1,i,j}}{\partial t} = \frac{\partial \bar{h}_i}{\partial x_{o,j}} + \sum_{\substack{k=1 \\ k\neq j}}^{n_x} \frac{\partial \bar{h}_i}{\partial x_{o,k}} \frac{\partial x_{o,k}}{\partial x_{o,j}} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_k} \frac{\partial x_k}{\partial x_{o,j}} + \sum_{k=1}^{n_u} \frac{\partial \bar{h}_i}{\partial u_k} \frac{\partial u_k}{\partial x_{o,j}} + \frac{\partial \bar{h}_i}{\partial t} \frac{\partial t}{\partial x_{o,j}} \tag{12}$$

Since t is an independent variable, u and x_o were assumed independent one from another as well as the initial conditions are also taken independently, equation (12) can be reduced to (13).

$$\frac{\partial y_{1,i,j}}{\partial t} = \frac{\partial \bar{h}_i}{\partial x_{o,j}} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_k} y_{1,k,j} \tag{13}$$

A similar procedure is now applied to $y_{2,i,j}$, as is shown in equations (14) and (15).

$$\frac{\partial}{\partial t} y_{2,i,j} = \frac{\partial}{\partial t}\left(\frac{\partial x_i}{\partial u_j}\right) = \frac{\partial}{\partial u_j}\left(\frac{\partial x_i}{\partial t}\right) \tag{14}$$

$$\frac{\partial}{\partial t} y_{2,i,j} = \frac{\partial \bar{h}_i}{\partial u_j} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_k} \frac{\partial x_k}{\partial u_j} + \sum_{\substack{k=1 \\ k\neq j}}^{n_u} \frac{\partial \bar{h}_i}{\partial u_k} \frac{\partial u_k}{\partial u_j} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_{o,k}} \frac{\partial x_{o,k}}{\partial u_j} + \frac{\partial \bar{h}_i}{\partial t} \frac{\partial t}{\partial u_j} \tag{15}$$

Since u_j and u_k were also assumed to be independent one from another, equation (15) can be reduced to (16).

$$\frac{\partial y_{2,i,j}}{\partial t} = \frac{\partial \bar{h}_i}{\partial u_j} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_k} y_{2,k,j} \tag{16}$$

By adding equations (15) and (16) to the set of differential equations (8), one can simultaneously obtain the state variables x and the Jacobian matrix. It is noteworthy to note that $2n_x n_u$ differential equations are needed to evaluate the Jacobian matrix. Since the number of degrees of freedom is usually not very large, the number of differential equations will not be too large.

5. EVALUATION OF THE HESSIAN MATRIX OF THE CONSTRAINTS

Equations (17) to (22) show the general block structure of the Hessian matrix associated with the i^{th} constraint of h_1. Recall that the Hessian matrix $\nabla^2 h_{1,i}$ is symmetric and so only the non-symmetrical elements are shown in (18) to (22). The first $n_x + n_z$ rows and columns of $\nabla^2 h_{1,i}$ will be composed of null vectors since h_1 does not depend on z and depends linearly on x^e.

$$\nabla^2 h_{1,i} = \left[\begin{array}{c|ccc} 0 & \multicolumn{3}{c}{0} \\ \hline & B_1 & B_3 & B_4 \\ 0 & B_2^T & B_2 & B_5 \\ & B_4^T & B_5^T & \dfrac{\partial^2 x_i}{\partial t^2} \end{array}\right] \tag{17}$$

$$B_1 = \begin{bmatrix} \dfrac{\partial^2 x_i}{\partial x_{o,1}^2} & \cdots & \dfrac{\partial^2 x_i}{\partial x_{o,n_x} \partial x_{o,1}} \\ & \ddots & \vdots \\ & & \dfrac{\partial^2 x_i}{\partial x_{o,n_x}^2} \end{bmatrix} \tag{18}$$

$$B_2 = \begin{bmatrix} \dfrac{\partial^2 x_i}{\partial u_1^2} & \cdots & \dfrac{\partial^2 x_i}{\partial u_{n_u} \partial u_1} \\ & \ddots & \vdots \\ & & \dfrac{\partial^2 x_i}{\partial u_{n_u}^2} \end{bmatrix} \tag{19}$$

$$B_3 = \begin{bmatrix} \dfrac{\partial^2 x_i}{\partial u_1 \partial x_{o,1}} & \cdots & \dfrac{\partial^2 x_i}{\partial u_{n_u} \partial x_{o,1}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 x_i}{\partial u_1 \partial x_{o,n_x}} & \cdots & \dfrac{\partial^2 x_i}{\partial u_{n_u} \partial x_{o,n_x}} \end{bmatrix} \tag{20}$$

$$B_4^T = \begin{bmatrix} \dfrac{\partial^2 x_i}{\partial t \partial x_{o,1}} & \cdots & \dfrac{\partial^2 x_i}{\partial t \partial x_{o,n_x}} \end{bmatrix} \tag{21}$$

$$B_5^T = \begin{bmatrix} \dfrac{\partial^2 x_i}{\partial t \partial u_1} & \cdots & \dfrac{\partial^2 x_i}{\partial t \partial u_{n_u}} \end{bmatrix} \tag{22}$$

As performed in section 4, the idea is again to evaluate several elements of the Hessian matrix, by differentiating them in time and appropriately apply the chain rule. The last column of the Hessian matrix (17) may be easily obtained and we will start with its characterization, which is done in equations (23) to (25).

$$\frac{\partial^2 x_i}{\partial t \partial x_{o,j}} = \frac{\partial}{\partial t} y_{1,i,j} = \frac{\partial \bar{h}_i}{\partial x_{o,j}} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_k} y_{1,k,j} \tag{23}$$

$$\frac{\partial^2 x_i}{\partial t \partial u_j} = \frac{\partial}{\partial t} y_{2,i,j} = \frac{\partial \bar{h}_i}{\partial u_j} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_k} y_{2,k,j} \tag{24}$$

$$\frac{\partial^2 x_i}{\partial t^2} = \frac{\partial \bar{h}_i}{\partial t} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_k} \frac{\partial x_k}{\partial t} + \sum_{k=1}^{n_u} \frac{\partial \bar{h}_i}{\partial u_k} \frac{\partial u_k}{\partial t} + \sum_{i=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_{o,k}} \frac{\partial x_{o,k}}{\partial t} \tag{25}$$

Since, the initial conditions are constants and by taking (8) into account, equation (25) is reduced to (26).

$$\frac{\partial^2 x_i}{\partial t^2} = \frac{\partial \bar{h}_i}{\partial t} + \sum_{k=1}^{n_x} \frac{\partial \bar{h}_i}{\partial x_k} \bar{h}_k(x_o, u, t) + \sum_{k=1}^{n_u} \frac{\partial \bar{h}_i}{\partial u_k} \frac{\partial u_k}{\partial t} \tag{26}$$

Now, we will show how to evaluate each term in blocks B_1 to B_3. For that purpose, let for each

$i = 1,\ldots,n_x$: $y_{B_1,i,k,j} = \dfrac{\partial^2 x_i}{\partial x_{o,k}\partial x_{o,j}}$, $k = 1,\ldots,n_x$,

$j = 1,\ldots,n_x$, $y_{B_2,i,k,j} = \dfrac{\partial^2 x_i}{\partial u_k \partial x_{o,j}}$, $k = 1,\ldots,n_u$,

$j = 1,\ldots,n_x$ and $y_{B_3,i,k,j} = \dfrac{\partial^2 x_i}{\partial u_k \partial u_j}$, $k = 1,\ldots,n_u$,

$j = 1,\ldots,n_u$. Each term of matrix B_1 can be evaluated by integrating $y_{B_1,i,k,j}$ in time accordingly to equation (29). Similarly, each term of matrices B_2 and B_3 are obtained by integrating $y_{B_2,i,k,j}$ and $y_{B_3,i,k,j}$ in time accordingly to equations (32) and (35).

Equation (29) is obtained by applying the chain rule to the derivative in time of $y_{B_1,i,k,j}$ as shown in (27) and by performing further simplifications as described next. Similar procedures are applied to $y_{B_2,i,k,j}$ and $y_{B_3,i,k,j}$, which is shown in the development that follows.

$$\frac{\partial}{\partial t} y_{B_1,i,k,j} = \frac{\partial}{\partial t}\left(\frac{\partial}{\partial x_{o,k}}\left(\frac{\partial x_i}{\partial x_{o,j}}\right)\right) = \frac{\partial}{\partial x_{o,k}}\left(\frac{\partial}{\partial t}\left(\frac{\partial x_i}{\partial x_{o,j}}\right)\right) = \frac{\partial}{\partial x_{o,k}}\left(\frac{\partial}{\partial t} y_{1,i,j}\right) \tag{27}$$

Thus, differentiating equation (13) in respect to $x_{o,k}$, one obtains:

$$\frac{\partial y_{B_1,i,k,j}}{\partial t} = \frac{\partial^2 \bar{h}_i}{\partial x_{o,k}\partial x_{o,j}} + \sum_{p=1}^{n_x}\left(\frac{\partial^2 \bar{h}_i}{\partial x_{o,k}\partial x_p} y_{1,p,j} + \frac{\partial \bar{h}_i}{\partial x_p}\frac{\partial y_{1,p,j}}{\partial x_{o,k}}\right) \tag{28}$$

Hence, $y_{B_1,i,k,j}$ is obtained by solving (29).

$$\frac{\partial y_{B_1,i,k,j}}{\partial t} = \frac{\partial^2 \bar{h}_i}{\partial x_{o,k}\partial x_{o,j}} + \sum_{p=1}^{n_x}\left(\frac{\partial^2 \bar{h}_i}{\partial x_{o,k}\partial x_p} y_{1,p,j} + \frac{\partial \bar{h}_i}{\partial x_p} y_{B_1,p,k,j}\right) \tag{29}$$

In a similar way we perform with $y_{B_2,i,k,j}$, as shown in the equations that follow.

$$\frac{\partial}{\partial t} y_{B_2,i,k,j} = \frac{\partial}{\partial t}\left(\frac{\partial}{\partial u_k}\left(\frac{\partial x_i}{\partial x_{o,j}}\right)\right) = \frac{\partial}{\partial u_k}\left(\frac{\partial}{\partial t}\left(\frac{\partial x_i}{\partial x_{o,j}}\right)\right) = \frac{\partial}{\partial u_k}\left(\frac{\partial}{\partial t} y_{1,i,j}\right) \tag{30}$$

$$\frac{\partial}{\partial t} y_{B_2,i,k,j} = \frac{\partial^2 \bar{h}_i}{\partial u_k \partial x_{o,j}} + \sum_{p=1}^{n_x}\left(\frac{\partial^2 \bar{h}_i}{\partial u_k \partial x_p} y_{1,p,j} + \frac{\partial \bar{h}_i}{\partial x_p}\frac{\partial y_{1,p,j}}{\partial u_k}\right) \tag{31}$$

$$\frac{\partial}{\partial t} y_{B_2,i,k,j} = \frac{\partial^2 \bar{h}_i}{\partial u_k \partial x_{o,j}} + \sum_{p=1}^{n_x}\left(\frac{\partial^2 \bar{h}_i}{\partial u_k \partial x_p} y_{1,p,j} + \frac{\partial \bar{h}_i}{\partial x_p} y_{4,p,k,j}\right) \tag{32}$$

Similarly we perform with $y_{B_3,i,k,j}$, as presented next.

$$\frac{\partial y_{B_3,i,k,j}}{\partial t} = \frac{\partial}{\partial t}\left(\frac{\partial}{\partial u_k}\left(\frac{\partial x_i}{\partial u_j}\right)\right) = \frac{\partial}{\partial u_k}\left(\frac{\partial}{\partial t}\left(\frac{\partial x_i}{\partial u_j}\right)\right) = \frac{\partial}{\partial u_k}\left(\frac{\partial}{\partial t} y_{2,i,j}\right) \tag{33}$$

$$\frac{\partial y_{B_3,i,k,j}}{\partial t} = \frac{\partial^2 \bar{h}_i}{\partial u_k \partial u_j} + \sum_{p=1}^{n_x}\left(\frac{\partial^2 \bar{h}_i}{\partial u_k \partial x_p} y_{2,i,j} + \frac{\partial \bar{h}_i}{\partial x_p} y_{3,p,k,j}\right) \tag{34}$$

By adding equations (29), (32) and (34) to the set of differential equations formed by (8), (15) and (16), one can simultaneously obtain the state variables *x*, the Jacobian and Hessian matrices of the constraints. In order to calculate the Hessian matrices for each one of the n_x constraints, one will need to solve $\dfrac{n_x(n_x+3)+n_u(n_u+3)}{2} + n_u n_x + 1$ differential equations, which indeed may become a large number if the number of state variables is large. It is also noteworthy to say that Hessian matrices related to real processes typically have a sparse structure and so several terms will be actually zero, which will also reduce the number of differential equations that will be solved.

6. APPLICATION TO A BATCH REACTOR

Because of the lack of space, in this section just a brief outline of the application of the proposed formulation to an academic simple problem will be presented. A thorough description of the application of the procedure to a problem of industrial interest will be presented elsewhere (Souza et al., 2006). Equation (35) corresponds to the optimization problem of minimizing the batch time of a reactor (t_f), where a first order reaction is taking place and the concentration of specimen A must be kept below a limit value. The optimal temperature profile ($T(t)$) and the initial concentration ($c_{A,o}$) are to be evaluated.

$$\begin{aligned} &\min_{c_{A,o},c_A^E,T(t),t_f} t_f \\ &s.t.\ -c_A^E + c_A(t_f) = 0 \\ &\quad 0 \le c_A^E \le c_{\max} \\ &\quad 0 \le T(t) \le T^{\max} \\ &\quad t_f \ge 0 \end{aligned} \tag{35}$$

where, $c_A(t_f)$ is the concentration of A at the end of the batch and is obtained by solving (36).

$$\frac{dc_A}{dt} = -kc_A \ ; k = k_o e^{-\frac{E_R}{T(t)}} \tag{36}$$

For any instant of time *t*, the Jacobian and the Hessian matrices of the equality constraint are given in equations (37) and (38) and the values for y_1 to y_5 are obtained for any instant *t* from the integration of equations (39) to (43). As for the initial conditions,

from the definitions presented in sections 4 and 5, it follows that y_1=1 and y_2 to y_5 equal to 0. So one will have to integrate equations (36) together with equations (39) to (43) to obtain the Jacobian and Hessian matrices as in (37) and (38).

$$\nabla h^T = \begin{bmatrix} -1 & y_1 & y_2 & -kc_A \end{bmatrix} \tag{37}$$

$$\nabla^2 h = \left[\begin{array}{c|ccc} 0 & \multicolumn{3}{c}{0} \\ \hline & y_5 & y_4 & -ky_1 \\ 0 & y_4 & y_3 & -kc_A\frac{E_R}{T^2} - ky_2 \\ & -ky_1 & -kc_A\frac{E_R}{T^2} - ky_2 & k^2 c_A \end{array}\right] \tag{38}$$

$$\frac{dy_1}{dt} = -ky_1 \tag{39}$$

$$\frac{dy_2}{dt} = -kc_A\frac{E_R}{T^2} - ky_2 \tag{40}$$

$$\frac{dy_3}{dt} = -kc_A\left[\left(\frac{E_R}{T^2}\right)^2 - 2\frac{E_R}{T^3}\right] - 2k\frac{E_R}{T^2}y_2 - ky_3 \tag{41}$$

$$\frac{dy_4}{dt} = -y_1 k\frac{E_R}{T^2} - ky_4 \tag{42}$$

$$\frac{dy_5}{dt} = -ky_5 \tag{43}$$

The above model describes an isothermal operation of the reactor, for which equation (36) has a trivial solution given by $c_A = c_{A,o}e^{-kt}$ and hence equations (37) and (38) must equal to equations (44) and (45). Table 1 gives the comparison of the Jacobian and Hessian matrices evaluated by equations (37) and (38) and by equations (44) and (45) with $c_{A,o}$=1000 mol/m^3, T=523 K, k_o=3370 s^{-1}, E_R=7000 K^{-1} and t_f=360 s.

$$\nabla h^T = \begin{bmatrix} -1 & e^{-kt} & -c_A tk\frac{E}{T^2} & -kc_A \end{bmatrix} \tag{44}$$

$$\nabla^2 h = \left[\begin{array}{c|ccc} 0 & \multicolumn{3}{c}{0} \\ \hline & 0 & -kt\frac{E}{T^2}e^{-kt} & -ke^{-kt} \\ 0 & =el. & c_A kt\frac{E}{T^2}\left(\frac{2}{T} - \frac{E}{T^2} + kt\frac{E}{T^2}\right) & =el. \\ & (2,3) & & (4,3) \\ & =el. & c_A k\frac{E}{T^2}(kt-1) & k^2 c_A \\ & (2,4) & & \end{array}\right] \tag{45}$$

Table 1: Comparison of the elements of the Jacobian and Hessian matrices

element	matrix	evaluated by (37) or (38)	evaluated by (44) or (45)	error (%)
(1,2)	∇h^T	0.15470	0.15474	0.03
(1,3)	∇h^T	-7.390	-7.389	0.01
(2,2)	$\nabla^2 h$	0	0	0
(2,3)	$\nabla^2 h$	-0.0073896	-0.0073895	0.001
(2,4)	$\nabla^2 h$	-8.0209e-4	-8.0208e-4	0.001
(3,3)	$\nabla^2 h$	0.19208	0.19203	0.03
(3,4)	$\nabla^2 h$	0.01779	0.01778	0.06

7. CONCLUSIONS

A general procedure was shown on how to develop analytical expressions for the evaluation of Jacobian and Hessian matrices for constraints described by ODEs. The developed expressions are also composed of ODEs and are obtained simultaneously with the integration of the process model.

ACKNOWLEDGMENT

Financial support from MACKPESQUISA under grant 1176, from FAPESP under grant 97/07963-4 and from CNPq under grant 303304/2004-9 is gratefully acknowledged.

REFERENCES

Agrawal, S.S., Beach, W., Rendon, G. T., Olvera, R. O. (1996) Implementation of advanced on-line control, optimization and planning systems in Mexican refineries. In: NPRA Computer Conference. Atlanta, USA, November, 11-13

Bartholomew_Biggs, M.C. and Hernandez, F.G. (1995) Using the KKT matrix in an augmented Lagrangian SQP method for sparse constrained optimization. *J.O.T.A.*, **85**(1), 201-220

Cuthrell, J.E. and Biegler, L.T. (1989) Simultaneous optimization and solution methods for batch reactor control profiles. *Computers Chem. Engineering*, **13**(1-2), 49-62

Ellis, R.C., Xuan, L.; Riggs, J.B. (1998) Modeling and optimization of a model IV fluid catalytic cracking unit. *AIChE Journal*, **44**, 2068-2079

Lucia, A., Xu, J., Layn, K.M. (1996) Nonconvex process optimization. *Computers Chem. Engng.*, **20**(2), 1375-1398.

Jakhete, R.; Rager, W.; Hoffman, D.W. (1999) Online implementation of composite LP optimizers FCCU/GPU complex. *Hydrocarbon processing*. **78**(2), 69

Schmid, C. and Biegler, L.T. (1994) Quadratic programming methods for reduced Hessian SQP. *Computers Chem. Engng.*, **18**(9), 817-832

Souza, G. F, Gonçalves, A. P. C., Odloak, D., Tvrzská de Gouvêa, M. Optimization of the operation of a reactor for the production of ethylene oxide/ working paper/ 2006

Ternet, D.J. and Biegler, L.T. (1998) Recent improvements to a multiplier-free reduced Hessian successive quadratic programming algorithm. *Computers Chem. Engng.*, **22**(7-8), 963-978

Tvrzská de Gouvêa, M. and Odloak, D. (1998) A new treatment of inconsistent quadratic programs in a SQP-based algorithm. *Computers Chem. Engng.*, **22**(11), 1623-1651

Zanin, A.C., Tvrzská de Gouvêa, M., Odloak, D. (2002) Integrating real-time optimization into the model predictive controller of the FCC system. *Control Engineering Practice*, **10**, 819-831

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

OPTIMAL MULTIPERIOD DESIGN AND OPERATION OF MULTIPRODUCT BATCH PLANTS

Marta S. Moreno[(1)], Jorge M. Montagna[(1) (2)] and Oscar A. Iribarren[(1)]

[(1)] INGAR – Instituto de Desarrollo y Diseño
[(2)] Universidad Tecnológica Nacional – Facultad Regional Santa Fe
Avellaneda 3657, S3002 GJC Santa Fe, Argentina

Abstract: New alternatives for the multiperiod design and operation planning of multiproduct batch plants are presented. Unlike previous works, this approach configurates the plant in every period considering the assignment of parallel units of different sizes operating either in or out-of-phase. The objective function maximizes the net profit considering incomes, investment costs, and both product and raw material inventory costs. The model takes into account batch units available in discrete sizes, and both raw material and product inventories accounting for seasonal variations for supplies and demands. Nonlinearities have been eliminated by an efficient scheme in order to get a MILP model to guarantee global optimality.

Keywords: Optimization, Batch, Design and Planning, Parallel Equipment, Integer Programming

1. INTRODUCTION

Continuous growth in complexity, competitivity, and uncertainty of the marketing environment of high added value chemicals and foodstuffs with a short life cycle has renewed the interest in batch operations and the development of optimization models. The main attraction of batch plants in this context is their inherent flexibility in utilizing the various resources available for the manufacture of relatively small amounts of several different products within the same facilities. Several excellent papers on designing and production planning of multiproduct batch plants have already been published (Grossmann, and Sargent, 1979; Ravermark, 1995). The goal is to determine the size and the number of batch units so they can meet production requirements in the provided time horizon.

Since such products usually have demand patterns that vary over time due to market or seasonal changes, multiperiod optimization models have been the object of a great deal of research effort in this area (Birewar and Grossmann, 1990; Voudouris and Grossmann, 1993; Varvarezos et al. 1992; Van den Heever and Grossmann, 1999). Multiperiod models for the design and operation planning in chemical plants involve designing plants that operate under variations in the model parameters along the time horizon. In general, this kind of problems is represented by mixed integer nonlinear programming (MINLP) models.

Multiproduct batch plants manufacture a set of products using the same equipment operating in the same sequence. Since products differ from one another, each unit is shared by all products but they do not use their total capacity for all of them. The unit with the minimum capacity limits the batch size while the limiting cycle time is fixed by stage with the longest processing time.

In order to reduce the investment cost, several alternatives are possible (Ravermark, 1995). The first one is the introduction of parallel units out of phase to reduce the cycle time if the unit has the longest operating time. Another option is to add a parallel unit in phase to increase the operating capacity of the stage.

The above referenced approaches for solving the multiperiod MINLP problem were restricted to equal periods in the design horizon (Voudouris, and Grossmann, 1993) or they did not include the design into the formulation (Birewar, and Grossmann,

1990). Also, no previous design work has considered the addition of parallel units in and out of phase, which can take different sizes, and has not addressed the issue of raw materials inventory.

This work is an attempt to expand the scope of multiperiod models for the design and production planning problems in multiproduct batch plants. In this paper, an optimization mixed integer linear programming (MILP) model which can handle seasonal changes of prices, costs and demands, discrete sizes of units, and inventories of both final products and raw materials will be proposed. Moreover, this multiperiod approach considers different period lengths and the possible alternatives to add parallel units that can have different sizes in every stage by using the concept of group introduced by Yoo, *et. al.* (1999) and extended by Montagna (2003). Also, this model takes into account flexible plant configurations where available units can be arranged in different structures for each product. In contrast to previous models that consider continuous sizes for the units, this model determines the optimal design selecting from available discrete sizes which corresponds to the real procurement of equipments. The major significance of the MILP model presented in this work is that it corresponds to a realistic design case that can be solved to global optimality with reasonable computation effort.

The remainder of this paper is structured as follows. The next section presents the problem description. In the subsequent section, a multiperiod model which incorporates all the elements of the design and planning problem is formulated. The non linear terms in the formulation are transformed into linear ones in order to obtain a MILP model by using a reformulation strategy. The application of this formulation to a specific example is illustrated for a plant that produces Oleoresins. Finally, the conclusions are presented in the last section.

2. PROBLEM STATEMENT

In a multiperiod scenario a multiproduct batch plant processes I products (i = 1, 2, …, I). Every product follows the same production sequence through all the J batch processing stages (j = 1, 2, …, J) of the plant. Each stage j may consist of one or more units k, which can have different sizes, operating either in phase to increase capacity, or out of phase to decrease the cycle time. The size of unit k at stage j is V_{jk} (k = 1, 2, …, K_j), where K_j is the maximum number of units that can be added at stage j. Also, the volume of each unit k at stage j is available in discrete sizes.

The configuration of the batch units must be determined at each stage for every product. K_j units of stage j can be grouped in different ways for each product i (Yoo, et al., 1999). It is possible to have groups in which all units operate in parallel and in phase. The different groups at the stage operate in parallel and out of phase.

Since this is a multiperiod problem, the time horizon H is discretized into T (t = 1, 2, …, T) specified time periods H_t not necessarily of the same length. Bounds on products demands, costs and availability of raw materials vary from period to period. It will be assumed that the plant operates in single product campaign (SPC) mode under zero wait (ZW) policy in each time period.

The objective is to maximize the benefit of the plant considering incomes from inventory and product sales, and capital costs. In the design of this plant, the problem lies in deciding the convenience of adding units at any time period, and selects the size of batch units V_{jk} among available discrete sizes v_{js}. At each time period t the model determines the number of groups and which of the existing units in a period are assigned to each of them. Moreover, it decides the amount of product to be produced q_{it}, the number of batches n_{it} and the total time T_{it} to produce product i.

Inventory considerations are an important aspect of plant operation. Actually, in practice, a plant may be faced with product demands and raw materials supplies that vary seasonally. So, products and raw materials can be maintained in stock until needed.

At the end of every period t, the levels of both final product IP_{it} and raw material inventories IM_{it} are obtained. Moreover, the total sales QS_{it}, the amount of purchased raw material C_{it}, and the raw material to be used for the production RM_{it} of product i in each period t are determined with this formulation. Semicontinuous and intermediate storage are not considered.

3. MATHEMATICAL FORMULATION

3.1 Assignment Constraints

Several variables are introduced to determine the plant structure. Since the units can be added at any time period, a binary variable w_{jkt} is used. The value of this variable is 1 if unit k is included in the plant structure at stage j in the period t; otherwise the value is zero. Each unit k at stage j can be added only in one period:

$$\sum_{t}^{T} w_{jkt} \le 1 \qquad \forall\, i, j \tag{1}$$

The units are included in a sequential manner in order to avoid alternative optimal solutions with the same value for the objective function:

$$\sum_{\tau=1}^{t} w_{j,k,t} \ge \sum_{\tau=1}^{t} w_{j,k+1,t} \qquad \forall\, i, k = 1,..., K_j - 1, t \tag{2}$$

Since the units can be grouped in different ways at each stage in every period, the binary variable y_{ijgt} is introduced. The value of this variable is equal to 1 if group g is generated for product i at stage j in time period t; otherwise, the value is zero. Group g is generated if at least one unit is assigned to it. Binary variable y_{ijkgt} is 1 if unit k of stage j is assigned to

group g for product i at period t; otherwise the variable is equal to zero (Montagna, 2003).

$$\sum_{g=1}^{G_j} y_{ijkgt} \le 1 \qquad \forall\, i, j, k, t \tag{3}$$

G_j is the maximum number of groups allowed at stage j. Group g exists at stage j in period t only if at least one unit is assigned to the group in that period:

$$y_{ijgt} \le \sum_{k=1}^{K_j} y_{ijkgt} \qquad \forall\, i, j, g, t \tag{4}$$

If unit k is assigned to the group in period t, the group must exist:

$$y_{ijkgt} \le y_{ijgt} \qquad \forall\, i, j, k, g, t \tag{5}$$

If the unit is assigned to group g at stage j for product i in period t, the unit must exist in that period:

$$y_{ijkgt} \le \sum_{\tau=1}^{t} w_{jkt} \qquad \forall\, i, j, k, g, t \tag{6}$$

If the unit exists at stage j in the period t, it must be included in a group:

$$\sum_{\tau=1}^{t} w_{jkt} = \sum_{g=1}^{G_j} y_{ijkgt} \qquad \forall\, i, j, k, t \tag{7}$$

Redundant assignation to a group with the same value for the objective function is avoided by the following constraint (Yoo, et al., 1999):

$$\sum_{k=1}^{K_j} 2^{K_j-k}\, y_{ijkgt} \ge \sum_{k=1}^{K_j} 2^{K_j-k}\, y_{ijk,g+1,t} \qquad \forall\, i, j, g = 1,\ldots,G_j - 1, t \tag{8}$$

This constraint order the different groups through a weight 2^{K_j-k} assigned to each unit k. The order of the group is obtained by adding the weights of all units in the group.

3.2 Design and Planning and Constraints

Unit k at each stage j can be configured in a different way for every product i manufactured in the plant. B_{it} is the batch size of product i in time period t. When B_{it} gets into a group of units, that is, units operating in phase, B_{it} is divided between the units that belong to the group. Thus, the sum of the units sizes included in the group g in every period t must be large enough to produce a batch of product i.

$$\sum_{k\in g} V_{jkt} \ge S_{ijt} \cdot B_{it} \qquad \forall\, i, j, g, t \tag{9}$$

where S_{ijt} is the size factor at stage j for product i, that can vary in each period taking into account seasonal effects.

In this work, the unit sizes V_{jk} are considered available in discrete sizes v_{js} which correspond to the real commercial procurement of equipments. To rigorously tackle this situation, the binary variable z_{jks} is introduced. It is one if unit k at stage j has size s; otherwise, it is zero. The variable V_{jk} is restricted to take values from the set $SV_j = \{v_{j1}, v_{j2}, \ldots, v_{jn_j}\}$, where n_j is the number of discrete sizes available for each stage. Using the previous definition, V_{jk} can be expressed in terms of discrete variables as:

$$V_{jk} = \sum_{s} v_{js} \cdot z_{jks} \qquad \forall\, j, k \tag{10}$$

If the unit k at stage j is added in some period t, it must take a size s for the volume from the available sizes at that stage:

$$\sum_{s} z_{jks} = \sum_{t=1}^{T} w_{jkt} \qquad \forall\, j, k \tag{11}$$

Only one of the available sizes at stage j must be selected if unit k at stage j exists:

$$\sum_{s} z_{jks} \le 1 \qquad \forall\, j, k \tag{12}$$

The amount of product i produced in time period t is

$$q_{it} = B_{it} \cdot n_{it} \qquad \forall\, i, t \tag{13}$$

where n_{it} is the batch number of product i in period t.

By combining Eq. (9) and Eq. (13) the constraints take the following form:

$$q_{it} \le \sum_{k\in g} V_{jkt} \cdot \frac{n_{it}}{S_{ijt}} + M_{ij} \cdot (1 - y_{ijgt}) \qquad \forall\, i, j, g, t \tag{14}$$

Eq. (14) is a Big-M constraint that guarantees that batches can be processed if group g exists; otherwise the constraint is redundant because of the large value of M_{ij}. The value of M_{ij} can be calculated by:

$$M_{ij} = K_j \cdot \max(s, v_{js}) \cdot \max(t, n_{it}^U / S_{ijt}) \qquad \forall\, i, j \tag{15}$$

In order to obtain the volumes that belong to each group, it is necessary to multiply the volume V_{jk} by the binary variable y_{ijkgt} which produces the equation:

$$q_{it} \le \sum_{k=1}^{K_j} \left(V_{jk} \cdot y_{ijkgt}\right) \cdot \frac{n_{it}}{S_{ijt}} + M_{ij} \cdot (1 - y_{ijgt}) \qquad \forall\, i, j, g, t \tag{16}$$

By substituting Eq. (12) into Eq. (16) new constraints can be formulated that restrict the volumes to discrete sizes:

$$q_{it} \le \sum_{k=1}^{K_j} \sum_{s} \left(\frac{v_{js}}{S_{ijt}} \cdot z_{jks} \cdot y_{ijkgt} \cdot n_{it} \right) + M_{ij} \cdot (1 - y_{ijgt}) \qquad \forall\, i, j, g, t \tag{17}$$

Constraint (17) is nonlinear because of the product of binary variables. In order to reformulate these constraints as linear ones, the cross product $n_{it}\, z_{jks}\, y_{ijkgt}$ can be eliminated by introducing the continuous

variable h_{ijkgst} that is equal to n_{it} if z_{jks} and y_{ijkgt} are one; otherwise the variable is equal to zero.

$$q_{it} \le \sum_{k=1}^{K_j} \sum_{s} \left(\frac{v_{js}}{S_{ijt}} \right) \cdot h_{ijkgst} + M_{ij} \cdot (1 - y_{ijgt}) \qquad \forall\, i, j, g, t \tag{18}$$

$$\sum_{s} h_{ijkgst} \le n_{it}^{U} \cdot y_{ijkgt} + M_{ij} \cdot (1 - y_{ijgt}) \qquad \forall\, i, j, k, g, t \tag{19}$$

$$h_{ijkgst} \le n_{it}^{U} \cdot z_{jks} + M_{ij} \cdot (1 - y_{ijgt}) \qquad \forall\, i, j, k, g, t \tag{20}$$

$$\sum_{g} \sum_{s} h_{ijkgst} = n_{it} \qquad \forall\, i, j, k, t \tag{21}$$

where n_{it}^{U} is the upper bound for n_{it}. The summation over the groups in Eq. (21) is performed in order to reduce the number of generated constraints because only one of the values is equal to n_{it}.

The inventory of final product i at the end of a period t, IP_{it}, depends on the inventory that is left from the previous interval, $IP_{i,t-1}$, the quantity produced and the total sales, QS_{it}.

$$IP_{it} = IP_{it-1} + q_{it} - QS_{it} \qquad \forall\, i, t \tag{22}$$

In the same way, the inventory of raw material is:

$$IM_{it} = DE_{i,t-1} \cdot IM_{it-1} + C_{it} - RM_{it} \qquad \forall\, i, t \tag{23}$$

The amount of raw material in the inventory IM_{i0} for each product at the beginning of the time horizon is assumed to be given. Idem for the initial product inventory, IP_{i0}.

The amount assigned to sales must be less than the amount of product in inventory plus the quantity produced during a period:

$$QS_{it} \le IP_{it-1} + q_{it} \qquad \forall\, i, t \tag{24}$$

The raw material necessary for the production of the product i is obtained from a mass balance:

$$RM_{it} = F_{it} \cdot q_{it} \qquad \forall\, i, t \tag{25}$$

where F_{it} is a parameter that accounts for the process conversion, e.g. ratio of solvent to solids, time of contact etc. In this presentation only one raw material is considered. However this condition can be easily extended in order to accounting for several raw materials.

The limiting cycle time is the maximum time between two successive batches of product i. It can be calculated by the division between processing time t_{ijt} and the number of groups out of phase for product i at stage j in every period:

$$TL_{it} \ge \frac{t_{ijt}}{\sum_{g=1}^{G_j} y_{ijgt}} \qquad \forall\, i, j, t \tag{26}$$

The total time for producing product i in time period t is defined as:

$$T_{it} = TL_{it} \cdot n_{it} \qquad \forall\, i, t \tag{27}$$

By multiplying Eq.(28) by the number of batches, the expression takes the form:

$$T_{it} \ge \frac{t_{ijt} \cdot n_{it}}{\sum_{g=1}^{G_j} y_{ijgt}} \qquad \forall\, i, j, t \tag{28}$$

Equation (28), however, is nonlinear. In order to obtain a linear expression, the following constraints are introduced:

$$\sum_{g=1}^{G_j} y_{ijgt} = \sum_{g=1}^{G_j} g \cdot u_{ijgt} \qquad \forall\, i, j, t \tag{29}$$

$$\sum_{g=1}^{G_j} u_{ijgt} = 1 \qquad \forall\, i, j, t \tag{30}$$

where the variable binary u_{ijgt} is 1 if at stage j there are g groups operating out of phase. Substituting y_{ijgt} for u_{ijgt} in Eq. (28), the expression gets the following form

$$T_{it} \ge \sum_{g=1}^{G_j} \left(\frac{t_{ijt} \cdot n_{it}}{g} \right) \cdot u_{ijgt} \qquad \forall\, i, j, t \tag{31}$$

This constraint is also nonlinear. To eliminate bilinear terms $n_{it}\, u_{ijgt}$, a new nonnegative continuous variable e_{ijgt} is defined to represent this cross product (Voudouris and Grossmann, 1992). Then the following linear constraints are obtained:

$$T_{it} \ge \sum_{g=1}^{G_j} \left(\frac{t_{ijt}}{g} \right) \cdot e_{ijgt} \qquad \forall\, i, j, t \tag{32}$$

$$\sum_{g=1}^{G_j} e_{ijgt} = n_{it} \qquad \forall\, i, j, t \tag{33}$$

$$e_{ijgt} \le n_{it}^{U} \cdot u_{ijgt} \qquad \forall\, i, j, g, t \tag{34}$$

where n_{it}^{U} is the upper bound for n_{it}.

Considering the case of SPC-ZW policy in the period t, all productions must be completed within the corresponding production horizon H_t:

$$\sum_{i}^{I} n_{it} \cdot TL_{it} \le H_t \qquad \forall\, t \tag{35}$$

Taking into account Eq. (27) the following expression is obtained:

$$\sum_{i}^{I} T_{it} \le H_t \qquad \forall\, t \tag{36}$$

3.3 Objective Function

The strategic objective in this formulation is to maximize the operating profit of the plant,

$$\Phi = \sum_t \sum_i p_{it} \cdot QS_{it} - \sum_t \sum_i \kappa_{it} \cdot C_{it} - CEQ - \sum_t \sum_i \varepsilon_{it} \cdot \left(\frac{IM_{it-1} + IM_{it}}{2} \right) \cdot H_t - \sum_t \sum_i \sigma_{it} \cdot \left(\frac{IP_{it-1} + IP_{it}}{2} \right) \cdot H_t \quad (37)$$

The first term of the objective function is the income corresponding to the product sales where the parameter p_{it} is the price of product i in each period. The second term is the cost of purchases with κ_{it} the price of raw material. The last two terms correspond to raw material and final product inventory costs, where ε_{it} and σ_{it} are inventory cost coefficients (Birewar, and Grossmann, 1990). Finally, the third term is the investment cost of the batch units and is obtained through the following equations:

$$CEQ = \sum_t^T \sum_j^J \sum_k^{K_j} \alpha_{jt} \cdot V_{jk}^{\beta_{jt}} \cdot w_{jkt} \quad (38)$$

where α_{jt} and β_{jt} are specific cost coefficients for each stage j in every period t. Eq.(10) is introduced into this expression to get:

$$CEQ = \sum_t^T \sum_j^J \sum_k^{K_j} \sum_s \alpha_{jt} \cdot v_{js}^{\beta_{jt}} \cdot z_{jks} \cdot w_{jkt} \quad (39)$$

$$CEQ = \sum_t^T \sum_j^J \sum_k^{K_j} \sum_s c_{jst} \cdot r_{jkst} \quad (40)$$

where the terms $c_{jst} = \alpha_{jt} \cdot v_{js}^{\beta_{jt}}$ represent the cost of standard batch vessels, and new variables r_{jkst} are introduced to eliminate the product of binary variables $z_{jks} w_{jkt}$ through the constraints:

$$r_{jkst} \geq z_{jks} + w_{jkt} - 1 \qquad \forall j,k,s,t \quad (41)$$

$$0 \leq r_{jkst} \leq 1 \quad (42)$$

4. MODEL RESOLUTION

To sum up, the multiperiod model of a multiproduct batch plant is defined by maximizing the objective function represented by Eq. (37) using Eq. (40) as the term of investment cost and subject to constraints Eqs. (1) – (8), (11), (12), (15), (18) – (25), (29), (30), (32) – (34), (36), (41), (42) plus the upper bounds that may apply. Bilinear terms have been eliminated through an efficient method in order to generate a MILP model which can be solved to global optimality.

5. EXAMPLES

5.1 Example 1

To illustrate the use of the MILP formulation presented in the previous section, let us consider optimizing the production of five oleoresins, sweet bay (A), oregano (B), pepper (C), rosemary (D), and thyme (E) oleoresins, manufactured in a multiproduct batch plant.

This plant consists of the following stages: 1) extraction in a four-stage countercurrent arrangement (2) expression, (3) evaporation, and (4) blending. All of these stages can be duplicated up to four units, so, the maximum number of groups that can exist at a stage is four, too.

In order to obtain parameter F_{it} necessary for Eq. (25), the following equations are used:

$$x_i^{n+1} \cdot [1 + E_i \cdot (1 - \eta_i)] = x_i^n \cdot (1 + E_i - \eta_i) + \eta_i \cdot x_i^1 \quad (43)$$

$$F_{it} = \frac{1}{\left(x_i^{n+1} - x_i^1 \right)} \quad (44)$$

where E_i is the extraction factor, η_i is the extent of the extraction, and x_i is the product concentration in the vegetable solid feed. Index n is the number of each stage for the n-staged countercurrent extraction.

A global horizon time of one year (6000 h working) has been considered, and it was divided into a set of equal time periods, namely from 1 to 6. Demands, costs, and prices differ from period to period. Table 1 contains some data for this example.

Table 1 Data for Example 1

	Size Factors (L/kg)				Processing Time (h)			
i	j_1	j_2	j_3	j_4	j_1	j_2	j_3	j_4
A	20	15	12	1.5	1.5	1	2.5	0.5
B	80	55	49	1.5	1.5	1	2	0.5
C	20	15	12	1.5	2.5	2	3	2
D	40	25	24	1.5	1.5	1	1.5	1
E	30	20	17	1.5	1.5	1	3	1
Sizes (liter, L) SV_j = {250, 500, 750, 1000, 1500}								

Table 2 shows the optimal unit assignment for this plant. To show the optimal solution, product B, the least convenient to produce, was chosen. The first diagram of Figure 1 shows that raw material for B is purchased during the two initial periods. In the second diagram, it can be noted that B is produced only during the first two periods, because the costs are lower mainly due the lower raw material price and the amount produced in these periods is stored as inventory for satisfying minimum demands in the subsequent intervals.

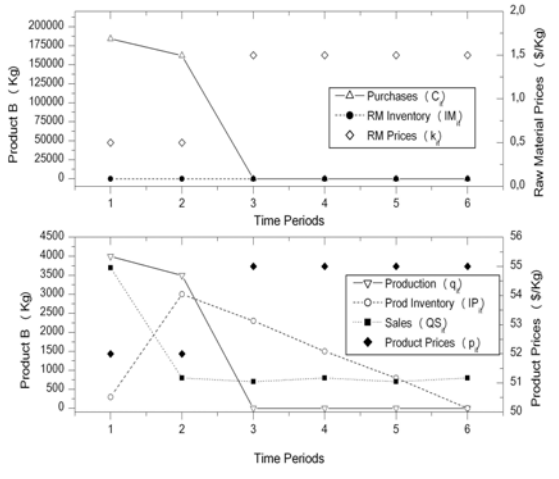


Figure 1. Results for Product B of Example 1

Table 2 Optimal unit assignment for Example 1

	Stage (L)			
Unit	j_1	j_2	j_3	j_4
k_1	1500	1000	750	250
k_2	-	-	750	-

5.2 Example 2

Consider now the production of 3 oleoresins (A, B, C) in 3 time periods where demands take higher values in subsequent periods to consider a possible market expansion. Table 3 shows the optimal unit assignment. Table 4 shows the different configuration for the units for each product in every period, respectively. In this table, units between parentheses are included in the same group. Also, it can be seen that all units are introduced in period 1 except units 2 and 3 that are added in stage 1 in period 2.

Table 3 Optimal unit assignment for Example 2

	Stage (L)			
Unit	j_1	j_2	j_3	j_4
k_1	5000	1500	5000	1500
k_2	5000	5000	5000	-
k_3	5000	5000	-	-

These problems were solved by using CPLEX through the modeling system GAMS on a Pentium IV Processor (3GHz). The total profit for example 1 and example 2 were \$1982822.84 and \$7766239.71 respectively. The information about the resolution of these examples is as follows. Example 1 has 12091 constraints, 8293 single variables, 1932 binary variables and the optimal solution was obtained after a CPU time of 111.70s. Example 2 has 3736 constraints, 2617 single variables, 636 binary variables and the optimal solution was obtained after a CPU time of 390.95s.
Due to the large amount of data for these examples, they are not presented in this paper. Readers interested in the data can contact the authors.

6. CONCLUSION

A new model for the optimal design and operation planning of multiproduct batch plants has been formulated as an MILP problem, which guarantees the global optimum solution.

This multiperiod MILP model involves discrete decisions for the structure selection and continuous decisions for the operation plan of the plant at each time period. Furthermore, this model allows considering all possible alternatives for the addition of equipments in parallel, which are available in discrete sizes.

Seasonal variations of products demands and raw materials availability are readily accounted for, and both raw materials and final product inventories are included in the formulation. The proposed model was applied to a plant that produces oleoresins.

REFERENCES

Birewar D.B., Grossmann I. E. (1990). Simultaneous Production Planning and Scheduling in Multiproduct Batch Plants. *Industrial Engineering Chemical Research.* 29, 570 – 580.

Grossmann I. E., Sargent R. W. H., (1979). Optimum Design of Multipurpose Chemical Plants. *Industrial Engineering and Chemical Process Design and Development.* 18, 343 – 348.

Montagna, J. M. (2003). The optimal retrofit multiproduct batch plants. *Computers and Chemical Engineering*. 27, 1277 – 1290.

Ravermark, D. (1995). Optimization models for design and operation of chemical batch processes. Ph.D. thesis, Swiss Federal Institute of Technology, Zurich.

Van den Heever, S. A., Grossmann, I. E. (1999). Disjuntive multiperiod optimization methods for design and planning of chemical process systems. *Computers and Chemical Engineering.*23, 1075 – 1095.

Vaselenak, J. A., Grossmann, I. E. and Westerberg, A. W. (1987). Optimal retrofit design in multiproduct batch plants. *Industrial Engineering Chemical Research.* 26, 718 – 726.

Voudouris V. T.; Grossmann, I. E. (1992). Mixed-Integer Linear Programming Reformulations for Batch Process Design with Discrete Equipment Sizes. *Industrial Engineering Chemical Research.* 31, 1315 – 1325.

Voudouris V. T.; Grossmann, I. E. (1993). Optimal Synthesis of Multiproduct Batch Plants with Cyclic Scheduling and Inventory Considerations. *Industrial Engineering Chemical Research.* 32, 1962 – 1980.

Yoo, D. J., Lee, H. Ryu, J., and Lee, I. (1999). Generalized retrofit design of multiproduct batch plants. *Computers and Chemical Engineering*. 23, 683 – 695.

Table 4 Arrangement of units for each product in every period

	Period 1			Period 2			Period 3		
	A	B	C	A	B	C	A	B	C
q_{it}	2500	500	2000	17000	23000	18000	175000	0	104166
j_1	(k_1)	(k_1)	(k_1)	(k_1 ,k_3)-(k_2)	(k_1,k_2,k_3)	(k_1,k_2,k_3)	(k_1,k_2,k_3)		(k_1,k_2,k_3)
j_2	(k_1,k_2,k_3)	(k_1,k_2,k_3)	(k_1 ,k_3)-(k_2)	(k_1,k_2,k_3)	(k_1,k_2,k_3)	(k_1,k_2,k_3)	(k_1,k_2,k_3)		(k_1,k_2,k_3)
j_3	(k_1,k_2)	(k_1)-(k_2)	(k_1,k_2)	(k_1,k_2)	(k_1,k_2)	(k_1,k_2)	(k_1,k_2)		(k_1,k_2)
j_4	(k_1)	(k_1)	(k_1)	(k_1)	(k_1)	(k_1)	(k_1)		(k_1)

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

IMPROVED TIGHTENED MILP FORMULATIONS FOR SINGLE-STAGE BATCH SCHEDULING PROBLEMS

Pablo A. Marchetti and Jaime Cerdá*

INTEC (UNL - CONICET)
Güemes 3450 - 3000 Santa Fe - ARGENTINA
** E-mail: jcerda@intec.unl.edu.ar*

Abstract: This work presents a set of improved MILP mathematical formulations for the scheduling of single-stage batch plants with parallel production lines. Minimization of the average weighted earliness and the makespan, i.e. the time needed to complete all processing tasks, are considered as alternative problem goals. For each objective function an enhanced model that incorporates specific tightening constraints is presented. These constraints improve each model's efficiency by increasing the corresponding objective function lower bound, thus accelerating the branch and bound node pruning process. Several problem instances with different number of batches demonstrate that the proposed approach reduces the computational effort by orders of magnitude. Sequence dependent setup times can also be effectively accommodated.

Keywords: Scheduling algorithms, Optimization, Manufacturing processes, Computer-aided engineering, Integer programming

1. INTRODUCTION

The problem of finding an efficient short-term schedule for a multiproduct batch plant is of major interest for most manufacturing companies. Several solution methodologies have been proposed for different kinds of scheduling problems. An extensive review of the state of the art can be found in Floudas and Lin (2004). Overall, exact solution methods have received most of the researchers attention. Among the different MILP models proposed in the literature, continuous-time models have shown a better computational performance for the scheduling of batch processes with sequence-dependent changeovers. In particular, the continuous time batch scheduling problem model of Méndez *et. al.* (2001) shows a good computational behaviour compared with other continuous approaches, like the time-slot formulation of Pinto and Grossmann (1995) or the unit-specific time event model of Ierapetritou and Floudas (1998). However, its performance is somewhat deteriorated by the big-M batch sequencing constraints, especially when the makespan objective function is considered. Big-M constraints produce an increase in the integrality gap (the difference between optimal values for the relaxed and MILP problems), which in turn makes the optimal solution harder to find by the MILP solver through a branch-and-bound algorithm.

This work presents a pair of improved formulations for single-stage batch plant scheduling problems that incorporate tightening constraints in order to reduce the integrality gap and enhance the branch-and-bound node pruning process. Based on the MILP approach by Méndez *et. al.* (2001), the proposed models account for release times, ready times, due dates, and sequence dependent setup times between batches. Minimization of the average weighted earliness and the makespan, i.e. the time needed to execute all processing tasks, are both considered as problem objectives. Specific tightening constraints are presented in order to improve the lower bound on each objective function, leading to different formulations for each problem.

This work is organized as follows. In section 2 the scheduling problem under consideration is properly defined. In section 3 the different mathematical models are presented. At first, the single-stage version of the MILP model by Méndez *et. al.* (2001) is reviewed. Common constraints related to assignment and sequencing decisions are included. Afterwards, specific constraints and binary variables that improve the model's efficiency for each objective function are introduced. These additional constraints are incorporated or just replace previous ones, and their main purpose is to increase the objective function lower bound. Section 4 shows the effectiveness of the proposed approach by tackling several problem instances with an increasing number of batches and different sequencing conditions for both objective functions. Conclusions are discussed in section 5.

2. PROBLEM STATEMENT

The problem of short-term scheduling of single-stage multiproduct batch plants with parallel production lines can be stated as follows. Given: (a) a single-stage multiproduct batch plant with multiple parallel units $j \in J$, (b) a set of single-batch orders $i \in I$ to be completed within the scheduling horizon, (c) the order release times rt_i and due dates dd_i for each $i \in I$, (d) the set of available processing units $J_i \subset J$ for each batch i, and the constant processing times pt_{ij} required at each unit $j \in J_i$, (e) the sequence-dependent setup times $\tau_{ii'j}$, (f) the equipment unit ready times ru_j, and (g) the specified time horizon H. The problem goal is to find a production schedule that completes all batch orders within their time limits, meeting assignment and sequencing constraints and optimizing a given schedule criterion, like the overage weighted earliness or the makespan.

3. IMPROVED MATHEMATICAL FORMULATIONS

3.1 The MILP approach of Méndez et. al. (2001)

PROBLEM CONSTRAINTS:

Assignment of batches to processing units. A single equipment item should be allocated to every batch. The binary variable Y_{ij} stands for the decision of allocating batch i to unit j.

$$\sum_{j \in J_i} Y_{ij} = 1 \qquad \forall i \in I \qquad (1)$$

Batch sequencing. If two batches $i, i' \in I$ can be assigned to the same processing unit $j \in J_i \cap J_{i'}$, then sequencing constraints dealing with setup or changeover times that prevent from task overlapping should be included. These sequencing decisions are handled through fewer binary variables by using the general precedence concept. Let the binary variable $X_{ii'}$ stand for the relative ordering of batches $i,i' \in I$, where $i<i'$, if both batches are assigned to the same unit j ($Y_{ij} = Y_{i'j} = 1$). Specifically, $X_{ii'} = 1$ if batch i is processed before batch i' in the common equipment unit j, or $X_{ii'} = 0$ if batch i takes place afterwards. Notice that this binary variable becomes meaningless if i and i' are assigned to different equipment units.

$$C_i + \tau_{ii'j} + su_{i'j} \le S_{i'} + H\left(1 - X_{ii'}\right) + H\left(2 - Y_{ij} - Y_{i'j}\right) \qquad \forall i,i' \in I,\ j \in J_{ii'}:\ i < i' \qquad (2)$$

$$C_{i'} + \tau_{i'ij} + su_{ij} \le S_i + H\,X_{ii'} + H\left(2 - Y_{ij} - Y_{i'j}\right) \qquad \forall i,i' \in I,\ j \in J_{ii'}:\ i < i' \qquad (3)$$

where: $J_{ii'} = J_i \cap J_{i'}$

Timing of batches. The starting time of batch i can be computed from its completion time by subtracting its processing time in the assigned unit.

$$S_i = C_i - \sum_{j \in J_i} pt_{ij}\, Y_{ij} \qquad \forall i \in I \qquad (4)$$

In addition, the starting time of a batch must be higher than either its release time or the sum of both the unit ready time and the batch setup time in the assigned equipment unit.

$$S_i \ge \sum_{j \in J_i} Max\left[rt_i,\, ru_j + su_{ij}\right] Y_{ij} \qquad \forall i \in I \qquad (5)$$

OBJECTIVE FUNCTIONS:

Makespan. The makespan represents how much time is required to complete all processing tasks. The definition of the makespan variable MK is incorporated in the model by the following inequality constraint:

$$C_i \le MK \qquad \forall i \in I \qquad (6)$$

Thus, the problem goal will be:

$$\text{Minimize } MK \qquad (7)$$

Average weighted earliness. An alternative problem goal is to minimize the average weighted earliness:

$$\text{Minimize } \frac{1}{|I|}\sum_{i \in I} \alpha_i \left(dd_i - C_i\right) \qquad (8)$$

where the parameter α_i stands for the weight of the earliness of batch i. Notice that minimizing the average or the overall weighted earliness are equivalent problems as long as $n = |I|$ is a constant parameter. Since due dates are also constant, this objective function involves maximizing the average completion time of the batches.

The previous objective function will be useful only if the completion time of each batch never exceeds its specified due date.

$$C_i \le dd_i \qquad \forall i \in I \qquad (9)$$

3.2 Tightening the Makespan lower bound

A simple analysis of the above equations (2)-(6) leads to the following conclusions: The makespan lower bound is defined as the maximum completion time of any batch by constraint (6). At the same time, constraint (4) defines the completion time of each batch based on its processing time and start time. Therefore, it can be expected that the optimal solution will minimize processing times and start times simultaneously, at least for the last batch completed, since both (4) and (5) depend on the unique processing unit allocated by (1). If the equipment availability is much higher than the batch requirements (i.e. more equipment units than batches to schedule are available), the above constraints will lower bound the objective function and the MILP solver will easily choose the best possible assignment solution.

Unfortunately, this is not the ordinary situation. In general, several batches will be allocated to each processing unit and the sequencing equations (2) and (3) will need to be taken into account. Since these constraints are of big-M type, they do not provide a good lower bound estimation on the objective function. Even if the assignment variables Y_{ij} and $Y_{i'j}$ were set to 1 for a given equipment j, the sequencing variable $X_{ii'}$ can take a fractional value during the branch-and-bound search causing that neither equation (2) nor (3) have any effect on the starting or completion times of the batches.

However, assignment variables partially or completely allocating units to batches, i.e. equal to 1 or a positive fractional value, constitute a valuable information to estimate a tight lower bound for the makespan. Usually, the processing time of a task is frequently larger than its setup time (either sequence dependent or independent). Based on this assumption, the overall workload assigned to each equipment unit can be estimated using the summation of processing times of all the tasks allocated to it. When sequence independent setup times are considered, they can be included in the summation, which happens to be a good estimation for the schedule makespan:

$$ru_j + \sum_{i \in I_j} \left(su_{ij} + pt_{ij} \right) Y_{ij} \leq MK \qquad \forall\, j \in J \quad (10)$$

Notice that constraint (10) determines a lower bound for the makespan based on the total processing time at each unit, whatever is the sequence of tasks selected. Consequently, the summation term is a valid estimation for the makespan, based only on assignment variables. Sequencing decisions neither appear in this equation nor influence its tightening effect.

If, instead, sequence dependent setup times are significant, they can still be included in the previous equation very easily. The summation in constraint (11) now includes the lowest possible sequence dependent setup for each batch, as defined by equation (12). Since τ_{ij}^{Min} is included for every batch allocated on the processing unit, and the first batch to be processed does not need any prior setup, the highest possible sequence dependent setup is subtracted in order to ensure optimality:

$$ru_j - \underset{i \in I_j}{Max} \left[\tau_{ij}^{Min} \right] + \sum_{i \in I_j} \left(\tau_{ij}^{Min} + su_{ij} + pt_{ij} \right) Y_{ij} \leq MK \qquad \forall\, j \in J \quad (11)$$

where:

$$\tau_{ij}^{Min} = \underset{i' \in I_j \,:\, i' \neq i}{Min} \left[\tau_{i'ij} \right] \qquad \forall\, i \in I,\ j \in J_i \quad (12)$$

Constraint (11) provides a good lower bound on the value of MK because the model will try to optimise the batch sequencing at each unit in order to minimize the setup times. However, only if setup times between batches at the same unit are of similar order of magnitude this inclusion will be useful as a tight estimation. If $\tau_{ij}^{Min} = 0$ for a given batch, no improvement on the lower bound is possible.

Although the above estimations are rather straightforward, they are quite useful to get a good lower bound estimation on the makespan using a formulation that allows sequence dependent setup times. It is desirable to also get simple estimations for other objective functions.

3.3 Tightening the lower bound on the Average Earliness

As mentioned before, the objective function defined by (8) maximizes the summation of batch completion times. In turn, constraint (9) gives an upper bound on the completion time of each batch, also defining a preliminary lower bound on the value of this objective function. Neither (4) nor (5) have any influence on it. Since start times are not primarily affected by the objective function, there is no direct model trend to choose any particular equipment unit for a given batch. In the makespan case it was likely to choose an equipment with minimum processing time. Thus, for a given batch $i \in I$ allocated to unit $j \in J_i$, it is clear that its completion time will not be deteriorated unless another batch i' is assigned to the same processing unit. But, as mentioned before, sequencing decisions must be made in order to change start or completion times, and these decisions are defined by constraints (2) and (3), which are of big-M type. Since changes on the objective function value during the branch-and-bound process depend on the relative ordering of the batches allocated to the same processing unit, and because batches are not assumed to be previously assigned to any equipment unit, estimating a lower bound on the average earliness will be a more complex task than before.

Nonetheless, valid estimations of batch earliness can still be inferred. Let us suppose that two processing tasks i and i' are both allocated to the same unit j with batch i preceding i' (i.e., $Y_{ij} = Y_{i'j} = 1$ and $X_{ii'} = 1$ if

$i<i'$). Therefore, let us define the parameter $\beta_{ii'j}$ to estimate the earliness deterioration caused by batch i' over batch i if both are allocated to the same unit j, and batch i is executed before i'. Notice that the latest start time (*LST*) of batch i' will be an upper bound on the completion time of batch i. Figure 1 shows three possible scenarios for the temporal relation between due dates and processing times of both batches. In Figure 1(a), batch i' is due before batch i. Since it was assumed that i precedes i', the completion time of batch i must be deteriorated by at least the sum of the processing time and the setup time of i'. Alternatively, if batch i' is due after batch i, the completion time of batch i will be bounded by the difference shown in Figure 1(b), only if such a difference is positive. Otherwise, the estimation will be null as in Figure 1(c). Therefore, the earliness of batch i is deteriorated for each batch i' executed on unit j after i by the amount $\beta_{ii'j}$. Summing up these individual deteriorations, it is possible to estimate a lower bound on the earliness of each batch i, as will be next shown.

To derive the proposed estimation, a different set of sequencing binary variables must be defined:

$$\Theta_{ii'j} = \begin{cases} 1 & \text{if batch } i \text{ is processed before batch } i' \text{ on unit } j \\ 0 & \text{otherwise} \end{cases}$$

$$\forall i,i' \in I,\ j \in J_{ii'} :\ i \neq i'$$

Since these binary variables are defined for every possible permutation of two distinct batches, and for each eligible equipment unit for both batches, an immediate conclusion is that this new approach will significantly increase the number of decision variables. With the model of Méndez *et. al.* (2001), a smaller set of sequencing variables is required because sequencing and assignment decisions were independent. Variable $X_{ii'}$ do not includes index j and is defined for every ordered pair of batches $i,\ i' \in I$, such that $i<i'$. As the number of variables $\Theta_{ii'j}$ is larger, it can be expected that the computational performance will not improve. However, as it will be shown, this is not true since a better lower bound for the objective function and, consequently, a lower CPU requirement is achieved.

Constraints (13) and (14) replace constraints (2) and (3) for the proposed relationships among sequencing decisions:

$$Y_{ij} + Y_{i'j} \leq 1 + \Theta_{ii'j} + \Theta_{i'ij}$$

$$\forall i,i' \in I,\ j \in J_{ii'} : i < i' \qquad (13)$$

$$C_i + \sum_{j \in J_{ii'}} \left(\tau_{ii'j} + su_{i'j}\right) Y_{i'j} \leq S_{i'} + H\left(1 - \sum_{j \in J_{ii'}} \Theta_{ii'j}\right)$$

$$\forall i,i' \in I :\ i \neq i' \qquad (14)$$

Constraint (13) produces that either $\Theta_{ii'j}$ or $\Theta_{i'ij}$ are set to 1 if batches i and i' are both allocated to unit j.

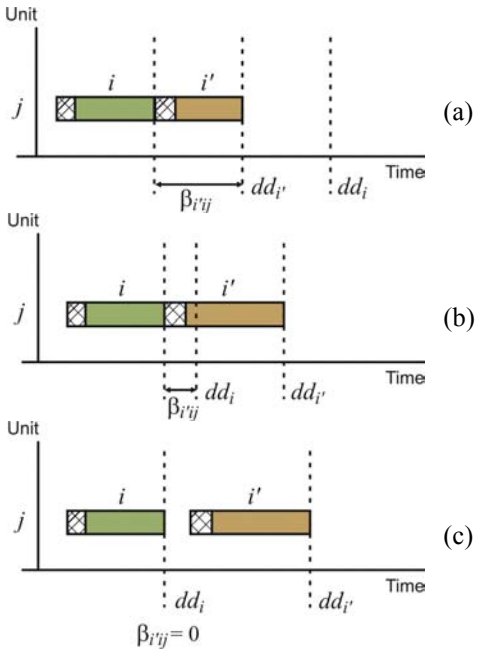


Fig. 1. Earliness deterioration caused by batch i' over batch i if both are allocated to the same unit j.

This constraint relates model's assignment decisions and model's sequencing decisions, which is the main difference with the previous approach. In turn, constraint (14) has the same effect that constraints (2) and (3), but here a summation over the available equipment units reduces the number of constraints by one order of magnitude.

Finally, the proposed constraint to tighten the lower bound on the overall earliness objective function is:

$$C_i + \sum_{i' \in I\,:\,i' \neq i} \sum_{j \in J_{ii'}} \beta_{ii'j}\, \Theta_{ii'j} \leq dd_i \qquad \forall i \in I \qquad (15)$$

Here constraint (15) replaces previous constraint (6), where $\beta_{ii'j}$ is the earliness estimation parameter defined as:

$$\beta_{ii'j} = \begin{cases} \tau_{i'j}^{Min} + su_{i'j} + pt_{i'j} & , \text{if } dd_{i'} \leq dd_i \\ dd_i + \tau_{i'j}^{Min} + su_{i'j} + pt_{i'j} - dd_{i'} & , \text{if } (dd_i < dd_{i'}) \text{ and } (dd_{i'} - \tau_{i'j}^{Min} - su_{i'j} - pt_{i'j} < dd_i) \\ 0 & , \text{otherwise} \end{cases} \qquad (16)$$

Constraint (15) causes that the upper bounds on the completion times are affected by the assignment decisions while batch-unit allocations are made. To reach this conclusion it is necessary to understand how constraints (13) and (15) work together. Constraint (13) assures that, once a batch i is assigned to equipment j ($Y_{ij} = 1$), any other batch i' assigned or partially assigned to the same unit ($0 < Y_{i'j} \leq 1$) will automatically increase binaries variables $\Theta_{ii'j}$ or $\Theta_{i'ij}$. Since these sequencing variables appear on the summation of constraint (15), it is expected that a variable $\Theta_{i'ij}$ related to the ordered pair ($i,\ i'$) on

unit j and featuring a lower $\beta_{ii'j}$ will take a nonzero value. Therefore, the lowest possible deterioration of both completion times will be chosen.

Tightening constraint (15) will have no effect until assignment decisions are at least partially made. If two batches have the same or almost the same due date, it is expected that both $\beta_{ii'j}$ and $\beta_{i'ij}$ will be nonzero. In this case, if i is already allocated to unit j ($Y_{ij} = 1$), the optimal relaxed solution will avoid the assignment of i' to the same unit j, if a deterioration of the earliness of one of the batches will happen. In this way, the model will avoid the assignment of new tasks to a unit if it is overloaded. In general, the lower bound proposed on the value of the objective function is useful during the node pruning process whenever $1 < Y_{ij}+Y_{i'j} \leq 2$, and consequently, i and/or i' are assigned or partially assigned to the same processing unit j.

4. COMPUTATIONAL RESULTS

The effectiveness of the above tightening constraints will be illustrated by finding the minimum-makespan (Example 1) and the minimum-earliness (Example 2) schedules for a single-stage multiproduct batch plant. Both sequence dependent and independent changeovers are considered, since sequence dependent setups have a significant influence on the model performance and they cannot be efficiently considered with other formulations.

The problem to be tackled involves a plastic compounding plant with a single stage and four extruders running in parallel. This problem was first studied by Pinto and Grossmann (1995) and Ierapetritou, Hené, and Floudas (1999) with up to 29 batch orders. Méndez and Cerdá (2003) expanded the number of batches to 40, in order to undertake an appropriate dynamic scheduling scenario. Order due dates and unit-dependent processing and setup times used on this section can be found in Méndez and Cerdá (2003).

Table 1. Product families

Family	Batches
F_1	O_1, O_2, O_3, O_5, O_{10}, O_{16}, O_{20}, O_{22}
F_2	O_4, O_8, O_9, O_{14}, O_{18}, O_{26}, O_{31}
F_3	O_7, O_{23}, O_{24}, O_{30}, O_{33}, O_{34}, O_{36}, O_{37}, O_{38}, O_{40}
F_4	O_6, O_{11}, O_{15}, O_{17}, O_{19}, O_{32}, O_{35}
F_5	O_{12}, O_{13}, O_{21}, O_{25}, O_{27}, O_{28}, O_{29}, O_{39}

Table 2. Sequence dependent setup times between families

	F_1	F_2	F_3	F_4	F_5
F_1	0.104	0.127	0.178	0.192	0.217
F_2	0.122	0.115	0.266	0.229	0.291
F_3	0.191	0.214	0.175	0.304	0.424
F_4	0.350	0.205	0.328	0.184	0.400
F_5	0.357	0.423	0.348	0.284	0.205

In order to address sequence dependent changeovers, batch orders are grouped into five product families F_1-F_5 as shown in Table 1, and sequence-dependent setup times between families are listed in Table 2. Hence, two versions of Examples 1 and 2 have been studied, one assuming sequence-independent changeovers and the other considering changeovers as sequence dependent. Both versions were solved using the approach of Méndez *et. al.* (2001) and the corresponding improved formulation, for an increasing number of batches ranging form 12 to 25/40, in order to reach the computational limit of each model.

Table 3. Makespan minimization results with the model of Méndez *et. al.* (2001)

n	Binary vars, Continuous vars, Constraints	Sequence independent setup times				Sequence dependent setup times			
		Objective Function	Relative Gap (%)	CPU time (sec.)	Nodes	Objective Function	Relative Gap (%)	CPU time (sec.)	Nodes
12	82, 25, 214	8.428	-	19.03	94365	8.645	-	8.36	39350
16	140, 33, 382	12.353	2.43	>3600	8893218	12.854	-	1188.50	3421982
18	161, 37, 444	13.985	-	2872.81	7166701	14.633	27.07	>3600	8708577
20	201, 41, 558	15.268	22.62	>3600	6282059	15.998	21.95	>3600	6570231

Table 4. Makespan minimization results with the proposed model

n	Binary vars, Continuous vars, Constraints	Sequence independent setup times				Sequence dependent setup times			
		Objective Function	Relative Gap (%)	CPU time (sec.)	Nodes	Objective Function	Relative Gap (%)	CPU time (sec.)	Nodes
12	82, 25, 218	8.428	-	0.05	12	8.645	-	0.05	15
16	140, 33, 386	12.353	-	0.03	1	12.854	-	0.09	44
18	161, 37, 448	13.985	-	0.11	27	14.611	-	40.36	116413
20	201, 41, 562	15.268	-	0.14	21	15.998	-	183.56	417067
22	228, 45, 622	15.794	-	0.20	49	16.396	-	167.09	359804
25	286, 51, 792	18.218	-	0.42	110	19.064 *	-	79.25	109259
29	382, 59, 1064	23.302	-	0.61	82	24.723 *	-	5.92	5385
35	532, 71, 1430	26.683	-	0.97	90				
40	625, 81, 1656	28.250	-	0.91	34				

* For a Relative Gap Tolerance of 0.01

Table 5. Overall earliness minimization results with the model of Méndez *et. al.* (2001)

n	Binary vars, Continuous vars, Constraints	Sequence independent setup times				Sequence dependent setup times			
		Objective Function	Relative Gap (%)	CPU time (sec.)	Nodes	Objective Function	Relative Gap (%)	CPU time (sec.)	Nodes
12	82, 24, 214	1.026	-	0.03	22	1.376	-	0.01	12
16	140, 32, 382	9.204	-	1.30	3668	11.647	-	2.70	8301
18	161, 36, 444	16.496	-	38.48	84843	18.773	-	55.77	123666
20	201, 40, 558	17.073	-	77.78	148101	19.131	-	81.38	159388
22	228, 44, 618	22.815	1.68	>3600	4385294	27.754	8.33	>3600	3616973
25	286, 50, 788	29.430	49.63	>3600	2720801	40.541	57.68	>3600	3435213

Table 6. Overall earliness minimization results with the proposed model

n	Binary vars, Continuous vars, Constraints	Sequence independent setup times				Sequence dependent setup times			
		Objective Function	Relative Gap (%)	CPU time (sec.)	Nodes	Objective Function	Relative Gap (%)	CPU time (sec.)	Nodes
12	191, 24, 245	1.026	-	0.02	1	1.376	-	0.03	1
16	351, 32, 437	9.204	-	0.30	78	11.647	-	0.56	404
18	408, 36, 508	16.496	-	1.19	785	18.773	-	1.20	853
20	519, 40, 639	17.073	-	1.63	807	19.131	-	3.27	2178
22	574, 44, 721	22.815	-	9.11	6598	27.754	-	90.75	75569
25	738, 50, 916	29.430	-	91.14	57741	37.216	15.25	>3600	2006319

All results were found on a Pentium IV PC (2.8 GHz) with ILOG OPL Studio 3.7, using the embedded CPLEX v. 9.0 mixed-integer optimizer. CPU time limit was defined on 1 hr. Except for the two cases indicated in Table 4, the solver default relative gap tolerance equal to 0.0001 was used. The time horizon limit $H = 30$ was used as the big-M parameter.

The results for the makespan minimization problem (Example 1) are shown in Table 3 for the model of Mendez *et. al.* (2001), and in Table 4 for the improved formulation. The proposed model is always faster for each problem instance being tackled. For sequence independent problems, 40 batches are scheduled in less than a second. For sequence-dependent problems almost optimal solutions are found in few CPU seconds, since the relative gap decreases significantly faster because of the tightening constraints.

For Example 2 the corresponding results are shown in Tables 5 and 6. Direct comparison for the 20 batches problem shows an improvement on the computational time of 47:1 for the sequence dependent and of 24:1 for the sequence independent cases. Since tightening constraints for this example do not have a notorious effect until assignments are made, the lower bound for the proposed formulation increases slower than before.

For both examples, the model of Méndez *et. al.* (2001) reaches good (most optimal) solutions in few seconds, but needs larger computational effort to prove their optimality, since the lower bound on the value of the objective function increases very slowly.

5. CONCLUSIONS

A pair of improved MILP formulations for single-stage batch scheduling problems that efficiently handle sequence dependent setup times has been proposed. Better computational results are achieved by incorporating tightening constraints that increase the lower bound on the objective function value. Problems of up to 40 batches have been solved in a very low CPU time.

REFERENCES

Floudas, C.A. and X. Lin (2004). Continuous-time versus discrete-time approaches for scheduling of chemical processes: A review. *Computers and Chemical Engineering* **28**, 2109-2129.

Ierapetritou, M.G. and C.A. Floudas (1998). Effective Continuous-Time Formulation for Short-Term Scheduling. 1 Multipurpose Batch Processes. *Industrial and Engineering Chemistry Research* **37**, 4341-4359.

Ierapetritou, M.G., T.S. Hené and C.A. Floudas (1999). Effective continuous-time formulation for short-term scheduling. 3 Multiple intermediate due dates. *Industrial and Engineering Chemistry Research* **38**, 3445-3461.

Méndez, C.A., G.P. Henning and J. Cerdá (2001). An MILP continuous-time approach to short-term scheduling of resource-constrained multistage flowshop batch facilities. *Computers and Chemical Engineering* **25**, 701-711

Méndez, C.A. and J. Cerdá (2003). Dynamic scheduling in multiproduct batch plants. *Computers and Chemical Engineering* **27**, 1247-1259.

Pinto, J.M. and I.E. Grossmann (1995). A continuous time mixed integer linear programming model for short term scheduling of multistage batch plants. *Industrial and Engineering Chemistry Research* **34**, 3037-3051.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

Constraint Logic Programming for Non Convex NLP and MINLP Problems

Kotecha P. R., Gudi R. D.*

Department of Chemical Engineering
Indian Institute of Technology Bombay, Powai, Mumbai – 400076, India.
* Author for correspondence: Email: ravigudi@che.iitb.ac.in

Abstract: This paper presents an algorithm to solve non-convex NLP and MINLP problems using CLP. In the proposed technique, the continuous variables are relaxed to take only integer values contained in the real domain of the variable. The merits of the CLP algorithm, *viz* powerful CP strategies are proposed to be exploited to get integer solutions to the relaxations. A lower bound to the objective function is obtained if the relaxed problem is feasible. This information is used in the successive stages wherein the continuous variables are corrected from their integer variable representation to obtain real solutions with desired accuracy. The proposed technique has been successfully demonstrated on two MILP, two non convex NLP and two non convex MINLP problems. The problems were also solved by traditional techniques and the superiority of the proposed method has been demonstrated.

Keywords: Constraint Logic Programming (CLP), Constraint Propagation (CP), Integer Programming (IP) Mixed Integer Linear Programming (MILP), Mixed Integer Non-Linear Programming (MINLP)

1. INTRODUCTION

Many of the problems arising in synthesis and design, and in planning and scheduling problems are MINLP models. This is due to the fact that MINLP provides much greater modeling flexibility for tackling a large variety of problems. While MILP methods have been largely developed outside process systems engineering, chemical engineers have played a prominent role in the development of MINLP methods. Some of the methods used to solve MINLP include Branch and Bound, Generalized Benders Decomposition (GBD), Outer Approximation (OA) and Extended Cutting Plane methods (ECP). In the above methods, the objective function and the constraints are assumed to be convex and differentiable. But, often times, some problems lead to formulations that do not satisfy these requirements of convexities. The trim loss problem in paper industry and the 10P3S problem (Munawar & Gudi, 2005) are some such problems. Ryoo and Sahinidis (1995) have reported a collection of twenty one non convex problems that arise in process synthesis.

There have been a number of attempts to handle non convex MINLPs. Tawarmalani and Sahinidis (2000) have developed a branch and bound method that branches on the continuous and discrete variables. This method, which relies on bounds reduction using underestimators, has been implemented in BARON. The SMIN-αBB and GMIN-αBB algorithms have been developed for twice-differentiable nonconvex MINLPs. The αBB method, which is a branch and bound procedure that branches on both the continuous and discrete variables according to specific options, was developed by using a valid convex underestimation of general functions as well as special functions. The branch-and-contract method for global optimization of process models which have bilinear, linear fractional and concave separable functions in the continuous variables and linear 0–1 variables, uses bound contraction and applies the outer-approximation algorithm at each node of the tree for the spatial search. Lee and Grossmann (2001) developed a two level-branch and bound method for solving nonconvex disjunctive programming problems. Munawar & Gudi (2005) have proposed a hybrid technique to solve MINLP that makes use of Differential Evolution (DE) and Non Linear Programming (NLP).

Finding its roots in computer science and artificial intelligence communities, Constraint Programming (Puget, 1994; Van Hentenryck, 1989) is an alternative approach to discrete and continuous problem solving. For decades, it has proved

successful in several applications, particularly in scheduling and logistics. Unlike mathematical programming, it does not use relaxations but it relies on methods of logical inference (primarily domain reduction and constraint-propagation) to reduce the domain of possible values for a discrete or continuous variable. The rich modeling language of CLP has contributed in a large way towards its success.

The methods described above for the solution of non convex MINLP essentially involve the Branch and Bound and Extended Cutting Plane methods. The successes of these methods are crucially dependent on the successful solution of the NLP sub-problems at each node. On the other hand, CLP methods are good at domain propagation but are restricted to the finite domain and do not handle continuous variables. In recent years, there has been substantial progress in the development of powerful constraint propagation engines, which could be exploited towards solution of problems represented in the finite domain. An alternative approach to solving MINLPs that relies on the use of CLP towards domain reduction could therefore be examined towards solution of such non-convex MINLP problems.

In this paper, a method has been proposed that uses CLP to solve non-convex MINLP problems. Contrary to the regular approach of relaxing on the integrality requirements as in the branch and bound algorithm, the proposed approach relaxes the continuous variables to discrete values. Since constraint propagation approaches are usually more suitable in the finite domain over other integer programming methods (Smith, *et al.,* 1997), we propose to use the powerful features of constraint propagation engine to reduce the finite domain space. The method involves solving a master problem obtained by relaxing the space of continuous variables to integer domain. If the relaxed problem is feasible, it ensures that a lower bound (for the maximization problem) is obtained for the problem. Also, the domain reduction inherently present in CP helps to specify tighter bounds on the continuous variables. These steps are followed by the solution of another sub problem in which the continuous variables are corrected from their integer variable representation to obtain real solutions with desired accuracy. In this sub problem, the bounds on the continuous variables are also tightened by inferring from the solution of the master problem. If the master problem is infeasible, the original problem itself is discretized and solved.

The remainder of the paper is organized as follows. The following section gives a review of CLP technique. Section 3 focuses on some theoretical aspects of CP relevant to the real domain. Section 4 discusses the results obtained by applying the proposed methodology on two MILP and two non convex NLP and MINLP problems. The second MILP problem is a planning and scheduling problem on a set of dissimilar parallel machines (Jain and Grossman, 2001). We solve this problem using CLP even when the start times, due dates, release dates and processing times are not integer parameters. This is particularly noteworthy considering the fact that ILOG Scheduler does not support continuous variables.

2. CONSTRAINT LOGIC PROGRAMMING

Constraint Programming (Hentenryck, 1989; Hooker, 2000) was originally developed to solve feasibility problems, but it has been extended to solve optimization problems as well. In finite domain CLP, each integer variable x_i has an associated domain D_i which is the set of possible values that this variable can take on in the optimal solution. The cartesian product of the domains D_1 x…..x D_n forms the solution space of the problem. This space is finite and can be searched exhaustively for a feasible or optimal solution, but to intelligently enumerate this search, CP is used to infer infeasible solutions and prune the corresponding domains. From this viewpoint, CP operates on, and narrows down, the set of possible solutions.

Constraint Programming is based on performing a tree enumeration. At each node the domains of the variables, which can be continuous, general integer, boolean and binary are reduced. If an empty domain is found the domain is pruned. Branching is performed whenever a domain of an integer, binary or boolean variable has more than one element, or when the bounds of the domain of a continuous variable do not lie within a tolerance. Whenever a solution is found, or a domain of a variable is reduced, new constraints are added to ensure that the node is not revisited. The search terminates when no further nodes need to be examined.

Traditional IP methods are very efficient for problems with good relaxations but suffer when the relaxation is weak or when its restricted modeling (linear inequalities) framework results in large models. CLP with its more expressive language results in smaller models that are closer to the problem description, and performs better for highly constrained problems; however, it lacks the global perspective of relaxations.

There have been a few attempts to integrate CLP and MILP so that their complementary strengths can be exploited. Some examples include the modified generalized assignment problem (Darby *et al.,* 1997), the template design problem, the progressive party problem (Smith *et al.,* 1997), and the change problem (Heipcke, 1999). These papers showed that MILP is very efficient when the relaxation is tight and the models have a structure that can be effectively exploited. CP works better for highly constrained discrete optimization problems where expressiveness of MILP is a major limitation. Hooker (2000) deals with the subject of MILP and CP integration in detail. Jain and Grossman (2001) have shown a decomposition method wherein a master MILP and a CLP subproblem work in cooperation and are able to address problems, that were otherwise intractable by both the methods.

3. DECOMPOSITION METHODOLOGY

The algorithms proposed in this section have been motivated by the high efficiency of CLP in reducing the domain of variables. A motivating example has also been presented at the end of the section. Although shown for an MINLP, this theory is also valid for solution of an MILP.

Consider a problem, which when modeled as an MINLP has the following structure,

$$\text{(M1)} \quad \max_{x,y} f(x,y) \tag{1}$$
$$\text{s.t} \quad G(\bar{x},y) \tag{2}$$
$$x \in \square \tag{3}$$
$$y \in \{0,1\} \tag{4}$$

where *G(x,y)* could represent both equality and inequality constraints. The above optimization problem has both continuous and binary variables. It is to be noted no restrictions are placed *on (1) and (2)* to be convex and linear in the discrete variables.

The master problem of (M1) is given by

$$\text{(M2)} \quad \max_{x,y} f(x,y) \tag{5}$$
$$\text{s.t} \quad G(\bar{x},y) \tag{6}$$
$$x \in \square \tag{7}$$
$$y \in \{0,1\} \tag{8}$$

Any solution obtained for M2 will provide a lower bound for the problem M1. In other words, M1 will never have a value that is lower than M2. Let f', the solution to M2, define this lower bound.

The sub problem (M3) is defined as

$$\text{(M3)} \quad \max_{\bar{x},y} f(\bar{x},y) \tag{9}$$
$$\text{s.t} \quad G(\bar{x},y) \tag{10}$$
$$f \geq f' \tag{11}$$
$$\bar{x} = \sum_{i=0}^{n} 10^{-i}(a_i\_x) \tag{12}$$
$$y \in \{0,1\} \tag{13}$$
$$a_0\_x \in \bar{\square} \tag{14}$$
$$a_i\_x \in \{0,1,2,3....9\} \quad i \neq 0 \tag{15}$$

The domain $\bar{\square}$ can be inferred from the solution of M2. For a maximization (minimization) problem, the lower (upper) bound of x will start from (end at) the solution of x in M2. The upper (lower) bound will essentially remain the same as in M2.

The sub-problem (M4) is defined as

$$\text{(M4)} \quad \max_{\bar{x},y} f(\bar{x},y) \tag{16}$$
$$\text{s.t} \quad G(\bar{x},y) \tag{17}$$
$$\bar{x} = \sum_{i=0}^{n} 10^{-i}(a_i\_x) \tag{18}$$
$$y \in \{0,1\} \tag{19}$$
$$a_0\_x \in \tilde{\square} \tag{20}$$
$$a_i\_x \in \{0,1,2,3....9\} \quad i \neq 0 \tag{21}$$

This problem is solved if and only if M1 gives an infeasible solution. The domain in equation (20) is larger than (14) because no inference can be made from the solution of M1 and M4 is a larger problem than M3. It should be noted that M4 can be solved for an optimal solution even without solving M2.

Table 1: Proposed Algorithm to solve MILP/MINLP problems

Algorithm

Step 1: Formulate M2 by relaxing the continuous variable space of M1 to Integer space
Step 2: Solve M2 using CLP.
If M2 is feasible,
go to Step 5
Step 3: Formulate M4 by discretizing M1 using (17) subject to (19) and (20)
Step 4: Solve M4.
If feasible,
Solution is Optimal for M1
else
M1 is Infeasible
Step 5: Formulate M3 by discretizing M2 using (11) subject to (13) and (14)
Step 6: Solve M3 to obtain optimal of M1

4. CASE STUDIES

This section discusses the application of the proposed methodology on some MILP, NLP and non-convex MINLP problems.

4.1 Case Study 1

The example discussed is an MILP with 3 continuous and 3 binary variables and will be hereon referred as Case Study 1.

(M1)

$$Min_{x,y} \quad f = x_1 + x_2 - x_3 + y_1 + y_2 - y_3$$
$$s.t \quad x_1 + x_2 - x_3 \leq 151.2$$
$$x_3 - x_1 \geq 2$$
$$x_2 - x_1 \geq 70$$
$$y_1 + y_2 + y_3 \geq 2$$
$$x_1 \geq 5.9$$
$$(5,0,0) \leq X \leq (100,100,100); \quad Y \in \{0,1\}^3$$

The formulation of model (M2) is done by relaxing the continuous spaces of x to integer space.

$$Min_{x,y} \quad f = x_1 + x_2 - x_3 + y_1 + y_2 - y_3$$
$$s.t \quad x_1 + x_2 - x_3 \leq 151.2$$
$$x_3 - x_1 \geq 2$$
(M2)
$$x_2 - x_1 \geq 70$$
$$y_1 + y_2 + y_3 \geq 2$$
$$x_1 \geq 5.9 \qquad x_2, x_3 \in \{0,1,2...,100\};$$
$$x_1 \in \{5,6...,100\}; \quad Y \in \{0,1\}^3$$

The optimal solution for (M2) is $(x_1^*, x_2^*, x_3^*, y_1^*, y_2^*, y_3^*; f^*) = (6,76,69,0,1,1;13)$.

This solution gives us a bound on the objective function i.e., the objective function of (M1) can never be greater than 13.

The model (M3) is formulated by inferring tighter bounds from the solution of (M2)

(M3)

$$
\begin{aligned}
Min_{x,y} \quad & f = \bar{x}_1 + \bar{x}_2 - \bar{x}_3 + y_1 + y_2 - y_3 \\
s.t \quad & \bar{x}_1 + \bar{x}_2 - \bar{x}_3 \le 151.2 \\
& \bar{x}_3 - \bar{x}_1 \ge 2 \\
& \bar{x}_2 - \bar{x}_1 \ge 70 \\
& y_1 + y_2 + y_3 \ge 2 \\
& \bar{x}_1 \ge 5.9; \quad Y \in \{0,1\}^3 \\
& a_0\_x_1 \in \{5,6\}; \quad a_0\_x_2 \in \{0,1,...76\}; \\
& a_0\_x_3 \in \{0,1,..69\} \\
& \left.\begin{matrix} a_1\_x_1, a_2\_x_1, a_1\_x_2, \\ a_2\_x_2, a_1\_x_3, a_2\_x_3 \end{matrix}\right\} \in \{0,1,2...9\} \\
where \quad & \bar{x}_1 = a_1\_x_1 + 0.1\, a_2\_x_1 + 0.01\, a_3\_x_1 \\
& \bar{x}_2 = a_1\_x_2 + 0.1\, a_2\_x_2 + 0.01\, a_3\_x_2 \\
& \bar{x}_3 = a_1\_x_3 + 0.1\, a_2\_x_3 + 0.01\, a_3\_x_3
\end{aligned}
$$

The optimal solution for (M3) is $(x_1^*, x_2^*, x_3^*, y_1^*, y_2^*, y_3^*; f^*)$ is $(5.95, 75.95, 69.3, 0, 1, 1; 12.6)$

The formulation for model M4 is given by

(M4)

$$
\begin{aligned}
Min_{x,y} \quad & f = \bar{x}_1 + \bar{x}_2 - \bar{x}_3 + y_1 + y_2 - y_3 \\
s.t \quad & \bar{x}_1 + \bar{x}_2 - \bar{x}_3 \le 151.2 \\
& \bar{x}_3 - \bar{x}_1 \ge 2 \\
& \bar{x}_2 - \bar{x}_1 \ge 70 \\
& y_1 + y_2 + y_3 \ge 2 \\
& \bar{x}_1 \ge 5.9; \quad Y \in \{0,1\}^3 \\
& a_0\_x_1 \in \{5,6,..,100\} \\
& a_0\_x_2 \in \{0,1,...100\} \\
& a_0\_x_3 \in \{0,1,..100\}
\end{aligned}
$$

Though the master problem was feasible, nevertheless, M4 can be formulated and solved for this case study. But as said earlier, it is more computationally expensive. This fact is substantiated by the Table 2.

Table 2: CLP Parameters for Case Study 1

Model		No. of Choice Points	No. of Failures
Method 1	M2	3	64
	M3	13	53
Method 2	M4	27	128

4.2 Case Study 2

The following planning and scheduling MILP model has been taken from Jain and Grossman (2001). The scheduling problem involves finding a least-cost schedule to process a set of orders I using a set of dissimilar parallel machines M. Processing of an order $i \in I$ can only begin after the release date r_i and must be completed at the latest by the due date d_i. Order i can be processed on any of the machines. The processing cost and the processing time of order $i \in I$ on machine $m \in M$ are C_{im} and p_{im}, respectively.

$$
\begin{aligned}
\min \quad & \sum_{i \in I} \sum_{m \in M} C_{im} x_{im} \\
\text{s.t} \quad & ts_i \ge r_i \quad \forall\, i \in I; \qquad \sum_{m \in M} x_{im} = 1 \quad \forall\, i \in I \\
& ts_i \le d_i - \sum_{m \in M} p_{im} x_{im} \quad \forall\, i \in I \\
& \sum_{i \in I} p_{im} x_{im} \le \max_i \{d_i\} - \min_i \{r_i\} \\
& y_{ii'} + y_{i'i} \ge x_{im} + x_{i'm} - 1 \quad \forall\, i, i' \in I, i' > i, m \in M \\
& ts_{i'} \ge ts_i + \sum_{m \in M} p_{im} x_{im} - U(1 - y_{ii'}) \quad \forall\, i, i' \in I, i' \ne i \\
& y_{ii'} + y_{i'i} \le 1 \quad \forall\, i, i' \in I, i' > i \\
& y_{ii'} + y_{i'i} + x_{im} + x_{i'm'} \le 2 \quad \forall\, i, i' \in I, i' > i \\
& \qquad\qquad m, m' \in M, m \ne m' \\
& ts_i \ge 0; \quad x_{im} \in \{0,1\} \quad \forall\, i \in I, m \in M \\
& y_{ii'} \in \{0,1\} \quad \forall\, i \quad , i' \in I, i' \ne i; \quad U = \sum_{i \in I} \max_{m \in M} \{p_{im}\}
\end{aligned}
$$

The main decisions involved in this scheduling problem are assignment of orders on machines, sequence of orders on each machine, and start time for all the orders. The binary variable x_{im} is an assignment variable, and it is equal to one when order i is assigned to machine m. Binary variable $y_{ii'}$ is the sequencing variable, and it is equal to one when both i and i' are assigned to the same machine and order i' is processed after order i. The continuous variable ts_i denotes the start time of order i.

Table 3: Data for Case Study – 2

Order	Cost		Processing time		Release date	Due date
	Mch 1	Mch 2	Mch 1	Mch 2		
1	10.6	6.25	10.2	14.5	2.1	16.9
2	8.26	5.45	6.25	8.10	3.5	14.2
3	12.47	7.06	11.98	16.14	4.8	25.5

Table 4: Results for Case Study – 2

Order	Machine	Start Time	Processing Time	Finish
1	2	2.1	14.5	16.6
2	1	3.5	6.25	9.75
3	1	9.75	11.98	21.73

Jain and Grossman (2001) have discussed 10 instances of this problem. They have formulated the same problem as a CLP model to be compatible with ILOG Scheduler. The processing times are assumed to be integers in their formulation. In our work, we allow these parameters to be real and test the ILOG Solver's capability to accommodate this change using our

formulation. Table 3 shows the scheduling data of 3 orders on 2 machines.

For this problem, the master problem M2 was infeasible and hence M4 was solved using CLP. Table 4 shows the results of this case study. The results were equivalent to those obtained when M1 was solved using a MILP solver such as CPLEX.

4.3 Case Study 3

The following is a pooling NLP problem that has been studied extensively in the literature (Ryoo and Sahinidis, 1995).

$$\begin{aligned}\min_{x,y}\quad & -9x_5 - 15x_9 + 6x_1 + 16x_2 + 10x_6 \\ s.t\quad & x_1 + x_2 = x_3 + x_4; \quad x_3 + x_7 = x_5 \\ & x_4 + x_8 = x_9; \quad x_7 + x_8 = x_6 \\ & x_{10}x_3 + 2x_7 \le 2.5x_5; \quad x_{10}x_4 + 2x_8 \le 1.5x_9 \\ & 3x_1 + x_2 = x_{10}\left(x_3 + x_4\right); \\ & \left(0,0,0,0,0,0,0,0,0,1\right) \le X \le (300,300,100, \\ & 200,100,300,100,200,200,3)\end{aligned}$$

The proposed CLP based strategy reaches the global optimum without getting stuck at any of the infinite local solutions (Ryoo and Sahinidis, 1995). The global optimum is given by $\left(x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*, x_7^*, x_8^*, x_9^*, x_{10}^*\right) = (0,100,0,100, 0,100,0,100,200,1)$.

The above problem when solved using the NLP solver CONOPT was dependent on initial guesses and has the possibility of converging at local optima.

4.4 Case Study 4

The following non convex NLP has been taken again from Ryoo and Sahinidis (1995)

$$\begin{aligned}\min_{x,y}\quad & x_1 + x_2 \\ s.t\quad & x_1^2 + x_2^2 \le 4; \quad x_1^2 + x_2^2 \ge 1; \\ & x_1 - x_2 \le 1; \quad x_2 - x_1 \le 1 \\ & -2 \le X \le 2\end{aligned}$$

The solution of this problem with GAMS modeler and CONOPT as solver did not give satisfactory results and it was found that the global optima reported was dependent on the initial guess provided. The global optima for this problem using the proposed CLP approach agreed with the reported value $X = \left(-1.414214, -1.414214\right)$ and $f = -2.828427$.

4.5 Case Study 5

The following problem is a MINLP with one binary and one continuous variable. This problem was proposed by Kocis and Grossmann (1988), and was also solved by Floudas et al. (1989), Ryoo and Sahinidis (1995) and Cardoso et al. (1997).

$$\begin{aligned}\min_{x,y}\quad & f = 2x + y \\ s.t\quad & 1.25 - x^2 - y \le 0 \\ & x + y \le 1.6 \\ & 0 \le x \le 1.6 \\ & y \in \{0,1\}\end{aligned}$$

The first nonlinear inequality constraint contains a non-convex term for the continuous variable x. The global optimum is (x, y; f) = (0.5, 1; 2).The master (M2) problem of this case study is infeasible and hence this problem was successfully solved to global optimality using the M4 transformation. Munawar and Gudi (2005) have solved this problem using GAMS Solvers viz. CONOPT2 and SNOPT. It has been shown that the optimum is strongly dependent on initial guesses. Such problems are not encountered when the above non-convex MINLP is solved using the proposed approach M4.

4.6 Case Study 6

The following problem is a non convex MINLP problem with three continuous and two discrete variables.

$$\begin{aligned}\max_{x,y}\quad & x_1 + y_1 + x_1^2 + y_1^2 + x_1x_2 + y_1y_2 + y_2^2 + x_1^2y_1^2 + x_3 \\ s.t\quad & x_1x_2y_1 \ge 10; \quad x_1 + y_1 \ge 1 \\ & x_2 + y_2 \ge 1; \quad y_1 + y_2 \le 1 \\ & y_1y_2 \le 1; \quad x_3\left(1 - x_3\right) = 0 \\ & x_1 \le 3.5; \quad x_2 \le 15.5 \\ & 0 \le x_1 \le 4; \quad 0 \le x_2 \le 16; \quad 0 \le x_3 \le 1; \\ & Y \in \{0,1\}^3\end{aligned}$$

Unlike in case study 3, this problem has a feasible solution for the master problem and is found to be $\left(x_1^*, x_2^*, x_3^*, y_1^*, y_2^*; f^*\right) = \left(3,15,1,1,0; 69\right)$.

This implies that the objective function cannot be less than 69. The sub-problem (M3) is solved and the global optima is determined to be $\left(x_1^*, x_2^*, x_3^*, y_1^*, y_2^*; f^*\right) = \left(3.5,15.5,1,1,0; 85.25,\right)$.

This problem was also solved using M4. Though the global optimum was obtained, it was at a higher computational effort, as we report in Table 5.

Table 5: CLP Parameters for Case Study 6

Model		**No. of Choice Points**	**No. of Failures**
Method 1	M2	0	26
	M3	31	537
Method 2	M4	271	3735

This problem was also solved using the GAMS modeler with the DICOPT Solver. As can be seen in Table 6, solution using DICOPT is dependent on the initial guess to reach the global optimum. But in the proposed algorithm, there is no need for any initial guesses and yet the solutions are guaranteed to be globally optimal.

Table 6: Results for Case Study 6 using (DICOPT)

Initial Guesses $\left(x_1^0, x_2^0, x_3^0, y_1^0, y_2^0; f^0\right)$	**Optima** $\left(x_1^*, x_2^*, x_3^*, y_1^*, y_2^*; f^*\right)$
No initial Guess	$(3.5, 15.5, 0, 1, 0; 84.25)$
$(3.5, 15.5, 0, 0, 1; 80,)$	$(3.5, 15.5, 0, 1, 0; 84.25)$
$(0, 10, 0.5, 1, 1; 80)$	$(0, 10, 0.5, 1, 1; 80)$
$(3.5, 10, 1, 1, 1; 80)$	$(3.5, 15.5, 1, 1, 0; 85.25,)$

5. CONCLUSION

In this paper, CLP has been used to solve non convex MINLP problems by transforming them into master problem that are of pure IP by nature. The enumeration strength of CP can be suitably exploited to generate solution to IP problem which can subsequently corrected in sub-problems. The solution of the master problem is used to tighten bounds and add additional constraints so as to reduce the computational burden. This method has been successfully tested on two MILP and two non convex NLP and two non convex MINLP problems. The superiority of the proposed method lies in the fact that it does not require any initial guess as required in the traditional techniques. It has also been demonstrated that the traditional techniques are not very robust and yield different optima, depending on the initial values.

ACKNOWLEDGEMENTS

We gratefully acknowledge the critical comments on an earlier version of this paper by Dr. K.P Madhavan, Prof. Emeritus, IIT Bombay. We also would like to thank Prof. Mani Bhushan, IIT Bombay for his comments on the paper.

REFERENCES

Cardoso, M. F., Salcedo, R. L., Feyo de Azevedo, S. and Barbosa, D., 1997, A simulated annealing approach to the solution of MINLP problems. *Comp. Chem. Eng.*, **21**(12): 1349.

Darby-Dowman, K., Little, J., Mitra, G., & Zaffalon, M. (1997). Constraint logic programming and integer programming approaches and their collaboration in solving an assignment scheduling problem. *Constraints—An International Journal,* **1**, 245–264.

Floudas, C. A., Aggarwal, A., Ciric, A. R., 1989, Global sdhbsd optimum search for nonconvex NLP and MINLP problems. *Comp. Chem. Eng.*, 13(10): 1117.

Heipcke, S. (1999b). Comparing constraint programming and mathematical programming approaches to discrete optimisation—the change problem. *Journal of Operational Research Society,* **50**, 581–595.

Hooker, J. N. (2000). *Logic-based methods for optimization: Combining optimization and constraint satisfaction*. New York: Wiley.

Hooker, J. N., & Osorio, M. A. (1999). Mixed logical, linear programming. *Discrete Applied Mathematics,* **96–97**, 395–442.

Hooker, J. N., Ottosson, G., Thorsteinsson, E. S., & Kim, H. J. (1999). On integrating constraint propagation and linear programming for combinatorial optimization. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence* **(AAM-99)** 136–141.

Jain, V., & Grossman, I.E. (2001). Algorithms for hybrid MILP/ CP models for class of optimization problems. *INFORMS Journal of Computing*, **13**, 258-276

Kocis, G. R. and Grossmann, I. E., (1988), Global optimization of nonconvex mixed-integer nonlinear programming (MINLP) problems in process synthesis. *Ind. Eng. Chem. Res.*, **27**: 1407.

Lee, S., & Grossmann, I. E. (2001). A global optimization algorithm for nonconvex generalized disjunctive programming and applications to process systems. *Computers and Chemical Engineering,* **25**, 1675– 1697.

Munawar, S.A., Bhushan, M., Gudi, R.D. and Belliappa, A.M., 2003, Cyclic scheduling of continuous multi-product plants in a hybrid flowshop facility. *Ind. Eng. Chem. Res.*, **42,** 5861-5882.

Munawar, S.A. and Gudi, R.D., "A nonlinear transformation based hybrid evolutionary method for MINLP solution", [To appear in *Chemical Engineering Research and Design* (2005)]

Puget, J. -F. (1994). A C++ implementation of CLP. In *Proceedings of SPICIS'94*. Singapore.

Ryoo, H. S. and Sahinidis, B. P., 1995, Global optimization of nonconvex NLPs and MINLPs with application in process design. *Comp. Chem. Eng.*, **19,** 551.

Smith, B. M., Brailsford, S. C., Hubbard, P.-M., & Williams, H.-P. (1997). The progressive party problem: integer linear programming and constraint programming compared. *Constraints—An International Journal,* **1**, 119–138.

Tawarmalani, M., & Sahinidis, N. V. (2000). Global optimization of mixed integer nonlinear programs: A theoretical and computational study. *Mathematical Programming.* Ser. A, 99(3), 563-591, 2004

Van Hentenryck, P. (1989). *Constraint satisfaction in logic programming*. Cambridge, MA: MIT Press.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

HEURISTICS FOR CONTROL STRUCTURE DESIGN

A. Heidrich and J. O. Trierweiler

Group of Integration, Modeling, Simulation, Control and Optimization of Processes (GIMSCOP)
Department of Chemical Engineering, Federal University of Rio Grande do Sul (UFRGS)
Rua Luiz Englert, s/n CEP: 90.040-040 - Porto Alegre - RS - BRAZIL,
Fax: +55 51 3316 3277, Phone: +55 51 3316 4072
E-MAIL: alencar@mp.rs.gov.br, jorge@enq.ufrgs.br

Abstract: A heuristic approach is proposed to solve material balance control problems for chemical plants. The heuristics are derived from structural and controllability analysis and validated through simulation cases. First, the plant analysis is decomposed into two parts: a reaction section and a separation section. Second, these two sections are combined to evaluate the heuristic procedure implementation to solve a plant-wide control problem. Some control structure designs based on the proposed heuristic procedure are tested in the Tennessee Eastman Case Study.

Keywords: plant-wide control, control structure design, decentralised control, snowball effect, heuristics.

1. INTRODUCTION

For the case where the control structure is well selected, decentralised control is a useful strategy to perform control targets. The advantages to apply decentralised control instead of centralized ones are: easy implementation, low cost, reliability and comprehensive operation.

The problems of decentralised control application are derived from interactions of coupled process. Mass and energy integration usually increase the coupling among the process variables, what makes more difficult to apply decentralised controllers. Therefore it is crucial for the success of a decentralized control strategy to select the right manipulated and controlled variables.

In this paper, four guidelines are proposed to develop good control structures for process control using decentralised controllers. The main idea is to identify which variables and variable ratios should be kept constant to eliminate the effect of the process disturbances in the process quality automatically.

The paper is structured as follows: First, an isothermal reactor is analysed to show how the residence time and inlet flowrates can affect the component material balance. Similar analysis is made for separation process and for a small process with a reactor, separation unit and a recycle stream. These examples are used to developed the guidelines, which are finally applied to propose new control structures for the Tennessee Eastman Case Study (Downs and Vogel, 1993).

2. DEVELOPMENT OF THE GUIDELINES

In this section are shown the model applied to develop the IO-controllability analysis of a reactor and a separator. The heuristics proposed in this paper are derived from these simple cases.

2.1 Reactor

The reactor considered is an isothermal CSTR, with the first order kinetic A → B. The corresponding mathematical model is given by:

$$\frac{dV_R}{dt} = u_1 - u_2 \quad (1)$$

$$\frac{d(V_R C_A)}{dt} = u_1 C_{A0} - u_2 C_A - k V_R C_A \quad (2)$$

The variables used in the model (1) and (2) are summarized in Table 1.

Table 1: Variables description for the reactor model.

Variable	Symbol	Unit
Reactor volume	V_R	m^3
Reactant A inlet composition	C_{A0}	$kgmol/m^3$
Reactant A outlet composition	C_A	$kgmol/m^3$
Flowrates	u_1 and u_2	m^3/h
Kinetic constant	k	h^{-1}

The RGA matrix is an effective tool to evaluate variables coupling. The system is decoupling when the main diagonal of the RGA matrix is close to 1 (Skogestad and Postlethwaite, 1996). If the controlled variables of the reactor are V_R and C_A, we have a 2X2 system and RGA matrix can be expressed by its 1,1-element, λ_{11}. If we choice the control pairs u_1-C_A and u_2-V_R (control structure CS1) or u_1-V_R and u_2-C_A (control structure CS2), we will have distinct values for the λ_{11}. These values can be analytically attained from linearization of (1) and (2), resulting the expressions (3) and (4).

$$\lambda_{CS1} = 1 - \frac{1}{\tau_R}\frac{1}{s} \quad (3)$$

$$\lambda_{CS2} = \frac{1}{\tau_R}\frac{1}{s} \quad (4)$$

In (3) and (4), λ_{CS1} and λ_{CS2}, are λ_{11} calculated for CS1 and CS2, respectively. The τ_R is the reactor residence time and s is the Laplace domain variable. From (3) and (4), we can conclude two remarks:

- **Remark 1:** The λ_{11} depends on the reactor residence time. If τ_R is not constant, it can change λ_{11}, keeping away from 1.
- **Remark 2:** The control structure CS1 is more appropriated than CS2 for high frequency, since $\lambda_{CS1} \to 1$.

From remarks 1 and 2 we see that it is interesting to propose a control structure, which can keep the reactor residence time constant as possible and to select the inlet flow rates to control the reactor composition.

2.2 Separation Unit

A simple model of one stage separation unit can be used to show how it is possible to control material balance without composition controllers.

For an ideal binary mixture this model permits to write the equation (5).

$$(\alpha - 1)(1 - \psi)x^2 + [1 + (\alpha - 1)(\psi - z)]x - z = 0 \quad (5)$$

From (5), α is the relative volatility, ψ is the vaporisation fraction, x is the liquid phase composition of the light component and z is the light component composition into the feed flowrate. Considering that there are no feed composition changes and knowing that the relative volatility is almost constant, x can be inferred by ψ. In the other words, the component material balance can be controlled by a fix ratio V/F (inlet vapour flowrate/feed flowrates).

A dynamic model proposed in the literature for one stage unit separation (Luyben, 1990) was implemented to show the composition response under feed flowrate changes. Considering hypothetical conditions, it is possible to show this behaviour when: a) the liquid phase composition has a feedback controller and b) a feedforward control strategy is implemented to fix the ratio V/F. These results are shown in the Figures 1 and 2. It is easy to see that composition variability is lower when the ratio V/F is direct controlled.

This idea can be extended to separation columns where the ratios inlet vapour flow rate/feed flow rate and reflux flow rate/feed flow rate are fixed by a feedforward strategy. This application and its effect on the plant economics are explored by self-optimising techniques for controlled variables selection procedures (Skogestad, 2000).

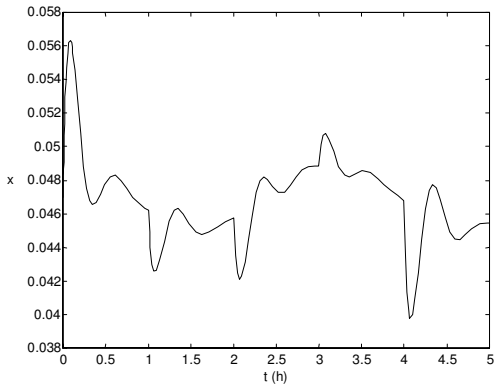


Fig. 1. Typical profile of the liquid phase composition change under flow rate disturbance when a feedback composition controller is used.

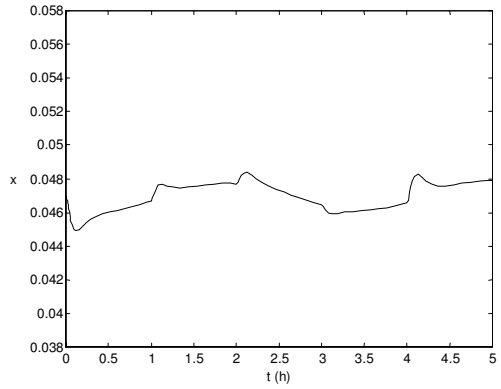


Fig. 2. Typical profile of the liquid phase composition change under flowrate disturbance when a feedforward strategy to fix the ratio *V*/*F* is applied.

A kind of the problem takes place when this control structure is implemented: under composition variations, it does not work. Solving this problem we can introduce a feedback composition controller through cascade configuration with feedforward strategy.

Base on these results two additional remarks can be written:

- **Remark 3:** We can implement feedforward control structures through constant flow rate ratios to control material balance under inlet flow rate changes.
- **Remark 4:** The feedback composition controllers can be introduced through cascade configuration with basic feedforward control structure.

2.3 Plant-Wide Control

In this section a hypothetical plant will be considered. The process flowsheet of this plant is shown in Figure 3.

The process is composed by liquid phase CSTR reactor, distillation column, and a recycle stream. The reactor is isothermal and the kinetic system is a single first order reaction, A → B. The process variables are described in the Figure 3.

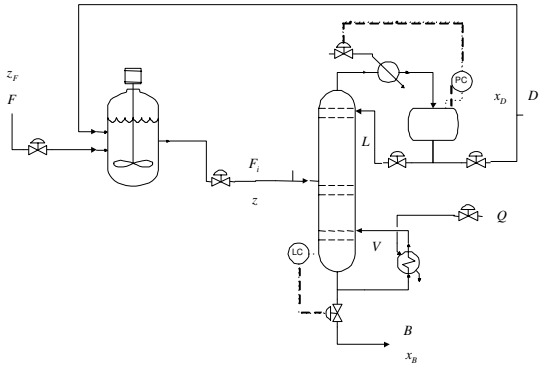


Fig. 3. Process flowsheet of the hypothetical plant.

In steady state we can write the material balance of this plant by equation (6).

$$F_i = \frac{F(x_D - x_B)kV_R}{kV_R x_D - F(z_F - x_B)} \tag{6}$$

The V_R in (6) is the reactor hold-up, expressed in molar base.

If the feed is the pure reactant A and considering an ideal separation, we have $z_F = 1$, $x_B = 0$ and $x_D = 1$. Thus, we can simplify (6), resulting (7).

$$F_i = \frac{FkV_R}{kV_R - F} = \left(\frac{k\tau_R}{k\tau_R - 1}\right)F \tag{7}$$

The expression (7) has been used to describe the snowball effect (Russel *et al*., 2002). The snowball effect is a usual behaviour of systems with recycle streams with constant reactor inventory (Luyben *et al.*, 1998). Figure 4 shows two typical curves from expression (7) when the process inlet flowrate is increased. This trends shows the behaviour of the recycle stream, F_i, when V_R or τ_R are fixed.

The high increasing of the recycle stream is verified when the reactor hold-up is controlled at its set-point. This means that high changes of the recycle streams are expected to keep controlled compositions under inlet flowrate variations.

The control of the reactor residence time is a way to minimize the snowball effect. Figure 4 shows a linear behaviour of the recycle stream when the feed flowrate is increased, when the residence time is constant, whereas increases exponentially when the inventory is maintained constant. Therefore, there are two advantages to maintain the residence time constant. First the outlet flow increase linearly and second the outlet composition suffer small variation. Base on equation (7) it can be easily concluded that to maintain residence time constant, we need just to keep the F/F_i ratio constant. In other words, if it is controlled the F/F_i ratio, it is possible to operate the plant without snowball effect and with automatic composition control. These conclusions agree with the remarks 1 and 3. Now we can apply the guidelines summarized in the remarks 1 to 4, to build a control structure for the hypothetical plant.

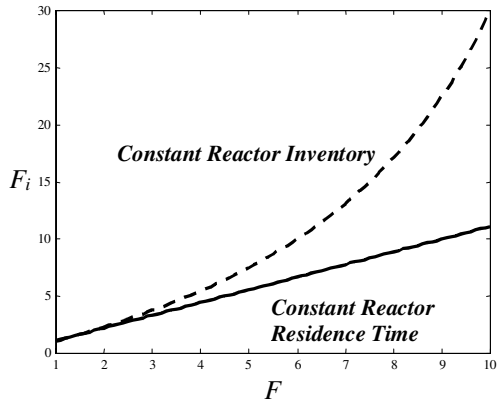


Fig. 4. Recycle stream behaviour.

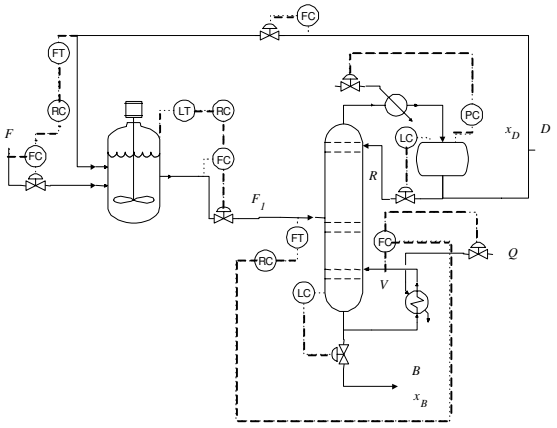


Fig. 5. CS1 flowsheet.

First, we introduce a feedforward ratio control to maintain reactor residence time constant. Second, flowrate ratios are implemented in the plant. The vapour flowrate to column/column feed flowrate and plant feed flowrate/recycle stream ratios are fixed through feedforward control. The recycle stream is chosen to set production rate. This control structure is called CS1 and it is shown in the Figure 5.

In Figures 7 and 8 the impurity composition of the column product shows a dynamic behavior when two kind of disturbances are applied: 10% production rate increasing and 10% feed composition increasing. We can see that CS1 has a good performance when the production rate is increased, but it does not work under composition disturbances. This problem is solved by introducing feedback composition controllers through a cascade configuration. This complete control structure, CS2, is shown in the Figure 6.

Based on all discussions made until here, we can write three relevant heuristics for plant wide control structure design:

- Use a control configuration (feedforward or feedback) to fix the reactor residence time. For this, the reactor hold-up (inventory) can change.
- Use a feedforward control configuration to fix flowrate ratios.
- The two heuristics above compose the basic design to control the inventory structure. The composition control is made through a cascade configuration

3. TENNESSEE EASTMAN PROBLEM

The Tennessee Eastman case study is described in (Downs and Vogel, 1993) and the process flowsheet is shown in Figures 9 and 11.

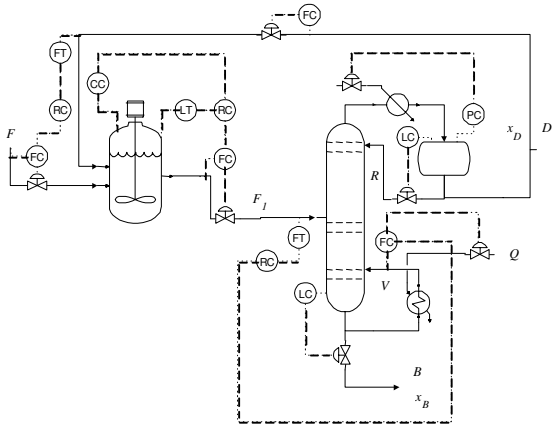


Fig. 6. CS2 flowsheet.

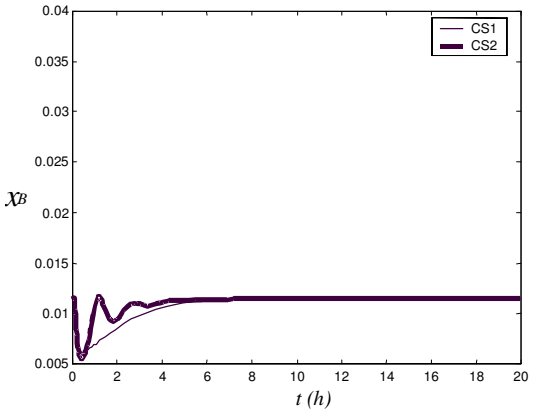


Fig. 7. Dynamic performances of control structures CS1 and CS2 when 10% of the production rate is increased.

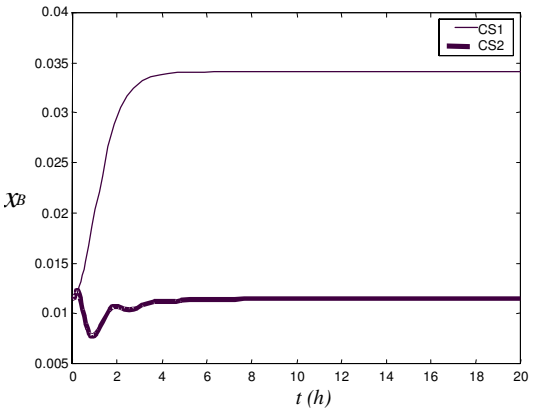


Fig. 8. Dynamic performances of control structures CS1 and CS2 when 10% of the feed composition is increased.

The kinetic system of the process is comprised by exothermic reactions given by the chemical reactions (8) to (11).

$$A(g) + C(g) + D(g) \longrightarrow G(liq) \quad (8)$$

$$A(g) + C(g) + E(g) \longrightarrow H(liq) \quad (9)$$

$$A(g) + E(g) \longrightarrow F(liq) \quad (10)$$

$$3D(g) \longrightarrow 2F(liq) \quad (11)$$

The main reactions are shown by (8) and (9). These reactions yield the main product constituted by G and H components, therefore the G/H ratio defines product quality. The reaction is gas-phase on the catalytic surface and the liquid-phase products are vaporized and partially condensed in a separator vessel. The gas-phase is compressed and recycled to reactor and the liquid-phase feeds the top of the separation column. A portion of the gas-recycle is purged to control inert inventory. This purge performs a meaningful loss of the plant.

The Tennessee Eastman Plant is an unstable system showing a RHP pole related to reactor temperature. The reactor is jacketed and cooling water cools the reactor content. Thus the low level of the reactor can be harmful to the system stability. When reactor level decreases below 40%, it takes place a high heat increasing (Farina et al., 2000) and below 10% the system exhibits a temperature runaway (Wu and Yu, 1997).

Next, we show two control structures applying the heuristics concepts above. Control performance is evaluated by variability of the product and reactor compositions ($\pm$5% is specified), constraints control and purge flow rate. The simulation examples apply 15% production rate increasing and model implementation is derived from Ricker's Tennessee Eastman Challenge Archive (http://depts.washington.edu/control/LARRY/TE/download.html).

3.1 Control Structure TE-CS1.

The control of the reactor residence time of the Tennessee Eastman plant is a not obvious task. Thus, it was implemented an indirect control applying an inventory-recycle ratio. The reactions take place in gas phase on catalytic surface, then the inventory considered to control must be the reactor vapor holdup. The ratio applied to perform this feedforward control is shown by Figure 9.

The flowrate ratios were applied as feed flow rates to recycle stream ratio. A similar structure with ratio control structure was proposed by Ricker (Ricker, 1996) where it was implemented for all flowrates a ratio control. The difference between the Ricker's control structure and the proposed here is the variable reactor inventory, which is used to keep the residence time constant.

The production rate is changed through recycle flowrate and other inventories are controlled by outlet flows. These alternatives have good io-controllability as shown by Farina et al., (2000).

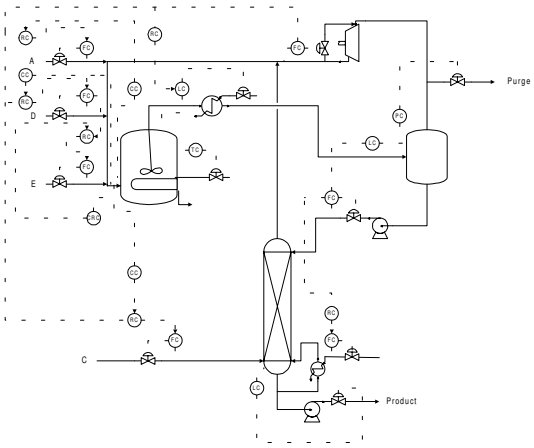
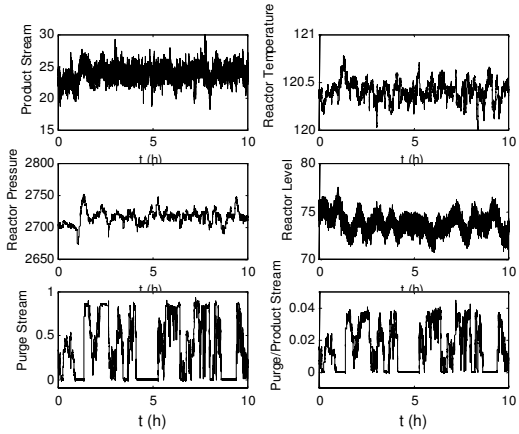
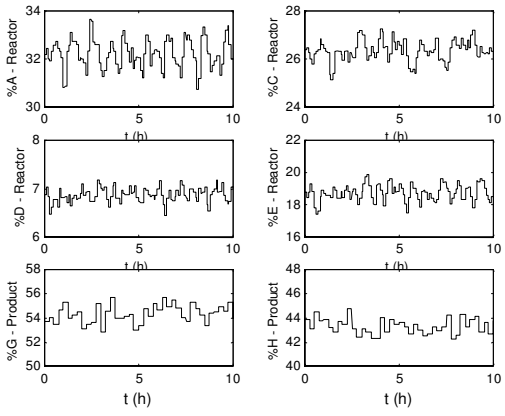


Fig. 9. Flowsheet of the control structure TE-CS1.

The feedback composition controllers are implemented through cascade configuration. The simulation results are shown in Figure 10.



(a)



(b)

Fig. 10. Simulation results of TE-CS1 for: (a) regulatory control and (b) composition control.

3.1 Control Structure TE-CS2.

In this CS, the reactor inventory is not directly controlled, since no reactor level control loop is used. The production rate is changed by separation drum effluent stream. Thus, the reactor inventory decreases when the plant production rate is increased. The control structure is sketched in Figure 11 and the simulation results are shown in the Figure 12. From

these results, there is a higher reactor level variation when we compare it with the first alternative (i.e., *TE-CS1*), but we can see a fast production rate change showing smooth variations.

From these control structures we can see that it is possible to attain the targets of the Tennessee Eastman Plant through energy stabilisation and material balance control, with an almost constant reactor residence time. These control structures also are reliable and comprehensive for the operation staff.

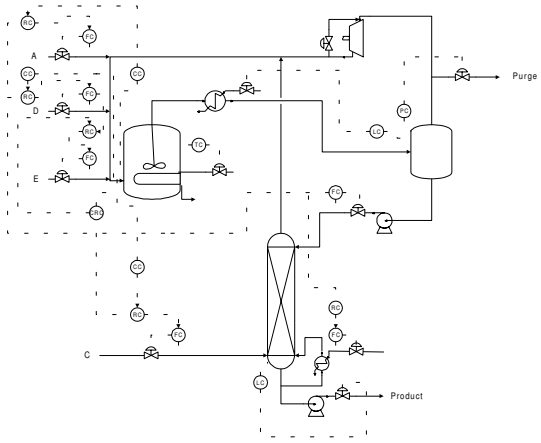
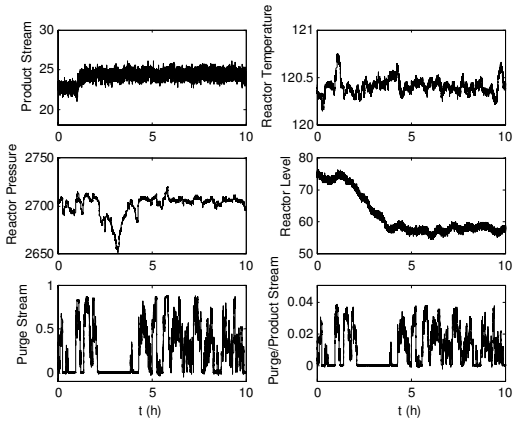
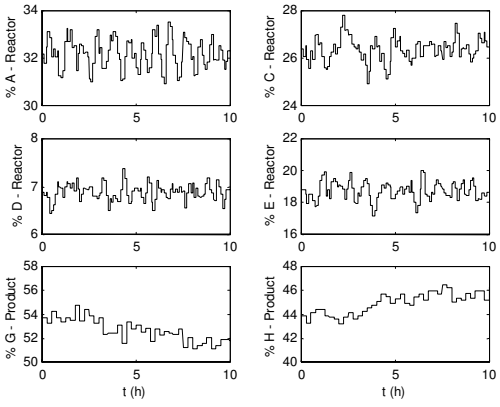


Fig. 11. Flowsheet of the control structure TE-CS2.



(a)



(b)

Fig. 12. Simulation results of TE-CS2 for: (a) regulatory control and (b) composition control.

4. CONCLUSION

This paper showed how it is possible to derive guidelines for control structure design based upon a model analysis of simple and general hypothetical plants. These heuristics, when they are applied in a real plant, produce very good control performance and are easy to apply. Finally, the proposed heuristic approach is used to develop two new control structures for the Tennessee Eastman Challenge Control Problem.

REFERENCES

Downs, J. J.; Vogel, E. F. (1993). A Plant-Wide Industrial Process Control Problem. *Computers & Chemical Engineering*, **v. 17**, p. 245-255.

Farina, L. A.; Trierweiler, J. O.; Secchi, A. R. (2000). A Systematic Comparison of Control Structures Proposed to Tennessee Eastman Benchmark Problem. *Preprints of ADCHEM*, p. 629-634.

Luyben, W. L.; Tyréus, B. D.; Luyben, M. L. (1998). *Plantwide Process Control*. McGraw-Hill.

Luyben, W. L. (1990). *Process Modeling, Simulation and Control for Chemical Engineers*. 2nd Edition, McGraw-Hill.

Ricker, N. L. (1996). Decentralized Control of the Tennessee Eastman Challenge Process. *Journal of Process Control*, **v. 6**, p. 205-221.

Ricker, N. L. (2004). Tennessee Eastman Challenge Archive. Available by: http://depts.washington.edu/control/LARRY/TE/download.html.

Russel, B. M.; Henriksen, J. P.; Jorgensen, S. B.; Gani, R. (2002). Integration of Design and Control through Model Analysis. *Computers & Chemical Engineering*, **v. 26**, p. 213-225.

Skogestad, S.; Postlethwaite, I. (1996). *Multivariable Feedback Control – Analysis and Design*. John Wiley & Sons.

Skogestad, S. (2000). Plantwide Control: The Search for the Self-Optimizing Control Structure. *Journal of Process Control*, **v. 10**, p. 457-507, 2000.

Trierweiler, J. O. (1997). *A Systematic Approach to Control Structure Design. Ph.D.* Thesis, Universität Dortmund.

Trierweiler, J. O.; Engell, S. (1997). The Robust Performance Number: A New Tool for Control Structure Design. *Computers & Chemical Engineering*, **v. 21**, p. 409-414.

Wu, K. L.; Yu, C. C. (1997). Operability for Process with Recycles: Interaction between Design and Operation with Application to the Tennessee Eastman Challenge Process. *Ind. Eng. Chem. Res.*, **v. 36**, p. 2239-2251.

Wu, K. L.; Yu, C. C.; Luyben, W. L.; Skogestad, S. (2002). Reactor/Separator Process with Recycles-2. Design for Composition Control. *Computers & Chemical Engineering*, **v. 27**, p. 401-421.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ALGORITHMS FOR REAL-TIME INTEGRATED OPTIMIZATION AND CONTROL: ONE LAYER APPROACH

Rezende, M. C. A. F., Maciel Filho, R., Costa, A. C., Rezende, R. A.

Chemical Engineering School, State University of Campinas (UNICAMP) Cidade Universitária Zeferino Vaz, CP 6066, CEP 13087-970 Campinas-SP, Brazil

Abstract: In this work, control and optimization algorithms are presented in order to be used in real-time process integration. The process considers as case study the o-cresol hydrogenation to obtain 2-methyl-cyclohexanol which is carried out in a three phase catalytic reactor. The real-time integration problem is postulate in one layer approach basis. Dynamic Matrix Control (DMC) is the control algorithm to be used and the optimization problem is solved by Genetic Algorithms. These algorithms showed to be efficient and robust to find out the optimal conditions and they can be used simultaneously to solve the problem in an one layer fashion.

Keywords: Real-time; Three-phase reactor; Control; Optimization; Genetic algorithm.

1. INTRODUCTION

Real-time process integration is an important task to feasible operation at high levels of operational performance. Conventional operational procedures, which require to have platforms specified by heuristic procedures and the controllers, which are tuned in a non-hierarchical way in relation to the operational specification, tend to fail for non-linear and complex systems. In this situation, the solution of the optimization problem coupled with the design and the controller tuning (Process Integration) can be carried out dynamically and at periods of time constrained by the process time constant (in real-time), becoming a necessary solution to reach the desired performance level.

The integration of chemical process in real-time is an interesting operation mechanism, because of the benefits that such approach may bring to the process in terms of profit and safety. Besides, it requires the development and application of several tools which may lead to a better understanding the steps and mechanisms taking place in the process.

One of the elements of real-time process integration is the advanced control, whose main function is to maintain the process in the desired set point, defined by an optimization strategy in the context of the real time process integration, and at the same time to avoid that the process variables violate their constraints. The controller, normally, is a linear multivariable predictive controller. Other element is the optimization algorithm used to solve simultaneously a large amount of equations and to send the optimal values (the set points) to the advanced controller.

The real-time optimization is mainly concerned with steady-state economics and uses nonlinear models while, at the advanced process control level, the objective functions used are not directly related to economics and the models employed are linear (Kookos, 2005).

An important point in real-time optimization is the necessity of a robust algorithm to converge to the optimal conditions, taking into account the need to obtain the response at relatively short time and with a lower computer burden.

Diehl et al. (2002) describe a new real-time and Nonlinear Model Predictive Control (NMPC) schemes based on the direct multiple shooting method. This approach shows CPU times in the range of a few seconds per optimization problem.

The real time optimization can be carried out through an optimization in one or two layers (Rezende et al., 2004). In the two layers approach, the low layer is responsible for the dynamic control. The high layer determines the optimum steady state for the process variables, which are used in the low layer as set points of controlled and manipulated variables (Zanin, 2001). In the one layer optimization approach, the problems of the control and economic optimization are solved simultaneously in a single algorithm (Tvrzská de Gouvêa, 1997).

In this work, it is addressed the problem of real-time integration for the multiphase reactor that produces 2-methyl-cyclohexanol, considering the most relevant aspects necessary to postulate and to solve the problem in a one layer fashion. A simple to use and easy to implement control structure, using SISO approach, for the desired product concentration is proposed and analyzed. For the optimization problem, responsible to generate the set points for the controllers, an algorithm is used coupled with the reactor rigorous model.

2. CASE STUDY

In this paper the o-cresol hydrogenation to obtain 2-methyl-cyclohexanol, which is carried out in a three phase catalyst slurry reactor, is considered as the representative of many important industrial processes. The hydrogenation of o-cresol is represented by:

$$3\,H_2(g) + C_6H_4OHCH_3(l) \rightarrow C_6H_{10}OHCH_3(l)$$

A scheme of a slurry reactor can be seen in Figure 1, for a co-current operating mode, in which the gas and liquid phase flow in the same direction of the solid catalyst slurry.

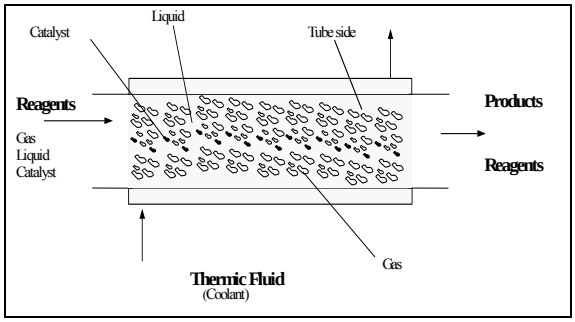


Fig. 1. A scheme of a three phase catalyst slurry reactor.

2.1 Mathematical Modeling.

The slurry type three phase system is represented by a non linear deterministic mathematical model developed by Vasco de Toledo and Maciel Filho (2004). This is a heterogeneous model constituted by material and energy balances for the three phases of the tube side, and the energy balance to the coolant fluid, with the following equations:

- Mass Balance for reactant A (hydrogen) in gas phase:

$$\varepsilon_g \frac{\partial A_g}{\partial t} = \frac{D_{eg}}{L^2}\frac{\partial^2 A_g}{\partial z^2} - \frac{u_g}{L}\frac{\partial A_g}{\partial z} - \left(K_{gl}\right)_A a_{gl}\,(A^* - A_l) \qquad (1)$$

Boundary Conditions:

$$\left.\frac{D_{eg}}{L}\frac{\partial A_g}{\partial z}\right|_{z=0} = u_g\left(A_g - A_{gf}\right) \qquad (2)$$

$$\left.\frac{\partial A_g}{\partial z}\right|_{z=1} = 0 \qquad (3)$$

- Mass Balance for reactant A (hydrogen) in liquid phase:

$$\varepsilon_l \frac{\partial A_l}{\partial t} = \frac{D_{el}}{L^2}\frac{\partial^2 A_l}{\partial z^2} - \frac{u_l}{L}\frac{\partial A_l}{\partial z} + \left(K_{gl}\right)_A a_{gl}\,(A^* - A_l) - \left(K_{ls}\right)_A a_{ls}\,(A_l - A_s^s) \qquad (4)$$

Boundary Conditions:

$$\left.\frac{D_{el}}{L}\frac{\partial A_l}{\partial z}\right|_{z=0} = u_l\left(A_l - A_{lf}\right) \qquad (5)$$

$$\left.\frac{\partial A_l}{\partial z}\right|_{z=1} = 0 \qquad (6)$$

- Mass Balance for reactant B (o-cresol) in liquid phase:

$$\varepsilon_l \frac{\partial B_l}{\partial t} = \frac{D_{el}}{L^2}\frac{\partial^2 B_l}{\partial z^2} - \frac{u_l}{L}\frac{\partial B_l}{\partial z} - \left(K_{ls}\right)_B a_{ls}\,(B_l - B_s^s) \qquad (7)$$

Boundary Conditions:

$$\left.\frac{D_{el}}{L}\frac{\partial B_l}{\partial z}\right|_{z=0} = u_l\left(B_l - B_{lf}\right) \qquad (8)$$

$$\left.\frac{\partial B_l}{\partial z}\right|_{z=1} = 0 \qquad (9)$$

- Energy Balance in the fluid phase:

$$\left(\varepsilon_g \rho_g Cp_g + \varepsilon_l \rho_l Cp_l\right)\frac{\partial T}{\partial t} = \frac{\left(\varepsilon_g \lambda_g + \varepsilon_l \lambda_l\right)}{L^2}\frac{\partial^2 T}{\partial z^2} - \frac{\left(\varepsilon_g \rho_g Cp_g u_g + \varepsilon_l \rho_l Cp_l u_l\right)}{L}\frac{\partial T}{\partial z} + h_s a_{ls}(T_s^s - T) - \frac{4U}{D_t}(T - T_r) \qquad (10)$$

Boundary Conditions:

$$\left.\frac{\left(\varepsilon_g \lambda_g + \varepsilon_l \lambda_l\right)}{L}\frac{\partial T}{\partial z}\right|_{z=0} = \left(\varepsilon_g \rho_g Cp_g u_g + \varepsilon_l \rho_l Cp_l u_l\right)\left(T - T_f\right) \qquad (11)$$

$$\left.\frac{\partial T}{\partial z}\right|_{z=1} = 0 \qquad (12)$$

- Energy Balance for the coolant:

$$\rho_r Cp_r \frac{\partial T_r}{\partial t} = -\frac{\rho_r Cp_r u_r}{L}\frac{\partial T_r}{\partial z} + \frac{4U}{D_t}(T - T_r) \quad (13)$$

Boundary Conditions:

$$T_r = T_{rf}, \qquad z = 0 \quad (14)$$

- Mass Balance for reactant A (hydrogen) in solid phase:

$$(1-\varepsilon)\varepsilon_s \frac{\partial A_s}{\partial t} = (K_{ls})_A a_{ls}(A_l - A_s) - \frac{(1-\varepsilon)\rho_s}{A_{ref}} R_W(A_s, B_s, T_s) \quad (15)$$

- Mass Balance for reactant B (o-cresol) in solid phase:

$$(1-\varepsilon)\varepsilon_s \frac{\partial B_s}{\partial t} = (K_{ls})_B a_{ls}(B_l - B_s) - \frac{\nu(1-\varepsilon)\rho_s}{B_{ref}} R_W(A_s, B_s, T_s) \quad (16)$$

- Energy Balance in solid phase:

$$(1-\varepsilon)\rho_s Cp_s \frac{\partial T_s}{\partial t} = h_s a_{ls}(T_s - T) + \frac{(1-\varepsilon)\rho_s(-\Delta H_R)}{T_{ref}} R_W(A_s, B_s, T_s) \quad (17)$$

where

a	surface area, m^{-1}
A	concentration of hydrogen, $kmol/m^3$
*A**	solubility of the component A, $kmol/m^3$
B	concentration of o-cresol, $kmol/m^3$
De	effective diffusivity, m^2/s
D_t	reactor diameter, m
h	heat transfer coefficient, kJ/m^2 s
K	mass transfer coefficient between the phases, cm/s
L	reactor length, m
Rp	radius particle, m
R_w	reaction rate, kmol/kg catalysts.s
T	absolute temperature
u	linear velocity, m/s
U	global heat transfer coefficient, kJ/m^2 s
ΔH_R	heat of reaction, kJ/kmol
ε	porosity
λ	heat conductivity, kJ/m s K
ρ	density, kg/m^3
ν	stoichiometric coefficient

Subscripts:

A	component A (hydrogen)
B	component B (o-cresol)
f	feed
g	gas phase
gl	gas-liquid
l	liquid phase
ls	liquid-solid
p	particle
r	coolant fluid
ref	reference value used to turn the equations dimensionless
s	solid

Superscripts:

s	catalyst surface

The model equations lead to a system of partial differential equations that are converted into a system of ordinary differential equations through discretization by orthogonal collocation. The resulting equations were integrated using DASSL software which is suitable for stiff systems.

3. ALGORITHMS FOR REAL-TIME INTEGRATED OPTIMIZATION AND CONTROL

3.1 Optimization procedure.

The real time optimization causes a lower profit when a local optimizer is used. This problem can be averted by using a global optimizer in the real time optimization procedure (Lacks, 2003). In this work, a global optimizer, based on the Genetic Algorithms, is chosen and proposed to be used in a posterior Real Time Integration.

The aim of the optimization is to find out an optimal steady state of the three phase reactor that produces 2-methyl-cyclohexanol using a rigorous model of this process. The optimal condition is the new set point in which, the reactor should be operated to have higher performance.

The objective function considered was the productivity of 2-methyl-cyclohexanol subjected to the conversion of o-cresol in the liquid phase, which is usually required for practical implementations:

$$\text{Productivity} = \frac{(B_{lf} - B_l) * u_l}{L} \quad (18)$$

As constraint, it was considered the operation of the unit under limits of o-cresol conversion larger than 90%, since the environment constraints require such levels of conversion.

$$\text{Conversion} = \frac{B_{lf} - B_l}{B_{lf}} > 0.90 \quad (19)$$

In order to proceed with the reactor optimization, the mathematical model (Eqs. (1-17)) is incorporated to the Genetic Algorithm code and the objective function is defined. The optimization is carried out in steady state so that the reactor is represented by equations 1 to 17 setting the derivatives in respect to time to zero.
The optimization problem can be written as:

Maximize: Productivity
Subject to: Model equations (Eqs. (1-17))
Conversion > 90%
$0.004195 \le u_l \le 0.011805$
$0.000608 \le A_{gf} \le 0.002392$
$0.009732 \le B_{lf} \le 0.038268$
$459.0 \le T_f \le 621$
$425.0 \le T_{rf} \le 575.0$
$1.08 \le u_g \le 252.0$
$0.003 \le u_r \le 0.007$
$0.00075 \le A_{gf} \le 0.00225$

In this work, the GA code used is the Fortran Genetic Algorithm Driver by David Carroll, version 1.7a (Carroll, 2004), with some modifications. This is a binary code that starts with a random population of chromosomes that are a set of solutions to the optimization problem. Each solution is evaluated by the fitness function that associates a value to the solution, determining the best ones. In this point, the genetic operators, that are the kernels of Genetic Algorithms, responsible to promote the evolution of the solutions are applied (Wang, 2005). This procedure is repeated along the iterations, also called generations, until a termination criterion is satisfied.

In this work the termination criterion is given by the number of generations, since the solutions are better along the generations. During the optimization, it was set the population size, the crossover and mutation probabilities, the maximum number of generations and number of children per pair of parents. In this code is possible to set elitism, niching and micro-GA techniques. The details of the tools can be found in Deb (1999). As genetic operators are used tournament selection, single-point crossover and jump and creep mutation. It was set elitism, two children per pair of parents and niching. In order to handle the constraint present at the reactor optimization problem, the constraint handling method proposed by Deb (2000) is incorporated to the Carroll's code. This method exploits the feature of the GAs algorithm of pairwise comparison in tournament selection (Deb, 2000; Costa and Maciel Filho, 2005).

Extensive simulations lead to the convergence to the optimal conditions. The results obtained with the optimization by GA are the productivity of 2-methyl-cyclohexanol of 1.64×10^{-4} Kmol/m^{3}s and the conversion of o-cresol of 90.18%. It shows an improvement of three times the productivity of 2-methyl-cyclohexanol and twice the conversion, compared to the steady state at previous work using the same model (Rezende et al., 2004).
Bearing in mind the high dimensionality and non linearity of the model, the genetic algorithm showed to be robust to converge to the optimal conditions.
before a main or secondary heading.

3.2 Control Procedure

For the layer of the control, a predictive controller based on the Dynamic Matrix Control (DMC), is implemented. The DMC algorithm development for monovariables systems (SISO) can be found at Rezende et al. (2004).
DMC makes use of a linear model, the convolution model, which is obtained through step disturbances in the input variables. In this work is presented the DMC algorithm developed to monovariable systems SISO, since a reliable control strategy was defined.
The DMC algorithm is based on the calculation of NC (Control Horizon) future values of the manipulated variables from a minimization of NP (Prediction Horizon) future values of the square of the difference between set point and output predicted by a convolution model with NM (Model Horizon) output values obtained from the step response to the manipulated variable.
The model horizon (NM), the prediction horizon (NP), the control horizon (NC) and the suppression factor (f) are parameters to be tunned in order to obtain a good performance of the controller.
In order to verify the controller performance, several sets of controller parameters were tested for load disturbances. The objective of the controller is to reach a suitable control for the o-cresol concentration at the reactor exit, able to work in a relatively wide range of operation conditions specified by the optimization algorithm.

Control of the o-cresol concentration at the reactor exit. The simulations to study the o-cresol concentration control at the reactor exit considered disturbances of ± 5% in the manipulated variable, step disturbance of ± 5% and alteration of the ± 5% in the set point. The set of operating variables is: ul = 0.0096 m/s, Agf = 0.001875 Kmol/m^3, Blf = 0.0300 Kmol/m^3, Tf = 648 K e Trf = 600 K. The required set point of the output reactor o-cresol concentration in the liquid phase is 0.00258 Kmol/m^3. A different set of parameters was tested in order to find out a set of parameters that allows for a good performance of the controller. This set of parameters is: NM = 4, NP = 3, NC = 1, and f = 0.0001. On-line concentration measurement can be obtained by near-infrared measurement with a good and robust performance in industrial environment. A sampling time of 100 s is used in this work, as this is the value normally found in industrial practice.
The controlled and manipulated variables profiles, observed in the Figures 2 and 3, show that the performance of the controller is satisfactory to this set of operating variables.
Figure 2 shows that the controlled variable reaches the set point around 1500s and remains on it along the period; the overshoot is small and oscillations were not observed.

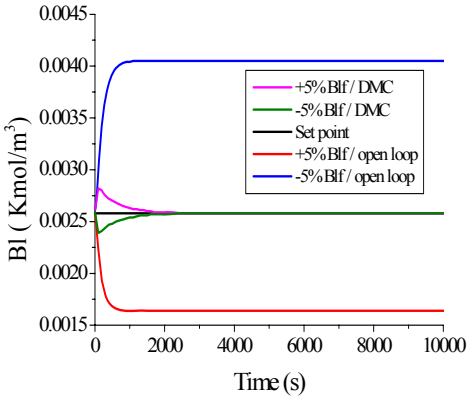


Fig. 2. Open and closed loop concentration response for disturbances of ± 5% in Blf.

Figure 3 shows that the manipulated variable reaches a steady state in approximately 2000s and does not present oscillations along the period. These results show the very good performance and robustness of the proposed control structure, for a relative large range of operational conditions.

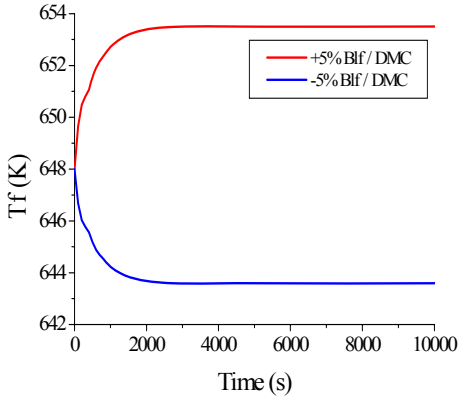


Fig. 3. Profile of manipulated variable (Feed temperature).

4. REAL TIME INTEGRATED OPTIMIZATION AND CONTROL

In order to proceed with the real-time integration, a robust and efficient algorithm to find out the optimal conditions in a relatively short time is required. This is essential especially for the one layer approach since the controller action depends on the convergence of the whole optimization/control problem. Figure 4 shows a scheme of the one layer approach.

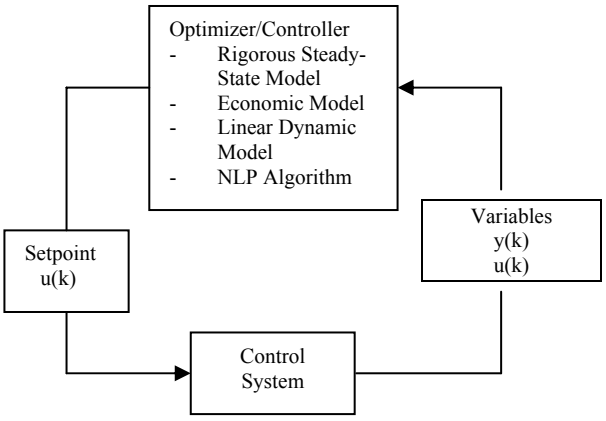


Fig. 4. Schematic diagram for the one layer approach.

The mathematical formulation of the control/optimization problem is described bellow (Zanin, 2001):

$$\min_{x_s,u_s,\Delta u(jT_a);j=1,\ldots NC} \sum_{i=1}^{NP} \left\| W_1\left(y_p(iT_a)-y_{sp}\right)\right\|_2^2 + \sum_{j=1}^{NC} \left\| W_2\Delta u(jT_a)\right\|_2^2 + W_3 f_{eco} \quad (20)$$

Subject to the following constraints:

$$h_p(x_s,u_s,d_s)=0 \quad (21)$$

$$h_e(f_{eco},x_s,u_s,d_s)=0 \quad (22)$$

$$u_s^{inf} \le u_s \le u_s^{sup} \quad (23)$$

$$x_s^{inf} \le x_s \le x_s^{sup} \quad (24)$$

$$-\Delta u^{max}(jT_a) \le \Delta u(jT_a) \le \Delta u^{max}(jT_a) \quad j=1,\ldots,NC \quad (25)$$

$$u^{inf}(jT_a) \le u_{at} + \sum_{i=1}^{j} \Delta u(iT_a) \le u^{sup}(jT_a) \quad j=1,\ldots,NC-1 \quad (26)$$

$$u_s = u_{at} + \sum_{j=1}^{NC} \Delta u(jT_a) \quad (27)$$

$$y_p(iT_a) = y_{pf}(iT_a) + \sum_{j=1}^{\min(NP,NC)} a_{dmc}((i-j+1)T_a)\Delta u(jT_a) \quad j=1,\ldots,NP \quad (28)$$

where:

a_{dmc}	matrix of the coefficients of the process linear model
ds	vector of the perturbations at the steady-state
f	weighting factor in the DMC algorithm
f_{eco}	economic objective function
h_e	economic model constraints
h_p	nonlinear model constraints
NC, NP	control horizon and prediction horizon
NM	model horizon
T_a	sampling period
u	vector of the manipulated variables
u_{at}	vector of the manipulated variables at the current time
u_s	vector of the manipulated variables at steady-state
x_s	vector of the nonlinear model variables in the steady-state
W_1	diagonal matrix of the weight of the dinamically controlled variables
W_2	diagonal matrix of the supression factor of the manipulated variables
W_3	weight of the economic parcel in the objective function
y_p	vector of the linear prediction of the dynamically controlled variables
y_{pf}	vector of the linear predicition of the dynamically controlled variables, based on passed control actions
y_{sp}	vector of the set points of the dynamically controlled variables
Δu	vector of the amplitude of the control actions

The challenge in the implementation of the optimizer is to be able to work with an integrated optimization and control procedure in a real time basis.

The process considered in this work is characterized by high dimensionality and non linearity of the model. Because of this, the solution of the optimization problem may bring difficulties related to the convergence. Optimization algorithms based on

the GA principles are robust and care has to be taken in the required computer time (Costa and Maciel Filho, 2005). For this particular problem the GA presented quite good performance with a reasonable computer time, around ten minutes in a Pentium 4, 2.8GHz, 512 MB RAM. It is an important aspect in real time applications of the GA optimization technique coupled with the high nonlinear and multivariable model, making this approach to be a good candidate to real time process integration.

5. CONCLUSION

This work presents the steps necessary to carry out the real-time process integration for the multiphase reactor that produces 2-methyl-cyclohexanol, through a one layer approach. The objective function of the control/optimization problem is composed of parcels from the dynamic control and the economic optimization. As controller of the process, it was used the DMC that showed to be efficient to control the reactor. The optimization problem was solved by Genetic Algorithm that revealed to be robust enough to lead to the convergence to the optimal conditions with a short computational effort. The results showed that the both techniques can be used simultaneously to deal with one layer real time integration.

REFERENCES

Carroll, D. (2004). Carroll's FORTRAN Genetic Algorithm Driver. In http://cuaerospace.com/carroll/ga.html accessed in November 16th , 2004.

Costa, C. B. B. and R. Maciel Filho (2005). Evaluation of optimisation techniques and control variable formulations for a batch cooling crystallization process. *Chemical Engineering Science*, **60**, 5312-5322.

Deb, K. (2000). An efficient constraint handling method for genetic algorithms. *Computer Methods in Applied Mechanics Engineering*, **186**, 311-338.

Deb, K. (1999). An introduction to Genetic Algorithms. *Sadhana-Academy Proceedings in Engineering Science*, **24**, 293-315.

Diehl, M., H. George Bock, J. P. Schloder, R. Findeinsen, Z. Nagy, and F. Allgower (2002). Real-time optimization and nonlinear model predictive control of process governed by differential-algebraic equations *Journal of Process Control*, **12**. 577-585.

Kookos, J. K (2005). Real-Time Regulatory Control Structure Selection Based on Economics. *Ind. Eng. Chem. Res.*, **44**, 3993-4000.

Rezende, M. C. A. F., A. C Costa. and R. Maciel Filho (2004). Control and Optimization of a Three Phase Industrial Hydrogenation Reactor *International Journal of Chemical Reactor Engineering*, **2**, A21.

Tvrzská de Gouvêa, M. T (1997). Uso de um algoritmo SQP na otimização de processos químicos em tempo real. São Paulo, SP. University of São Paulo. *Ph.D. Thesis*.

Vasco de Toledo, E. C. and R. Maciel Filho (2004). Detailed Deterministic Dynamics Models for Computer Aided Design of Multiphase Slurry Catalytic Reactor. *European Symposium on Computer-Aided Process Engineering 14*, **18**, 823-828.

Wang, Y. Z (2005). A GA-based methodology to determine an optimal curriculum for schools. *Expert Systems with Applications*, **28**, 163-174.

Zanin, A. C (2001). Implementação industrial de um otimizador em tempo real. São Paulo, SP. University of São Paulo. *Ph.D. Thesis*.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

STEAM AND POWER OPTIMIZATION IN A PETROCHEMICAL INDUSTRY

Eduardo G. de Magalhães[1], Tiago Fronza[2] , Keiko Wada[2], Argimiro R. Secchi[2]

[1]*Process Team – Engineer Unity - COPESUL*
[2]*Group of Integration, Modeling, Simulation, Control and Optimization of Processes (GIMSCOP)*
Departamento de Engenharia Química - Universidade Federal do Rio Grande do Sul

Abstract: The rational use of utilities (electric energy, steam, and water) represents nowadays the great challenge to assure the competitiveness and sustainability of industries. The proposed work presents the minimization of the cost of steam and power generation in a petrochemical company, which has production of electric energy and steam by co-generation system. In this case, there is a balance to achieve between work and heat supply and this cannot be readily defined by heuristics or localized control loops. The application of an optimization model is proposed based on needs of energy demands, and readily exposing the scenery that minimizes the power and high pressure steam level production. It was observed a potential of economy of 46 t/h in the generation of steam in boilers and could be achieved a reduction of 6 MW of electric power consumption.

Keywords: Steam, Power, Optimization, Turbines, Mixed Integer Programming, Utilities.

1. INTRODUCTION

The optimization of utility system has been explored regarding conceptual graphic tools that allows an steam network analysis and offer a better understanding of the interactions and could accelerate the application of an algorithm method (Strouvalis et al., 1998); aiding the decision of when is convenient to a factory to generate energy with an existent co-generation system or buy outer energy and heat, using MILP routine (Bojic & Stojanovic, 1998); helping the management of energy in a multi-period basis regarding a three/four level steam network; handling annual budging planning, investment decisions, electricity contract optimization, shutdown maintenance scheduling and fuel/water balance problems in a petrochemical plant with a site-model (Hirata et al., 2004); achieving benefits from an complex refinery co-generation system avoiding loss of energy in letdown valves and helping energy management problems basically using a solver tool from a common commercial spreadsheet (Milosevic & Ponhöfer, 1997).

The proposed work presents the minimization of the steam and power generation in a petrochemical company, which has a production of electric energy and steam by co-generation system. In the case study, the steam network is composed by four levels of pressure that supply either thermal (heat exchangers), process (strippers) or power (pumps, compressors and electric energy) demands. The application of an optimization model is proposed in a way that based on the definition of needs of electric energy generation, process loads and steam heat or separation demands, it can readily expose the scenery that minimizes the power and high pressure steam level production.

2. MOTIVATION AND VIABILITY

A petrochemical industry and its associated second generation industries in a petrochemical site consume steam in various areas of their productive processes. These applications can be related to machine drives, stream heating, or separation processes (strippers, etc). These applications also demands different temperature and pressure conditions, needing to operate with four steam pressure levels, such as Super High Pressure Steam (VS - 113 kgf/cm^{2}g and 525°C); High Pressure Steam (VA - 42 kgf/cm^{2}g and 400°C); Medium Pressure Steam (VM - 18 kgf/cm^{2}g and 315°C) Low Pressure Steam (VB - 4.5 kgf/cm^{2}g and 225°C). In the case study of this work, the petrochemical industry generates VS in the Process Unities furnaces (70% in mass) and in the Utility

Unity Boilers (30% in mass). Power is produced by two steam turbo generators and a heavy duty gas turbine – power can also be purchased from the off-site supplier. The normal production of VS is about 1150 t/h and the other levels of steam are produced by the extraction of turbines that generate work with the feed of VS and also by pressure letdown valves with desuperheater systems to complement the needs of steam in the headers. The pressures in the VA and VM headers are controlled by acting in the relation of extracted and exhausted of the machines or by letdown valves, and this control does not necessarily generate optimized scenery.

The optimization means the minimization of cost of energy, which is defined here as the sum of cost of VS produced in the auxiliary boilers, cost of power produced or imported and cost of using letdown valves. Nevertheless, the optimization of a steam and power system in Rankine cycle with such dimension and complexity is not an easy problem, because of the several and different applications involved and the connections of the pressure levels (Milosevic & Ponhöfer, 1997; Eastwood & Bealing, 2003). The potentiality in optimization is apparently huge, due to dimension of the scale of production in a petrochemical company (industry of intensive capital) and due to the continuous regime of production. Depending on the model, an annual economy of 2 to 5% of the energy bill could be achieved, besides the environmental advantages of reduction of emissions and withdrawn of superficial water (river).

For steam, the most basic procedure used to administer the commitment between the demands of several steam levels and the generation of VS is the relationship of extraction and condensation in the turbo-machines (huge process compressors and turbo-generators). This is done to increase the readiness of VA or VM (in agreement with each machine) by extraction or to use all the useful energy accomplishing work (by expanding VS to exhaust steam), without extracting smaller pressure steam whose low demand would cause need of steam relief. This extraction and condensation relationship is not free – there is a balance among these steam rates for a given load (electrical or mechanical) demanded by each machine. Within this relationship and regarding the operational and mechanical limits of the system and their equipments, however, it could be achieved an optimized distribution for each scenery of steam demands for production of energy (electric, thermal or work). Other optimization form is to alternate the operation among different drives from same equipments (for example, pumps with electric motor and steam turbines drives). The total optimization of the system, however, is not the target of the current control loop of the steam system and neither is possible of being achieved by the operation people in a practical and fast way. However, the application of a computational tool that can show the best scenery (smaller cost of generation of VS and power, avoiding use of pressure letdown valves and the use of relieves) is very useful and can be implemented from definitions of each scenery inputs. The implementation of this tool is proposed in this paper, with optimization of a real scenery as an example.

3. METHODOLOGY

The method involves two main tools: a Steam Model and an Optimization Model. The first is necessary to collect the non-freedom degrees, process dependant variables and operational definitions – this is made importing data from the DCS system and by some manual inputs. This tool was constructed in a commercial available spreadsheet, in the most rigorous form possible, in order to minimize balance errors. Even if some steam measurement is not available in DCS, it was tried to evaluate this value indirectly, with mass and/or thermal balances, suppliers' performance curves, project data, and so on. High precision balance is the key to minimize errors in the data to be exported to the optimization model. This data has to be in a form in which is included eventual errors, to guarantee that the steam balance is closed.

The second tool imports data from the Steam Model and performs the optimization from a constructed model of the steam and power system, equipments and network constraints. The results can be applied as a guideline for engineers and operators in day-to-day routines or as a project tool, showing best configurations for alternative selections.

The "real scenery" refers to a specific state of the steam network of the petrochemical industry, selected in a random way. In the case study, it refers to the situation of June 6 (2005), 06:00 PM, when it was being generated about 1205 t/h of VS and it was observed openings in VS/VA pressure letdown valves (42.5 t/h), VA/VM (50 t/h) and VM/VB (62.4 t/h). This situation can be observed in Figure 1. Some steam consumers do not possess flow measurement, and this is the case of most of the steam flows of VB level; nevertheless, there are in the company evaluations of these normal daily demands, and the application of these values leads to a balance that reflects reasonably the situation. So, using the available data, the steam material balance was defined. So, the modelling of the steam network is made respecting the thermal demands and the requested power of the machines in this day and time. The data exported from the model then will be submitted to an appropriate optimization routine and the result will be compared with the steam balance observed in real conditions, as a way of verifying the potential earnings.

4. DEFINITIONS

Generating equipments: these are the steam sources of the several existent steam levels and they can be variable or fixed. Some of these sources are also consumers of steam of higher level, generating by extraction a lower class steam. Fixed steam sources

are related to equipment, in which the steam consumption is fixed and dependent on the process loads, but also produce steam of smaller pressure in the outlet. Then, according to Figure 1:

- VS is generated by the process furnaces (fixed generation) and auxiliary boilers (variable generation);
- VA can be generated by the turbines 12-TBC-01/21, 47-TG-01/02, 112-TBC-01 and by the letdown valves 10-PV-51 and 46-PV-12 - variable generation;
- VM can be generated by the turbines 14-TBC-01/21, 112-TBC-01 and for the letdown valves 10-PV-52 and 46-PV-13 (variable generation), as well as by other fixed generations (as example: 14-TBC-02/22, 48-B-01 B/C/D);
- VB can be generated by the letdown valves 10-PV-13, 110-PV-04 and 46-PV-14 (variable generation), as well as by other fixed generations (as example: 12-TBB-11, 114-TBC-01);

Consumers: these represent the several steam levels demands. These demands can be:

- Thermal: Heating of another fluid with steam. As the steam leaves the system definitively (as condensate or exhaust steam), these are not considered steam generator equipments;
- Process: Steam injection directly in other equipments (as strippers, ejectors). As the steam leaves the system definitively, these consumers are not steam generators equipments;
- Power: the power consumers can also be steam generator equipments, when extracted steam is produced.

Letdown Pressure Valves: these are control valves that, allied with desuperheater systems, have the function of adjust the pressure and temperature of some steam level, sending excesses to the lower level or supplying the next lower level in order to increase its pressure. The use of letdowns reduces the efficiency of the system and should be avoided.

Relief Valves: these are existent control valves in the levels of VS and VB that are used to limit the maximum pressure of these headers, discharging steam to atmosphere.

External Clients: these are all the others industries that surround the petrochemical company and consume utilities (in this case, steam) produced in the company. External clients are considered fixed consumers – the steam is process dependant and the steam leaves the company system definitively.

Table 1 presents the syntax of the abovementioned groups.

Variable Class	Description
VS	Super High Pressure Steam Flow (t/h)
VA	High Pressure Steam Flow (t/h)
VM	Medium Pressure Steam Flow (t/h)
VB	Low Pressure Steam Flow (t/h)
V	Steam Flow (t/h)
CV	Vacuum Steam Flow (t/h)
CM	Medium Pressure Steam Condensate Flow (t/h)
CB	Low Pressure Steam Condensate Flow (t/h)
PO	Power (MW)
AD	Desuperheating Water (steam temperature control)
CO	Cost (R$)
z	Binary variable for switchable drivers

Indices	Equipments Groups Description
h	VS generators
i	VA generators
j	VM generators
k	VB generators
l	CV generators
o	VS power consumers
p	VA power consumers
q	VM power consumers
r	VB process & heat consumers and exports
t	VS relieves
u	VB relieves
oc	VS process & heat consumers and exports
pc	VA process & heat consumers and exports
qc	VM process & heat consumers and exports
ps	Power Sources (internal and external)
ms	Motor of Switchable drivers equipment
ts	Turbine of Switchable drivers equipment
ld	Letdown valves (VS/VA, VA/VM, VM/VB)

Table 1: Syntax definitions for the variables applied in the model optimization.

4. OPTIMIZATION MODEL FORMULATION

As mentioned before, the objective function to be minimized is cost of energy. This is defined in the form:

$$\text{Min}\left(\sum_{h=1}^{Nh} CO_h \times VS_h + \sum_{ps=1}^{Nps} CO_{ps} \times PO_{ps} + \sum_{ld=1}^{Nld} CO_{ld} \times V_{ld}\right)$$

The equalities constrains are defined by the steam header balances, turbines model equations and power balances.

- Material Balance in the Control Envelope (Company Steam Network):

$$\sum_{h=1}^{Nger} VS_h = \sum_{l=1}^{Nger} CV_l + \sum_{pc=1}^{Ncons} VA_{pc} + \sum_{qc=1}^{Ncons} VM_{qc} + \sum_{r=1}^{Ncons} VB_r + \sum_{oc=1}^{Ncons} VS_{oc} + \\ + \sum_{t=1}^{Naliv} VS_t + \sum_{u=1}^{Naliv} VB_u$$

[VS Generations] = [Steam Condensations] + [steam exportation] + [steam injections in processes] + [losses] + [relieves]

- Material Balance in each Steam Header:

$$\sum_{h=1}^{Nger} VS_h = \sum_{o=1}^{Ncons} VS_o + \sum_{oc=1}^{Ncons} VS_{oc} + \sum_{t=1}^{Naliv} VM_t$$

$$\sum_{i=1}^{Nger} VA_i = \sum_{p=1}^{Ncons} VA_p + \sum_{pc=1}^{Ncons} VA_{pc}$$

$$\sum_{j=1}^{Nger} VM_j = \sum_{q=1}^{Ncons} VM_q + \sum_{qc=1}^{Ncons} VM_{qc}$$

$$\sum_{k=1}^{Nger} VB_k = \sum_{r=1}^{Ncons} VB_r + \sum_{u=1}^{Naliv} VB_u$$

- Material Balance and Performance Curve for Two-Stage Turbines. For a generic extraction-condensation turbine:

Material Balance:
$VS_{turbine} = VA_{turbine} + CV_{turbine}$

Performance Equation, for given rotation and power:
$VA_{turbine} = a.VS_{turbine} + b$

Each degree of freedom turbine was modeled from real data, defining the parameters ‘a’ and ‘b’.

- Power Balance:

$$PO = \sum_{ps=1}^{Nps} PO_{ps} =$$

= [Process Dependant POWER] + [Switchable drivers motor POWER]

Two of the power sources (ps) available are turbine generators, which are also modeled to be optimized and the power variable is a second degree of freedom.

- Additional Material Balances: The optimization model also consider material balances in steam letdown valves and condensate flash drums.

- Inequality Constrains: Establish physical conditions and project or operation limits. This applies to turbines, valves and headers. For example, in generic turbine and VS/VA letdown valve:

$0 < VS_{turbine} < 195$
$5 < VS/VA_{-valve} < 310$

- Binary Variables: With only 1 or 0 value, this variable is used to permit selection between available drives for same equipment. Thus, it is possible to optimize the steam and power situation. The balances involved are:

a. Power consumed by motors of switchable drives equipment:

$$\text{[Switchable drivers motor POWER]} = \sum_{ms=1}^{Nms} z_{ms} PO_{ms}$$

b. Steam consumed by turbines of switchable drivers (SD) equipment:

$$\text{[VA/VB SD turbine]} = \sum_{ts_{p,k}=1}^{Nts_{p,k}} z_{ts_{p,k}} VA_{ts_{p,k}}$$

$$\text{[VA/VM SD turbine]} = \sum_{ts_{p,j}=1}^{Nts_{p,j}} z_{ts_{p,j}} VA_{ts_{p,j}}$$

$$\text{[VM/VB SD turbine]} = \sum_{ts_{q,k}=1}^{Nts_{q,k}} z_{ts_{q,k}} VM_{ts_{q,k}}$$

c. Demand of equipment that has switchable drivers: The number of these equipments that is operating is obtained from the Steam Model. Thus, the following condition must be attended:

$$\text{[Number of SD operating equip.]} = \sum_{ms=1}^{N_{ms}} z_{equip} + \sum_{ts=1}^{N_{ts}} z_{equip}$$

Analyzing the objective function, the following form can be observed:

$$f(x) = \sum_{i=1}^{r} c_i x_i$$

with $x_i \geq 0$; i=1,2,…,r
and,

$$\sum_{i=1}^{r} a_{ji} x_i + \sum_{k=1}^{s} b_{jk} y_k = b_j$$

with $y_k \in Y = \{0,1\}$; j=1,2,…,m ; k = 1,2,…,s
and,

$$\sum_{i=1}^{r} a_{ji} x_i \geq b_j \text{ with } j = m + 1,\ldots,p$$

This is a Mixed Integer Linear Programming problem (MILP). Today there are a large amount of commercially available solvers for MILP methods. In this work, the software GAMS was applied for the Optimization Model. GAMS is an optimization platform that allows, through specific language, to formulate the problem and to solve it through the application of an optimization routine. In this problem, the solver OSL was used by applying the *branch and bound* method.

The solution for the studied scenery can be found summarized in Figure 2 and Table 2. As can be observed, savings of 46 t/h of VS can be achieved if the extraction / condensation ratio of the turbines (mainly the utilities turbogenerators) were better explored. In this case, the turbines consumption decreased, but the VA extraction increased, with decreasing in the condensate generation, leading to a more efficient condition regarding to the cycle. It was also achieved a decrease in the electric power consumption, due to the possibility of changing electric motors for steam turbines, in some pumps, compressors and fans.

The time is a relevant point if we should consider shift-to shift adjustments. In the actual stage of development of these tools, the models aren't completely automatic, since some information still should be manually inserted by people. It takes about twenty minutes to gather the additional information for the plant, input in the steam model, check the consistency by analyzing the steam headers balance, and convert the final set of date from spreadsheet format to GAMS input format. Finally, the execution time the optimization model registered was 0,015 seconds. Even without full automatic sequence, the process can be done in a work shift, with enough time to adjust the system.

It must be remembered that the formulation consider the same load for process compressors, but allow the model to select the best way to generate energy, given the price per MW in each power generator or offsite purchasing – in this case, the power load of the turbogenerators was decreased from 27 to 18 MW, due to the power consumption reduction and also due to an increase of purchasing offsite power. It also must be underlined that the optimized scenery leads to a condition where the use of letdowns was reduced to its minimum, except the VM/VB, which was reduced to 70% to its actual value.

It also must be stressed that this tool should only be applied to a steady state conditions. In practice, if there's any transient, the operators should wait for the control loops set the system to a stationary condition again. When there are no more oscillations in the system, so operators and engineers are encouraged to explore the system for a more optimized condition. There's no need to worry with transient conditions, since the steam network runs constant most part of the time. Nevertheless, to force optimization changes in a transient condition is a risky condition, reducing the steam supply reliability.

	Case	
Steam and Energy Rates	**Actual**	**Optimization**
VS Generation (t/h)	1207.6	1161.2
VS/VA letdown stations (t/h)	41.4	20.0
VA/VM letdown stations (t/h)	49.8	20.0
VM/VB letdown stations (t/h)	62.4	43.1
VB Relieves (t/h)	0.0	0.0
Power Demand (MW)	58.0	52.0
VA/VM turbines from changeable drives equipaments (t/h)	32.4	58.4
VA/VB turbines from changeable drives equipaments (t/h)	46.9	67.9
VM/VB turbines from changeable drives equipaments (t/h)	9.0	8.8
VS for two stages compressors and generators (t/h)	1091.8	1066.8
VA from two stages compressors and generators (t/h)	466.0	506.5

Table 2: Summary of main comparison parameters of actual and optimized cases for studied scenery (June 6th, 2005)

5. CONCLUSION

The approach for solving the optimization problem of a specific scenery shows that it's possible to improve earnings from a better adjust of the steam system, without any further investments. The steam model, which was one of the most challenging development tools in this work, was well succeeded in supplying the optimization model with rigorous precision collected data and it can be made with any chosen scenery. With simple sequence of steps, an engineer or operator can collect the data, and export these figures to the optimization model, which will immediately give the best way to operate the system. This can be implemented in day-to day or shift-to-shift conditions, as a tool to orient the personnel in a readily form.

Next step will be the development of a better interface between the steam model and the optimization model, in order to make the process of data collection and optimization run in a more automatic form.

REFERENCES

Bojic, M., B. Stojanovic (1998). MILP Optimization of a CHP Energy System, *Energy Convers. Mgmt*, **39** (7) 637-642.

Eastwood, A., C. Bealing (2003). *Optimizing the Day to Day Operation of Utility Systems*, Linhnhoff March, Northwich, UK.

Hirata, K., H. Sakamoto, L. O´Young, K.Y. Cheung (2004). Multi-Site Utility Integration – an Industrial Case Study, *Computers and Chemical Engineering*, **28**, 139-148.

Milosevic, Z. (1997). Refinery Improves Steam System with Custom Simulation / Optimization Package, *Oil & Gas J.*, Aug. **25**.

Rodríguez-Toral, M.A., W. Morton, D.R. Mitchell (2001). The Use of New SQP Methods for the Optimization of Utility Systems, *Comput. Chem. Eng.*, **25**, 287-300.

Strouvalis, A.M., S.P. Mavromatis, A.C. Kokossis (1998). Conceptual Optimization of Utility Networks using Hardware and Comprehensive Hardware Composites, *Comput. Chem. Eng.*, **22**, 175-182.

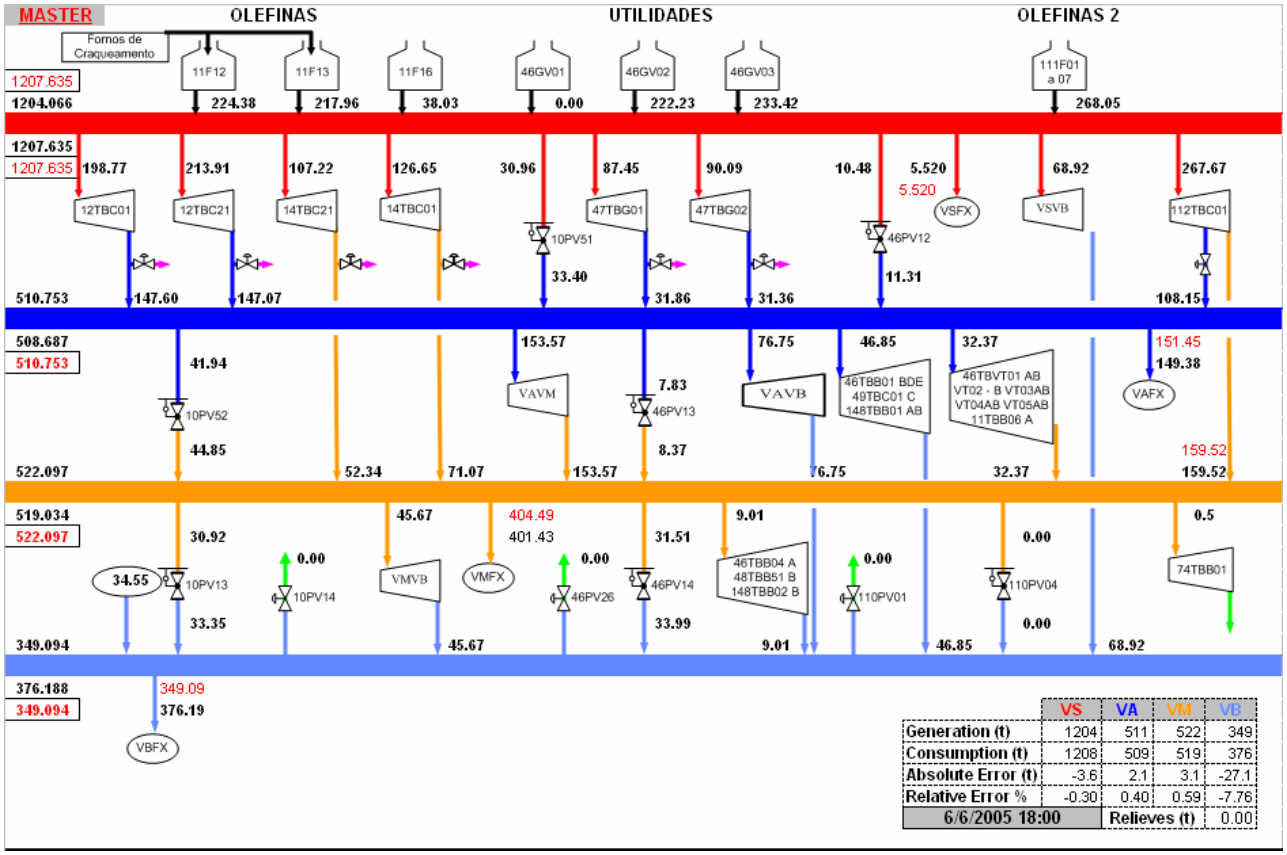


Fig. 1: Real scenery of 6/6/5. The red, blue, orange and cyan coloured levels indicate respectively the VS, VA, VM and VB headers and distribution lines. The consumers and generator are grouped with definition. It can be observed the increase of steam after each letdown valve because of the desuperheater water injection (temperature control). The oval VB "supplier" represents the steam generated from VM condensate flash vessels. The red figures refers to the corrected values in order to close the balance (distribution of errors), a necessary condition to implement optimization. It was observed and total electric power demand of 58 MW.

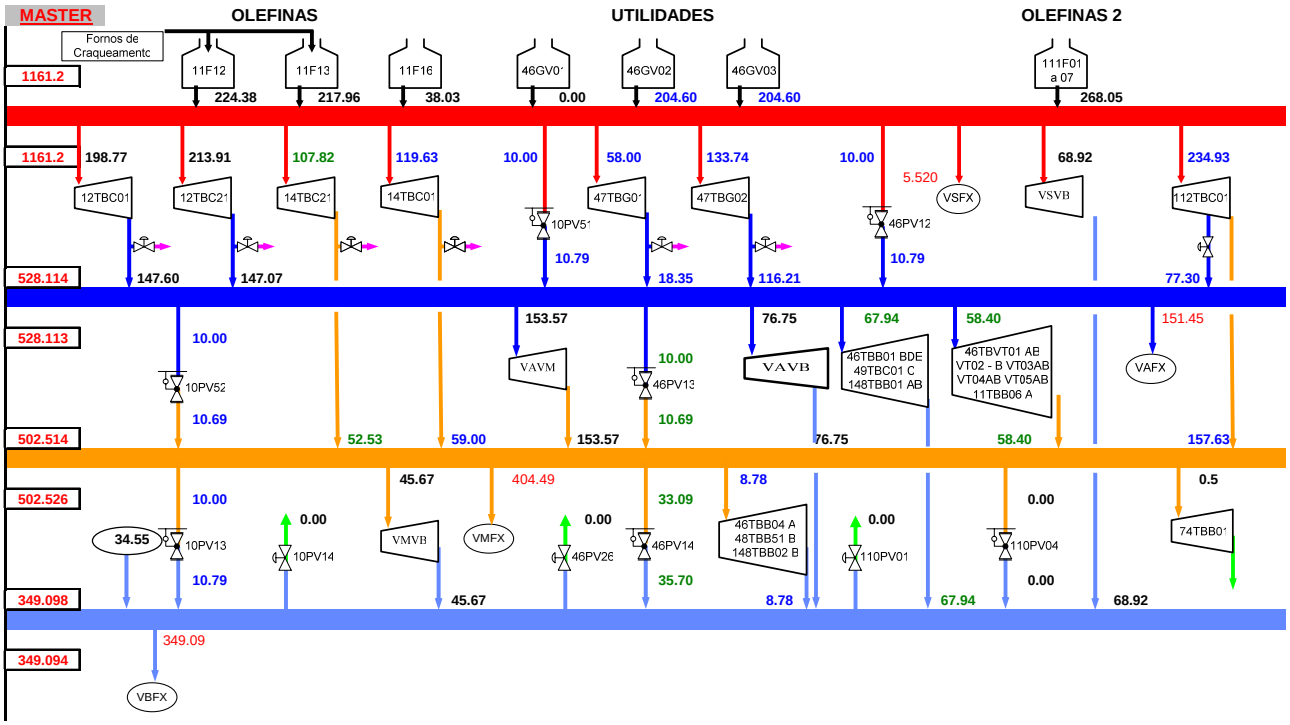


Fig. 2: Optimized scenario for 6/6/5 date conditions. The blue font represents the steam generation or consumption reduction and green are the steam increases. Along with these results, it was observed a reduction in the demanded power to 52 MW. This was possible exchanging electric motors for steam turbines.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

MULTIPERIOD OPTIMIZATION MODEL FOR SYNTHESIS, DESIGN, AND OPERATION OF NON-CONTINUOUS PLANTS

Gabriela Corsano, Jorge M. Montagna, Pio A. Aguirre, and Oscar A. Iribarren

INGAR – Instituto de Desarrollo y Diseño
Avellaneda 3657, S3002 GJC Santa Fe, Argentina

Abstract: In this paper, a general multiperiod nonlinear optimization model is presented, which incorporates synthesis, design, and operation, and takes into account the corresponding benefits and costs in each time period. The model is formulated as a non linear programming (NLP) model in which plant structure decisions are modeled in terms of a superstructure embedded in the overall model. This approach is novel since it involves new decision variables, integrates algebraic and differential equations, and solves a NLP problem even when discrete decisions are involved. The proposed model is applied to a Brandy production plant with high detail level in the operations description. The optimal solution is found and different tradeoffs between process and design variables are assessed.

Keywords: Structured Programming, Synthesis, Design, Process Models, Optimization, Nonlinear Programming

1. INTRODUCTION

Multiperiod plants are process plants where costs, demands and resources typically vary from period to period due to market or seasonal changes. Models for multiperiod optimization have an objective, e.g. maximize total profit or minimize cost, which is subjected to constraints that represent mass balances, process performance equations or design equations. Some constraints can be valid for all periods or for an individual period. These models typically involve both continuous and discrete variables, and consequently most mathematical formulations for this problem result in a mixed integer nonlinear programming (MINLP) model (Voudouris and Grossmann, 1992; Paules, and Floudas, 1992; Varvarezos et. al. 1992; Van den Heever and Grossmann, 1999).

MINLP problems are usually solved through methodologies that successively solve Mixed Integer Linear (MILP) approximations to the model and NLP problems for fixed configurations, i.e. certain decisions as regards the value of binary variables (Viswanathan and Grossmann, 1990). For the case of a non-convex problem, the drawback of this mechanism is the fact that successive linearizations usually cut part of the feasible region. In this way, some solutions to the problem are lost (Grossmann, 2002). In addition, many solutions of plant configurations, which are found through MILP, correspond to non-feasible structures, which are not suitable for meeting production requirements.

In order to overcome the aforementioned difficulties, a general nonlinear programming (NLP) model is proposed in this paper, where plant structure decisions are simultaneously considered with the process and design variables. In this way, both discrete variables and the complexity of solving a MINLP are avoided. The structured plant is obtained for all periods, and therefore different tradeoffs between process and design variables are analyzed in each time period.

A study case that considers the seasonal production of Brandy is presented in this work. The proposed model presents a high detail level that is rarely found in the literature. The mass balances for some units in each period are given in terms of dynamic equations written as algebraic equations and included in the overall model. Design equations require process performance variables, operative conditions, and several raw materials and energy resources to be taken into account in order to obtain a real scenario for this process production.

The paper is organized as follows. The next section presents the problem description. In Section 3, the general formulation proposed for the optimal synthesis, design and operation of a multiperiod plant is formulated. The Brandy process production is described in Section 4; and the results of its model optimization is presented in Section 5. Some comments and results analysis are also presented in this last section. Finally, conclusions are presented in Section 6.

2. PROBLEM STATEMENT

A non-continuous plant involves two types of units: batch (j = 1, …, N_j) and semicontinuous (k = 1, …, N_k). In addition, the process considered in this paper is monoproduct, i.e., only one product is produced. For each time period t (t = 1, …, T), the product is manufactured in each unit. In this model, the plant structure is the same for all periods. The number of periods, the total time horizon HT and the time horizon for each period H_t is a data problem. For some batch stages, the number of units in series is unknown beforehand and the stage configuration is decided including the superstructure model presented by Corsano et al. (2004) in the overall model. In this way, the use of binary variables is avoided.

The plant receives raw materials and energy resources r (r = 1, …, N_r) of another plant (mother plant) that seasonally produces them within the same industrial complex. Therefore, in some time periods, the non-continuous plant must buy material and energy resources from another industrial complex. Resources obtained from the mother plant have no cost.

Batch blending, batch splitting and recycles are allowed as novel components for this type of models, decisions taken in this work as optimization variables. The transfer policy adopted between batch stages is Zero-Wait (ZW).

The objective is to maximize the total benefit considering incomes from product sales and operative and investment costs.

3. MATHEMATICAL FORMULATION

Given T periods of time over the horizon time HT, the model considers:

Objective Function: Maximization of annualized net profits given by the total expecting selling price minus the investment and operative cost is considered

$$Max \sum_{t=1}^{T} p_t Nb_t B_t - \left(\sum_{j=1}^{N_j} \alpha_j V_j^{\beta_j} + \sum_{k=1}^{N_k} \alpha_k V_k^{\beta_k} + \sum_{t=1}^{T} \sum_{r=1}^{N_r} c_{rt} F_{rt}^{trans} + Res \right) \quad (1)$$

where p_t is the expected net profit in period t, Nb_t the number of batches produced in period t, B_t the product batch size in period t, V are the batch (j) and semicontinuous (k) unit size, F_{rt}^{trans} is the amount of resource r transported from a plant other than mother plant in period t and c_{rt} is its cost that considers supply and transportation costs. α and β are the cost coefficients and Res the disposal cost that varies according to the effluent.

Mass balances at each unit of the plant: some material balances are given by differential equations like

$$\frac{dC_{xjt}}{d\tau} = g(\tau, x, t) \quad (2)$$

which are discretized and included in the global model as algebraic equations. We adopt the trapezoidal method to discretize the differential equations. The performance of this method for this kind of models was analyzed in Corsano et al. (2004). The difference finite equations according to the trapezoidal method are

$$C_{xjt}^{n+1} = C_{xjt}^{n} + \frac{h}{2}\left(g\left(\tau_n, C_{xjt}^{n}\right) + g\left(\tau_{n+1}, C_{xjt}^{n+1}\right)\right) \quad (3)$$

where C_{xjt} is the concentration of component x (biomass, substrate, product, etc.), at stage j in period t. τ represents the time variable and $h \geq 0$ defines the discretization grid points by $\tau_n = \tau_0 + nh$ and $n \geq 0$.

Mass balances between units of the same plant: blending of batches is considered in this model, so a batch unit size depends on the previous unit batch size and the batch size of the feeding to this unit. The model considers *global material balances*:

$$VS_{jt} = \sum_{r=1}^{N_r} f_{rjt} + VS_{j-1,t} \quad (4)$$

and *component material balances:*

$$VS_{tj} C_{xjt}^{ini} = \sum_{r=1}^{N_r} C_{xjt}^{r} f_{rjt} + C_{x,j-1,t}^{fin} VS_{j-1,t} \quad (5)$$

where superscripts *ini* and *fin* represent the initial and final concentration respectively and VS_{jt} represent the batch volume at stage j in period t. f_{jrjt} is the amount

of r consumed at stage j in period t; and C_x^r represent the concentration of x in r.

Interconnection constraints between mother plant and multiperiod plant:

$$F_{rt} + F_{rt}^{trans} \geq \sum_{j=1}^{N_j} \frac{f_{rjt}}{CT_t} + \sum_{k=1}^{N_k} \frac{f_{rkt}}{CT_t} \quad \text{for each } r \qquad (6)$$

where F_{rt} is the amount of resource r produced by the mother plant in period t and CT indicates the plant cycle time. F_{rt}^{trans} represents the amount of resource r that must be transported from another plant in period t. Resources obtained from the mother plant have no cost, and as a consequence, only transported resources costs are considered in the objective function.

Design equations: for each batch units

$$V_j \geq S_j B_t \qquad \text{for each period } t \qquad (7)$$

and semicontinuous units

$$V_k \geq S_k \frac{B_t}{\theta_{kt}} \qquad \text{for each period } t \qquad (8)$$

where B is the product batch size (kg); S represents size or duty factor of batch and semicontinuous units which depends on process variables; and θ_{kt} is the processing time of unit k in period t. Note that these constraints are in " $\geq$ " form because some units can be sub-occupied in some period.

Constraints of production rate of the plant:

$$\frac{Nb_t B_t}{CT_t} = Q_t \qquad (9)$$

$$Q_t^{\min} \leq Q_t \leq Q_t^{\max} \qquad (10)$$

where Q_t is the production rate in period t which is bounded by $Q_t^{\min}$ and $Q_t^{\max}$; and CT_t is the cycle time of the plant on period t.

Timing constraints: as the plant produces only one product, the ZW transfer policy indicates that

$$CT_t \geq T_{jt} \qquad \text{for all } j \text{ batch units} \qquad (11)$$

$$CT_t \geq \theta_{kt} \qquad \text{for all } k \text{ semicontinuous unit} \qquad (12)$$

where

$$T_{jt} = \theta_{k't} + t_{jt} + \theta_{k''t} \qquad (13)$$

T_{jt} represents the time for which batch unit j will be occupied, which contemplates the material loading ($\theta_{k't}$) and unloading ($\theta_{k''t}$) time if this unit is located between semicontinuous units. t_{jt} is the processing time of unit j. It is worth noting that in this approach, variables t_{tj} and $\theta_{tk'}$ are assumed to be involved in detailed submodels, some of them written as differential equations and included in the actual model.

The product in each period must be produced within the period horizon time, so

$$Nb_t\, CT_t \leq H_t \qquad \text{for each } t = 1, \ldots, T \qquad (14)$$

and

$$\sum_{t=1}^{T} H_t \leq HT \qquad (15)$$

4. STUDY CASE: A BRANDY PRODUCTION PLANT

A Brandy production plant that receives material and energy resources from a neighboring Sugar plant is considered. Besides producing this alcohol, the Brandy plant generates a non-distilled remainder called vinasses or distillery broth that represents another contribution of sugaring substrate for fermentations stages. Four stages for Brandy production are considered: biomass fermentation, alcohol fermentation, centrifugation and distillation. The main objective of the first stage is biomass production. This stage operates in batch form and it is fed with molasses and filter juices from the sugar plant, vinasses, and water. The first biomass fermentor is fed with a broth containing biomass prepared in laboratory: the inoculums. At this stage, large amounts of air are supplied. The alcohol fermentor is also a batch item and it is fed with the product of biomass fermentors, molasses, filter juices, vinasses, and water. Brandy production occurs at this stage without air supply. The fermented broth is centrifuged in a disk stack centrifuge that operates in a semicontinuous mode. The objective of this stage is to separate the biomass from the liquid that contains the brandy. The solids can be recycled to a Yeast production plant. In this work, yeast production is not considered. The last stage of the process is the batch distillation. The batch distiller model is a combination of two batch items, namely the distiller feed vessel and the distillate tank, and three semicontinuous items: the heating surface to evaporate, the cooling area to condense the steam and the column itself. An analytical model presented by Zamar et al. (1998) for batch distillation is adopted. This model relates both the minimum and operational reflux values as well as the minimum and operational number of stages.

For biomass and alcohol fermentations, the superstructure model presented by Corsano et al. (2004) is included in the overall model in order to find the optimal synthesis and design of these stages. In this paper, only duplication of units in series is considered.

The sugar plant produces molasses, filter juices, electricity and vapor that are used for Brandy production. Molasses and filter juices serve as sugaring substrates for biomass and alcohol fermentations. In addition, water and vinasses are added to the fermentation feed. The electricity generated in the sugar plant is used in the centrifuge of the plant, whereas fermentors and the distillation column consume the steams.

For the Sugar plant, two seasons are distinguished: harvest and no-harvest date. During the harvest date, the Sugar plant provides molasses, filter juices, electricity, and vapor to the Brandy plant. In addition, if necessary, molasses, vapor, and electricity can be imported from other plants, allocating operative costs to the total annual cost due to the purchase and transportation of these products.

During the no-harvest date, vapor and electricity are imported from other power stations. The molasses that are not consumed during the harvest date can be stored, while filter juices cannot, since they are degraded in a short time. The model considers an additional cost for molasses inventory and for importations and transportation of electricity and vapor. Again, if needed, molasses can be imported from another complex. Figure 1 shows the flowsheet for Brandy production plant.

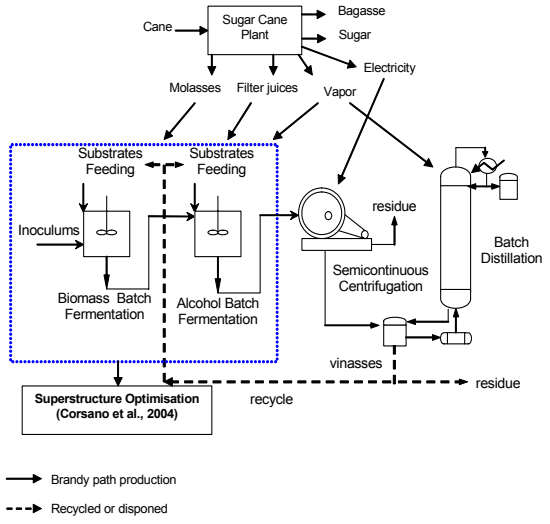


Fig. 1. Flowsheet for Brandy Production Plant integrated to a Sugar plant

The produced vinasses have a substrate concentration variable that depends on the processing time of the last alcohol fermentor in the series, that is, there is a tradeoff between processing time of this unit and the substrate concentration of the vinasses. A longer processing time implies a smaller substrate concentration because the substrate is consumed in fermentation stages. Unused vinasses are discarded and a disposal cost is added in the objective function (*Res* in equation (1)).

For this model, we consider a total time horizon of 7500 hours divided in two periods: harvest with 3000 hours and no-harvest with 4500 hours. Table 1 shows the adopted cost for material and energy resources imported in each period and the amount produced for the Sugar plant in harvest period. Production rates for both periods are lower and upper bounded by 0.5 t h^{-1} and 2 t h^{-1} respectively.

5. RESULTS AND ANALYSIS

The model was implemented and solved in GAMS (Brooke et al., 1998) in a Pentium IV, 1.60 Ghz. The code CONOPT2 was employed for solving the NLP problems. The number of equations and variables is about 3000 and 3200 and the CPU time needed for resolution is 340 sec.

Table 1. Material and energy imported resources cost

	Harvest Date	No-Harvest Date	Sugar plant production
Molasses	10 \$ t^{-1}	35 \$ t^{-1}	36 t h^{-1}
Stored molasses	-	5 \$ t^{-1}	-
Vapor	3.53 \$ t^{-1}	8.5 \$ t^{-1}	4.6 t h^{-1}
Electricity	0.02 \$ kwh^{-1}	0.04 \$ kwh^{-1}	260 kwh
Inoculums	1 \$ kg^{-1}	1 \$ kg^{-1}	-
Water	0.05 \$ t^{-1}	0.05 \$ t^{-1}	-

The optimal solution obtained for the Brandy production plant considering two different periods of time consists of a plant with one biomass fermentor and three alcohol fermentors in series. Table 2 shows the optimal design variables and the processing time of each unit in each period.

The cycle time of the plant is equal to 11.2 h for the harvest date and 13.7 h for no-harvest date, and the number of batches at each period is 268 and 328 respectively. Production rate in each period is equal to 2 t h^{-1} (upper bound). Total profit is 3575.8 \$ h^{-1}. Table 3 shows the resources used in each period and the resources bought in the no-harvest period. In harvest period, all resources used the in Brandy plant come from the Sugar plant. The table also shows the cost for the purchased resources.

As shown in Table 3, molasses used in the no-harvest period are the totally stored molasses, so that no molasses are imported from other sugar complexes. Water included in the table corresponds to the consumed water in fermentation stages, but its

reported costs are the sum of the cost for water in fermentation and the cost for cooling water in distillation column.

Table 2. Optimal design variables and processing times in each period

	Unit Size	Processing times: Harvest Date (h)	Processing times: No-Harvest Date (h)
Biomass	81.9 m^3	11.2	13.7
Fermentor			
Alcohol Ferm. 1	294 m^3	11.2	13.7
Alcohol Ferm. 2	329 m^3	8.8	12.1
Alcohol Ferm. 3	372.2 m^3	8.8	4.11
Centrifuge	70.6 Kwh	2.3	9.4
Distillation		8.8	4.3
Stages Number	9		
Reflux Ratio	5.2		
Distillate Tank	34.8 m^3		
Still Vessel	277.1 m^3		
Condenser Area	117.6 m^2		
Evaporator Area	69.6 m^2		
Column	2.9 m^2		

Table 3. Resources used in each period and costs

	Harvest Date	No-Harvest Date	Cost ($\$ h^{-1}$)
Molasses	26.4 t h^{-1}	9.6 t h^{-1}	48.15
Vapor	2.12 t h^{-1}	1.27 t h^{-1}	10.8
Electricity	70.6 Kwh	11.1 Kwh	0.41
Inoculums	3.9 kg h^{-1}	0.7 kg h^{-1}	4.64
Water	0.1 t h^{-1}	0.02 t h^{-1}	17.3[1]
Vinasses	6.3 m^3 h^{-1}	6.4 m^3 h^{-1}	3.62[2]

The optimal substrate concentration in vinasses is 2.1 g l^{-1} for harvest date (o period?) and 49 g l^{-1} in no-harvest date. Having this substrate concentration variable allows a better performance in molasses utilization. Since molasses are more expensive in the no-harvest period, vinasses substrate concentration is increased in order to attain more concentrated blending to fermentation stages. In order to obtain a higher vinasses substrate concentration, fermentation stages have idle time due to the existing tradeoff between these two processing variables (as previously mentioned). Figure 2 shows the substrate concentration in each fermentor for both periods; and as it can be noted, the substrate in the last fermentor for no-harvest date is not totally consumed in order to attain higher substrate concentration in vinasses.

[1] Water cost includes distillation column cooling water and fermentation fresh water cost.
[2] Vinasses cost represents the disposal vinasses cost.

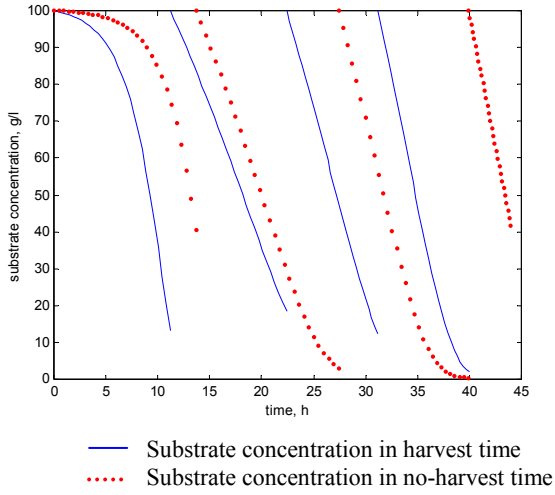


Fig. 2. Substrate concentration in fermentation stages of each period

Total produced vinasses in no-harvest period are recycled to fermentation stages, while about 30% of the produced vinasses in harvest time are discarded.

If more vinasses were used in fermentation stages, the unit sizes would be increased and therefore the investment cost of fermentation stages would be also increased. So, there is another trade-off between vinasses use and fermentation investment cost.

In no-harvest period, some units are sub-occupied. This means that the batch size is smaller than the unit size. This occurs with the three alcohol fermentors, where only about 65% of the units are used.

Simultaneously optimizing synthesis, design, and operation allows obtaining solutions that differ from those obtained in the usual industrial practice, and thus research in this direction is worth being explored. In general, these problems are dealt with by separate: first the plant configuration problem, then the sizing problem and last the operation and scheduling optimization. This leads to sub-optimal solutions. Therefore, simultaneous optimization enables obtaining more accurate solutions and analyzing the tradeoff between different process and design variables.

6. CONCLUSION

A general formulation for the simultaneous synthesis, design and operation for a non-continuous multiperiod plant was proposed and modeled as a NLP problem. Integration between units as well as variable feed blends, unit sizes and operation times were taken into account.

There are no previously published works dealing with simultaneous optimization of the plant structure, design and process variables for a multiperiod plant formulated as a NLP problem. The NLP formulation

avoids difficulties that arise with resolution methodologies of MINLP problems applied to non convex programs.
The model was applied to a Brandy production plant with two time periods: harvest and no-harvest. The optimal number of units in series of the fermentation stages was determined simultaneously with the optimal values of the process variables and the optimal sizing of the downstream stages.

A model with a high level of detail was presented. Operations have been represented through discretized differential equations that describe mass balances (in this case, mass balances of batch fermentors). Furthermore, constraints on feeds to each processing unit, recycles, and equations of interconnections between stages are considered. It is a level of detail that has been posed by few authors.

The model solution allowed analyzing different tradeoffs between process and design variables: the presence of idle times in the fermentation stages and vinasses substrate concentration, the vinasses disposal and the unit size of fermentation stages, molasses use and vinasses substrate concentration.

Duplication in series of biomass fermentors is an industrial practice, while duplication in series of alcohol production fermentors is not. And for the particular case of Brandy production from sugar plant residuals, vinasses recycles are rarely used. In our opinion, a strong point of this research report is that constructing a model for simultaneously optimizing the plant structure and process variables allowed envisaging that plant structures and process variables figures different from those of the current industrial practice may be worth exploring.

Furthermore, with the specific results presented in this work, this approach shows the capabilities of integrated formulations that simultaneously consider synthesis, design and operation decisions applied to a multiperiod context. This is a powerful tool for managers to analyze different scenarios, assessing the joint effect of all the involved elements.

ACKNOWLEDGEMENTS

The authors wish to acknowledge financial support provided by CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) under Grant PIP 2706 and PIP 5914.

REFERENCES

Brooke, A., Kendrick D., Meeraus A., Raman, R. (1998). GAMS, A User Guide. Scientific Press, Calif.

Corsano, G., Iribarren, O. A., Montagna, J. M., Aguirre, P.A. (2004). Batch Fermentation Networks Model for Optimal Synthesis, Design and Operation. *Ind. Eng. Chem. Res.*, **43**, 4211 - 4219.

Grossmann, I. E. (2002). Review of Nonlinear Mixed-Integer and Disjunctive Programming Techniques. *Optimization and Engineering*, **3**, 227 - 252.

Paules, G. E. IV, Floudas, C. A. (1992). Stochastic programming in process synthesis: a two-stage model with MINLP recourse for multiperiod heat-integrated distillation sequences. *Computers and Chemical Engineering*, 16(3), 189–210.

Van den Heever, S. A., Grossmann, I. E. (1999). Disjuntive multiperiod optimization methods for design and planning of chemical process systems. *Computers and Chemical Engineering*.23, 1075 – 1095.

Vaselenak, J. A., Grossmann, I. E. and Westerberg, A. W. (1987). Optimal retrofit design in multiproduct batch plants. *Industrial Engineering Chemical Research.* 26, 718 – 726.

Viswanathan, J., Grossmann, I. E. (1990). A Combined Penalty Function and Outer-Approximation Method for MINLP Optimization. *Comp. Chem. Engng.*, **14** (7), 769 - 782.

Voudouris V. T.; Grossmann, I. E. (1992). Mixed-Integer Linear Programming Reformulations for Batch Process Design with Discrete Equipment Sizes. *Industrial Engineering Chemical Research.* 31, 1315 – 1325.

Zamar, S.D., Salomone, H.E., Iribarren, O.A. (1998) Shortcut Method for Multiple Task Batch Distillations. *Ind. Eng. Chem. Res***,** *37*, 4801 – 4807.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DYNAMIC PENALTY FORMULATION FOR SOLVING HIGHLY CONSTRAINED MIXED-INTEGER NONLINEAR PROGRAMMING PROBLEMS

Cláudia Martins Silva[1], Evaristo Chalbaud Biscaia Jr.

Programa de Engenharia Química/COPPE - Universidade Federal do Rio de Janeiro

Abstract: This contribution presents a heuristic approach for solving nonconvex mixed-integer nonlinear programming (MINLP) problems with highly constrained discontinuous domains. A new fuzzy penalty strategy is proposed to make stochastic algorithms capable of solving optimization problems with a large number of difficult-to-satisfy constraints. The method consists of a dynamic penalty formulation based on the magnitude and frequency of the constraint violation, applied according to a hierarchical classification of the constraints. The new strategy is introduced to a multi-objective optimization algorithm based on evolutionary strategies. The performance of the proposed methodology is investigated on the basis of a multi-enterprise supply chain optimization problem.

Keywords: Nonlinear programming, multi-objective optimization, discrete-time system, heuristic search, integer programming, algorithms, hierarchical decision making, planning

1. INTRODUCTION

Mathematical programming techniques have been widely applied to solve process systems engineering problems. A variety of practical problems such as optimization for integrated process design and control, dynamic allocation and location-allocation problems, design of multi-product batch plants, etc, have been modeled. These problems often involve hybrid discrete-continuous systems and are therefore formulated as mixed-integer optimization problems. Continuous variables usually describe process states, while discrete ones are related to the structure of the process. Discrete variables may be restricted to binary values, when defining the assignments of equipments and sequencing of tasks.

The basic formulation of mixed-integer optimization problems, when represented in algebraic form is:

$$\text{Min } Z = f(x, y) \quad \text{s.t.} \begin{cases} h_i(x, y) = 0 & i \in I \\ g_j(x, y) \le 0 & j \in J \\ x \in X, y \in Y \end{cases}$$

where $f(x, y)$ is the objective function, $h(x, y)$ are the equality relationships that describe the performance of the system (material balances, production rates) and $g(x, y)$ are inequalities that define specifications or constraints for feasible scheduling. I and J are the index sets of equalities and inequalities, and x and y are continuous and discrete variables. Optimization problems are classified according to the type of variables and important properties of the functions, like linearity, convexity and differentiability. Mixed-integer programming problems are commonly regarded as steady-state models. Dynamic models give rise to multi-period optimization problems, in case of discrete time models and optimal control problems, in case of continuous time. Powerful

[1] To whom all correspondence should be addressed.
E-mail: cmartins@peq.coppe.ufrj.br

methods for solving large-scale mixed-integer linear programming (MILP) are well established and have been applied to practical problems for the last few decades. Methods for mixed-integer nonlinear programming (MINLP) problems, on the other hand, have become available recently. Some reviews on optimization methods have been published (Biegler and Grossmann, 2004, Grossmann, 2002). Most common optimization algorithms are based on branch and bound and on decomposition methods. Such algorithms, however, are not guaranteed to locate the global optimum in case of nonconvexity of objective functions or constraints, as it may give rise to multiple local optima (Stein *et al.*, 2004). Relaxation of integer variables as continuous ones and subsequent rounding of the solutions may lead to inaccuracy and infeasible solutions. Decomposition of the original problem to a set of sub-problems may require the objective functions and constraints to be differentiable, which restricts its applicability for a large number of real-life problems (Cheung *et al*., 1997). Moreover, algorithms based on classical nonlinear optimization theory may not be capable of solving large-scale applications, due to their high computational effort requirement (Stein *et al.*, 2004).

Evolutionary algorithms (EAs) have received considerable attention over the last decade, as they have shown to be robust for solving highly nonlinear, nondifferentiable and multimodal optimization problems. Some studies have confirmed the capability of EA-based methods to solve MINLP problems involving local optima and nonconvexities (Ryoo and Sahinidis, 1995, Ostermark, 1999, Cheung *et al*., 1997, Lin *et al*., 2004). Ostermark (1999) has successfully tested EA on a set of complex problems that could not be solved by the GAMS/MINOS package. Hybrid stochastic algorithms have been also employed to solve MINLP problems. Cheung *et al.* (1997) employed a modified grid search method with a genetic algorithm. Lin *et al.* (2004) proposed a migration operation and a population diversity measure to avoid clustering. Ko and Evans (2005) applied a genetic algorithm-based heuristic to solve a set of NP-hard problems. Stochastic methodologies have been also used to treat multi-objective optimization problems. Guillén *et al*. (2005) solve a supply chain design problem as a multi-objective stochastic MILP model. Chan *et al.* (2005) develop a hybrid genetic algorithm based on analytic hierarchy process to solve multi-factory supply chain models. Zhou and Hua (2000) use goal programming and analytic hierarchy process to address sustainable supply chain optimization and scheduling of continuous process industries. Azapagic and Clift (1999) use life cycle assessment in environmental management to solve a multi-objective optimization system.

Besides the inherent complexity of MINLP, the problem of finding any feasible solution may be itself NP-hard. Different approaches are employed to deal with constrained optimization problems. Some methods reject the infeasible solutions while others adopt repair operations. Modifying nearly-feasible solutions, however, may disrupt the schema excessively or incur undue computational overhead. The most promising methods make use of penalty functions (Ostermark, 1999). By penalizing infeasible individuals, these methods turn such individuals into mediocre ones. This procedure prevents the propagation of the infeasible solutions to future generations, since mediocre individuals have little chance to survive. Such strategy transforms constrained problems into unconstrained ones.

In this contribution, a new penalty function method based on fuzzy logic theory has been specially developed to treat problems in which feasible regions are very difficult to reach. The approach was first developed for multi-objective optimization, but it can be extended to any stochastic optimization algorithm. It comprises a dynamic penalty function based on the constraint classification and the intensity and frequency of violation. The optimization is encouraged to solve the constraints according to pre-established priority, until the feasible region is reached. The proposed formulation is illustrated on a numerical example of a multi-enterprise supply chain network. A multi-product, multistage and multi-period production and distribution-planning model is addressed. A multi-objective optimization algorithm based on evolutionary strategies is applied to determine the best configuration of the supply chain network. The proposed method has successfully attained a compromise solution among all participant enterprises, providing a balanced satisfaction for all objectives. The results of a hypothetical case study confirmed the ability of the proposed method in solving complex MINLP problems.

2. EVOLUTIONARY ALGORITHMS

Evolutionary algorithms are robust stochastic methods for global and parallel optimization. These methods are founded on the principles of natural genetics, in which the fittest species survive and propagate while the less successful tend to disappear. The evolution process consists of performing a population of individuals with operators to generate the next generation. The basic operators simulate the processes of selection, crossover and mutation, which happen according to pre-established probabilities. Selection is based on the survival potential, expressed by the fitness function. Crossover involves random exchange of characters between pairs of individuals, in order to produce new ones. Mutation is an occasional change in individual's characters randomly chosen. It introduces diversity to a model population. Evolutionary methods are able to deal with ill-behaved problem domains, such as the ones presenting multimodality, discontinuity, time-variance, randomness and noise.

Evolutionary algorithms are regarded to be suitable to solve multi-objective optimization problems as they work on a population of individuals. Multi-objective optimization is a special extension of the optimization theory in which multiple opposing

targets must be accomplished simultaneously. The search process aims to find solutions that are the best on all objectives. The optimal solution constitutes a family of points, called Pareto optimal front, that equally satisfy the set of objective functions. An important characteristic of the Pareto set is that no improvement can be obtained in any objective without deteriorating at least one of the other objectives. As all objective functions are optimized at the same time, the solution constitutes a compromise between the conflicting aims.

3. THE PROPOSED STRATEGY

This work focuses on the solution of nonconvex MINLP optimization problems that involve discontinuous domains and a large number of constraints. Such problems are difficult to solve as they present multiple local optima dispersed in a discontinuous search space. Even stochastic algorithms can be easily trapped in a local optimum surrounded by an infeasible region. Additional difficulty emerges in case of discrete problems with binary variables. Any change in these variables may interrupt the search progress. Also, constraints involving these variables are easily violated, which makes feasible regions hard to be found.

In order to face these drawbacks, a heuristic strategy is proposed to provide multi-objective stochastic algorithms with an efficient tool to handle these difficulties. The strategy consists of a penalization procedure that incorporates the constraints into the objective function by means of a penalty function. This function associates a certain value with the extent each constraint is violated by each individual. The procedure is formulated as follows:

$$\text{Minimize} \quad F(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}, \boldsymbol{y}) + P(\alpha, \boldsymbol{x}, \boldsymbol{y})$$

$$\text{where} \quad P(\alpha, \boldsymbol{x}, \boldsymbol{y}) = \alpha_{\mathrm{k}}\, \mathrm{SVC}(\boldsymbol{x}, \boldsymbol{y})$$

$\boldsymbol{x} \in \boldsymbol{R}^{\mathrm{n}}$, $\boldsymbol{y} \in \boldsymbol{Z}^{\mathrm{m}}$, α is a predefined constant related to the k-th rank and $P(\alpha, \boldsymbol{x}, \boldsymbol{y})$ is the dynamic penalty function. SVC $(\boldsymbol{x}, \boldsymbol{y})$ is the sum of violated constraints, which incorporates the distance from the feasible set and the frequency of constraint violation:

$$\mathrm{SVC}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{p} D_i(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j=1}^{q} D_j(\boldsymbol{x}, \boldsymbol{y})$$

Inequality constraints:

$$D_i(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} 0, & \text{if } g_i(\boldsymbol{x}, \boldsymbol{y}) \geq -\varepsilon, \quad i = 1, \ldots, p \\ \left| g_i(\boldsymbol{x}, \boldsymbol{y}) \right| & \text{otherwise} \end{cases}$$

Equality constraints:

$$D_j(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} 0, & \text{if } -\varepsilon \leq h_j(\boldsymbol{x}, \boldsymbol{y}) \leq \varepsilon, \; j = 1, \ldots, q \\ \left| h_j(\boldsymbol{x}, \boldsymbol{y}) \right| & \text{otherwise} \end{cases}$$

A procedure based on the fuzzy logic theory is adopted to generate a hierarchical sequencing on the optimization process. The constraints are first arranged into classes, according to the level of difficulty offered to search evolution. Three classes are suggested: rank 1 - binary variables; rank 2 - discrete or integer variables; rank 3 - continuous variables. A finite value is associated with each rank, differing by at least one order of magnitude. The aim is to encourage the hardest constraints to be satisfied first. Constraints involving binary variables are usually the most demanding and must be strongly penalized. This strategy is essential to the process to succeed. As the degree of freedom continuously reduces during optimization, it will probably fail if the hardest constraints are left to the end.

The proposed strategy also introduces a modified mutation operator, which is applied when the search is trapped in an infeasible region. This operator consists of a Monte Carlo-based mutation, which simulates different procedures for early and final iterations, due to the distinct degrees of freedom. A wide mutation, called extensive Monte Carlo-based mutation, is first performed, alternating discrete and continuous variables. In final stages, a local Monte Carlo-based mutation is applied. In this case only variables directly associated to the violated constraints are mutated within a certain range. An EA-based algorithm developed in a previous work (Silva and Biscaia, 2003) is employed to solve MINLP problems. Some adaptations are required to introduce the proposed strategy into a general genetic algorithm, as suggested as follows:

1) create a random initial population;
2) evaluate the individuals and apply the penalty function method;
3) rank the individuals and calculate their fitness;
4) apply selection, crossover, mutation operators;
5) if progress fails for N iterations, apply extensive Monte Carlo-based mutation in a best individual;
6) repeat steps (2)-(5);
7) if iterative process stagnates for M attempts of step (5), apply a local Monte Carlo-based mutation in a best individual;
8) repeat steps (2)-(7) until no constraint is violated or a limit number of attempts is reached;
9) if all constraints are satisfied, register non-dominated individuals in the Pareto set filter;
10) if a limit number of attempts is reached repeat (1)-(9).

Other modifications of the original algorithm include a rounding procedure to operate in discrete variable space, reformulation of the mutation operator to perform changes in a random number of the characters and grouping of decision variables associated to each constraint. High mutation probabilities are also used to increase the algorithm's exploitation ability and improve the convergence.

4. PROBLEM STATEMENT

An example problem dealing with a multi-enterprise supply chain is considered in this work. It consists of a centralized three-echelon structure including manufacturing, storage and market. The structure comprises two retailers, two warehouses and one

plant. The distribution channels consist of a smaller-scale distributor with fast delivery service and a larger-scale distributor with a slower delivery service. The larger-scale service implies lower operating costs, but has a transportation lead-time of one week. Delayed shipment problem is considered in the distribution system. The plant batch manufactures two different products. The production has a fixed cost associated and can be conducted in regular time or overtime, to satisfy customer demand. If the production line is idle, a fixed idle cost is added to the total manufacturing cost. The raw material purchasing cost is included in the manufacturing cost.

The overall problem aims to determine: a) production schedule, including production rates for all time intervals; b) transportation of products; c) sale quantity; d) costs and revenue of each enterprise and e) inventory level of each enterprise. Given: a) product sale prices; b) costs of unit manufacturing, transport, handling and inventory; c) manufacturing data in regular time and overtime; d) transportation data - capacity level and lead time; e) inventory capacity and safe inventory quantity and f) forecasted customer demand over a time horizon.

The objective is to determine the configuration of the supply chain that maximizes the profit of each enterprise, the customer service level and safe inventory level, taking into account a fair distribution of these targets among all the participants.

4.1 Model Formulation

The optimization problem is formulated as a multi-objective mixed-integer nonlinear programming (MOMINLP) problem. The mathematical formulation for the supply chain model was originally proposed by Chen *et al*. (2003). All parameters and system information are presented in the above-mentioned reference.

Objective functions:

Overall profit:

$$\max \sum_t Z_{rt} = \sum_i USR_{ir} S_{irt} - \sum_d \sum_i USR_{idr} S_{idrt} - \sum_i UIC_{ir} I_{irt} - \sum_i UHC_{ir} \left(\sum_d S_{idr,t-TLTdr} + S_{idrt} \right) \quad (1)$$

$$\max \sum_t Z_{dt} = \sum_r \sum_i USR_{idr} S_{idrt} - \sum_p \sum_i USR_{ipd} S_{ipdt} - \sum_i UIC_{id} I_{idt} - \sum_i UHC_{id} \left(\sum_d S_{ipd,t-TLTpd} + S_{idrt} \right) - \sum_k \sum_r \left(FTC_{kdr} Y_{kdrt} + UTC_{kdr} Q_{kdrt} \right) + \sum_{k'} \sum_p \left(FTC_{k'pd} Y_{k'pdt} + UTC_{k'pd} Q_{k'pdt} \right) \quad (2)$$

$$\max \sum_t Z_{pt} = \sum_d \sum_i USR_{ipd} S_{ipdt} - \sum_i [FMC_{ip} \gamma_{ipt} + FIC_{ip} (\beta_{ipt} - \alpha_{ipt}) + UMC_{ip} FMQ_{ip} \alpha_{ipt} + OMC_{ip} OMQ_{ip} o_{ipt}] - \sum_i UIC_{ip} I_{ipt} - \sum_i UHC_{ip} (FMQ_{ip} \alpha_{ip,t-1} + OMQ_{ip} o_{ip,t-1} + \sum_d S_{ipdt}) \quad (3)$$

Average customer service level:

$$\max \frac{1}{T} \sum_t \frac{100}{I} \sum_i \frac{S_{irt}}{FCD_{irt} + B_{ir,t-1}} \quad (4)$$

Average safe inventory level:

$$\max \frac{1}{T} \sum_t \frac{100}{I} \sum_i \left(1 - \frac{D_{irt}}{SIQ_{ir}} \right) \quad (5)$$

$$\max \frac{1}{T} \sum_t \frac{100}{I} \sum_i \left(1 - \frac{D_{idt}}{SIQ_{id}} \right) \quad (6)$$

$$\max \frac{1}{T} \sum_t \frac{100}{I} \sum_i \left(1 - \frac{D_{ipt}}{SIQ_{ip}} \right) \quad (7)$$

Constraints:

Inventory balance - Retailer:

$$I_{irt} = I_{irt,t-1} + \sum_d S_{idr,t-TLTdr} - S_{irt}$$

$$I_{irT} \geq SIQ_{ir}$$

Backlog level - Retailer:

$$B_{irt} = B_{irt,t-1} + FCD_{irt} - S_{irt}$$

$$B_{irT} = 0 \;\; I_{irt} \geq 0 \;\; B_{irt} \geq 0 \;\; S_{irt} \geq 0$$

Maximum inventory capacity - Retailer:

$$\sum_i I_{irt} \leq MIC_r$$

Safe inventory - Retailer:

$$SIQ_{ir} - I_{irt} \leq D_{irt} \leq SIQ_{ir}$$

$$D_{irT} = 0 \;\; D_{irt} \geq 0$$

Inventory balance - Distribution center:

$$I_{idt} = I_{idt,t-1} + \sum_p S_{ipd,t-TLTpd} - \sum_r S_{idrt}$$

$$I_{idT} \geq SIQ_{id} \;\; I_{idt} \geq 0 \;\; S_{idrt} \geq 0$$

Maximum inventories- Distribution center:

$$\sum_i I_{idt} \leq MIC_d$$

Shortage in safe inventories - Distribution center:

$$SIQ_{id} - I_{idt} \leq D_{idt} \leq SIQ_{id}$$

$$D_{idT} = 0 \;\; D_{idt} \geq 0$$

Output transportation - Distribution center:

$$\sum_k Q_{kdrt} = \sum_i S_{idrt}$$

$$TCL_{k-1,dr} Y_{kdrt} \leq Q_{kdrt} \leq TCL_{kdr} Y_{kdrt}$$

$$\sum_k Y_{kdrt} \leq 1$$

$$\sum_r \sum_k TCL_{kdr} Y_{kdrt} \leq MOTC_d$$

Input transportation - Distribution center:

$$\sum_{k'} Q_{k'pdt} = \sum_i S_{ipdt}$$

$$TCL_{k'-1,pd} Y_{k'pdt} \leq Q_{k'pdt} \leq TCL_{k'pd} Y_{k'pdt}$$

$$\sum_{k'} Y_{k'pdt} \leq 1$$

$$\sum_p \sum_{k'} TCL_{k'pd} Y_{k'pdt} \leq MITC_d$$

Inventory balance - Plant:

$$I_{ipt} = I_{ipt,t-1} + FMQ_{ip} \alpha_{ip,t-1} + OMQ_{ip} o_{ip,t-1} - \sum_d S_{ipdt}$$

$$I_{ipT} \geq SIQ_{ip} \;\; I_{ipt} \geq 0 \;\; S_{ipdt} \geq 0$$

Maximum inventory - Plant:

$$\sum_i I_{ipt} \leq MIC_p$$

Shortage in safe inventory constraints - Plant:

$$SIQ_{ip} - I_{ipt} \le D_{ipt} \le SIQ_{ip}$$

$$D_{ipT} = 0 \;\; D_{ipt} \ge 0$$

Manufacturing - Plant:

$$\sum_i \beta_{ipt} = 1 \;\; \alpha_{ipt} \le \beta_{ipt}$$

$$\gamma_{ipt} \ge \beta_{ipt} - \beta_{ip,t-1} \qquad o_{ipt} \le \alpha_{ipt}$$

$$\sum_i \sum_t o_{ipt} \le MTO_p$$

$$\sum_i \sum_n o_{ip,t-n+1} \le N - 1$$

5. RESULTS AND DISCUSSION

According to the problem description, three-levels of enterprises are integrated in a multi-objective optimization problem. Planning horizons varying from 3 to 8 weeks are tested. The multi-objective optimization problem consists of 12 objective functions. A population size of 25 individuals, crossover probability of 90% and mutation probability of 30% are used to solve the problem. Some of the results obtained for a three-week planning horizon are shown in Figure I. Each line on the graphics represents an optimal result. Table 1 presents the best results obtained for each objective function in some of the optimization cases. For the sake of space, the complete Pareto set is omitted.

Table 1. Best results

i	t = 3	t = 5	t = 6	t = 8
1	$5.84x10^5$	$1.22 x10^6$	$9.69x10^5$	$9.25x10^5$
2	1.0	0.83	0.91	0.86
3	0.97	0.61	0.74	0.78
4	$5.68x10^5$	$1.09x10^6$	$1.06x10^6$	$1.45x10^6$
5	0.97	0.84	0.97	0.68
6	0.94	0.73	0.71	0.76
7	$1.68x10^5$	$4.61 x10^5$	$3.05x10^5$	$6.26x10^3$
8	1.0	0.99	0.67	0.73
9	$7.63x10^5$	$2.14 x10^6$	$1.99x10^6$	$1.61x10^6$
10	1.0	0.96	0. 94	0.69
11	$1.05x10^6$	$1.52 x10^6$	$1.09x10^6$	$2.54x10^6$
12	1.0	0.76	0.77	0.72

High values for all objective functions are obtained, which indicates that the proposed strategy leads to an unbiased search process. A balanced exploration process is mandatory to obtain a good compromise solution for all objectives and satisfy a fair distribution. The results obtained for the most relevant objective functions, which represent each enterprise profit, do not differ in order of magnitude for each case study. Hence, any of the solutions in the Pareto set would be satisfactory to all participant enterprises.

Table 2 presents the number of variables and constraints of each optimization case, as well as the number of generations required. The three-week period problem was solved in 112 seconds on a Pentium IV 2.4 GHz. The eight-week problem, on the hand, took around 12 hours to perform 8,790 iterations. It should be highlighted that this computational effort is required to find the feasible region. In a previous work, the original version of the algorithm was used to solve the same problem (Silva and Biscaia Jr., 2005). The maximum planning horizon the algorithm was able to solve was three weeks.

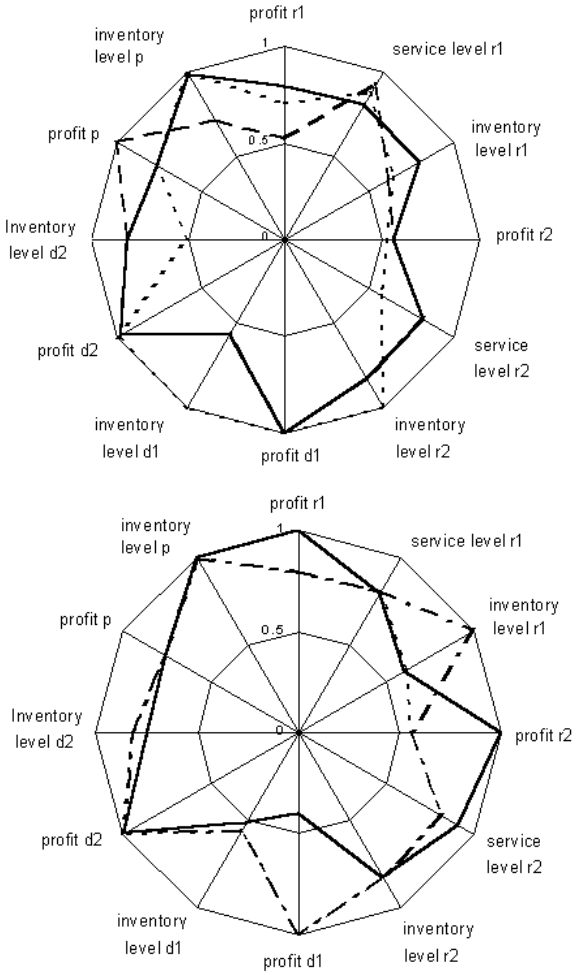


Fig. I. Optimal solutions

Table 2. Number of variables, constraints and iterations

	t = 3	t = 5	t = 6	t = 8
# var	122	254	320	452
# constr	172	344	430	602
# iter	1,066	3,190	4,090	8,790

6. CONCLUSIONS

In this contribution, a dynamic penalty formulation based on the fuzzy logic theory is proposed to solve highly constrained MINLP problems in which feasible regions are very difficult to be achieved. The strategy includes a constraint classification, which induces a hierarchy search progress; a penalty function, which incorporates different levels of penalization into the fitness function according to the constraint classification, the intensity and frequency

of constraint violation; and a mutation operator, to prevent the search to stagnate. A problem involving a large number of difficult-to-satisfy constraints is presented to evaluate the performance of the algorithm. A multi-product, multistage and multi-period production and distribution-planning model, formulated as a multi-objective mixed-integer nonlinear programming (MOMINLP) problem, was selected. A compromise solution among all participant enterprises of the supply chain is achieved, ensuring a fair distribution profit. The results confirm the efficiency of the proposed approach to solve nonconvex MINLP problems involving large search spaces, number of constraints and objective functions.

NOTATION

Indices

i	products	r	retailers
d	distribution centers	p	plants
t	periods		
k	transportation capacity level from d to r		
k'	transportation capacity level from p to d		

Parameters

USR{i, pd, dr, r}	unit sale revenue of i
UICi{i, p, d, r}	unit inventory cost of i
UHC{i, p, d, r}	unit handling cost of i for p, d, r
UTC{k, dr}	kth-level unit transportation cost
FTC{k, dr}	kth-level fixed transportation cost
FTC{k', pd}	k'th-level fixed transportation cost
UMC{i, p}	unit manufacturing cost of i
OMC{i, p}	overtime unit manufacturing cost
FMC{i, p}	fixed manufacturing cost for changing plant to make i
FIC{i, p}	fixed idle cost to keep plant idle
FCD{i, r, t}	forecasted customer demand for i
TLT{pd, dr}	transportation lead time
SIQ{i, p, d, r}	safe inventory quantity
MIC{i, p, d, r}	maximum inventory capacity
TCL{k, dr}	kth transportation capacity level
MITC{d}	max. input transportation capacity
MOTC{d}	max. output transportation capacity
FMQ{i, p}	fixed manufacturing quantity of i
OMQ{i, p}	overtime fixed production quantity
MTO{p}	maximum total overtime in manufacturing period

Binary Variables

Y{k, dr, t}	kth transportation capacity
α{i, p, t}	manufacture in regular-time
β{i, p, t}	set up plant to manufacture i
γ{i, p, t}	change plant over to manufacture i
o{i, p, t}	manufacture with overtime workforce

Integer variables

S{pd, dr, r, t}	sales quantity of i
Q{k, dr, t}	kth-level transportation quantity
Q{pd, dr, t}	total transportation quantity
I{i, p, d, r, t}	inventory level of i in p, d, r
B{i, r, t}	backlog level of i in r at end of t
D{i, p, d, r, t}	shortage in safe inventory level
TMC{p, t}	total manufacturing cost of p
TPC{d, r, t}	total purchase cost of d, r
TIC{p, d, r, t}	total inventory cost of p, d, r
THC{p, d, r, t}	total handling cost of p, d, r
TTC{d, t}	total transportation cost of d
PSR{p, d, r, t}	product sales revenue of p, d, r
Z{p, d, r, t}	net profit of p, d, r

Continuous variables

SIL{ p, d, r,, t}	safe inventory level of p, d, r
CSL{r, t}	customer service level of r

REFERENCES

Azapagic. A. and R. Clift (1999). The Application of Life Cycle Assessment to Process Optimization. *Comp. Chem. Eng.,* **10**, 1509.

Biegler, L. T. and I. E. Grossmann (2004). Retrospective on Optimization. *Comp. Chem. Eng.,* **28**, 1169.

Chan. F.S, S.H. Chung and S. Wadhwa (2005). A Hybrid Genetic Algorithm for Production and Distribution. *Omega,* **33**, 345.

Chen, C. L., B.W. Wang and W. C. Lee (2003). Multiobjective optimization for a Multienterprise Chain Network. *Ind. Eng. Chem. Res.,* **42**, 1879.

Cheung, B.K.S., A. Langevin and H. Delmaire (1997). Coupling Genetic Algorithm with a Grid Search Method to Solve Mixed Integer Nonlinear Programming Problems. *Comp. Math. Applic.*, **34**, 13.

Grossmann, I. E. (2002). Review of Nonlinear Mixed-Integer and Disjunctive Programming Techniques. *Optimization and Eng.*, **3**, 227.

Guillén, G., F. D. Mele., M. J. Bagajewicz, A Espuna and L. Puigjaner (2005). Multiobjective Supply Chain Design under Uncertainty. *Chemical Engineering Science,* **60**, 1535.

Ko, H. J. and G. W. Evans (2005). A Genetic Algorithm-Based Heuristic for the Dynamic Integrated Forward/Reverse Logistics Network for 3PLs. *Comp. & Oper. Res.* (in press).

Lin, Y. C., K.S. Hwang and F. S. Wang (2004). A Mixed-Coding Scheme of Evolutionary Algorithms to Solve Mixed-Integer Nonlinear Programming Problems. *Comp. Math.*, **47**, 1295.

Ostermark, R. (1999). Solving a Nonlinear Non-Convex Trim Loss Problem with a Genetic Hybrid Algorithm. *Com. & Oper. Res.*, **26**, 623.

Ryoo, H.S. and N.V. Sahinidis (1995). Global Optimization of Nonconvex NLPs and MINLPs with Applications in Process Design. *Comp. Chem. Eng.*, **19**, 551.

Silva, C. M. and E.C. Biscaia Jr. (2003). Genetic Algorithm Development for Muli-Objective Optimization Batch FreeRadical Polymerization Reactors. *Comp. Chem. Eng.,* **27**, 1329.

Silva, C. M. and E.C. Biscaia Jr. (2005). Multi-Enterprise Supply Chain Optimization by Means of Evolutionary Strategies, *Proceedings of ENPROMER 2005*, Angra dos Reis.

Stein, O., J. Oldenburg and W. Marquardt (2004). Continuous Reformulations of Discrete-Continuous Optimization Problems. *Comp. Chem. Eng.,* **28**, 1951.

Zhou. Z., S. Cheng and B. Hua (2000). Supply Chain Optimization of Continuous Process Industries with Sustainability Considerations. *Comp. Chem. Eng.,* **24**, 1151.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

APPLICATION OF GENETIC ALGORITHMS TO THE OPTIMIZATION OF AN INDUSTRIAL REACTOR

Igor R. de S. Victorino* **and R. Maciel Filho**

Laboratory of Optimization, Design and Advanced Control (LOPCA). Faculty of Chemical Engineering.

State University of Campinas (Unicamp)

P.O. Box 6066, 13081-970, Campinas, SP, Brazil

email:igor_rsv@yahoo.com.br or maciel@feq.unicamp.br, Tel.:+55-19-37883971; Fax: +55-19-3788396

Abstract: Genetic Algorithms (GAs) have shown great potential and ability to solve complex problems of optimization in diverse industrial fields, including chemical engineering process. In this paper, the main objective is to develop and implement a GA code in an industrial reactor of Cyclic Alcohol (CA) production for the optimization of operational parameters. The intention is to show that this technique is suitable for the maximization of Cyclic Alcohol production, obtaining good results with operational improvements (reduction of catalyst, reduction of the temperature of the process). The results show that the best performance of the process was achieved with the application of GAs. The developed procedure works very well in all the considered conditions which cover the most usual operating range for the considered process.

Keywords: Global optimization, Genetic algorithms, Chemical process.

1. INTRODUCTION

Several works have been carried out having as objective to optimize, through Genetic Algorithms (GAs), the diverse parameters involved in kinetic models of chemical processes (Moros, et al.,1996; Simant and Deb, 1997; Hongqing et al., 1999). In this work the objective is to find the best operating conditions of a Cyclic Alcohol (CA) reactor, which involves the hydrogenation of a specific Benzylic Alcohol (Main Reactant – MR or BA). The optimization of this unit was chosen for several factors among which: the reactor of Cyclic Alcohol presents a complex behaviour and existence of a great energy expense associated to the pressures and temperatures variations in the operation of the process. As the reactor is a non linear multivariable distributed parameter system leading to a system of differential equations, the optimization problem is a hard task and conventional optimization methods have show severe limitations, especially in terms of convergence. Bearing this in mind in this work is proposed an optimization procedure based on Genetic Algorithms method.

2. GENETIC ALGORITHMS (GAS)

These algorithms are a procedure of optimization developed based on the principles of natural selection (Holland, 1992; Goldberg, 1989). The GA initiates with a population of represented random solutions in some series of structures. After this first stage, a series of operators, are applied repeatedly, up to convergence is achieve. In fact the optimization procedure based in such approach can be considered as an global optimization method with the advantage to do no be dependent upon the initial value to achieve the convergence. Most probably the more significant disadvantage is the computer time and burden required. These operators are: coding, reproduction, crossover and mutation. These two last operators are used to create new and better populations. This procedure continues until a termination criterion defined in accord to the need to achieve the goal in the optimization problem. The determination of the parameters is made through the development of an objective function that represent the problem in a suitable way. The application of the GA follows some steps as: coding, determination of the population size, selection (reproduction), crossover and mutation.

2.1 Coding

The coding stage is very important for the success of the genetic code application in the solution of the optimization problem (Goldberg, 1989). The target is to create a parameter representation which allows to its modification through the division in some position. The formed parts are separated sequences in conditions to be matched with others. A codified parameter should be seen as a chromosome in genetics, in other words a modifiable carrier of information. In some GA algorithm, the coding method is based on the representations of binary series number; other forms of coding can be used as

representations in real numbers and whole numbers. In this paper the binary approach is adopted.

2.2 Population Size

Wehrens and Buyders, 1998 mentioned that for each case, population sizes range can vary, but for most of the cases is used between 20-500. In general, when many parameters are optimized larger populations are used. For the CA optimization problem the population size is considered to be about 20 and 500 generations.

2.3 Selection – Reproduction

The reproduction is normally the first procedure applied in the population, and it is a choice of good individuals (series) in order to form one mating pool. Some types of reproduction are found in literature (Goldberg and Deb, 1991). The main idea is to select individuals that possess values above of the average of a current population. The more traditional methods of selection are the proportional selection, roulette wheel and based in rank. The main feature in the stage of selection is the prevention of individuals (series) that promote values of the undesirable evaluation function (fitness) considering the objective of the problem. In this work was considered tournament selection form. This method is the most popular forms of selection in evolutionary algorithms (EAs). In its simplest form, a group of n individuals is chosen randomly from the current population, and the individual with the best fitness is selected (Bäck et al., 2000). This selection performs tournaments by first sampling individuals uniformly and randomly from the population and then selecting the best of the sample for some genetic operation. This sampling process needs to be repeated many times, creating a new generation.

2.4 Crossover

Crossover is applied in the series originated from mating pool (after the stage of reproduction). In the same way that the reproduction operator, the idea is to find some operators of crossover applied in GA (Syswerda, 1989). In the majority of the operators two series (individuals) are chosen randomly from the mating pool. After this stage it is made a recombination of the construction tablets (parts of the series of the relatives) that correspond to the favorable sub-solution. The uniform crossover was used with crossover probability 0.8.

2.5 Mutation

The main target of this genetic operator is to promote new solutions (individuals) that cannot be generated for another form. The mutation introduces an element of the random research (sometimes called exploration). The intention of such procedure is to focus in promising regions of the search space (exploitation). The occurrence of this operator is determined by the researcher through a mutation probability. This value is around 0.01 and it is inside a recommended range by a tray and error procedure (Goldberg, 1989). Usually, this value is smaller than the adopted one for the crossover and the criterion for a good value is to prevent too much random search.

3. DESCRIPTION OF THE PROCESS

The process is a multiphase catalytic reactor, where hydrogenation reactions take place. A typical process of industrial interest is the hydrogenation of ortho-cresol (Vasco de Toledo et al., 2001). A series of parallel and consecutive reactions may happen, so that the reactor has to be operated in a suitable way to achieve high conversion as well as high selectivity.

The reactor is constituted of a series of tubes, cooled by pressured water which flows in a jacked around the tubes. The reactants flow inside the tubes, while the thermal fluid flows through the annular regions. The deterministic mathematical model used to describe the reactor is based on the work by Santana, 1995 and Vasco de Toledo et al., 2001. The reactor model is a set of differential equations, considering two regions of each reactional module: tubular and annular. In the sequence the mass and energy balances for the BA and CA respectively are presented. For the other components of the reactional system, the equations are similar and can be found in the studies of Santana, 1995.

3.1 Model Equations

The equations are customized to the situation of the reactor-CA from the general expressions for the modelling of described mass and energy by Froment and Bischoff, 1990.

Mass Balance for Benzylic Alcohol - Tubular Region:

$$\frac{dX_{BA}}{dz} = \frac{\pi D_{1i}^2}{4}\frac{1}{F_{BA_0}} R_{eBA} \quad (1)$$

Mass Balance for Benzylic Alcohol – Annular Region:

$$\frac{dX_{BA}}{dz} = \frac{\pi\left(D_{4i}^2 - D_{3e}^2\right)}{4}\frac{1}{F_{BA_0}} R_{eBA} \quad (2)$$

Mass Balance for Cyclic Alcohol - Tubular Region:

$$\frac{dX_{CA}}{dz} = \frac{\pi D_{1i}^2}{4}\frac{1}{F_{CA_0}} R_{eCA} \quad (3)$$

Mass Balance for Cyclic Alcohol – Annular Region:

$$\frac{dX_{CA}}{dz} = \frac{\pi\left(D_{4i}^2 - D_{3e}^2\right)}{4}\frac{1}{F_{CA_0}} R_{eCA} \quad (4)$$

Energy Balance – Reactants and Products - Tubular Region:

$$\frac{dT}{dz} = \frac{1}{\sum F_i C_{pi}} \begin{bmatrix} U_2 \pi D_{3e}(T_R - T) + U_3 \pi D_{4i}(T_S - T) + \\ (-\Delta H_1)\frac{\pi(D_{4i}^2 - D_{3e}^2)}{4} R_{eBA} + \\ (-\Delta H_2)\frac{\pi(D_{4i}^2 - D_{3e}^2)}{4} R_{eCA} \end{bmatrix} \quad (5)$$

Energy Balance – Reactants and Products – Annular Region:

$$\frac{dT}{dz} = \frac{1}{\sum F_i C_{pi}} \begin{bmatrix} U_2 \pi D_{3e}(T_R - T) + U_3 \pi D_{4i}(T_S - T) + \\ (-\Delta H_1)\frac{\pi(D_{4i}^2 - D_{3e}^2)}{4} R_{eBA} + \\ (-\Delta H_2)\frac{\pi(D_{4i}^2 - D_{3e}^2)}{4} R_{eCA} \end{bmatrix} \quad (6)$$

Energy balance for the coolant:

Annular Region – I

$$\frac{dT_R}{dz} = -\frac{U_1 \pi D_{1i}}{Q_R C_{pR}}(T_R - T) \quad (7)$$

Annular Region – II

$$\frac{dT_R}{dz} = -\frac{U_2 \pi D_{3e}}{Q_R C_{pR}}(T_R - T) \quad (8)$$

In the previous equations there appear three global coefficients of heat transference, correspondent to the diverse circuits of the reaction medium mixture, U_1, U_2 and U_3 (coefficients tube-coolant, annular-coolant and annular-heating system respectively).

The considered main reaction is the hydrogenation of Benzylic Alcohol to CA. The kinetic model considered by Coussemant and Jungers, 1950 was applied in this work and all the data and calculations related to the global coefficient of heat exchange, pressures, physical properties prediction of the components are described with details by Santana, 1995.

These equations are written to each part of the reactor (tubular and annular region) as well as for each phase of the system, since the reactor is a multiphase one. Moreover, equations for predicting the heat coefficients must be present as well as a way to describe evaporation that may occur, depending upon the operating conditions. Each of these equations must be applied to each tube for both regions, namely, the tubular and annular. Since the reactor is essentially a tubular one usually operating at high flow rates, axial dispersion is neglected. Thus steady-state process model presents a set of ordinary differential equations if radial dispersions is neglected, which is, together with the hypothesis that the solid-liquid phase is a single pseudo-homogenized fluid, a reasonable simplification that can be made in order to reduce the complexity of the process model.

3.2 Kinetic equations

The work developed by Coussemant and Jungers, 1950 does not consider some stages and is represented in accord to with the main equations that occur in the process, described below:

$$BA + 2H_2 \xrightarrow{Ra/Ni} CA + H_2 \xrightarrow{Rb/Ni} C + H_2O$$

The intermediate stages with CEX (cycloalkene) formation are not considered in the model. The formation of alcohols is explained by admitting a mechanism of adsorption in individual small sites of the catalyst.

3.3 Kinetics of the Main Reaction

The considered main reaction is the hydrogenation of Benzylic Alcohol to CA. The kinetic model considered by Coussemant and Jungers, 1950 is used they studied this nickel process using a catalytic reactor of the type autoclave. It was found evidence of formation in intermediate stages of Cyclohexanone, and that to a pressure raised enough the reaction possesses order zero in relation to hydrogen. The global Benzylic Alcohol conversion (R_{BA}) to CA is described by the following relation:

$$R_{BA} = -\frac{dC_1}{dt} = k_1 \frac{b_1 C_1}{b_3 + (b_1 - b_3)C_1 + \frac{b_2 - b_3}{K - 1}(C_1 - C_1^K)} \quad (9)$$

The rate of reaction R_{BA} is express in mol-BA/mim.g-catalyst, and the temperature T, in the expressions of the kinetic constants must be in K.

3.4 Kinetic of Secondary Reaction

The considered secondary reaction is the dehydration of the CA with water formation and Cycloalkene, which is immediately hydrogenated, with consequent formation of C (Cycloalkane - undesirable product). The rate of formation of C (R_{CA}) from dehydration of CA is described in relation (10), as follows:

$$R_{CA} = k_3 \frac{\sqrt{C_2}}{\sqrt{C_2} + bC_3} \quad (10)$$

The parameters b, b_1, b_2, b_3, k_1, K, C_1, C_2 and C_3 are described in Coussemat and Jungers, 1950.

3.5 Effective Rates of Reaction

The rate of reaction of a catalytic process is directly associated with the catalyst concentration, being expressed by the equation (11):

$$r_i = C_{cat} R_i \quad (11)$$

where r_i must be expressed in mol-i/min.m^3, whereas the catalyst used in the hydrogenation processes is considered as highly active. It has a certain level of activity related to presence of the metal on the catalyst.

Thus, in the formularization of the expressions for the rates of the considered reactions, a F_i factor that attempts to quantify the effectiveness of the catalyst for the two reactions (hydrogenation of BA and dehydration of the Cycloalkene), was introduced so

that each one of the reaction effective rates is expressed in the form of equation (12):

$$R_{ei} = F_i r_i \quad (12)$$

the factor F_i can be seen as a numerical constant whose value can vary in a range of (0 and 1), where the null value represents activity absence (absence of the reaction) and unitary value meaning the maximum of the catalytic activity (full activity). Intermediate values can characterize different states of the activity of the catalyst wheels is function of the reactor severity (operation temperature).

Other details of components physical properties and other considerations are described in Santana, 1995.

4. OPTIMIZATION STRATEGIES

The optimization using the mathematical model takes into account the real operational conditions of the reactor. The chosen parameters to implement the optimization are those with more sensitivity in the production process. The objective is to maximize the production of CA (Q_{CA}), using as main variables the outflows of coolant fluid (Q_{ri}'s), the feed reactants temperature (T_0) and the outflow of catalyst (Q_{cat}), in a total of eight variables. Table 1 shows the valid parameter limits to be optimized. The genetic code developed by Carroll, 1996 was coupled with the reactor model. The genetic code possesses the following characteristics: binary code; uses the elitism; search in niches and selection by tournament. The presented values in the tables are in the normalized form. In the industrial reactor all the flows are measured in kg/h (Q_{CA}, Q_{MR}, Q_C, Q_{cat} and Q_{ri}'s respectively) and the temperature is in Celsius degrees.

Table 1 Limits of validity of the parameters to be optimized (normalized values)

Parameters	Lower limits of variable	Upper limits of variable
Q_{r1}	0.01	1.00
Q_{r2}	0.01	1.00
Q_{r3}	0.01	1.00
Q_{r4}	0.01	1.00
Q_{r5}	0.01	1.00
Q_{r6}	0.01	1.00
T_0	y*	0.84
Q_{cat}	0.0000	x*

The value of y^* is related to the inferior limit of the initial temperature of the reactants mixture and products in the entrance of the reactor (normalized values), being 0.60 for the Level 1 of production and 0.68 for the two other production Levels (2 and 3 respectively). In the Levels 2 and 3, smaller values than 0.68 supply discontinuous values for solution of the reactor model. This is not appropriate to be used in the optimization.

The value of x^* refers to the maximum catalyst flow (Q_{cat} normalized) (upper limits) that also depends of the operational level of production that is analyzed. For the Level 1 the maximum value is 0.6000, the Level 2 the value is 0.8000 and last (Level 3) assumes the value of 1.0000. Values above the upper limits of each level also lead to discontinuity in the model solution, and hence were not used.

4.1 Objective Function

The optimization is performed through the development of an objective function. In this work the objective function is related to the productivity of the main product (Cyclic Alcohol) and considers the the following restrictions presented in Table 2. The restrictions are related to the product of interest (CA), the main reactant (MR) and secondary product (C) without interest, as can be observed in Table 2.

Table 2 Production Levels to be optimized considering the respective restrictions (normalized values)

Level 1	Level 2	Level 3
$0.0100 \le Q_{ri} \le 1.0000$	$0.0100 \le Q_{ri} \le 1.0000$	$0.0100 \le Q_{ri} \le 1.0000$
$0.60 \le T_0 \le 0.84$	$0.68 \le T_0 \le 0.84$	$0.68 \le T_0 \le 0.84$
$0.0000 \le Q_{cat} \le 0.6000$	$0.0000 \le Q_{cat} \le 0.8000$	$0.0000 \le Q_{cat} \le 1.0000$
$Q_{CA} - 0.6554 \ge 0$	$Q_{CA} - 1.0000 \ge 0$	$Q_{CA} - 0.9621 \ge 0$
$0.1833 - Q_{MR} \ge 0$	$1.0000 - Q_{MR} \ge 0$	$0.2111 - Q_{MR} \ge 0$
$0.4851 - Q_C \ge 0$	$0.7551 - Q_C \ge 0$	$1.0000 - Q_C \ge 0$
i = 1, 2, 3....6		

The three levels of CA production are considered, as shown in the Table 3 (industrial operational values).

Table 3 Operating conditions for three industrial production levels (Levels 1, 2 and 3) (normalized values)

Parameters	Level 1	Level 2	Level 3
Q_{r1}	0.2520	0.0360	0.0390
Q_{r2}	0.2590	0.0380	0.0000
Q_{r3}	0.2760	0.2740	0.0850
Q_{r4}	0.0360	0.0660	0.3490
Q_{r5}	0.0520	0.1190	0.1190
Q_{r6}	0.0290	0.1400	0.0500
T_0	0.6320	0.6920	0.6920
Q_{cat}	0.4340	0.7520	0.8280
Q_{CA}	0.6554	1.0000	0.9622

4.2 Parameters of Control of the Genetic Algorithms

In accordance to Table 4 were selected the control parameters of the genetic algorithms in the process optimization. The parameters to be optimized were codified in the binary form, as great part of published works.

Table 4 Control parameters of genetic algorithms utilized in the optimization

Size Population	Parameters	Crossover (UC)	Mutation Rate (JM)	Generations
20	8	80%	1%	500
UC is Uniform Crossover				JM is Jump Mutation

The parameters to be optimized were codified with the binary form, based and adapted of many published literature works (Carroll, 1996; Deb, 1998; Goldberg, 1989).

The control parameters of the genetic algorithms can be varied and tested in the same way. In this work it was decided to use these values only to verify the application of the optimization method. In future works these parameters will be modified, besides the coding form.

5. RESULTS AND CONCLUSIONS

In the sequence it is presented in the Table 5 (results optimized) and Figures 1 to 3 (evolution of optimization in 500 generations) the results obtained by optimization.

Table 5 shows the results of the parameters before and after the optimization. It may be verified that in the production Levels 1, 2 and 3 there were increase of the CA production and reduction of mass flows of catalyst with an increase of the amount of coolant fluid used in the process. Figures 1 (Level 1), 2 (Level 2) and 3 (Level 3) indicate improvements in the productivity. The results had been presented of normalized form. Taking into consideration the operation Levels 1, 2 and 3 there were increase of the CA production (increase of 0.0078 – Level 1, 0.0140 – Level 2 and 0.0179 – Level 3 – all values are normalized) with an reduction in the value for the catalyst flow (reduction of 0.1252 – Level 1, 0.1884 – Level 2 and 0.2585 – Level 3).

Table 5 Analysis of the performance of the CA production before and after the optimization for the production Levels 1, 2 and 3 (normalized values)

Parameters	Level 1		Level 2		Level 3	
	Before	After	Before	After	Before	After
Q_{r1}	0.2520	0.1158	0.0360	0.7895	0.0390	0.8430
Q_{r2}	0.2590	0.2524	0.0380	0.4920	0.0000	0.2744
Q_{r3}	0.2760	0.0124	0.2740	0.2434	0.0850	0.9258
Q_{r4}	0.0360	0.7095	0.0660	0.0516	0.3490	0.1193
Q_{r5}	0.0520	0.7036	0.1190	0.1550	0.1190	0.4813
Q_{r6}	0.0290	0.3337	0.1400	0.0700	0.0500	0.6179
T_0	0.0158	0.0176	0.0173	0.0171	0.0173	0.0184
Q_{cat}	0.0217	0.0154	0.0376	0.0282	0.0414	0.0285
Total Q_{CA}	0.6554	0.6632	10,000	10,140	0.9622	0.9801
Total Q_{ri}'s (Coolant)	0.9040	2.1273	0.6730	1.8016	0.6420	3.2617

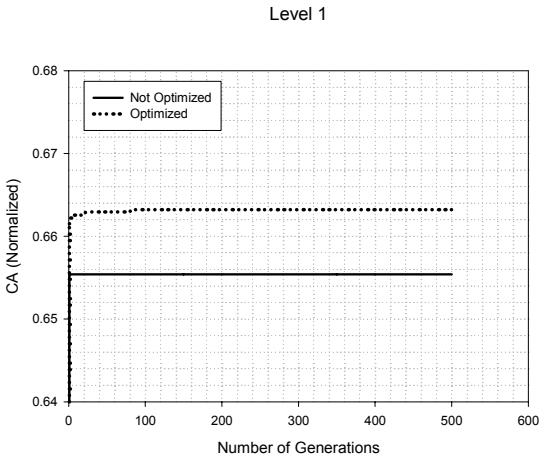


Fig. 1. Profile CA productivity (mass rate normalized) for production Level 1 with the optimization evolution.

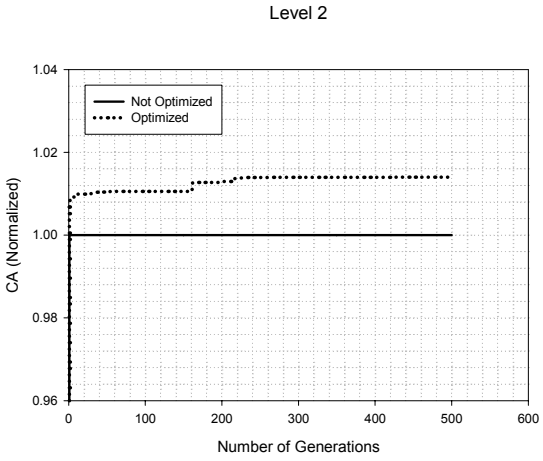


Fig. 2. Profile CA productivity (mass rate normalized) for production Level 2 with the optimization evolution.

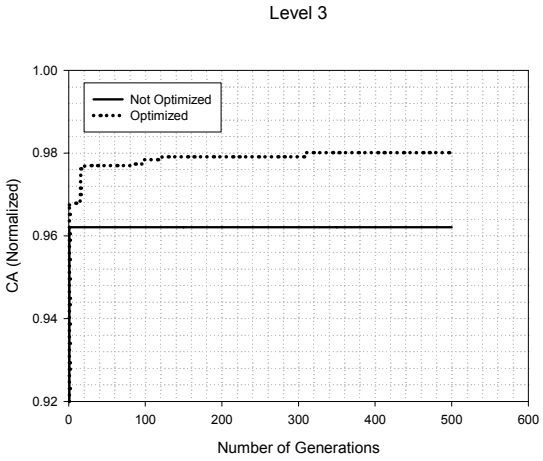


Fig. 3. Profile CA productivity (mass rate normalized) for production Level 3 with the optimization evolution.

The GA procedure revealed to be very efficient and robust for all the considered situations. Several testes with different population sizes, crossover and mutation values allow to conclude that the optimization by GA works well without be so dependent of its design values as well as the initial value. Optimization of the same problem by

conventional methods (as SQP) was not possible to be obtained in all the cases considered in this work.

In relation to the GA used in this study an attention has to be verified in some parameters this code. The population size used was of 20 and not of 50 or 100 as recommended (Carroll, 1996) because the computational time is very high. The crossover rate of 80% is satisfactory to supply good results. There are not significant changes when the number of generations is increased, therefore a number around 500 generations is enough to achieve the optimization. The mutation rates didn't follow the determined rules for the code. The values used for jump and creep mutation were: 0.01 and 0.02 respectively. These values allowed good efficiency, unlike what was usually recommended (Carroll, 1996). The GA code coupled to the reactor model showed to be a very efficient technique for reactor optimization. Similar problems or other systems can be studied for verification of his efficiency.

ACKNOWLEDGEMENTS

The authors are grateful to the Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP and to the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq for their financial support.

REFERENCES

Bäck, T., Fogel, D. B. and Michalewicz, T. (2000). editors. Evolutionary Computation 1: Basic Algorithms and Operators. Institute of Physics Publishing, 2000.

Carroll, D. L. (1996). “Chemical Laser Modeling with Genetic Algorithms”. AIAA Journal, Vol. 34, No. 2, February.

Coussemant, F. and Jungers, J. C. (1950). “La Cinétique de L´Hydrogénation Catalytique des Phénols”. Bull. Soc. Chim. Bel., vol. 59, pp. 295-326.

Deb, K. (1998). “Genetic algorithms in search and optimization: The technique and applications”. Proceedings of International Workshop on Soft Computing and Intelligent Systems, Calcutta, India: Machine Intelligence Unit, Indian Statistical Institute, pp. 58 – 87.

Froment, G. F. and Bischoff, K. B. (1990). “Chemical Reactor Analysis and Design”. John Wiley and Sons, 2ed., 664pp, New York.

Goldberg D. E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Publishing Company, INC.

Goldberg, D. E. and Deb, K. (1991). A Comparison of Selection Schemes Used in Genetic Algorithms, Foundation of Genetic Algorithms, edited by G. J. E. Rawlins, pp. 69-93.

Holland, J. H. (1992). Adaptation in Natural and Artificial Systems. University of Michigan Press, 2nd.

Hongqing C., Jingxian Y., Lishan K., Yuping C. and Yongyan C.. (1999). The Kinetic Evolutionary Modeling of Complex Systems of Chemical Reactions. Computers e Chemistry, 23, pp. 143-151.

Moros, R., Kalies, H., Rex, H. G. and Schaffarczyk, St. (1996). A Genetic Algorithm for Generating Initial Parameter Estimations for Kinetic Models of Catalytic Processs. Computers Chem.Engineering, Vol. 20, No. 10, pp. 1257-1270.

Santana, P. L. (1995). Mathematical Modeling for three phase reactor: deterministic, neural and hybrid models, PhD Thesis, School of Chemical Engineering, Unicamp, São Paulo, Brazil (in Portuguese).

Simant R. U. and Kalyanmoy D. (1997). Optimal Design of an Ammonia Synthesis Reactor Using Genetic Algorithms. Computers Chem. Engineering, Vol. 21, No. 1, pp. 87-92.

Syswerda, G. (1989). Uniform Crossover in Genetic Algorithms. In J. D. Schaffer (Ed.), Proceedings of The Third International Conference on Genetic Algorithms, pp. 2-9.

Vasco de Toledo, E.C.., Santana, P.L., Wolf-Maciel, M.R., Maciel Filho, R. (2001). Dynamic modelling of a three-phase catalytic slurry reactor, Chem. Eng. Sci., **56**, 6055-6061.

Wehrens, R and Buyders, M. C. (1998). Evolutionary Optimisation: A tutorial. Trends in Analytical Chemistry, Vol. 17, No. 4.

Session 6.3

Process Monitoring

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

A NOVEL MODULAR NONLINEAR NETWORK FOR FAULT DIAGNOSIS AND SUPERVISED PATTERN CLASSIFICATION

B. Bhushan[*] and J. A. Romagnoli[**]

** Department of Chemical Engineering*
The University of Sydney NSW, 2006, Australia
Email: bharat@chem.eng.usyd.edu.au
*** Department of Chemical Engineering, Louisiana State University,*
Baton Rouge, LA 70803
Email: jose@lsu.edu

Abstract: A novel modular network is proposed in this work for supervised pattern classification. The parameters of the hidden layer are determined using polygonal line algorithm. No further training of the network is required. Firstly, an abnormality is detected and responsible sensors identified using polygonal line based radial basis function network algorithm. Furthermore, the proposed strategy is applied for fault diagnosis. A continuous pilot plant is selected as the case study to show the efficiency of the proposed strategy. The result shows that, the proposed framework is a promising direction towards fault detection and diagnosis in real time, non-linear systems.

Keywords**:** Fault detection, Fault identification, Fault diagnosis, Nonlinear PCA, Polygonal lines, Modular network

1. INTRODUCTION

The advent of faster and more reliable computer systems has revolutionized the manner in which industrial processes are monitored and controlled. Once thought of as just data logging and storage units, these computer systems now perform sophisticated computer-based control strategies, and real-time simulation and optimization. These advances have resulted in the generation of a large amount of process data, yet the task of interpreting and analysing these data is daunting.

Fault detection and diagnosis is the primary module for any process monitoring framework. Principal component analysis (PCA) and projection to latent structure (PLS) are one of the most used multivariate statistical process control (MSPC) techniques. The major drawback of this method is that, it assumes linear correlation between data which is not always true in case of process data that are generally nonlinearly correlated. However, the philosophy behind these approaches is to reduce the dimensionality of the problem by forming a new set of latent variable to obtain an enhanced understanding of the process behaviour.

Many methodologies have been proposed for nonlinear principal component analysis (NLPCA). Kramer (1991) proposed a NLPCA based on five layer auto associative neural networks. Dong and McAvoy (1996) proposed NLPCA based on principal curves and neural networks. The principal curve method was used to calculate the associated score and corrected data point for each original data point. But, since principal curve method does not produce a nonlinear principal component in the sense of principal loading, Dong and McAvoy (1996) developed an alternative approach based on multi layer perceptron to model the calculated data. Two three layer neural networks were trained separately to map the data to lower dimensional feature space and remapping the data back to the sample space. The number of hidden layer nodes was decided using cross validation scheme. A methodology based on

"well – defined" architecture of radial basis function (RBF) network and polygonal line (PL) has been suggested by Bhushan and Romagnoli (2005) for dimensionality reduction and fault detection. The data of the normal operating region is used to fit the polygonal lines and the output generated has been used for determining the architecture of the network and to train the model. Online data are projected to the RBF-PL model and an abnormality is indicated whenever the prediction is significantly different from the projected measurements. Furthermore, the measured variables that makes significant contribution towards the deviation in the model prediction is identified. However, this information is insufficient for the operator to find the root cause, since the operator needs to infer the root cause which is difficult in case of process with large number of variables.

Vedam and Venkatasubramanian (1999) proposed an integrated approach based on PCA and signed digraphs (SDG) for fault detection and diagnosis. Fault detection is performed using PCA. Whenever an abnormality is detected, the contribution of measured variables is presented as an input to the SDG to perform fault diagnosis.

Leonard and Kramer (1992) suggested a decomposition strategy based on modular neural network approach for solving large scale fault diagnosis problems. Though, RBF networks are many time faster than similar back propagation networks (BPN); it still required large computational resources. Two decompositions are proposed for this work: decomposition in time, reducing the dimensionality of the input space; and decomposition among the fault classes, reducing the size of the training set for each subnet.

In this work, a novel modular network is suggested to accomplish the diagnosis task. The advantage of this methodology over others is that it uses the same PL algorithm of fault detection to decide the architecture and related parameters of each module of the network. Furthermore, there is no additional training required of the network and hence it is computationally very less expansive. The output of the proposed network can be the partial belongingness of the input pattern to more than one fault classes and the strength of the fault.

The remaining part of the paper is organized as follows. In section 2, a brief introduction of RBF-PL methodology for fault detection and identification is given. The proposed network for classification is explained in section 3. Section 4 contains the results and discussion on the application of the entire strategy to a real time pilot plant environment. Finally, section 5 contains the conclusion and future direction of the work.

2. FAULT DETECTION AND IDENTIFICATION

Polygonal line algorithm proposed by Verbeek, *et al*., 2002 is used for fitting the data. Each data point is projected orthogonally onto the PL. Thus for each data points there are corresponding lengths $t_1, t_2, \ldots, t_n$ along the curve where n is the number of data points in d-dimensional space.

In analogy to PCA, this length represents the non-linear scores of the data points. Thus the sample vector can be represented as

$$X = f_1(t(X)) + E_1 \quad (1)$$

where t is the non-linear principal component score and $\boldsymbol{E_1}$ is the residual vector.

The next non-linear component score can be found by projecting the data points of $\boldsymbol{E_1}$ on the PL constructed using $\boldsymbol{E_1}$. These steps are repeated until all the information is extracted. It is found that the first few nonlinear principal components explain most of the variance of the dataset. Though, this method is quite effective in reducing the data dimensionality, it is to be noted that $\boldsymbol{f}$ has no parametric form**,** and it is quite cumbersome and memory expensive to use it for online application. A RBF network is trained to model the relationship. This network is further used for fault detection and identification. A non-parametric approach based on kernel density estimation (KDE) is used to determine the confidence limit. The detail of this methodology can be found in Bhushan *et. al.* (2005).

3. SUPERVISED CLASSIFICATION AND FAULT DIAGNOSIS

It should be noted here that each segment of the PL in the sample space represents a localised region around which the data is concentrated. We propose that each of these regions in the input space can be represented by a multidimensional Gaussian function (Figure 1).

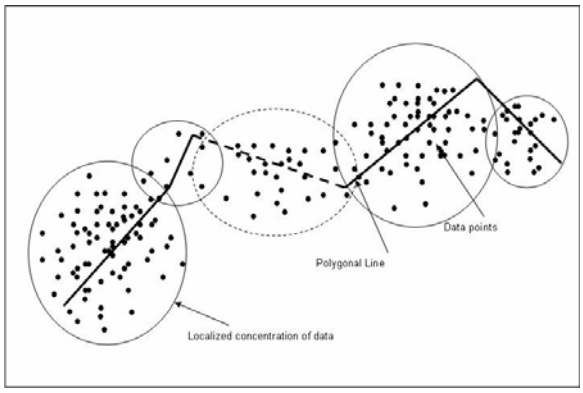


Fig. 1: Schematic representation of the region covered by each segment of PL.

The Gaussian function with equal spread in all the directions is defined as:

$$\varphi(x;c,\sigma) = exp\left(-\frac{||x-c||^2}{2\sigma^2}\right) \qquad (2)$$

where $\| \boldsymbol{x} - \boldsymbol{c} \|$ is the Euclidean distance of $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$ from the vector centre $\boldsymbol{c} = (c_1, c_2 \ldots c_d)$ and σ is the spread. When the spread of the data points is not uniform, a multidimensional Gaussian function takes the form

$$\varphi^{\,i}(x,c^i,\sigma^i) = exp\left(-\left[\frac{(x_1-c_1^i)^2}{2\sigma_1^{i\,2}} + \ldots + \frac{(x_d-c_d^i)^2}{2\sigma_d^{i\,2}}\right]\right) \qquad (3)$$

where c^i is the centre of the region and is defined as the mean of the data contained in the region i and σ^i represents the standard deviation of the dataset in the region.

Since it is the supervised classification, the class of each training data set is known in advance. The training data set is grouped according to its class. Each group is presented to the PL algorithm. The number of segments required to fit the polygonal line is found out. However, segments which have just been used to construct the PL and do not contain any data points are neglected.

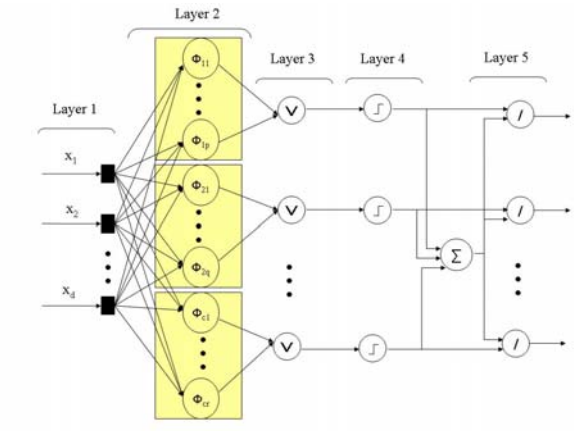


Fig. 2: Proposed modular fault diagnosis network.

Figure 2 shows the proposed modular fault diagnostic system. The system contains five layers. Nodes at layer one are input nodes representing the input variables and the last layer is the output nodes. The number of nodes in the output layer is same as the number of fault classes. A node at layer two represents a region in the domain of a specific fault, in other words, it is one of the segment of the PL which fitted that fault class. The nodes of a fault class constitute one module in layer two and there will be as many modules as the number the fault classes. The number of nodes at layer three is the same as the fault classes and each node of this layer is linked to the nodes of only one module of layer two. The nodes in this layer calculate the maximum strength of the input data in that particular class. Each node in layer four is linked with only one node of layer three and decides whether the strength of the belongingness is strong enough to be considered as a fault or not. Finally, the nodes in layer five decide the contribution of each fault class to the abnormality.

The function of a node in each layer of the proposed network is described in detail next.

Layer 1: The input vector is presented to the nodes of this layer. The number of nodes is same as the dimension of the input vector and each element is linked to one of the node of this layer. The nodes in this layer transmit the input data to the next layer without any change.

The output from this layer y_j^1 is defined as

$$y_j^1 = x_j^1, \quad j = 1, \ldots, d \qquad (4)$$

Where y_j^1 denotes the output of node j in layer one and x_j^1 denotes the input to node j at layer one.

Layer 2: The output of each node of layer one is presented to each node of this layer. This is one of the most important layers of this network. As mentioned earlier, the parameters of each node are determined using the data of that region.

$$c_{j,c}^k = \frac{\sum_{i=1}^{I_k} x_{j,c}^i}{I_k} \qquad c = 1,2,\ldots,C;\ j = 1,2,\ldots,d \quad (5)$$

where $c_{j,c}^k$ is the centre of the region k in class c, I_k represents the data points in region k and C is the number of classes or modules. The spread of k^{th} node in class c is defined as

$$\varsigma_{j,c}^k = \delta.\sigma_{j,c}^k \quad c = 1,2,\ldots,C;\ j = 1,2,\ldots,d \qquad (6)$$

where $\sigma_{j,c}^k$ is the standard deviation of the input vectors contained in region k of class c in j^{th} dimension and δ is a user defined parameter which ensures the optimum receptive field covered by each region.

A Gaussian membership function is constructed with $c_{j,c}^{k}$ and $\varsigma_{j,c}^{k}$ as the centre and the width respectively. Each node in this layer is represented by one such function and all the nodes generated by using segments of the PL of a group constitute a module. Therefore, we have as many modules as the number of classes. The output from this module defines the belongingness of a input vector to a particular region.

$$y_c^{2,k}(x, c_c^k, \varsigma_c^k) = exp\left(-\left[\frac{(y_1^1 - c_{1,c}^k)^2}{2\,\varsigma_{1,c}^{k\,2}} + \ldots + \frac{(y_d^1 - c_{d,c}^k)^2}{2\,\varsigma_{d,c}^{k\,2}} \right] \right) \tag{7}$$

where $y_c^{2,k}$ is the output from node representing region k of class c in layer two.

Layer 3: The output from each module is fed to not more than one node of this layer. Hence the number of nodes is same as the number of fault classes. The output of the node from this layer represents the strength of the belongingness of the input data in a particular class. An input vector may belong to more than one region of the same class; however, the belongingness of the vector in a class is dictated by the maximum membership value. Therefore, the output of each node from this layer is defined as:

$$y_c^3 = max(y_c^{2,1}, y_c^{2,2}, \ldots, y_c^{2,k}) \tag{8}$$

where *k* represents the number of regions in class *c*.

Layer 4: The nodes in this layer decide whether the output from previous layer is strong enough to assign the data in to a particular class or not. This task is accomplished by a function defined as follows:

$$y_c^4 = \begin{cases} y_c^3 & if \;\; y_c^3 \geq \lambda \\ 0 & if \;\; y_c^3 < \lambda \end{cases} \tag{9}$$

where λ is a user defined parameter.

Layer 5: This layer is the decision making layer. It gives an idea to the operator which fault is more severe if there are multiple faults. The output from this layer is defined as:

$$y_c^5 = \frac{y_c^4}{\sum_{k=1}^{C} y_k^4} \tag{10}$$

It should be noted that once the network is build there is no further requirement of the training since there is no weight adjustment required.

4. APPLICATION TO PILOT PLANT ENVIORNMENT

To test the overall strategy in real time, a general-purpose pilot plant facility is used. The process contains two CSTRs, a mixer, a feed tank and a number of heat exchangers.

Each CSTR consists of a reaction vessel, a steam jacket, a cooling coil and a stirrer. Material from the feed tank is heated before being fed to the first reactor and the mixer. The effluent from the first reactor is then mixed with the material in the mixer before being fed to the second reactor. The effluent from the second reactor is, fed back to the feed tank and the cycle continues. The pilot plant is well instrumented to provide many possible control scenarios and configurations.

Nine variables [Fin (feed flow rate in), Tin (temperature of feed in), Tc,in (temperature of cooling water in), Ts,in (temperature of steam in), Lvl (level of the reactor), Fout (feed flow rate out), Tout (temperature of feed out), Tc,out (temperature of cooling water out), Ts,out (temperature of condensate)] related to the first CSTR are considered for this study. Once the plant reached its normal operating condition, 100 training data points at 5 second interval were collected. All variables in the training data set were normalized in order to give equal weights to each. The training data set were exposed to PL algorithm (kmax = 24) and the nonlinear scores were found by projecting the data point onto the polygonal line. The residual was calculated and was exposed again to PL algorithm and so on. The PL algorithm fitted the training dataset into nine segments yielding a RBF mapping network with nine input nodes, nine hidden layer nodes and two output nodes. The centre and spread of the hidden layer nodes were calculated using the centre and standard deviation of the segments and hence only weights of the output layer were to be calculated. GA with 100 generations was used to first get near an optimum solution followed by BFGS Quasi-Newton algorithm for faster convergence. The total time taken for training was 12.80 seconds. The demapping layer with two input nodes, nine hidden layer nodes and nine output layer nodes was trained using the similar strategy, however all the parameters were trained and hence the training time was 35.46 seconds.

KDE is used for finding the two warning limits at 95% and 99%. 95% and 99% warning limit was found to be 9.19592 and 10.8476 respectively. SPE along with these warning limits were constructed to facilitate fault detection. Any violation to the 95% warning limit, fault identification algorithm is triggered to identify the measurements responsible for out of control signal followed by the fault diagnosis model to decide the root cause.

For this paper, two different fault scenario were planned. Firstly, process condition under normal operation was achieved that was similar to when data were collected to train the network. Secondly, to simulate a process upset scenario, the feed flow rate was increased from 0.6 l/min to 1.0 l/min. This change affected many other variables including the reactor level and the effluent flow rate. For simulating the single sensor failure condition, a random bias of mean 6 (30% of the actual) and standard deviation 2 was added to the feed temperature. 100 values at 5 second interval for all these conditions were captured.

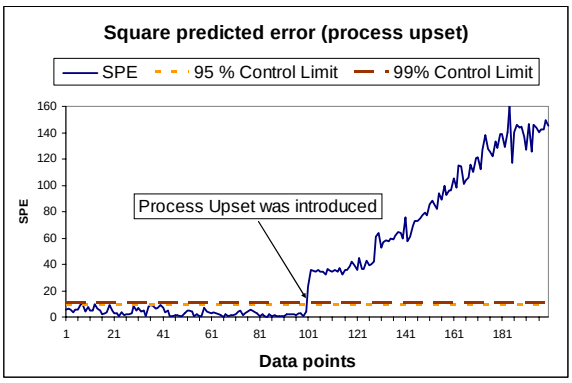
The data of these three conditions (normal, process upset and sensor failure) were fed to the PL algorithm which fitted it into 9, 22 and 12 regions respectively. Therefore, the structure of the fault diagnosis network was 9-43-3-3-3. The value of δ and λ used are 1.5 and 0.05 respectively. The result of the fault detection and diagnosis are as follows:

1) Process Change (Flow rate increased from 0.6 l/min to 1.0 l/min)

Figure 3(a) and 3(b) shows the SPE and contribution plot for process upset respectively. It should be noted that as soon as feed flow rate was increased, SPE crossed both the warning limits and was well above this condition throughout the period this condition prevailed.

However, the contribution plot shows that though the contribution by feed flow rate was high in the beginning, in the later part, prime contribution was due to the level measurement which is readily expected for this process. Also, this change in the process condition affected the effluent flow rate. From the process point of view, these three variables are related to each other and they are also being reflected in the result.

(a)



(b)

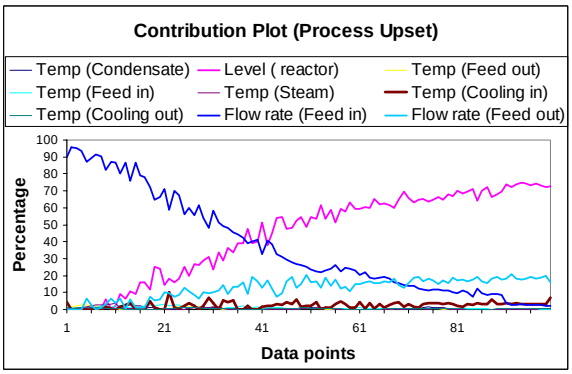


Figure 3: (a) Square predicted error plot (b) contribution plot in case of process upset.

Table 1: Results of the fault diagnosis in case of process upset

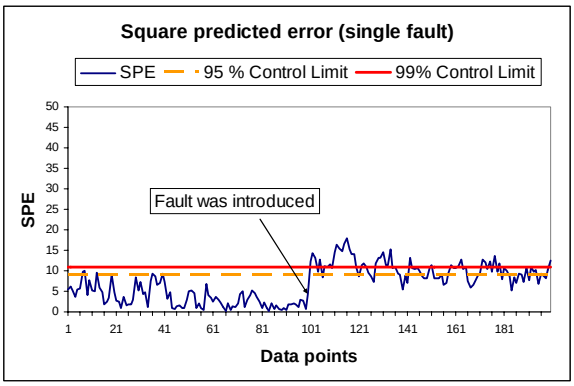
Normal	Process Upset	Sensor Failure
0.00	1.00	0.00
0.00	1.00	0.00
0.00	1.00	0.00
0.00	1.00	0.00
0.00	1.00	0.00
0.00	1.00	0.00
0.00	1.00	0.00
0.00	1.00	0.00
0.00	1.00	0.00
0.00	1.00	0.00

Table 1 shows the result of the proposed fault diagnostic system. The result for first ten points is only shown however it detected correctly for all the testing data. In the result, 0 indicates that this condition is not prevailing in the process whereas 1 indicates that according to the knowledge of the network this condition is 100% present.

2) Sensor failure (a random noise of 30% of the actual with std. dev of 2 was added to feed temperature)

In case of sensor failure, SPE is above the warning limits in most cases (Fig. 4(a)), though in some cases the magnitude of SPE is not very high because of the presence of noise in the measurements the normal feed temperature is quite close to the value in case of the faulty sensor (maximum normal feed temperature: 28.83 0C, minimum feed temperature in case of sensor fault: 31.13 0C). The contribution plot (Fig. 4(b)) clearly identify feed temperature sensor as the sensor which has the highest contribution in fault.

(a)



(b)

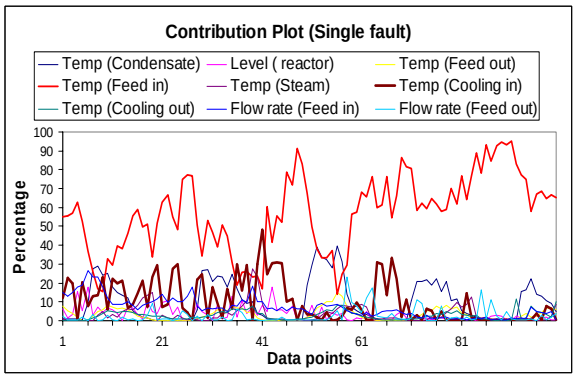


Figure 4: (a) Square predicted error plot (b) contribution plot in case of single fault.

The result of the fault diagnostic module is shown in table 2. It should be noted that a module for this fault was already included in the network; hence it did diagnose all the conditions correctly.

Table 2: Results of the fault diagnosis in case of Sensor failure

Normal	Process Upset	Sensor Failure
0.00	0.00	1.00
0.00	0.00	1.00
0.00	0.00	1.00
0.00	0.00	1.00
0.00	0.00	1.00
0.00	0.00	1.00
0.00	0.00	1.00
0.00	0.00	1.00
0.00	0.00	1.00
0.00	0.00	1.00

5. CONCLUSION AND FUTURE WORK

In this work, a novel modular network is proposed for supervised classification. The key feature of this network is its simplicity and less computational complexity. First, the training data were separated into different classes, and each set was fed to the PL algorithm for finding the optimum number of regions in that class. Each region was used to find out the parameters of the network. There was no further training required. This work is integrated with non linear PCA based on RBF and PL for simultaneous fault detection and diagnosis. The proposed methodology is used for monitoring the condition of a continuous pilot plant. Two different types if abnormalities were simulated to test the capability of the framework. The results show that the fault was detected and diagnosed in both of the cases correctly. However, the model needs to be validated for controller faults and multiple faults which will be a part of future work.

6. REFERENCES

Dong, D., and T.J. McAvoy (1996). Nonlinear Principal Component Analysis-Based on Principal Curves and Neural Networks. *Computers Chem.. Engng.*, **20**, 65-78.

Bhushan, B., J A Romagnoli and D. Wang (2005). Fault Detection Using Radial Basis Function Network and Polygonal Line. *16th IFAC world congress*, Prague, Czech Republic.

Kramer, M. A. (1991). Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. *AIChE Journal*, **37**, 233-243.

Leonard, J. A. and M. A. Kramer (1993). A Decomposition Approach to Solving Large-Scale Fault Diagnosis Problems with Modular Neural Networks. *IFAC Symposia Series*,1, 237-242.

Looney, C.G. (1996). *Pattern Recognition Using Neural Networks*. Oxford University Press, Oxford.

Verbeek, J. J.,, N. Vlassis and B. Krose (2002), A k-segments algorithm for finding principal curves. *Sample Recognitions Letters*, **23**, 1009-1017.

Venkatasubramanian, V., R. Rengaswamy, S. N. Kavuri and K. Yin (2003), A Review of Process Fault Detection and Diagnosis Part III: Process History based Methods. *Computers and Chemical Engineering,* 27, 327-346.

Vedam, H. and V. Venkatasubramanian (1999), PCA-SDG based Process Monitoring and Fault Diagnosis. Control Engineering Practice, 7, 903-917.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

BLOCK DIAGRAM PROPOSAL OF PROTECTION SYSTEM FOR A PWR NUCLEAR POWER PLANT

Francisco Joailton de Lima[(1)], Claudio Garcia[(2)]

Telecommunications and Control Engineering Department
Escola Politécnica of the University of São Paulo
Av. Prof. Luciano Gualberto, Trav. 3, 158 - Butantã
City: São Paulo - SP - Zip code: 05508-900 - Brazil
[(1)]francisco.lima@poli.usp.br, [(2)]clgarcia@lac.usp.br

Abstract: This text presents a block diagram proposal of protection system for a PWR nuclear power plant. It describes the plant operation, as well as it defines what a protection system is. Some of the main inherent definitions to protection systems are shown and the system operation is explained in a global form. The block diagram proposed for the protection system can be used as a project foundation for the protection automation. *Copyright © 2006 IFAC*

Keywords: Fault tree, nuclear reactor, protection system, reliability, safety.

1. INTRODUCTION

There are approximately 440 nuclear power plants (NPP) in the world, according to figure 1. However, this kind of energy generation is still considered as a threat to human life and with a great potential to cause grave accidents, although modern NPP comply to strict project, building and operation criteria, which make them very safety.

The two greatest accidents that contributed to this insecurity representation of the nuclear power plants occurred at the TMI – Three Mile Island, Pennsylvania, in the United States on March, 28th , 1979 (Chairman at all, 1979) and at Chernobyl in the former Union of Soviet Socialist Republics - where today lies Ukraine - on April, 26th, 1986 (Edwards, 1987). In the first accident (TMI), despite having happened heating and deterioration of the reactor core, there was little emission of radioactive material into the environment. In the second accident (Chernobyl) occurred an explosion of the reactor and a great quantity of radioactive matter was spread into the atmosphere.

Besides other safety devices existing in a nuclear power plant such as the containment building, where the reactor is located which - if ever existed at the Chernobyl NPP - could have slowed down the amount of radioactive matter disseminated into environment, there is the protection system for a nuclear power plant that is the scope of this study.
At present, The Brazilian Navy is working on the LABGENE (Electrical Core Generation Lab) project at the Navy Technological Center in São Paulo (CTMSP), which consists of the development and construction of a nuclear power plant to generate electrical energy that will act as a basis and developing laboratory to another projects of nuclear reactors in Brazil (INFOREL, 2004).

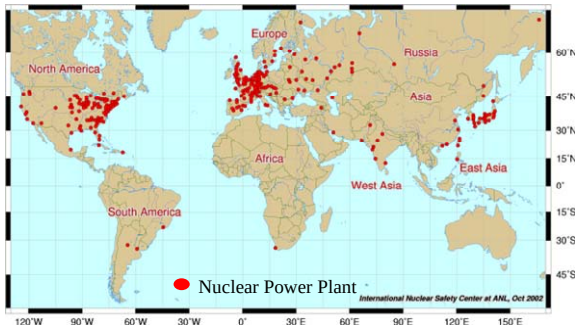


Fig. 1. Maps of Nuclear Power Reactors: WORLD MAP (Nuclear Plants, 2005).

As far as we know, articles presenting the block diagram of the protection system, as proposed here, have not been published yet. Thus, this is one of the reasons that led us to write it.

2. PWR NUCLEAR POWER PLANT

The acronym PWR stands for the English term “Pressurized Water Reactor”. This name is derived from the fact that in this kind of plant its cooling system is obtained from a pressurized water circuit.

The operating principle of the power plant illustrated in figure 2 is the following: the fission heat in the reactor core is used to increase the water’s temperature (coolant) of the primary circuit, the steam generator absorbs the heat from this water, transferring it into the secondary circuit in the form

of steam; the produced steam runs a turbine that transmits mechanical energy to an electrical generator, which, then, converts the mechanical energy to electricity.

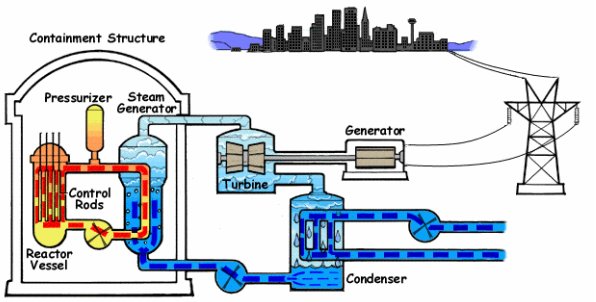


Fig. 2. Pressurized Water Reactor (Nuclear Reactors, 2005).

3. PROTECTION SYSTEM

Below are presented some definitions and is proposed block diagram of the protection system for a PWR nuclear power plant.

3.1 Definitions

3.1.1 The Plant Protection System

The protection system of a nuclear power plant has the function to take the necessary measures to avoid accidents which may lead to overheating and degeneration of the reactor core as well as contamination of the environment caused by radioactive matter. Moreover, it makes sure the plant will operate within the safety limits stated in the project. These actions consist on turning off the reactor, bringing to an end the fission of the fuel element and/or activating the protection system such as emergency diesel-generators, valves, emergency cooling-system, etc.

3.1.2 Reliability

A very important requirement to the protection systems is their reliability, which can be defined as the probability of a system to work accurately (according to its project specifications) within a time interval $[t_0 \ ; t]$, in which it was working properly at the starting point t_0.

3.1.3 Redundancy

Consists in the use of more than one equipment to execute the same function, that is, in case of failure of the first device, the next can be able to warrant the continuity of that function in the system, increasing the system reliability. This procedure is applied to protection systems, e.g. when three sensors are used to measure a certain temperature variable. If two of these measures are above a pre-established value ("set point"), the system acts by disconnecting, alarming and/or activating the security system. When that happens, it is said that occurred a voting of 2 out of 3.

3.1.4 Fault Tree

Fault Tree is a tool used for the analysis of the reliability in protection systems. A fault tree represents a system or a subsystem through a diagram that has a top event which occurs from a combination of other events. This combination is represented by symbols that interconnect those events by means of logical operations such as "AND", "OR", etc (McCormick, 1981). A qualitative analysis can be performed by checking in the fault tree which basic events and paths lead to the occurence of the top event. On the other hand, the quantitative analysis is applied when it is possible to determine the probability of the top event to occur because of the probabilities of the basic events. The diagram permits us to visualize the fault sequences which must happen in order to the top event to occur. After the building of the fault tree, it is possible to insert the probabilities of each represented event to occur. In this way, it is possible to calculate the probability of the top event to occur.

3.2 Block Diagram Proposal

Figure 3 presents the suggested block diagram for a protection system in a nuclear plant. The blocks to the left of the dotted line are directly linked to the security function, while the ones to the right are used for audits and interface with the operator in the form of alarms and indications. Each block that comprises the diagram presented in figure 3 will be described below.

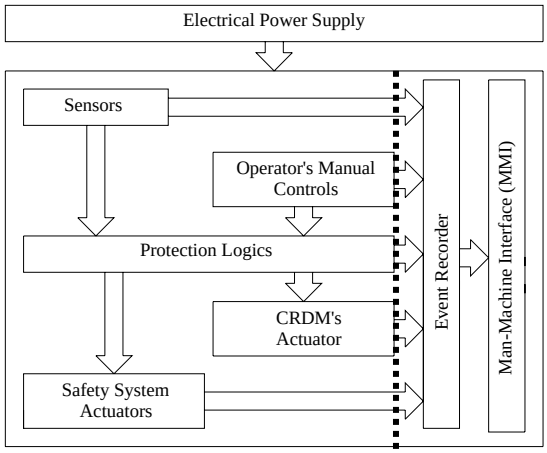


Fig. 3. Block Diagram Proposal of the Protection System.

3.2.1 Electrical Power Supply

It is the common component to all other system blocks, responsible for the supply of electrical energy. This block is normally comprised of more than one source of electrical energy, such as energy produced by the plant itself, batteries, diesel-generators and, whenever available, the electrical wiring of the energy supplier company, originated

from other type of generating source (hydroelectric, or thermal powers, for instance).

3.2.2 Sensors

The sensor is an equipment that responds to a certain physical phenomenon and conveys this response to another component of the system that will apply it to control a process. In the case of the protection system, the main used sensors are to measure temperature, pressure, level, voltage, neutron flow (or neutron detectors) and radioactivity. According to the values given by the sensors and operator manual controls, the system protection processes the protection logics. This fact reveals the fundamental relevance of those components without which it is not possible to operate or manage the power plant in a safe way.

3.2.3 Operator's Manual Controls

These are the controls originated by the nuclear power plant operator. The manual command of turning off the reactor due to any abnormal condition recognized by the plant operator has the highest priority over any automatic action of the protection system, because it cuts off the power supply to the CRDM'actuator (Actuator of the Control Rod Drive Mechanism) which, as a result, shuts the reactor down through the fall of the control rods into the reactor core (Glasstone, 1994). In case of operator omission, the protection system starts to operate when an abnormal condition is detected.

The increase of the reactor power occurs gradually since its initial cold condition until the nominal power of operation, usually called the start-up procedure. During this procedure, the operator generates the range change controls for the detection of the neutron flow according to the plant operation conditions (neutron flow, temperature and pressure); such ideal conditions are guaranteed to be within the safaty limits by the protection system. If any limit condition is exceeded, the protection system shuts the reactor down and, at the same time, starts the safety systems necessary to remove the residual heat and sustain the reactor integrity.

3.2.4 Protection Logics

At this block, system decisions are made in conformity with the inputs originated from the sensors and the operator manual controls. The protection logics are intended to guarantee that the plant operation safety boundaries are not exceeded.
Below are listed the protection logics which lead to the scram of water-cooled reactors (Glasstone, 1994) and/or the activation of safety systems:

- Quick increment of the neutron flow during the start of the reactor;
- High flow of neutrons at the range power, indicating over power;
- Abnormal pressure and temperature;
- Loss of coolant water;
- Damage of the steam line;
- High level of water in the pressurizer in PWR;
- Low level of water in the BWR reactor vessels ("BWR – Boiling Water Reactor");
- Low level of water in the steam generator in PWR;
- High level of radiation in the steam; and
- Low voltage or power loss (safety buses).

3.2.5 Actuator of the Control Rod Drive Mechanism

The actuator of the Control Rod Drive Mechanism is the main component for reactor control and scram. This component enables to remove or to insert the control and safety rods (rods that absorb radiation) in the interior of the nucleus, increasing or decreasing, respectively the nuclear power (McCormick, 1981). This device has a solenoid that, when it is turned off, makes possible the abrupt fall of the rods inside the nucleus by action of the gravity and springs, what provokes the reactor shutdown. This shutdown form is fail-safe because the electric power failure the turns off the reactor.

For effect of plant control, the actuator position indicator is important because by manipulating the rods position, the nuclear power can be controlled. However, for the protection system, it will only be important the information of turned on or off actuator.

3.2.6 Safety System Actuators

When the protection system identifies an abnormal situation, besides shutting the reactor down, it takes actions to assure the safety of the plant. The action of the protection system, when an abnormal situation is detected, consists of one or more of the following actions:

- Turns on diesel-generator to assure the power supply;
- Activates the Emergency Cooling System to assure the nucleus cooling and integrity. This system consists in injection of boron water through circulation pumps or nitrogen pressure;
- Isolates damaged steam line;
- Turns off circulation coolant pumps to minimize coolant loss, in the case of LOCA - Loss-of-coolant accident; and
- Isolates the containment (the place where the reactor is installed), avoiding the radioactive material release to the atmosphere.

3.2.7 Event Recorder

All of the blocks of the protection system are interconnected with the event recorder, where the information is stored during the operation of the plant. The recording of those information will make it possible to trace subsequently the reactor scram or the activation of safety system cause, since an appropriate physical media is employed (hard disk, magnetic storage, optical tape recorder, memory,

etc), chosen in function of the hardware specification in the project.

3.2.8 Man-Machine Interface (MMI)

This block reads the data recorded by the event recorder and it introduces them to the operator in alarm or report form. The alarms will be used on-line during the operation of the plant and they are displayed in visual form and/or audible in the panel of the protection system, while the reports will be used later for evaluation and supervision of the system, as well as for finding out the causes of reactor scram or activation of safety systems, and they are displayed in printed form in agreement with the information requested by the user.

4. CONCLUSION

Starting from the reliability level required to the protection system that the project is aimed at, having the tools such as the fault tree or Markov model, and also after the detailing of the block diagram, the study will seek to establish the necessary redundancy for the components of the protection system.

The components of the protection system connected to the safety function are of the 1E class (IEEE Std 603, 1998). This classification is given to equipments or safety systems that are essential for the scram of the nuclear reactor, insulation of the containment and the cooling of the reactor, in order to avoid radioactive matter emission to the environment.

Although digital electronics has evolved over the years, there is still prudence in using different kinds of software in safety systems. As an illustration, digital protection was not considered to be used in China before the year 1990. The first digital system of the kind only appeared in 2002 for a 10 MW reactor (Li, F.; Yang, Z.; An, Z.; Zhang, L, 2002).

The great challenge for the safety system project designer will be to detail each component of the block diagram proposal, organize ways to test the integrated system and guarantee the reliability. The expectation is to employ this work in the development of the safety system at LABGENE in the future.

REFERENCES

Chairman, John G. Kemeny at all. *Report of The President's Commission on The Accident at Three Mile Island (TMI).* Washington, DC. 1979.

Edwards, M. Chernobyl-One Year After. National Geographic, May 1987.

Glasstone, Samuel; Sensonske, Alexandre. *Nuclear Reactor Engineering: Reactor Systems Engineering.* 4th ed. New York: Chapman & Hall, 1994. v.2.

IEEE Std 603-1998. *IEEE Standard Criteria for Safety Systems for Nuclear Power Generating Stations.* New York: Institute of Electrical and Electronics Engineers. 1998.

INFOREL – Relações Internacionais, Notícia e Informação. Brasília. *Interview with the Brazilian Navy Minister in December 10, 2004.* Available at: <http://www.inforel.org>. Access in February 3, 2005.

Li, F.; Yang, Z.; An, Z.; Zhang, L. (2002). *The first digital reactor protection system in China.* Nuclear Engineering and Design, n. 218, p. 215–225.

McCormick, Norman J. *Reliability and Risk Analysis – Methods and Nuclear Power Applications.* Department of Nuclear Engineering – University of Washington Seattle, Washington. California: ACADEMIC PRESS, 1981. 446p.

Nuclear Plants around The World. Available at: <http://www.nucleartourist.com/ >. Access in February 21, 2005.

Nuclear Reactors. *U. S. Nuclear Regulatory Commission – Pressurized Water Reactors.* Available at: <http:// www.nrc. gov/ reactors/ pwrs.html>. Access in May 22, 2005.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PERFORMANCE ASSESSMENT OF MODEL PREDICTIVE CONTROL SYSTEMS

Oscar A. Z. Sotomayor, Darci Odloak

LSCP – Department of Chemical Engineering
Polytechnic School of the University of São Paulo
Av. Prof. Luciano Gualberto, trav.3, n.380, 05508-900 São Paulo-SP, BRAZIL
oscar@pqi.ep.usp.br, odloak@usp.br

Abstract: This paper aims at to propose a benchmark MPC controller to be used in the performance assessment of existing industrial MPC systems. The basic questions are how the performance could be evaluated in a realistic basis and how to judge the performance of a controller that is already in operation by comparing it with another controller that could be really implemented in the same system. Here, it is assumed that the ideal controller will inherit the structure, input constraints and tuning parameters of the controller whose performance is to be evaluated. This means that the design of the ideal controller is standard and there is no need to tune the performance assessment algorithm. It is proposed a controller that preserves closed loop stability for any adopted tuning parameters. This is requisite for any performance evaluation procedure that is expected to operate in an on-line scheme. The proposed controller is compared by simulation with other benchmark controllers proposed in the control literature.

Keywords: Performance assessment, Model predictive control, Controller performance monitoring, Constrained control systems, Industrial process control.

1. INTRODUCTION

Model predictive control (MPC) strategies, such as the ones based on dynamic matrix control (DMC), have become the standard control alternative for advanced control applications in the process industries (Qin and Badgwell, 2003). It is a market, which is growing at a compound annual rate of approximately 18% (Automation Research Corporation, 2000), and substantial benefits are generated directly from the ability of MPC to ensure that the plant operates at its most profitable constraints. But, as most control algorithms, after some operation time, MPC is seldom performing as when it was commissioned. It is common to find MPC applications delivering only 50% of the expected benefit when the assessment is made 2-3 years after commissioning (Treiber et al., 2003). MPC controller design and tuning involve many uncertainties related to approximate process models, estimation of disturbances and assumptions about operation conditions. A surprising high percentage of the implemented MPC controllers suffers degradation in terms of the achieved performance as a result of changes in process dynamics, sensor/actuator failure, estimator bias, equipment fouling, feedstock variability, changes in product specifications, etc. Therefore, to sustain the benefits of MPC systems over a considerable period of time, the performance needs to be monitored and assessed on a constant basis. This task has proven to be a much greater challenge (Hugo, 2000; Shah et al., 2001) than initially expected and it requires the presence of effective tools to establish the root causes of the poor control quality and to define the need to retune if necessary.

Practical applications of controller performance assessment (CPA) have triggered an increasing interest of academia and industry in the development of a benchmark MPC controller. Several CPA techniques have been proposed in the literature during the last years, some of them being incorporated into commercial software packages. But, in general, all the CPA techniques explicitly or implicitly involve comparison of the current controller quality with a theoretical benchmark, i.e. an ideal controller that could never be implemented.

CPA techniques can be divided into two major categories (Qin, 1998): stochastic and deterministic methods. Stochastic CPA methods evaluate the closed-loop performance for zero-mean changes, such as random disturbances, measurement noise, etc. These techniques utilize stochastic measures such as variance to evaluate the performance of the controller. In this area, the most notable work is by Harris (1989) that proposed the use of the minimum variance controller (MVC) as a benchmark to assess the performance of SISO feedback controllers. On the other hand, deterministic CPA methods are concerned with non-zero mean changes in the set-point or load disturbances and utilize deterministic measures such as settling time, integral square error (ISE), rise time, etc. Aström (1991) discussed some alternatives to evaluate the performance of PID controllers. Although MVC benchmarks bring up important aspects of the controller performance, deterministic methods are more informative and present a more practical way of assessing controller performance. Results from statistic and deterministic CPA methods usually cannot be best achieved simultaneously (Qin, 1998).

In this paper it is proposed a systematic CPA framework for MPC systems with focus on set-point tracking. The methodology utilizes the available process model to determine the ideal control system performance under constraints. The work is motivated by the fact that the major disturbances in chemical engineering processes are not stochastic but deterministic such as set-point moves and sudden load changes on the system (MacGregor et al., 1984) and for the demand for new methodologies to evaluate MPC performance (Patwardhan et al., 2002).

The paper is organized as follows. In Section 2, we review the most relevant CPA techniques for MPC. Section 3, presents our proposed CPA method. Section 4 shows a case study based on the Shell standard control problem. Finally, conclusions and future directions are pointed in Section 5.

2. REVIEW OF CPA TECHNIQUES FOR MPC

In the literature on CPA of feedback control systems, the MVC benchmark has been used as a measure of performance in the first level of the control structure where SISO controllers are to be evaluated. This benchmark is reasonable because the objective of most SISO controllers is to keep the process output at their set-point. However, MPC controllers have much more sophisticated objectives than merely keeping outputs at their set-points. They are usually implemented as part of a hierarchical control structure, where in an upper layer an optimization algorithm continuously updates a set of optimal economic reference values and passes this set to the MPC (Qin and Badgwell, 2003). The MPC must move the plant from one reference point to another subject to operational constraints. Of course, MPC turns to be essentially a nonlinear controller, especially when operating at the constraints. In this case, the use of MVC or a linear controller benchmark is not well suitable as some inherent limitations imposed by constraints are neglected and minimum variance performance will be unachievable by the MPC (Zhang and Henson, 1999). Examples highlighting the limitations of the MVC benchmark to assess the DMC controller are shown by Hugo (1999).

Patwardhan et al. (1998) discussed the use of the best historical values of objective function as a practical benchmarking technique. This approach requires *a priori* knowledge of an example case where the performance was good during a certain time period of time according to some expert assessment. Huang and Shah (1999) proposed the linear quadratic Gaussian (LQG) benchmark as an alternative to the MVC. The LQG is more general than the MVC and it can be designed, using the available process model, in a similar form as the MPC. This benchmark is translated in a tradeoff curve that displays the minimal achievable performance in terms of the input and output variances. However, the LQG cannot handle constraints and it still represents an unattainable standard for commercial MPC algorithms. Zhang and Henson (1999) suggested the use of the on-line comparison between expected and actual process performance. The expected performance is obtained when the MPC controller is applied to the process model instead of the actual plant and it is neglected the effects of unmeasured disturbances. Ko and Edgar (2001) presented a benchmark based on the constrained finite-horizon MVC controller, which is obtained using the knowledge of the process and noise models. The main utility of this approach lies in quantifying the effects of constraints on the MPC performance. When the constraints become inactive, the proposed method naturally becomes the unconstrained MVC.

Patwardhan et al. (2002) suggested the design case as a benchmark to evaluate the statistical performance of MPC. The methodology is very straightforward to be implemented on-line. The cost function used in the design of the CPA can be obtained from the MPC controller. The achieved cost function can be computed with little effort through appropriate weighting of the measured input and output data. The technique can explicitly handle constraints and it is a

true index that represents whether the controller is performing as it was designed or not. However its application is limited, as most of the commercial MPC algorithms do not return the design value of the cost function. Grimble (2003) presented a multistep linear quadratic Gaussian predictive control (LQGPC) cost function as benchmark to evaluate MPC. The cost function involves the unconditional expected value of the tracking error and weighted control signal components at present and future time steps, whose values are obtained from the solution of appropriate Riccati and Lyapunov equations. The results highlight the relationship between MPC and LQG and the way that the performance of MPC should be assessed. Julien et al. (2004) proposed a MPC benchmark for assessing univariate MPC controllers. By using routine operating data and knowledge of the process time-delay, two performance curves are constructed. One represents the operation of the installed MPC, while the other corresponds to the operation of a hypothetical MPC. If the gap between these operating curves is significant, it may indicate that a re-design of the MPC is necessary.

Schäfer and Cinar (2004) presented an integrated methodology for CPA and diagnosis of MPC systems. They use a LQG benchmark to evaluate performance and a ratio between design and achieved costs for diagnosis of causes of poor performance. Finally, Huang and Georgakis (2005) proposed the minimum (settling) time optimal control (MTOC) benchmark. MTCO-FB is an ideal benchmark for unmeasurable disturbance regulation, while MTCO-FF is an ideal benchmark for set-point tracking. This last, serves also as a reference to determine whether extra sensors and feedforward will yield significant control improvement.

3. PROPOSED TECHNIQUE FOR CPA OF MPC SYSTEMS

Following Zhang and Henson (1999), we propose a CPA technique that involves an on-line performance comparison between expected and actual MPC. But, in this case, the expected performance is obtained with a particular MPC, called here "ideal MPC", that is used to control the nominal process model. The Proposed benchmark is represented schematically in figure 1.

From figure 1, we observe that the optimal set-point (y_{sp}) provided by the upper optimization layer is applied to both MPC systems, actual and benchmark. Estimated disturbances ($\hat{d}$) are used to correct model prediction to asymptotically remove offset. The performance of these systems is measured using adequate indeces, which are compared to determine the performance status of the actual MPC.

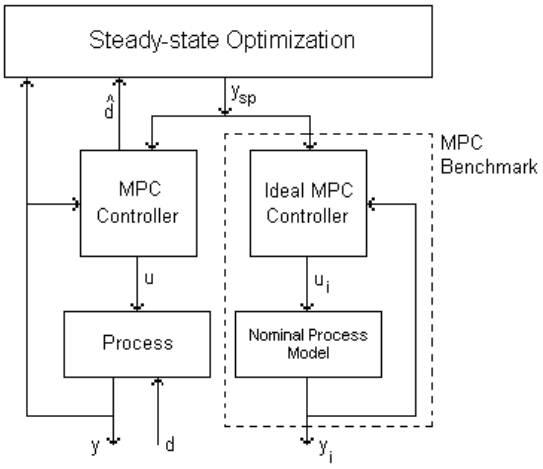


Fig. 1. Scheme of the proposed CPA for MPC.

In the sequel, our ideal MPC controller and performance index will be discussed.

3.1 "Ideal MPC" Controller

For a more realistic comparison, an "ideal MPC" should preserve some characteristics of the implemented MPC, i.e. full utilization of the available process model, incorporation of constraints and computation based on the receding horizon control philosophy. Also, as the ideal controller may utilize tuning parameters that are different of the tuning parameters of the implemented controller, we require that the ideal controller be at least nominally stable. A MPC controller with nominal stability and that tolerates input saturation was proposed by Rodrigues and Odloak (2005). It is assumed that input saturation can occur in the transition from one set-point to another and that the system remains stabilizable during the time the input remains saturated. Consideration of input saturation is usually necessary in a process that operates near the optimal economic conditions.

It can be shown that an unconstrained MPC law can be formulated as:

$$\Delta u_{(m\cdot nu\text{x}1)}(k) = K_{MPC} E^{o}_{(p\cdot ny\text{x}1)}(k) \qquad (1)$$

where $\Delta u(k) = u(k) - u(k-1)$ is the vector of future increment control actions, $K_{MPC} \in \Re^{(m\cdot nu)\text{x}(p\cdot ny)}$ is the time-invariant feedback control gain matrix, $E^{o}(k)$ is the vector of predicted unforced errors, k is the sampling instant, m is the control horizon, p is the prediction horizon, and nu and ny ($nu > ny$) are the number of manipulated and controlled variables, respectively. The development of our "ideal MPC" is to follow a two-step procedure (Rodrigues and Odloak, 2005):

1) *Off-line step*. Compute a bank of stable unconstrained MPC controllers (K_{MPC}),

corresponding to all possible configurations of manipulated inputs and stabilizable outputs. Let *nc* be the number of configurations and let us designate as $K_{MPC,j}$ the gain of the controller corresponding to configuration *j* (*j*=1,…,*nc*).

2) *On-line step*. At each sampling period, compute the predicted unforced error (E^o) and find the solution of the following optimization problem:

$$\min_{\beta_0,\cdots,\beta_{nc}} \quad J_k = \sum_{i=1}^{p} e^T(k+i)Qe(k+i) + \sum_{i=0}^{m-1} \Delta u^T(k+i)R\Delta u(k+i) \tag{2}$$

Subject to:

$$\Delta u(k) = \left[\beta_0 K_{MPC} + \cdots + \beta_{nc} K_{MPC,nc}\right] E^o(k) \tag{3}$$

$$\sum_{j=0}^{nc} \beta_j = 1 \tag{4}$$

$$\beta_j \geq 0, \qquad j = 0,1,\cdots,nc \tag{5}$$

$$u_{\min} \leq u(k+j) \leq u_{\max}, \qquad j = 0,1,\cdots,m-1 \tag{6}$$

where Q is the output error weighting matrix and R is the input increment weighting matrix. Note that the input increment constraints are not included in the above problem. Only the first component of the computed Δu is used. The successive application of this control law produces an asymptotically stable closed-loop system.

3.2 Measure Performance Index

Various dimensionless performance indices have been proposed in the literature. In this work, the controller performance is represented by:

$$J(k) = \sum_{j=1}^{N} \left(y_{sp}(k-j) - y(k-j)\right)^T Q \; \mathrm{x} \left(y_{sp}(k-j) - y(k-j)\right) \tag{7}$$

where $y_{sp}(k)$ is the set-point, $y(k)$ is the value of the controlled variable and N is the length of the past data operation window. The performance measure index $\eta(k)$, which is selected, to bear some similarity with the one proposed by Harris (1989), is the ratio of the performance provided by the "ideal MPC" to the actual performance provided by the present MPC system:

$$\eta(k) = 1 - \frac{J_{\text{ideal}}(k)}{J_{\text{act}}(k)} \tag{8}$$

The index defined in Eq. (8) gives numerical bounds for controller performance $0 \leq \eta \leq 1$, where $\eta = 0$ indicates excellent performance and $\eta = 1$ indicates poor performance. In the subsequent section, the proposed approach is applied to evaluate the MPC performance for a simulated industrial process.

4. CASE-STUDY

4.1 The Shell Standard Control Problem

The Shell standard control problem (SSCP) is a well-known process control problem developed with the intention of providing a standardized simulation protocol for the evaluation of control systems. This system is an industrial heavy oil fractionator process, as shown in figure 2 (Prett and Morari, 1987).

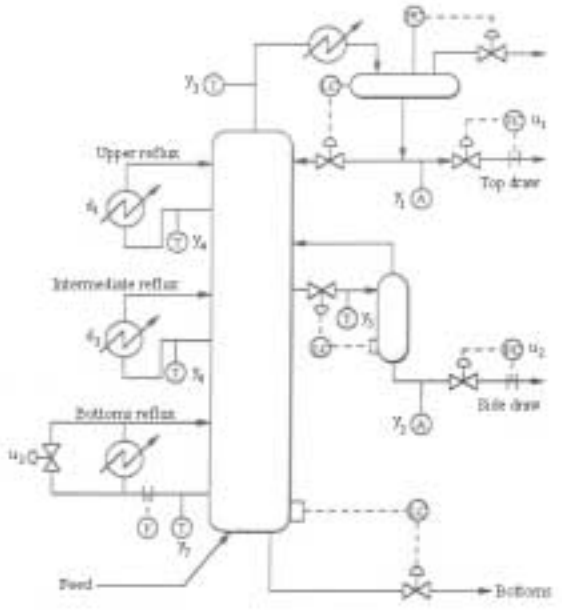


Fig. 2. Layout of the heavy oil fractionator process

The SSCP embodies a number of scenarios that can occur in controlling the process unit. It is represented by a 5x7 MIMO system, which is highly constrained, with very strong interactions, unmeasured disturbances, mixed fast and slow responses, severe uncertainties, large time-delays and simultaneous and conflicting control and economic objectives. The process input/output relations are modeled by transfer functions of first-order plus time-delay. The full process model and its associated uncertainty can be found in Prett and Morari (1987) and Maciejowski (2002).

4.2 MPC Control Design

Let us consider only a part of the SSCP and study the servo problem of the subsystem in which the controlled variables are the top draw composition (y_1) and side draw composition (y_2), and the manipulated variables are the top draw (u_1), side

draw (u_2) and bottom reflux duty (u_3). The transfer function of this subsystem is:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \dfrac{(4.05+2.11\delta_1)e^{-27s}}{50s+1} & \dfrac{(1.77+0.39\delta_2)e^{-28s}}{60s+1} & \dfrac{(5.88+0.59\delta_3)e^{-27s}}{50s+1} \\ \dfrac{(5.39+3.29\delta_1)e^{-18s}}{50s+1} & \dfrac{(5.72+0.57\delta_2)e^{-14s}}{60s+1} & \dfrac{(6.90+0.89\delta_3)e^{-15s}}{40s+1} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \quad (9)$$

where δ_1, δ_2 and δ_3 represent uncertainties in the gain parameters and they can vary between –1 and +1. Here, they are assumed to be $\delta_1 = \delta_3 = 0.5$, $\delta_2 = -0.5$.

The control objective is set-point tracking and, for this purpose, a conventional QDMC (quadratic dynamic matrix control) is proposed. This controller is designed based on the nominal process model (i.e., $\delta_1 = \delta_2 = \delta_3 = 0$) and on the minimization of the following cost function:

$$\min_{\Delta u(k),\cdots,\Delta u(k+m-1)} J_k = \sum_{i=1}^{p} e^T(k+i)Qe(k+i) + \sum_{i=0}^{m-1} \Delta u^T(k+i)R\Delta u(k+i) \quad (10)$$

Subject to:

$$\begin{aligned} -0.5 \le u_i \le 0.5, \quad & i = 1,2,3 \\ |\Delta u_i| \le 0.1, \quad & i = 1,2,3 \end{aligned} \quad (11)$$

The tuning parameters used in the QDMC and in the "ideal MPC" are: $p = 75$, $m = 10$, $Q = \mathrm{diag}(1,1)$, $R = \mathrm{diag}(1.5, 0.15, 1.5)$ and sampling time $T = 4\,\mathrm{min}$. The controllers are implemented in the hierarchical control structure as shown in figure 1. However, in this study, it is considered that the optimization layer can be ignored in the simulations, i.e. the set-points are assumed to be known.

4.3 MPC Performance Assessment

The hierarchical control structure of the MPC system is not included in the problem considered by the performance assessment method, since it does not really matter if the set-point for the MPC controllers comes from an operator or a computer program.

Thus, in our example performance assessment is carried out for the set-point changes shown in figure 3. It is also shown the responses of the system outputs for the QDMC based on the nominal model when controlling the true system that contains uncertainties as described before. In figure 3 are also shown the output responses for three different benchmark controllers. $QDMC_2$ is the benchmark corresponding to the same QDMC controller of the previous case but controlling the nominal model. This scenario corresponds to the design case. The second benchmark controller corresponds to the LQG (Dorato et al., 1995) optimally tuned to control the nominal process. Finally, the third benchmark controller is the proposed controller with the same tuning parameters as the QDMC except the constraint in the control moves that is not included in the control problem. Figure 4 illustrates the responses for the inputs of the system.

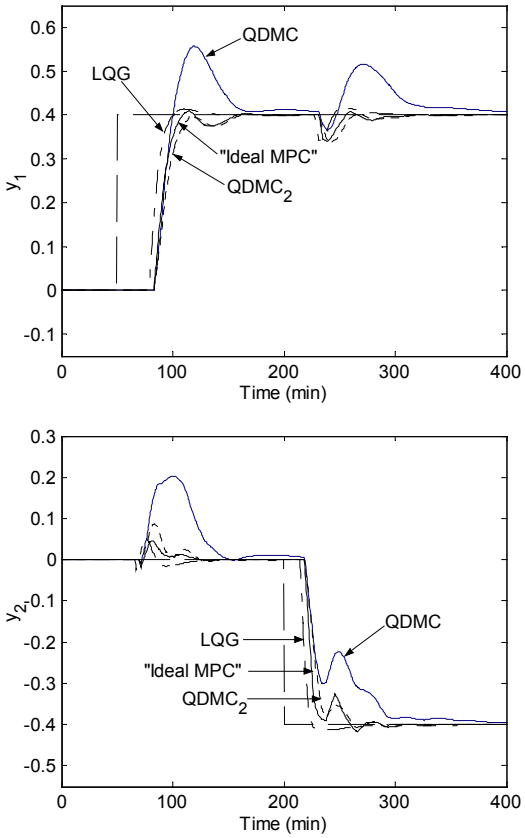


Fig. 3. Output responses of the Shell system

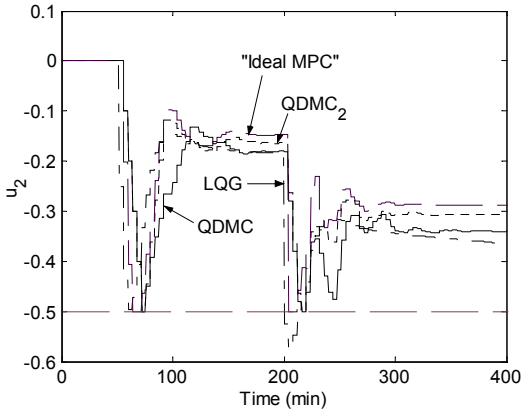


Fig. 4 Input u_2 responses of the Shell system

Table 1 shows the numerical values of the controller performance defined by Eq. (7) and also shows the performance indices defined by Eq. (8) for the QDMC controller applied to the uncertain system in

terms of each of the benchmark controllers described before. From figure 3, it is clear that the LQG produces the best nominal performance but it is not realistic as it does not satisfy the input constraints as shown in figure 4, where the minimum bound of input u_2 is not satisfied during part of the simulation time. Table 1 shows that $QDMC_2$ produces a more conservative index that the proposed benchmark which gives a better indication of the performance of the implemented controller.

Table 1. Control performances and indices

System	J	η
QDMC	3.4273	---------
$QDMC_2$	2.6710	0.2207
"Ideal MPC"	2.4845	0.2751
LQG	2.0597	0.3990

5. CONCLUSIONS

In this work, it is proposed a new MPC benchmark controller whose purpose is the realistic evaluation of the performance of MPC controllers implemented in industry. The proposed benchmarks has as main characteristics to consider input constraints and guaranteed nominal stability, which is usually not attended by other proposed benchmark controllers. The approach was tested by simulation in a typical system of the process industry with satisfactory results.

Acknowledgements: The authors gratefully thank the finantial support from FAPESP under grants 02/08119-2 and 03/11150-1.

REFERENCES

Aström, K.J. (1991). Assessment of achievable performance of simple feedback loops, *International Journal of Adaptive Control and Signal Processing*, 5(1), 3-19.

Automation Research Corporation (2000). *Real time process optimization and training worldwide outlook*, http://www.arcweb.com/.

Dorato, P.C., C. Abdallah, V. Cerone (1995). *Linear quadratic control: an introduction*, Prentice Hall.

Grimble, M.J. (2003). Performance assessment and benchmarking LQG predictive optimal controllers for discrete-time state-space systems, *Transactions of the Institute of Measurement and Control*, 25(3), 239-264.

Harris, T.J. (1989). Assessment of control loop performance, *Canadian Journal of Chemical Engineering*, 67(5), 856-861.

Huang, B., S.L. Shah (1999). *Performance assessment of control loops – theory and applications*, Springer-Verlag: London.

Huang, L., C. Georgakis (2005). On the assessment of controller performance for constrained systems, *7th Italian Conference on Chemical and Process Engineering* (ICheaP-7), Giardini di Naxos-Taormina.

Hugo, A. (1999). Performance assessment of DMC controllers, *1999 American Control Conference* (ACC'99), San Diego-CA.

Hugo, A. (2000). Limitations of model predictive controllers, *Hydrocarbon Processing*, 79(1), 83-88.

Julien, R.H., M.W. Foley, W.R. Cluett (2004). Performance assessment using a model predictive control benchmark, *Journal of Process Control*, 14(4), 441-456.

Ko, B.-S., T.F. Edgar (2001). Performance assessment of constrained model predictive control systems, *AIChE Journal*, 47(6), 1363-1371.

MacGregor, J.F., T.J. Harris, J.D. Wright (1984). Duality between the control of processes subject to randomly occurring deterministic disturbances and arima stochastic disturbances, *Technometrics*, 26(4), 389-397.

Maciejowski, J.M. (2002). *Predictive control with constraints*, Pearson Education Limited: England.

Patwardhan, R.S., S.L. Shah, K.Z. Qi (2002). Assessing the performance of model predictive controllers, *Canadian Journal of Chemical Engineering*, 80(5), 954-966.

Patwardhan, R.S., S.L. Shah, G. Emoto, H. Fujii (1998). Performance analysis of model-based predictive controllers: an industrial case study, *1998 AIChE Annual Meeting*, Miami-FL.

Prett, D.M., M. Morari (1986). *The Shell process control workshop*, Butterworths: Stoneham-MA.

Qin, S.J. (1998). Control performance monitoring – a review and assessment, *Computers and Chemical Engineering*, 23(2), 173-186.

Qin, S.J., T.A. Badgwell (2003). A survey of industrial model predictive control technology, *Control Engineering Practice*, 11(7), 733-764.

Rodrigues, M.A., D. Odloak (2005). Robust MPC for systems with output feedback and input saturation, *Journal of Process Control*, 15(7), 837-846.

Schäfer, J., A. Cinar (2004). Multivariable MPC system performance assessment, monitoring, and diagnosis, *Journal of Process Control*, 14(2), 113-129.

Shah, L.S., R.S. Patwardhan, B. Huang (2001). Multivariate controller performance analysis: methods, applications and challenges, *VI Chemical Process Control* (CPC-6), Tucson-AZ.

Treiber, S., R. Sttelmaier, M. Starling (2003). Sustaining benefits of process automation, *2003 AIChE Spring Nat. Meeting*, New Orleans-LA.

Zhang, Y., M.A. Henson (1999). A performance measure for constrained model predictive controllers, *5th European Control Conference* (ECC'99), Karlsruhe, Germany.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

TOWARDS AN INTEGRATED CO-OPERATIVE SUPERVISION SYSTEM FOR ACTIVATED SLUDGE PROCESSES OPTIMISATION

Bassompierre C. (1,2), Cadet C. (1), Béteau J.F. (1), Aurousseau M. (2), Guillet A. (2)

1 LAG – Laboratoire d'Automatique de Grenoble, ENSIEG, BP 46
2 LGP2 – Laboratoire de Génie des Procédés Papetiers, EFPG, BP 65
38402 Saint Martin d'Hères Cedex, France

Abstract: This paper deals with the design of an integrated co-operative supervision system to optimise activated sludge wastewater treatment. For this purpose a biological wastewater treatment pilot plant, suitable for industrial up scaling and composed of a biological reactor, a settler, sludge recirculation and aeration systems, has been carried out. The reactor configuration is adjustable to treat both urban and industrial wastewater. The pilot plant scale allows to reproduce industrial hydrodynamics and is illustrated by RTD (Residence Time Distribution) measurements. The supervisory system includes instrumentation and industrial supervisor. The integrated co-operative supervision approach emphases the importance of process behavior anticipation. Information for the human operator will be extracted from a suitable model and a state estimator, which results are presented.

Keywords: Water pollution – Biotechnology – Process equipment – Supervision – Modelling – Observers

1. INTRODUCTION

Activated Sludge Process is a biological treatment commonly used for urban and industrial wastewater. Nevertheless performances improvement requires an efficient supervisory system which is still a research subject. On one hand, numerous laboratory-scale pilot plants have been developed for many years but their small size weakens hydrodynamics influence on the treatment efficiency. Therefore results can be hardly implemented on real processes. On the other hand, some experiments (such as toxicity, foaming …) cannot be carried out on industrial plant because working conditions must be maintained. Consequently, a medium pilot plant scale seems to be judicious. In addition the pilot plant configuration has to be adjustable so as to reproduce most of full scale encountered configurations: urban/industrial effluent, adjustable reactor configuration, etc.

A modular pilot plant has been realized and is shown on figure 1. It may be used for a wide range of experiments as for example: toxicity tests, fungi treatment tests instead of bacteria use, hydrodynamics influence studies, biological modelling and control strategies validations.

At the same time on real processes, human operator tasks are various and complex: monitoring, data interpretation, local control loops tuning, process behavior prediction, dysfunction detection, fault diagnosis. To perform his duty, the operator faces many difficulties due to imprecise and incomplete information and, for some processes, to numerous alarms. In this context, having at his disposal an efficient on-line tool for decision aid becomes essential (Rosen, *et al.*, 2003).

Fig. 1. The activated sludge pilot plant.

The purpose of the work described in this paper is to develop a co-operative supervision system for operator decision support. This system, which makes use of automatic control methodologies and recent research results such as modelling or estimation of unmeasured variables, will be tested on the pilot plant. One reference plant is an urban reference simulator (Copp *et al.*, 2002). The activated sludge

process of an industrial paper mill, partner of the project, is the other reference plant.

In the following part 2, design and realisation of the pilot plant is described and first experimental results are shown. Part 3 details the integrated supervisory approach, including model and observer design and validation.

2. PILOT PLANT REALISATION AND RTD EXPERIMENTS

2.1 Pilot plant design and sizing

All the components of an industrial biological treatment process are included in the pilot plant:

- a bioreactor in which the biomass assimilates pollution. A movable aeration system creates aerobic compartments to treat both urban (aerobic and anoxic phases) and industrial paper mill (aerobic only) effluents. Movable baffles allow to adapt hydrodynamic behavior to different process configurations (channel aerator, carrousel etc.)
- a settler, in which the biomass agglomerates in falling flocks (sludge) that lets the superficial water purified. Real settling conditions are obtained thanks to its industrial shape (a slight slope cone with a scraper surmounted by a cylinder)
- a sludge recirculation system which allows to maintain a constant biomass concentration in the process.

The pilot plant has been sized according to the urban and industrial reference plants, which characteristic data are shown on Table 1. The three first parameters characterize the working conditions, which have to be realizable by the pilot plant. Thus the input flow range and the reactor volume have been dimensioned as a compromise between a realistic size and the previous constraints.

Table 1. Characteristic data of reference and pilot plants

Data	Urban plant	Industrial plant	Pilot plant
Volumic load to be treated C_v ($kg_{BOD}.m^{-3}.d^{-1}$)	1	0.97	0.8 – 1.3
Residence time (h) based on the total input flow	3.6	8.72	2 – 12
Recirculation flow to input flow ratio	1	1.2	1 – 2
Input flow ($m^3.d^{-1}$)	20000	5000	0.25 – 1.5
Reactor Volume (m^3)	6000	4000	0.250

[1] C_V (Volumic load) is the ratio of input biodegradable substrate mass per day to reactor volume.
[2] COD (Chemical Oxygen Demand) is the measurement of the oxygen quantity needed to treat all organic matter.
[3] BOD_5 (Biological Oxygen Demand) is the measurement of the dissolved oxygen needed to treat biodegradable organic matter in 5 days.

2.2 Realization

Additional elements complete the pilot plant (Figure 2). A steered and cooled storage tank allows an effluent supply autonomy of five days whereas a buffer tank allows pH and nutrient adjustments of the effluent. The whole process is secured by an independent electrical device for human and material safety.

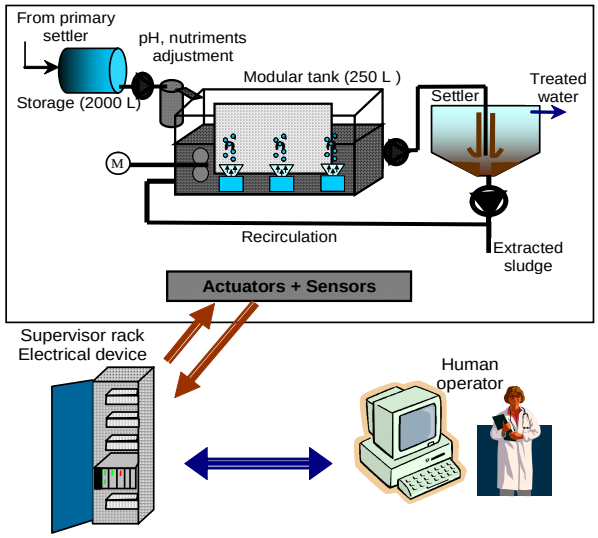


Fig. 2. Schematic representation of the pilot plant.

The pilot plant is provided with many measurements:

- on-line sensors: pH, temperature, oxygen concentration, redox potential and conductimetry;
- off-line measurements with an automatic sampling device: Chemical Oxygen Demand[2], Biological Oxygen Demand[3], nitrogen compounds concentrations for effluent and treated water.

Control loops on the liquid level and on the dissolved oxygen concentration in the bioreactor are implemented through the supervisor.

The industrial tool PC3000 by Eurotherm™ has been chosen for the plant supervision. All the sensors and actuators are linked to the input/output modules of the supervisor rack. The process control files (discret events control, continuous control loops, dysfunctions detection) are implemented on the motherboard. The displayed data on the operator synoptic screen are updated each second. The operating modes proposed to the operator are detailed in the sequel.

2.3 Operating modes

On figure 3, the synoptic screen (which is under construction) presents the global scheme of the pilot plant (reactor, settler, pipes system). Besides each continuous actuator (pumps and valves), both measured value and manual setpoint are displayed. Measurements, set points and on-off actuators are given on the table under the scheme.

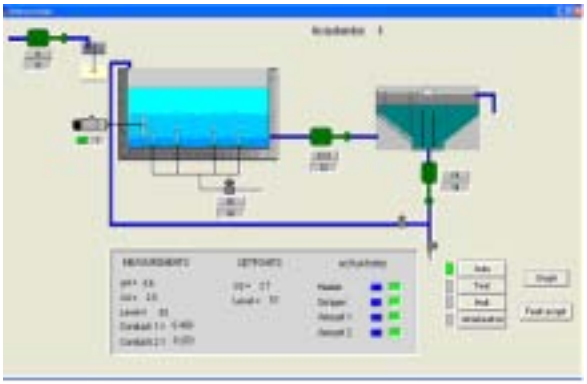


Fig. 3. Pilot plant synoptic

The different operating modes are available at the right bottom of the screen and can be activated at any moment by the operator.

- The auto mode assigns the set points to their nominal values and switch-on the required control loops. A dysfunction routine scans continuously in case of dysfunction (sensors measurements going through security thresholds, actuators faults). The auto mode is then interrupted, a color code advises the operator on the screen, and the dysfunction mode automatically runs.
- The test mode is proposed to test manually the PID tuning, to disconnect the control loops and to handle manually dysfunction problems.
- The halt mode allows the pilot plant emptying for cleaning or effluent change.

2.4 RTD experiments

Residence time distribution (RTD) experiments on the bioreactor have been compared with similar experiments on an industrial process to verify the hydrodynamic scaling-up. The pilot bioreactor configuration is a carrousel and the flows are adjusted to fit industrial configuration.

Tracing experiments have been carried out by injecting pulses of a KCL solution. The concentration is collected by a conductimetry sensor at the reactor output. The resulting RTD function corresponds to the ratio of instantaneous tracer concentration to the integral concentration recovered at the output.

Because of the normalisation of RTD data, the very long residence time and the unity area of the curve, RTD scale (figure 4) is very small. The duration of the experiment is more than five hours, which is considerable and consequently the expected final zero value of concentration is not reached and may take a rather longer time. On the first points of the experimental graph, the rough disturbance corresponds to a weak short-circuit due to experimental protocol, which is of course not visible on the model.

The hydrodynamic model (figure 4) is identified by using the software DTSPRO® 4.2. After many tests and though the curve shape is not exactly reproduced,

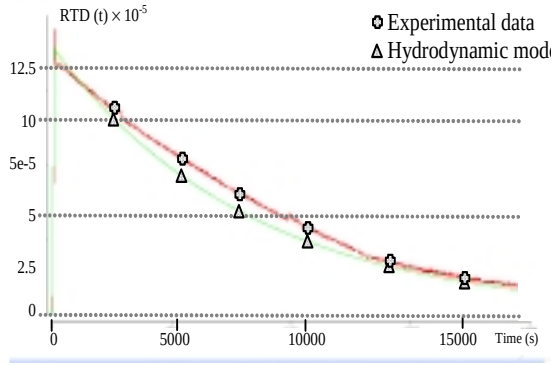


Fig. 4. RTD evolution (adimensionnal form).

the more accurate model is proposed. It is composed of a plug flow reactor (very small size thus it can be considered as neglected) followed by a well-mixed reactor. The industrial data lead also to a well-mixed reactor, which validates our methodology (experimental results are confidential and cannot be published).

These results are first experiments and need therefore to be improved and extended with many other ones. However, they are sufficiently good to consider that the bioreactor design and technological choices are the right ones to create similar hydrodynamic behaviors between pilot and industrial plants.

In a further work, the effect of hydrodynamics on the treatment efficiency will be studied. Simply by baffles placements, the experimental RTD may be adjusted. The correlation between the RTD and the number of perfectly-mixed reactors, proposed by Potier (Potier, *et al.*, 1998), will be easily verified on the pilot plant, thanks to its modular conception.

3. INTEGRATED SUPERVISION

In part 2.3, the common supervisory tasks have been described. The information that is available for the operator is strictly limited to direct on-line measurements or off-line backward measurements. Finally all the data interpretation, which consists in deducing the biological treatment quality in the reactor and in predicting the treatment efficiency, is entirely in charge of the human operator. Moreover, the simulation of possible different actions in case of dysfunction or disturbances is not possible. The interest of a real co-operative supervisory tool to help the operator in his every day decisions is developed in the following part.

3.1 Interest of a co-operative supervision system

Here is a scenario example of a process dysfunction. The chosen event is the decrease of biomass concentration in the bioreactor, due to a dysfunction on the recirculation system (pump fault and flow sensor fault) or to biomass disease. The process events are described in figure 5 (mid-column): due to the dysfunction, the first consequence is the decrease of biomass concentration in the bioreactor. The

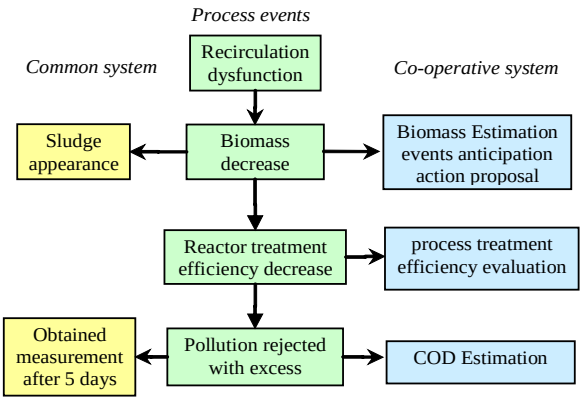


Fig. 5. Dysfunction scenario and comparison of common and co-operative systems

treatment efficiency is then also decreasing, and at the same time the reactor pollution increases. At last, the treated water that is rejected in the river is more and more polluted and if nothing is done, exceeds the limits authorized by the law.

With typical supervision tools (left column), the supervisor gives only basic measurements. To detect the previous described dysfunction the operator may:

- go to the plant (which may be far away from the supervision station), notice sludge appearance (colour, form) and conclude thanks to its experience ;
- wait results of off-line measurements which are usually done every day for COD or even every week for BOD_5 measurement.

In addition, if the biomass decrease is fast and strong the operator may realize the seriousness of the situation far too late to act efficiently and the biomass may be totally washed out.

In comparison, the co-operative supervision system (figure 5, right column), gives not only the basic measurements but also estimates useful unmeasured data and proposes some interpretations on the reactor operating conditions. In addition, to previous informations, the operator can know:

- the on-line estimation of unmeasured data as the biomass concentration and the COD of treated effluent,
- the anticipation of the process behavior and the data tendency,
- the treatment efficiency evaluation,
- the proposal of control actions and their efficiency.

The operator may also introduce his own remarks as sludge appearance, which ideally may be used by the supervisor. In addition, some elements as the cost of the proposed actions may be also included (low cost or efficiency choice). The simulation of new plant configurations and new equipments may also propose information on the associated treatment cost saving, which may facilitate sensor investments.

Through this example, the lack of available information on the process is obvious for common supervisory systems. In these conditions, it is difficult for the human operator to detect a dysfunction and above all to anticipate the phenomenon.

Consequently, the development of an on-line aid in decision task is decisive for a good process operating. Such a tool allows the operator to better understand the process behavior and to detect dysfunctions from the initial stages.

3.2 Co-operative integrated supervision design

For usual industrial processes, the aim of supervision tools is to detect and to localize faults. Different methods are available to solve this problem such as causal graph (Leyval, *et al.*, 1994) or fuzzy set theory and multi-criteria decision (Gentil, *et al.*, 1994; Montmain and Gentil, 1996). Nevertheless, the main difficulty of supervision for wastewater treatment process is to detect dysfunctions on biological and hydrodynamic behaviors and to react as rapidly as possible. Thus, other tools have to be fitted.

The structure of the proposed tool for decision aid will be based on a succession of steps. Each one is sufficiently interesting to be used alone. Proposed steps are the followings:

1) *A model* is developed, and validated with data. The model may be use for simulation purposes and is the basis of all the information treated in the next points.

2) *Unmeasured data estimation*: a model-based nonlinear observer is designed to compensate the lack of sensors. At this step, all the information needed for a complete understanding of the process is available.

3) *Reliability data indicator*: sensor precision, estimation quality.

4) *Tendency extraction*: to anticipate data variations and then to predict the process behavior. Based on the results proposed by Charbonnier, *et al.*, (2004), the method uses a data segmentation algorithm and a classification tool.

5) *Action proposal*: depending on present and future behavior, some actions will be proposed to operator to improve the treatment process by respecting a trade-off between depollution, operating cost and biomass activity.

Different tools mentioned above have been developed and tested in simulation. The next two sections will focus on process modelling and estimation of unmeasured data.

3.3 Model design

For activated sludge processes, a reference model has been proposed by the European group COST Action 624 (Copp *et al.*, 2002) where the biological model corresponds to the Activated Sludge Model (ASM) n°1 developed by the task group of the International Association on Water Quality (IAWQ) (Henze, *et al.*, 1986). Due to the number of state variables and parameters, this model is difficult to handle for

estimation purpose, whereas a simplified model is more suitable, always respecting a trade-off between complexity and precision. Thanks to previous works ((Cadet, *et al.*, 2003), (Cadet, *et al.*, 2004)), the design of a new reduced model, for which simplifications are mainly based on biological considerations, is presented. It has been design for industrial effluents (treatment of carbonaceous matter).

1) *Simplification hypothesis.*

- *Hydrodynamics.* The bioreactor corresponds to one well-mixed reactor (instead of five) which provides a large reduction of the number of states;

- *State variables considerations.* Only the components necessary for the main reactions are kept and leads to 5 state variables (instead of 13):

- 1 fraction of nitrogen (S_{NH})
- 1 fraction of organic matter (XS_S)
- 2 types of micro-organisms (heterotrophic biomass X_{BH}, autotrophic biomass X_{BA})
- 1 fraction of dissolved oxygen (S_O)

- *Biological processes.* Only four processes are considered: the biomasses decays, the carbon oxidation and nitrification. Thus, reduced process rates expressions are:

$$
\begin{aligned}
r_{XS_S} &= -\frac{1}{Y_H}.\gamma_H.\frac{XS_S}{K_S+XS_S}.\frac{S_O}{K_{OH}+S_O}.\frac{S_{NH}}{K_{NH}+S_{NH}}.X_{BH} \\
&\quad +(1-f_p).(b_H.X_{BH}+b_A.X_{BA}) \\
r_{X_{BH}} &= \gamma_H.\frac{XS_S}{K_S+XS_S}.\frac{S_O}{K_{OH}+S_O}.\frac{S_{NH}}{K_{NH}+S_{NH}}.X_{BH}-b_H.X_{BH} \\
r_{X_{BA}} &= \gamma_A.\frac{S_{NH}}{K_{NH}+S_{NH}}.\frac{S_O}{K_{OA}+S_O}.X_{BA}-b_A.X_{BA} \\
r_{S_{NH}} &= -i_{XB}.\gamma_H.\frac{XS_S}{K_S+XS_S}.\frac{S_O}{K_{OH}+S_O}.\frac{S_{NH}}{K_{NH}+S_{NH}}.X_{BH} \\
&\quad -\left(i_{XB}+\frac{1}{Y_A}\right).\gamma_A.\frac{S_{NH}}{K_{NH}+S_{NH}}.\frac{S_O}{K_{OA}+S_O}.X_{BA} \\
r_{S_O} &= -\left(\frac{1-Y_H}{Y_H}\right).\gamma_H.\frac{XS_S}{K_S+XS_S}.\frac{S_O}{K_{OH}+S_O}.\frac{S_{NH}}{K_{NH}+S_{NH}}.X_{BH} \\
&\quad -\left(\frac{4{,}57-Y_A}{Y_A}\right).\gamma_A.\frac{S_{NH}}{K_{NH}+S_{NH}}.\frac{S_O}{K_{OA}+S_O}.X_{BA}
\end{aligned}
\qquad (1)
$$

where γ_H, γ_A and Y_H are reaction rate factors that have to be tuned. These parameters have been selected according to sensitivity functions. Other rate factors keep their reference value (Copp *et al.*, 2002). The obtained reduced model is composed of 5 state variables and 3 parameters to estimate.

Validation. Three urban wastewater databases are available, corresponding to dry, rain and storm weather, which allow simulating realistic operating conditions. The parameters have been adjusted with a quasi-Newton algorithm for dry weather conditions and the model is then validated with different data (rain and storm weathers). Some results in storm conditions at the output of the reactor (aerobic state variables) are shown on figure 6. The reduced model presents an excellent accuracy. During the first day, the steady state point of the reduced model is remarkably close to the reference steady state point. The following 7 days, with dry weather, the model accuracy is excellent. From the day 9, the storm event induces rough disturbances on model inputs which influence hardly output state variables.

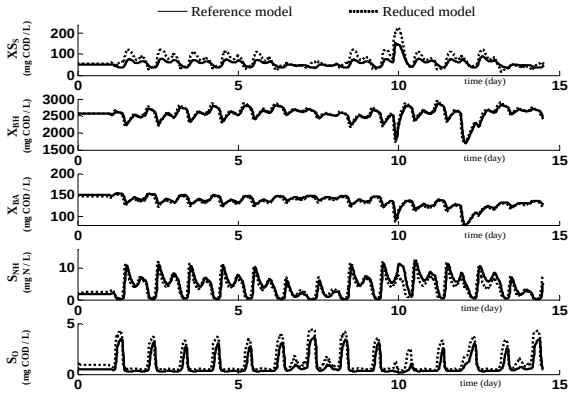


Fig. 6. Reduced model validation on storm weather.

More precisely, the difference on dynamic behavior of the carbon fraction is due to the influence of nitrogen concentration on carbon oxidation processes which is not take into account in the reference model. The biomasses and the dynamic variations of oxygen and nitrogen state variables are very close to the reference model.

Consequently, this model is suitable for our purpose: on one hand, its accuracy is excellent so as to expect good performances for the observer based on that model and, on the other hand, it is sufficiently simple to be validated on a real treatment plant.

3.4 Observer design

A comparative study of the two observers that are most commonly used on actual activated sludge processes, asymptotic observer (Bastin and Dochain, 1990) and extended Kalman filter, has been presented in a previous work (Cadet and Plouzennec, 2003). Several drawbacks (such as convergence rate determined by experimental conditions, model linearization) lead to study an other approach: the Moving Horizon State Observer. Only the principle is presented here, more details may be found in (Alamir, 1999).

1) Moving horizon state observer principle

Considering the nonlinear model:

$$
\begin{cases}
\dot{x}(t) = f\left(x(t)\right) \\
y(t) = h\left(x(t)\right)
\end{cases}
\qquad (2)
$$

where $x \in \Re^n$ is the state vector, and $y \in \Re^p$ is the measured output. The functions *f* and *h* are supposed to be continuously differentiable. We denote *z* the estimate of *x*.

The method transforms the state estimation of a dynamical system into static optimization problem updated with respect to time. The new problem is to estimate the initial conditions *z(t-T)* which have lead the process to its present state *x(t)* since the beginning of the moving horizon *(t-T)* while producing the measured output. The criterion to be minimized is the integration of the squared difference between the outputs given by the model from the initial conditions at *(t-T)* and the corresponding measured outputs.

The state vector *z(t-T)* is adjusted by resolving the model equations (2) modified by adding a correction term. This term decreases with the criterion *J(t)*. Thus an asymptotic observer with a fixed gain γ can be established:

$$\begin{cases} \dot{z}(t-T) = f\left(z(t-T)\right) - \gamma G(t)^T \left[G(t)G(t)^T\right]^{-1} \sqrt{J(t)} \\ \hat{\dot{x}}(t) = f\left(\hat{x}(t)\right)\Big|_{z(t-T)} \end{cases} \quad (3)$$

where $G(t) = \dfrac{\partial J(t)}{\partial z}$.

In order to lighten the computations while keeping a good final precision, the estimation is updated with a period (t_{n+1}- t_n).

2) *Simulation results with respect to reduced model.* Realistic measured state variables have been chosen: S_{NH} and S_O for which on-line sensors exist. For testing purpose, we used the reduced model to generate measurements to be used by the observers. The input flow is decreased of 10% (simulation of an upstream machine dysfunction) between day 0,05 and day 1. This perturbation is unmeasured and therefore not applied to observer inputs. As shown in figure 7, the observer presents a good convergence time even if it is slower for the heterotrophic biomass. There is no bias for steady state conditions. Therefore, the moving horizon observer gives an accurate estimation of the process states.

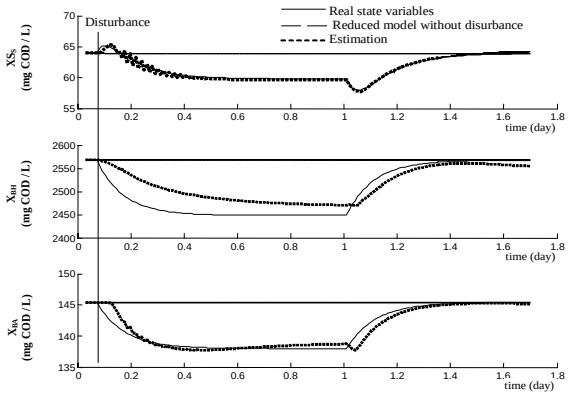


Fig. 7. Observer convergence with dry weather data.

4. CONCLUSION

In this paper, the conception and realization of an activated sludge pilot plant has been described. The pilot plant behavior satisfies the fixed purposes, like modular configuration, and allows scaling-up results. The design of an on-line aid decision tool, dedicated to activated sludge processes, has been developed. First results show the very good accuracy of the reduced model which is sufficiently simple to allow validation on a real effluent. Simulation observer results confirms that the Moving Horizon State Observer approach is relevant for this application. Results are now sufficient to plan further developments: validation on real effluent and other supervisory steps realization to contribute to a better control of activated sludge processes.

REFERENCES

Alamir, M. (1999). Optimization based non-linear observers revisited. *International Journal of Control*. **72(13)**, pp.1204-1217.

Bastin, G. and D. Dochain (1990). On-line estimation and adaptative control of bioreactors. Elsevier Science Publishers B.V., Amsterdam, The Netherlands.

Cadet, C. and P. Plouzennec (2003). Synthesis of model based observers for activated sludge processes. *Proceedings of 4th International IMACS Symposium on Mathematical Modelling, MATHMOD 2003*, Vienna, Austria.

Cadet, C., N. Lucas, J.F. Béteau, A. Guillet, M. Aurousseau and N. Simonin (2003). Modelling of wastewater treatment of a pulp and paper mill for monitoring. *Environmental. Engineering and management Journal*. **2(3)**, pp.163-174.

Cadet, C., J.F. Béteau and S.C. Hernandez (2004). Multicriteria control strategy for cost/quality compromise in wastewater treatment plants. *Control Engineering Practice*. **12**, pp.335-347.

Charbonnier, S., C. Garcia-Beltan, C. Cadet and S. Gentil (2004). Trends extraction and analysis for complex system monitoring and decision support. *Engineering Applications of Artificial Intelligence*. **18**, pp. 21-36.

Copp, J.B. (2002). *The COST simulation benchmark : description and simulator manual* (COST Action 624 & 682). Luxembourg: Office for Official Publications of the European Communities.

Gentil, S., M. Le Cœur and J. Montmain (1994). Integrated quantitative and symbolic reasoning for fault detection. *Proceedings of Intelligent System Engineering workshop, ISE'94.* Hamburg, Germany.

Henze, M., Jr. Gredy, W. Gujer, G. Marias and T. Matsuo (1986). Activated *Sludge Model n°1*. IAWQ Scientific and Technical Report n°1.

Leyval, L., S. Gentil and S. Feray-Beaumont (1994). Model based causal reasoning for process supervision. *Automatica*. **30(8)**, pp.1295-1306.

Montmain, J. and S. Gentil (1996). Operation support for alarm filtering. *Proceedings of CESA'96, IEEE-IMACS Multiconferences on Computational Engineering in Systems Applications*, Lille, France.

Potier, O., M.N. Pons, J.P. Leclerc and C. Prost (1998). Etude de l'hydrodynamique d'un réacteur canal à boues activées en régime variable. *Récents progrès en génie des procédés*. **2(61)**, pp.367-372.

Rosen, C., J. Rottorp and U. Jeppsson (2003). Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation. *Water Science and Technology* **47(2)**, pp.171-179.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

QUANTIFYING CLOSED LOOP PERFORMANCE BASED ON ON-LINE PERFORMANCE INDICES

M. Farenzena and J. O. Trierweiler[#]

Group of Integration, Modeling, Simulation, Control and Optimization of Processes (GIMSCOP)
Department of Chemical Engineering, Federal University of Rio Grande do Sul (UFRGS)
Rua Luiz Englert, s/n CEP: 90.040-040 - Porto Alegre - RS - BRAZIL,
Fax: +55 51 3316 3277, Phone: +55 51 3316 4072
E-MAIL: {farenz, jorge}@enq.ufrgs.br

Abstract: This article aims to construct an "inference model" (IM) that assesses the closed loop performance and robustness for SISO controllers, with no need of intrusive tests (i.e. set-point changes or open-loop step tests). The IM is generated for a large set of plants, disturbances, and tuning parameters. The possible inputs for the IM are 9 standard assessment measurements (e.g., FCOR, standard deviation, etc) on-line available, commonly present in commercial tools. Three IMs were developed for the following targets: the closed loop and open loop rise time ratio (Rt_R), Gain Margin (GM), and normalized integral of square error (ISE). These values are obtained by intrusive tests. Four different classes of inferential models (i.e., Neural Networks, Neuro Fuzzy, PLS, and QPLS) are compared. The best results are obtained by Neural Network IM. The results obtained show that the methodology is very promising. *Copyright © 2006 IFAC*

Keywords: Performance Monitoring, Neural Network, Performance Assessment.

1 INTRODUCTION

Process control increases the plant performance by the reduction of the variability of the key variables. After the variability reduction, the process can achieve a new operating point nearer the restrictions, where the profit is higher (Marlin, 1995).

Even if the actual controller performance is good, some factors can deteriorate the performance during the operation:

- Equipment fouling
- Sensor/actuator problems
- Seasonal influence
- Feed changes
- Operating point changes

Assess on-line the performance of each controller is essential to keep the plant in a profitable operating point. However, quantify the performance of each controller in a typical refinery of petrochemical plant is a difficult task, due to the large amount of loops (usually about 1000 to 2000 loops).

The aim of this work is build an inferential model (IM) to the closed loop performance and robustness indexes. The inputs of this IM are the indexes commonly present in commercial tools, described in section 2.

[#] Author to whom the correspondence should be addressed.

This paper is organized as follows: section 2 provides an overview about the main performance indexes. Section 3 shows the main limitations of these techniques. In section 4 an inferential model to determine the closed loop performance based on on-line indexes is introduced. In section 6, different techniques to build the inferential model to predict closed loop performance is shown, which is tested through a case study in section 7.

2 ON-LINE INDEXES

This section shows the most used indexes to assess closed loop performance available in the commercial software. These indexes will be used as the input of the inferential model proposed in this work.

2.1 Performance Index based on Minimum Variance Controllers

Harris (1989) proposed an index that assesses the performance of controllers using the minimal variance controller as benchmark. The performance index, proposed by Harris ($h(d)$) is calculated by the following relation:

$$h(d) = \frac{s_{MV}^2}{s_y^2} \qquad (1)$$

where s_{MV}^2 is the variance produced by the minimum variance controller and σ_y^2 is the actual loop variance.

The values of η always are between 0 and 1. Increasing values of η indicates the performance becomes better. The actual variance is easy to determine with a window of closed loop data. However, the minimal variance for a given control loop is more difficult. It depends both on the plant and the kind of disturbance. Several methodologies to estimate the minimum variance are described in Huang and Shah, (2001).

The main advantage of minimal variance based index is only closed loop data must be provided to assess the performance.

2.2 Performance Index based Controlled Variables (CV) monitoring

Another quite simple possibility is to use indexes based on the error of the controlled variable. The most used are:

- Standard deviation of CV (StdCV)
- ISE of CV (ISECV)
- Percentile error of CV (E%CV), i.e. the mean absolute error of controlled variable divided by the mean value of CV.

All these indexes are calculated using only closed-loop data. No invasive tests are needed.

2.3 Index related to Manipulated Variables (MV)

Indexes that quantify the work of the manipulated variable are also used to estimate the performance. The most used are:

- Minimal variance calculated over the manipulated variable (ηMV)
- Travel of manipulated variable (MV) (TMV)
- Number of inversions of MV (IMV)

2.4 Commercial Software

The indexes discussed in the last sections together with other ones are available in almost all commercial software for performance assessment (e.g., ProcessDoctor of Matrikon, PlantTriage of ExperTune, TriPerfeX of TriSolutions, among others). These modern tools give to engineer a large amount of indexes, which are not so conclusive and easy to interpret. To overcome this problem, in section 4 it is proposed a novel approach, but before it is interesting to analyze the main limitations the current methodologies.

3 LIMITATIONS CONCERNING THE PREVIOUS METHODOLOGIES

This section shows the limitation of performance index based on Minimum Variance Controllers available in commercial software. Here the Harris Index are calculated using the FCOR algorithm (Huang and Shah, 2001) to estimate the minimum variance. Despite the fact that this index aims to generate a unique and absolute grade to quantify the performance of the controllers, the estimation of minimal variance is not error free.

The main limitations of conventional on-line index are:

- Only performance is assessed, no information about robustness is provided.
- The scale is not absolute (i.e. there is no guaranty that a loop with $h(d) = 0.6$ have a better performance than other loop for other process variable with $h(d) = 0.4$)
- The spam of the scale is deficient. In some cases, the difference between a poor and a good tuning is very small.

To illustrate these limitations, consider two different systems with the following transfer functions:

$$G_1 = \frac{1}{2s+1}e^{-s} \quad and \quad G_2 = \frac{1}{(3s+1)^2}e^{-0.5s} \qquad (2)$$

Three different PI controllers were designed for each plant, using Frequency Domain Methodology (Trierweiler et al. 2000):

- with the closed loop performance 6 times faster than open loop (6X),
- equal closed and open loop performance (1X), and
- closed loop 4 times slower than open loop (0.25X).

Table 1 shows the Harris's Index estimated calculated by the FCOR algorithm for the two plants with the three controllers. This simple example clearly shows the scale and resolution problems of FCOR.

To overcome these limitations, commercial software try to assess the performance based on heuristics that consider the performance indexes and the characteristics of the loop (flow, temperature, pressure, etc.), (Thornhill et al., 1999).

Table 1: FCOR for two different plants with three different controllers

Controller	Plant 1	Plant 2
6X	0.8823	0.7338
1X	0.7638	0.3306
0.25X	0.5724	0.0729

4 DEVELOPING AN INFERENTIAL MODEL

The main reason of failure of conventional performance indexes is the absence of a common and absolute target to quantify the performance and robustness. These absolute index are common in the literature, however they are infeasible to determine in a real plant, because invasive tests are necessary.

In this work, a set of representative indexes are determined using a set of plants with different characteristics and several control loop performance measurements calculated for several controller and plant pairs. In these plants, invasive tests are made and the absolute indexes are calculated. After that, using on-line measured indexes, a curve that gives the absolute indexes is fitted, using different techniques. Figure 1 shows a schematic representation of the inferential model proposed in this work.

The main contribution of this work is build an inferential model for performance and robustness indexes that need intrusive tests to be calculated, using only indexes that can be quantified on-line.

The performance and robustness indexes to be inferred are:

- Ratio between closed loop and open loop rise time (Rt_R);
- Integral square error (ISE) for a unitary setpoint change and unitary load disturbance normalized by the ISE of a controller with the same performance of open loop;
- Gain margin (GM).

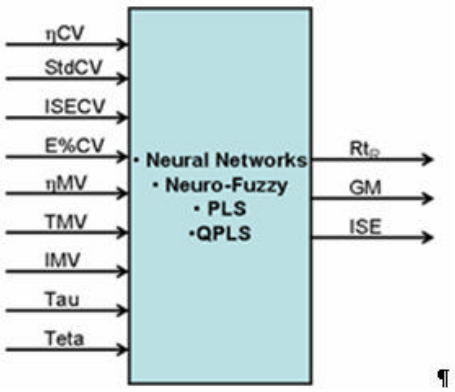


Figure 1: Schematic representation of the inferential model proposed

To develop the inferential model the following four different techniques will be will be tested:

- Neural Networks
- Neuro-Fuzzy (ANFIS)
- PLS
- QPLS

5 THE PLANTS AND CONTROLLERS

5.1 The Plants

Initially a set of plants to be analyzed are determined, emphasizing different effects: dynamics, model order, RHP zeros, pure time delay, RHP-poles, and integrating processes. These plants are very similar to used by Åström, and Hägglund (1995) to develop the Kappa-Tau PID tuning method.

5.2 The Controllers

The controllers are tuned using the Frequency Domain Methodology (Engell and Müller, 1993, Trierweiler *et al*., 2000). For each plant, the desired performance is determined and the methodology gives the parameters for each plant, when the desired performance is achievable. In all cases, the controller used is Proportional-Integral (PI). Table 2 shows the desired performance for the controller, when it is achievable.

The system is affected by white noise and a periodic load disturbance with variable frequency and constant magnitude (unitary). The indexes are calculated in each scenario (each plant with each

controller performance) with four different periods for the periodic disturbances (10, 20, 30 and 50 time units).
About 80% of points are used to train each fit method and 20% of points are used to test the curves and quantify the correlation among the points. The points of the test set were selected randomly from the total set.

Table 2: Set of desired performances for the controllers

N	Closed loop/open loop performance ratio	Overshoot (%)
1	10	10
2	8	10
3	6	20, 10, 5
4	5	10
5	4	20, 10, 5
6	2	10
7	1.5	10
8	1	5
9	0.75	5
10	0.5	5
11	0.25	5

5.3 Variable selection

A subset of original variables is selected, using a Genetic Algorithms (GA). The implemented algorithm has the traditional operators (cross-over, reproduction, and mutation), with binary codification (Han and Yang, 2004). A maximum set of 5 variables could be selected. The objective function, to be minimized, is the predictive sum of squares (PRESS) (Qi and Zhang, 2001). The five variables selected by the GA are the same for three developed inferential models (i.e., Rt_R, GM, and ISE):

- Estimated the CV minimal variance
- Process dead time (estimated)
- Process time constant (estimated)
- Standard deviation of CV
- Travel of manipulated variable (MV)

6 RESULTS

This section shows the results of the developed inferential models obtained by different techniques for the set of plants.

The results will be presented with more details for the ratio between closed loop and open loop rise time.

The target to be achieved is normalized using logarithm, because the most controllers are faster than closed loop, and some are slower. This normalization gives more importance to the faster controllers, and makes the interpolation more representative.

To quantify the quality of each method, the correlation coefficient (R^2) between y_p (predicted value) and y (target) is calculated.

6.1 Neural Network Inferential Model

The first approach to build the inferential model was based on neural networks (Hagan, 1995). The neural network used has 2 layers, the first using hyperbolic tangent sigmoid neurons with variable number and the second with one linear neuron. The train method used is Levenberg-Marquardt backpropagation (Hagan, 1995). The first test aims to estimate the ratio between closed loop and open loop rise time. Table 3 shows the performance of several neural networks with different number of neurons in the hidden layer. The results are calculated considering the validation data.

Table 3: Relationship between neural networks performance and number of neurons

Neurons	R^2 for the validation data
10	0.91
20	0.96
30	0.96
40	0.96
50	0.97
60	0.97

The best relationship between prediction quality versus number of neurons is 20. Figure 2 shows the relationship between the predicted and real values for the closed/open loop ratio for each controller.

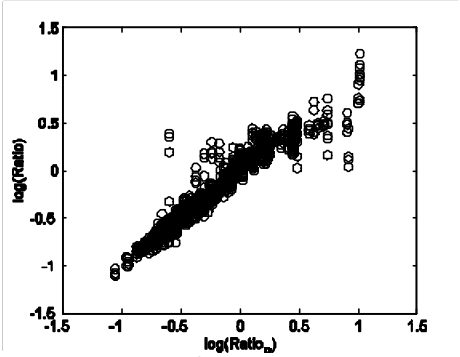


Figure 2: Estimation of Rt_R rise time using neural networks for the validation data

Table 4: Relationship between neural networks performance for gain-margin (GM) and Integral of Square Error (ISE) with different neural-networks

Neurons	R^2 for GM	R^2 for ISE
10	0.95	0.96
20	0.96	0.98
30	0.97	0.99
40	0.97	0.99
50	0.98	0.99
60	0.98	0.99

As shown in Figure 2 and Table 3, the prediction of the neural networks is very good. Similar results are also obtained for the gain margin and the integral of square error (ISE) as it is shown in Table 4.

6.2 Neuro-Fuzzy (ANFIS) Inferential Model

The second class of inferential model, which has been tested, is Neuro-Fuzzy. The architecture used is ANFIS proposed by Jang (1993). The system has the same five inputs used by the neural networks. The output to be fitted is closed loop and open loop rise time ratio (Rt_R). The ANFIS has 2 membership functions, for each input. The architecture of each membership function is *S-Functions* (Kasabov, 1998). Figure 3 shows the result using neuro-fuzzy approach.

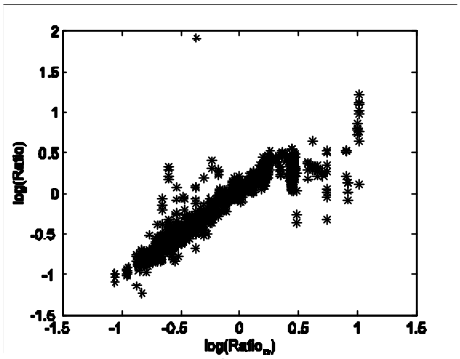


Figure 3: Interpolation of the Rt_R time using neuro-fuzzy

Figure 4 shows that the predictive capacity of the Neuro-Fuzzy models is less effective than by neural networks. The R^2 obtained is also lower (0.91). Similar results are obtained for gain-margin and integral square error (ISE). The correlation factors obtained are 0.89 and 0.96, respectively.

6.3 Partial Least Squares (PLS) Inferential Model

PLS is a very robust technique to interpolate correlated and noise data BAFFI *et al*. (1999). In this section, the latent variables are defined as linear functions of input variables. All inputs are provided to the algorithm, and different number of latent variables is used, as shown in Table 5:

Table 5: Correlation factor for the inferential models for Rt_R using linear PLS models

Latent variables	R^2 for the validation data
1	0.76
2	0.81
3	0.81
4	0.82
5	0.82
6	0.82

Figure 4 shows the best result obtained with linear PLS. The correlation index (R^2) (cf. Table 5) and Figure 4 show that the inferential quality is very poor. Similar results are obtained for GM and ISE, as shown in Table 6.

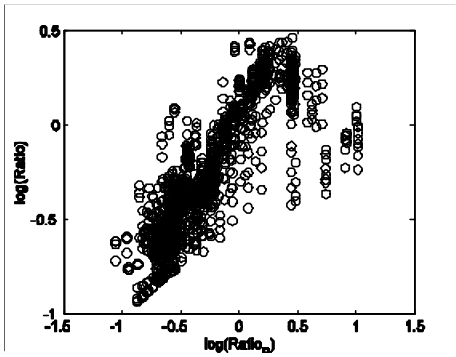


Figure 4: Interpolation of Rt_R using linear PLS

Table 6: Results obtained by inferential linear PLS models developed for gain-margin (GM) and Integral of Square Error (ISE) with different number of latent variables

Latent variable	R^2 for GM	R^2 for ISE
1	0.74	0.73
2	0.87	0.82
3	0.89	0.83
4	0.90	0.84
5	0.90	0.84
6	0.90	0.84

6.4 Quadratic Partial Least Squares (QPLS) Inferential Model

QPLS is an alternative to improve the results of PLS models (BAFFI *et al*., 1999). Table 7 shows the obtained results with QPLS approach.

Table 7: Correlation factor for the inferential models for closed loop and open loop rise time ratio using QPLS models

Latent variable	R^2
1	0.83
2	0.84
3	0.86
4	0.87
5	0.88
6	0.88

Table 8: Results obtained by inferential QPLS models developed for gain-margin (GM) and Integral of Square Error (ISE) with different number of latent variables

Latent variable	GM	ISE
1	0.90	0.86
2	0.90	0.88
3	0.90	0.89
4	0.91	0.90
5	0.91	0.91
6	0.92	0.91

The QPLS models gave a good result, however inferior than neural networks. Similar results are

obtained for the inferential models developed for GM and ISE (cf. Table 8).

7 APPLYING TO THE CASE STUDY

The example shown in section 3 is used to verify the prediction quality of the neural network inferential model to predict Rt_R. Table 9 compares the results for each case. The true values are in the first column while the corresponding predictions are in the other columns for each plant.

Table 9: Prediction of the best inferential model (IM) for two plants

True Values	Plant 1	Plant 2
6X	5.3	5.7
1X	0.87	1.3
0.25X	0.37	0.24

Table 9 shows that the IM gives representative results for Rt_R. These results are much superior and conclusive than of the obtained by the performance index based on minimum variance controllers. With these results, a control engineer can easily quantify the performance for each controller for the plant.

8 CONCLUSIONS

The work presented in this paper built an "inference model" that can really and conclusively quantify closed-loop performance. This novel approach is based on measurements that can be easily assessed on-line, without intrusive tests. The indexes determined are closed loop and open loop rise time ratio, gain margin (GM) and integral square error (ISE).

A set of plants are generated, considering different processes and a set of controllers with different performances are tuned for each plant. The input indexes are calculated and invasive tests are made to determine the output indexes. The set of input – output indexes are fitted using different techniques. The best results are obtained using neural networks, for the three output indexes. Neuro-Fuzzy and QPLS have also given good results. Linear PLS gave the worst results.

Based on the results obtained, we can affirm that the IM proposed can not only assess the performance of industrial controllers, but quantify the real closed-loop performance and robustness, with absolute indexes, with no need of intrusive tests.

ACKNOWLEDGMENT

The authors thank CAPES, PETROBRAS and FINEP for supporting this work.

REFERENCES

Åström, K. J.; Hägglund, T.; (1995). PID Controllers: Theory, Design, and Tuning. 2ª ed. Research Triangle Park: Instrument Society of América.

Baffi, G.; Martin E. B.; Morris A. J.; (1999), Non-Linear Projection To Latent Structures Revisited: The Quadratic PLS Algorithm. Computers & Chemical Engineering, V.23 , 395-411.

Engell, S. E.; Müller, R.; (1993) Multivariable Controller Design by Frequency Response Approximation. *Proc. of the 2nd European Control Conference*, 3, 1715-1720.

Hagan, M.T.; Demuth, H.B.; Beale, M.; (1995). Neural Networks Desing. PWS Publishing Company.

Han, S. H.; Yang, H.; (2004). Screening Important Design Variables For Building A Usability Model: Genetic Algorithm-Based Partial Least Squares Approach. Industrial Ergonomics, V.33, 159-171.

Harris, T. J.; Seppala, C. T.; Desborough, L. D.; (1999). A review of performance monitoring and assessment techniques for univariate and multivariate control systems. Journal of Process Control, Vol 9, pp. 1-17.

Huang, B.; Shah, S.; (2001) Performance Assessment of Control Loops. Theory & Applications, Springer, 1999, ISBN 1-85233-6390; DM 139, Journal of Process Control, Volume 11, Issue 4, Pages 441-442.

Jang, J.-S.R. (1993).ANFIS: adaptive-network-based fuzzy inference system.IEEE Transaction on System Man and Cybernet, 23(5), 665–685.

Kasabov, N. K.; (1998). Foundations on Neural Networks, Fuzzy Systems, and Knowledge Engineering. The MIT Press, Cambridge, Massachusetts.

Marlin, T. E. (1995). Process Control, McGraw-Hill.

Qi, M; Zhang, G. P.; (2001). An Investigation of Model Selection Criteria For Neural Network Time Series Forecasting. European Journal of Operational Research V.132, 666-680.

Thornhill, NF.; Oettinger, M.; Fedenczukc (1999). Refinery-wide control loop performance assessment. Journal of Process Control V.9, 109-124

Trierweiler, J. O.; Müller, R.; Engell, S. (2000). Multivariable Low Order Structured-Controller Design by Frequency Response Approximation. Brazilian Journal of Chemical Engineering, 17, n.º 4, 793-807.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

VARIABILITY MATRIX: A NEW TOOL TO IMPROVE THE PLANT PERFORMANCE

M. Farenzena, J. O. Trierweiler[#]

Group of Integration, Modeling, Simulation, Control and Optimization of Processes (GIMSCOP)
Department of Chemical Engineering, Federal University of Rio Grande do Sul (UFRGS)
Rua Luiz Englert, s/n CEP: 90.040-040 - Porto Alegre - RS - BRAZIL,
Fax: +55 51 3316 3277, Phone: +55 51 3316 4072
E-MAIL: {farenz, jorge}@enq.ufrgs.br

Abstract: This work introduces a novel methodology to quantify the profit gain due to reduction in the product variability. The base of the proposed approach is the variability matrix (VM), which relates how the loop variance of main loops is changed when the variance of the other loops are changed. Based on the potential reduction on the main loop variance, it is possible to quantify the economic impact produced by improving the tuning of given control loop. Based on the VM, it is possible to select the control loops responsible for the major impact in the variability of the products and which should be the vocation of the loop: good performance of robustness. The VM concept is applied to a simple distillation process. This example shows how the plant profitability can be improved by utility reduction and by selling products more impure. *Copyright © 2006 IFAC*

Keywords: Process Control**,** Variability, Control System Design, Economic Design.

1 INTRODUCTION

Reduce process variability allows the process arrives near the restriction, increasing its performance. In the most of processes the operating point with the high efficiency is normally in a corner of the operating window. Achieve this point means reduce energy and increase production (Seborg et al., 1989).

To keep the process near the maximum efficiency operating point, the performance of control system should be ensured.

Usual refinery or petrochemical plants has hundreds or thousand loops and guarantee the performance of all loops is impossible without a systematic procedure, which would determine and sort the loops following their economic impact to the process profitability. Usually, several control loops can be improved reducing the loop variability. But the main question is which loop should be attack firstly. Many times, some loops can have a great potential to reduce its variance, but they have a very small impact in the plant profitability. On the other hand, some other loops can have a little potential to reduce its variance, since they are already well tuned. But if these loops could be a little better, they would

[#] Author to whom the correspondence should be addressed.

contribute more significantly to the final economical result. Therefore, to select and order, which loops should be firstly improved, it is necessary to quantify the corresponding economic impact.

In this paper the novel concept of variability matrix is introduced as a tool to quantify the economic impact of improving the control loop performance.

The base of the proposed approach is the variability matrix, which relates how the loop variance of main loops is changed when the variance of the other loops are changed. Based on the potential reduction on the main loop variance, it is possible to quantify the economic impact produced by improving the tuning of given control loop.

The article is structured as follows: in the section 2 some definition is introduced using a distillation process. In section 3, a new methodology to relate the decrease of the variability of a given product in profit will be defined. In section 4, the concept of variability matrix is formally presented. In section 5 the new methodology is applied to the distillation introductory exempla. Finally, the conclusions concerning to this work are shown in section 6.

2 DEFINITIONS

To quantify the economic impact, it is interesting to classify the control loops into two following categories

1. **Main or Primary Loops** are the loops that directly control the products specification. Its performance improvement causes the reduction in product variability, which can be directly translated into profitability.

2. **Auxiliary or Secondary Loops**: Loops that do not directly control the product quality, but can indirectly affect the product variability.

To exemplify these definitions, Figure 1 shows a typical distillation process with a vessel to smooth the feed of the second column (V1). The objective is obtaining all products with high purity.
The primary loops are the three cascades that control the composition of both products (AC1, AC-3, and AC-4). The other loops are called secondary loops.

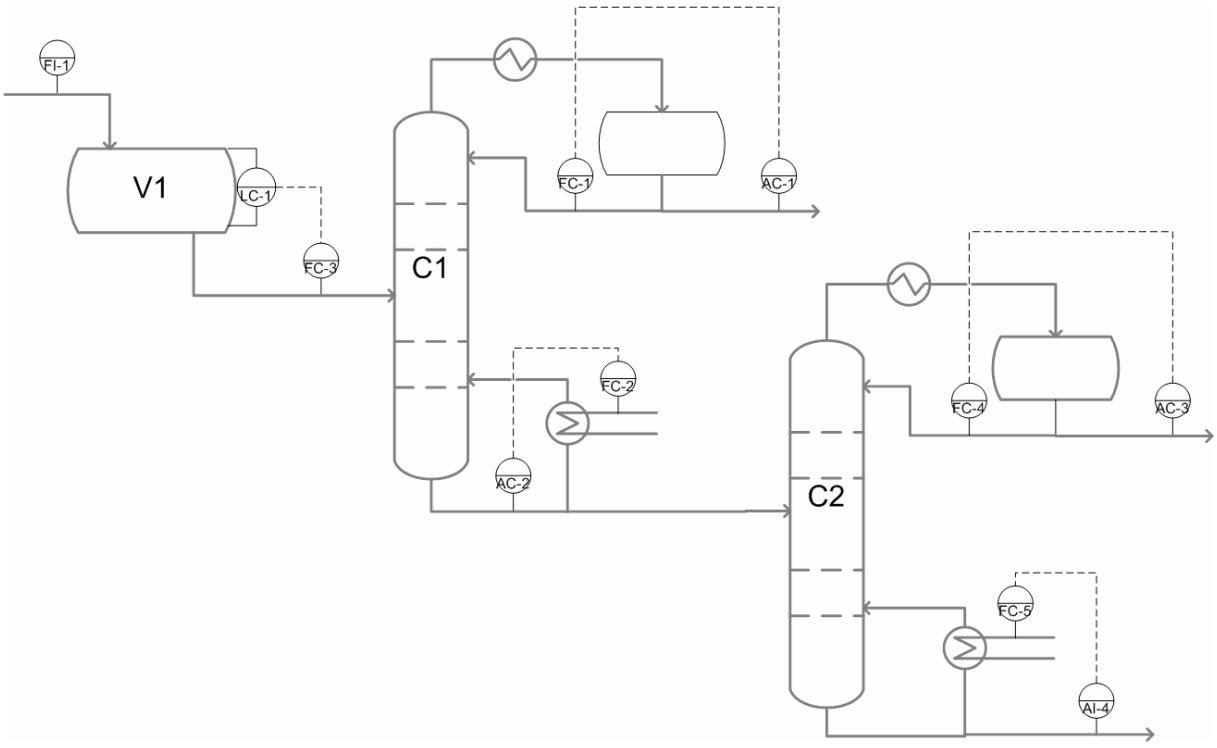


Figure 1: Schematic representation for a typical distillation process

The impact of each loop and its magnitude in the variability of the products will be shown in the section 4, where the concept of the matrix of variability is introduced.

Some of control loops should have a good performance to ensure the products specification (e.g. AC-2) and others should only smooth de disturbances, to stabilize the feed of the columns (e.g. the level of V1).

3 ASSESSING THE PROFIT OF THE PRIMARY LOOPS

In the literature there are some methodologies to estimate the profit as function of the operating point. The most famous is called Taguchi (Taguchi, et al., 1989) and relates the profit as a quadratic function of product purity where the vertex is in the specification and the profit decreases with a quadratic constant. The form of the curve that relates the profit as function of process performance is purely heuristics. Besides, the parameters obtaining is also a difficult task, and does not use explicitly the energy reduction or production increase.

In this article a simple methodology based on first principles (mass balance) is introduced to quantify the profit of the reduction of the primary loops. In the next section, this methodology is applied to translate the how can a reduction of variability of the secondary loops into final product variability.

When the process arrives closer to the specification, there are three ways of increasing the profitability:

1. Reduction of energy consume: when the specification of the products is lower, the energy spent is also lower, in the most of process.
2. Sell impurity as product: with the product near the specification a part of impurities can be sold with the same price of expensive product.
3. Increasing of the unity production: If the equipment restricts the increase of the plant production, the variability reduction allows the production to increase. This analysis is relevant but complex, and will not be considered in this article.

3.1 Process Performance

The key factor to quantify the profit due to variability reduction is measure the process performance. The performance is specific for each process (e.g. product composition, conversion, profit).

When the performance curve has a high slope, increase controller performance will improve the plant performance. On the other hand, if the curve is flat, a low variability controller will achieve almost the same process performance of a controller with poor performance.

With a process histogram, the mean process performance can be easily calculated. The process performance is multiplied by the frequency that the process assumes that process variable interval (Marlin, 1991). Figure 2 shows two different performance curves.

Depending on the performance curve, the process variable can have a narrow or broad distribution: when the performance is the same in all operating point (curve 2) the simple reduction of variability of the loop does not represent increase profit. On the other hand, when the performance has abrupt slope, small changes in variability will cause significant increase in the performance (values near 100 of curve 1).

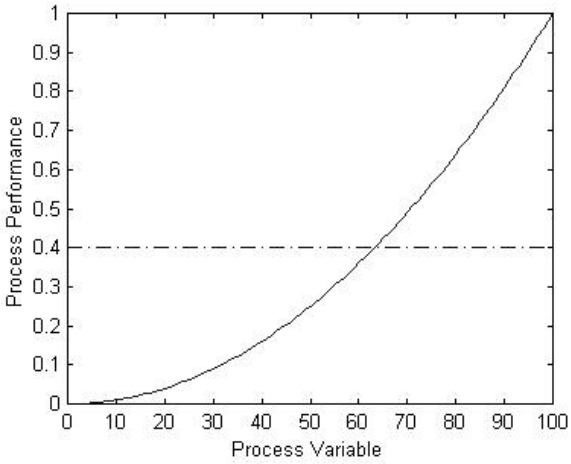


Figure 2: Schematic representation of performance curves

3.2 Reduction of Energy Consume

When the process achieves an operating point of higher efficiency, the energy consumed usually decreases, because the higher profit operating point is in a corner of the operating window (Marlin, 1995). To quantify the profit given by energy reduction (P_{ER}), the difference in the utilities streams (ΔU) are calculated:

$$\Delta U = G(0)^{-1} \Delta Y \qquad (1)$$

Where G(0) is the static gain matrix for the system and ΔY is the difference in the operating points specification for the controlled variables variability reduction.

The profit given by energy reduction (P_{ER}) can be estimated multiplying the difference in the utility streams (ΔU) by the energy cost (E_C).

$$P_{ER} = \left(G^{-1} \Delta Y\right)' E_C \qquad (2)$$

Suppose a system with two manipulated variables, with the following static gain matrix:

$$G(0) = \begin{bmatrix} -0.01 & 0.005 \\ 0.005 & -0.01 \end{bmatrix} \qquad (3)$$

After the process variability reduction, the difference in the specification of products is 0.005 and 0.01. The energy cost are (US$) 300 and 200 per energy unity. The profit due to energy reduction can be calculated as follows:

$$P_{ER} = \left(\begin{bmatrix} -0.01 & 0.005 \\ 0.005 & -0.01 \end{bmatrix}^{-1} \begin{bmatrix} 0.005 \\ 0.01 \end{bmatrix} \right)' \begin{bmatrix} 300 \\ 200 \end{bmatrix} = 733 \tag{4}$$

3.3 Sell a product more impure

When the variability of a given product becomes smaller, the mean composition can be reduced to a new value closer the specification, allowing mix in the final product a greater part of impurities, less valuable than the product.

To estimate the profit (P_{IP}), we need to determine the amount of impurities (ΔF) can be mixed in the product. A simple mass balance can be done, based on the actual flow (F), the final and initial contamination of a given stream (y_F and y_I) and the increased flow (ΔF), considering that the amount of product is constant (Fy_I).

$$\begin{aligned} Fy_I &= (F + \Delta F) y_F \\ \Delta F &= F\left(\frac{y_I}{y_F} - 1 \right) \end{aligned} \tag{5}$$

Multiplying the increased flow (ΔF) by the price difference between the main product of the stream and the contaminant (P_E - P_C), the increased profit can be estimated.

$$P_{IP} = F\left(\frac{y_I}{y_F} - 1 \right)(P_E - P_C) \tag{6}$$

4 ASSESSING THE PROFIT OF THE SECONDARY LOOPS – THE VARIABILITY MATRIX

4.1 Definition

To determine the influence of each loop in the final product we need to translate the variance reduction of all control loops into the primary loops. The variability matrix is a matrix where the elements (i,j) quantify how the change in the variance of the control loop (j) produces a change in the variance of the main loop (i), i.e.,

$$VM(i,j) = \frac{\Delta \text{variance}(cv_i)}{\Delta \text{variance}(cv_j)} \tag{7}$$

The structure of the variability matrix is the following:

- Rows: The rows show the influence of each loop in the same final product. The number of rows is the same as the products or main loops.
- Columns: Shows the influence of a given loop in each final product. The number of columns is the same as the number of control loops implemented in the plant. The first columns correspond to the main loops.

Figure 3 shows a schematic representation of variability matrix.

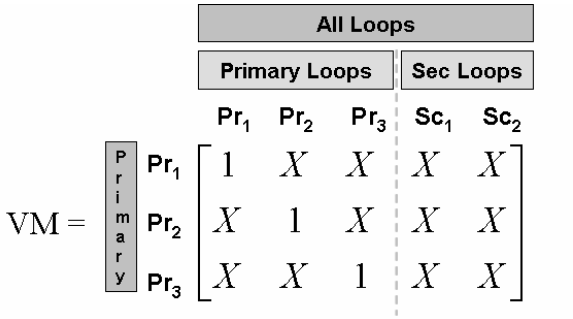


Figure 3: Schematic representation of variability matrix

4.2 Similarities to a gain matrix

The variability matrix has similar properties to static gain matrix. Its values can be positive or negative. A positive value of the element VM(i,j) means that increasing the variance of the auxiliary loop j will increase the variance of the main loop i. For the control loop j "faster is better".

On the otherwise, if VM(i,j) is negative, it will decrease the variance of the main loop i by increasing the variance of the loop j. This situation is typical for buffer tanks, which typically should reduce the propagation of the disturbance by a variation on the tank level. Here, for the control loop j, "slower is better" is true. These loops should have a poor performance, being responsible for smooth variations in the process. The limit in this case is the safety of the unity. In this case, usual PID tuning methodologies are not adequate. We suggest in this case a methodology to tuning level controllers for tanks shown in Smith (2002).

The procedure to construct the variability matrix is analogous to build the static gain matrix. Any identification technique can be applied to identify it. For that, it should be used as inputs the variance of the primary and secondary loops and as outputs the variance of the main loops. As by standard identification procedure, it is necessary to make a perturbation in the inputs. In this case the perturbation can be produced by a change in controller parameters or through the addition of known perturbation in the control loop, which will change the control loop variance.

5 EXAMPLE

This section will show the use of these new concepts to reduce the variability of a given plant, showing the profit before and after the tests.

5.1 The plant

Consider the example shown in the Figure 1. This hypothetical plant is feed by three different products (A, B, and C).

The plant has two distillation columns with internal trays. Before the columns, the system has a vessel to smooth the variations in the unity feed, which are the main disturbance. Both columns have total condenser.

The total feed of the unity is 6000 ton/day. In the top of the first column (C1) the less valuable component (A) is removed from the unity. In the bottom of the column, the products B and C are removed. This stream feeds the second column (C2). The products B and C are removed in the top and bottom of C2, respectively.

The main disturbance of the unity is the feed flow of V1 that has a periodical oscillation. The columns have on-line analyzers that can provide their composition to the control system.

We assume that the inventory control is properly tuned. The control structure is build as shown in Table 1:

Table 1: Control structure for the distillation process

Loop	CV	MV
AC-1	x_B in top stream of C1	FC-1
AC-2	x_A in bottom stream of C1	FC-2
LC-1	Level of V1	FC-3
AC-3	(x_A + x_C) in top stream of C2	FC-4
AC-4	x_B in bottom stream of C2	FC-5

Where x_i is the mass fraction of the component *i*. All these variables are controlled using a proportional-integral controller (PI).

The objective of this unity is split all the three components. Table 2 shows the value of each product (US$/ton), their specification and price.

Table 2 shows that the product A is the less valuable while the product C is the more valuable. The specification is the same for all the three products.
To ensure the products specification, the mean of each controlled specification must be 3 standard deviation from the restriction.

Table 2: Values and feed of each product

Prd	Feed (ton/day)	Specification	Price (US$/ton)
A	1000	$x_B < 0{,}05$	400
B	2000	$x_{A+C} < 0{,}05$	900
C	3000	$x_B < 0{,}05$	1200

Table 3 shows the utility consumed in the unity and their costs US$/ton:

Table 3: Utility consumed and price

Utility	Consume (ton/day)	Cost (US$/ton)
C1 - Condenser	2500	20
C1-Reboiler	2000	80
C2 - Condenser	1300	200
C2-Reboiler	1000	150

The transfer matrix G1 for the column C1 is given by

$$G_1 = \begin{bmatrix} \dfrac{-0.0001}{8s+1}e^{-2s} & \dfrac{0.000085}{12s+1}e^{-3s} \\ \dfrac{0.00007}{15s+1}e^{-3s} & \dfrac{-0.0001}{10s+1}e^{-20s} \end{bmatrix}. \qquad (8)$$

Whereas G2 is the transfer matrix of the column C2 and is given by

$$G_2 = \begin{bmatrix} \dfrac{-0.00015}{8s+1}e^{-6s} & \dfrac{0.0001}{12s+1}e^{-3s} \\ \dfrac{0.00009}{15s+1}e^{-3s} & \dfrac{-0.00021}{10s+1}e^{-5s} \end{bmatrix}. \qquad (9)$$

The transfer function G3 is the vessel V1

$$G_3 = \frac{0.5}{40s}. \qquad (10)$$

All the five controllers are tuned using the methodology based in frequency domain (Engell and Müller, 1993) using a desired performance 2 times faster than open loop.

Table 4 shows the mean and the standard deviation for each product, with the mentioned tuning.

Table 4: Mean and standard deviation for each stream

Product	Mean	Standard deviation
A (C1 – Top)	0.957	0.0022
B (C2 – Top)	0.974	0.0081
C (C2 – Bottom)	0.963	0.0042

We consider that the mean composition of A in bottom of C1 is maintained constant during all tests. The variability matrix (VM) for the system is given by

$$VM = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & V1-LC & C1-BOT \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & -0.0001 & -0.0155 \\ -0.41 & 1 & -0.08 & -0.0006 & 1.25 \\ -0.29 & -0.05 & 1 & -0.0002 & 0.5 \end{bmatrix} \end{array} \quad (11)$$

The variability matrix shows that the increase of the variability of the product A will decrease the variability of the others products (B and C). The same can be said of the product B and C, if the variability of one product decrease, the variability of the other will increase. But nothing will occur with the variability of A. The control of the V1 level could reduce the variability of all products, if a robust tuning is applied.

Based on VM, we can see that the V1 level causes increase of variability in all products. A new set of parameters is used, based on the methodology shown in Smith (2002) to retune the level control. The impact of new tuning is shown in Table 5.

Table 5: Mean and standard deviation for each stream with the level of V1 more robust

Product loop	Mean	Standard deviation
C1 – Top	0.955	0.0018
C2 – Top	0.965	0.0048
C2 – Bottom	0.959	0.0029

Table 5 shows that the variability of all products are reduced, allowing the system arrive nearer the restriction.

The loop that control the composition of the product A also cause impact in all other loops, as shown by (11). A new adjust more robust is made (3 times slower than open loop). Table 6 shows the new operating point, with the new parameters:

Table 6: Mean and standard deviation for each stream with the C1-Top more robust

Product loop	Mean	Standard deviation
C1 – Top	0.966	0.0053
C2 – Top	0.956	0.0019
C2 – Bottom	0.956	0.0019

Table 6 shows that the variability of product A becomes higher, however the B and C standard deviation becomes lower, allowing the system arrives closer to the restrictions.

Now, the indexes to quantify the profit (eqs. 2 and 6) of the new tuning will be applied. Table 7 shows the profit for each controller that is retuned for the reduction in the utilities and sell products more impure (US$/day).

Table 7 shows a visible increase in the profit, with energy reduction and product more impure. This example shows that key loops are responsible for high variations in the unity and the VM is a powerful tool to detect these loops, guiding the control engineer to achieve a more profitable operating point for his plant.

Table 7: Profit increase for energy reduction and product more impure

Increase profit	Energy reduction	Product more impure	Total (US$/day)
Level Retuning	33800	12200	46000
x_A loop retuning	7560	16800	24360
Total	41300	29000	**70360**

6 CONCLUSIONS

In this article a new tool called Variability Matrix (VM) is introduced. The VM show the influence of a given loop in the variability of all products of the plant. This information allows identifying the loops that cause the variability in each product, determining what would be the best performance of each controller (fast or robust).

Besides, this article shows also a methodology to quantify the gain caused by the variability reduction, due to the utility reduction and sell a product more impure (closer to the specification).

This new methodology is applied to a hypothetical distillation process, showing the potentialities of the VM. Based on VM, two loops are retuned, showing a visible increase in the plant profitability.

ACKNOWLEDGMENT

The authors thank CAPES, PETROBRAS and FINEP for supporting this work.

REFERENCES

Engell, S. E.; Müller, R.; (1993) Multivariable Controller Design by Frequency Response Approximation. *Proc. of the 2nd European Control Conference*, 3, 1715-1720.

Marlin, T. E. (1995). Process Control, McGraw-Hill.

Seborg, D. E., Edgar, T. F., & Mellichamp, D. A. (1989). Process dynamics and control. New York: Wiley.

Skogestad, S.;Postlethwaite, I. (1996). Multivariable Feedback Control _Analysis & Design. John Wiley&Sons.

Smith, C. A.;(2002). Automated Continuous Process Control. John Wiley & Sons, Inc.

Taguchi, G.; Elsayed, E.A.; Hsiang, T.; (1989) Quality Engineering in Production Systems, McGraw-Hill, New York.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ASSESSMENT OF ECONOMIC PERFORMANCE OF MODEL PREDICTIVE CONTROL THROUGH VARIANCE/CONSTRAINT TUNING

Fangwei Xu [*], **Biao Huang** [*], **Edgar C. Tamayo** [**]

[*] *Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2G6*
[**] *Syncrude Canada Ltd., Fort McMurray, AB, Canada, T9H 3L1*

Abstract: Multivariate controller performance assessment (MVPA) has been developed over the last several years, but its application in advanced model predictive control (MPC) has been very limited mainly due to issues associated with comparability of variance control objective and that of MPC. MPC has been proven as one of the most effective advanced process control (APC) strategies to deal with multivariable constrained control problems with an ultimate objective towards economic optimization. Any attempt to evaluate MPC performance should therefore consider constraints and economic performance. This work is to establish a link between variance control and MPC in terms of economic performance. We show that the variance based performance assessment may be transfered to economic assessment of MPC. Algorithms for economic performance assessment and tuning are developed through linear matrix inequalities using routine operating process data. The proposed algorithms are illustrated via an industrial MPC application example.

Keywords: performance assessment, model predictive control, economic performance assessment

1. INTRODUCTION

Model predictive control (MPC) has been proven as one of the most effective advanced process control (APC) strategies to deal with multivariable constraint control problems However, less efforts have been made on the performance evaluation of existing MPC applications, especially on economic performance.

Although MVPA has been developed over last several years, its application on MPC evaluation has been very limited mainly due to issues associated with comparability of MVC objective and that of MPC strategy. One of the most important incentives of MPC applications is to deal with multivariable constrained control problems with an ultimate objective on economic optimization. Any attempt to evaluate MPC performance should therefore consider constraints and economic benefits.

In traditional economic benefit analysis, the back off approach has been applied in the benefit analysis of improved process control. The benefit potential is achieved against the base case operation by reducing the variance of quality variables and pushing the average values closer to the optimum point or constraint limit (Muske, 2003). The base case operation should be a period of

typical closed-loop operation with the existing control system. Benefit analysis for different base case conditions should be done separately since they may lead to different economic benefit values (Muske, 2003). An appropriate back off away from the constraint limit should be introduced and the optimal operation is too conservative if the constraint limit is never violated. Many different rules have been discussed for allowable constraint violation. A reasonable rule should be adopted in terms of base case condition and desired specifications, e.g., 5 percentage of violation. Once the base case operation and the optimal operation condition are both identified, the economic benefit potential is readily obtained when the economic objective function is explicitly established.

In the latest generation MPC algorithms, a separate steady state optimization is performed at each control cycle in order to drive steady state inputs and outputs to their optimal economic targets For example, industrial model predictive control integrates a linear program (LP) and/or a quadratic program (QP) for economic optimization. Since this LP or QP reflects economic objective explicitly, it can be utilized to evaluate the economic performance of MPC applications. This work is to establish a link between variance control and MPC in terms of economic performance. We show that the variance based performance assessment may be transfered to economic assessment of MPC. Algorithms for economic performance assessment and tuning are developed through linear matrix inequalities using routine operating process data. The proposed algorithms are illustrated via an industrial MPC application example.

The remainder of this paper is organized as follows. In Section 2 several different scenarios are described in the form of constrained quadratic optimization problems. Section 3 presents and explains a systematic approach for the purpose of economic performance assessment. The QP problem is reformulated as LMI in Section 4. An industrial MPC application is evaluated for economic performance in Section 5, followed by concluding remarks in Section 6.

NOTATION

a_{ki}	quadratic coefficient of i^{th} output variable
a_{kj}	quadratic coefficient of j^{th} input variable
b_{ki}	linear coefficient of i^{th} output variable
b_{kj}	linear coefficient of j^{th} input variable
u_{dkj}	target value of j^{th} input variable
u_{kj0}	sampled value of j^{th} input variable
y_{dki}	target value of i^{th} output variable
y_{ki0}	sampled value of i^{th} output variable
K_{ij}	the steady state gain value
N_u	the number of input variable
N_y	the number of output variable
R_{uj}	changing ratio of j^{th} input variable
R_{yi}	changing ratio of i^{th} output variable
U_{holkj}	half of constraint range of j^{th} input variable
U_{qorj0}	quarter of range of j^{th} input variable
U_{Hkj}	high limit of j^{th} input variable
U_{Lkj}	low limit of j^{th} input variable
Y_{holki}	half of constraint range of i^{th} output variable
Y_{stdi0}	standard deviation of i^{th} output variable
Y_{Hki}	high limit of i^{th} output variable
Y_{Lki}	low limit of i^{th} output variable

2. PROBLEM DESCRIPTION

For illustration, assume a multivariable process consists of only two controlled variables with interaction, where y_1 is a quality variable that has direct impact on profit and y_2 is a constrained variable (Figure 1). Because of disturbances, there is variability on both y_1 and y_2. Assuming the optimal operating condition of y_1 is located on its upper limit, it is clear from the figure that the actual average operating condition (dash line) is not at its optimal operating condition, leading to lost profit. The base case operation is defined by its current mean values and variances. A reasonable percentage of constraint violation, 5%, is adopted such that 95% of operation falls within the range of ± 2 times standard deviation. Since the benefit potential is calculated against that of base case operation in all scenarios, in the following we list the problem formulations of optimal operations for different scenarios with quadratic economic objective function in the steady state optimization.

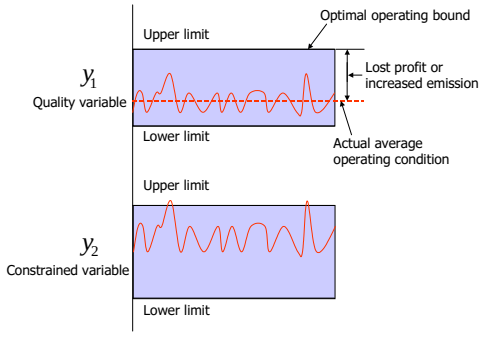


Figure 1. Base case operation

2.1 Assessment of ideal yield

In the ideal scenario strict steady state operation is considered and there is no variability on both y_1 and y_2 shown in Figure 1. Under this scenario the

operation of y_1 can be pushed closest to its optimal operating point, upper limit in this example. In this case the operating points of y_1 and y_2 are the decision variables, and the corresponding optimization problem can be formulated as follows.

$$\min_{\bar{y}_i, \bar{u}_j} J = \frac{1}{N_L} \sum_{k=1}^{N_L} J_k \tag{1}$$

subject to

$$\begin{aligned} Y_{Lki} \le \bar{y}_i \le Y_{Hki}, \quad i = 1, \cdots, N_y \\ U_{Lkj} \le \bar{u}_j \le U_{Hkj}, \quad j = 1, \cdots, N_u \end{aligned} \tag{2}$$

$$\sum_{j=1}^{N_u} [K_{ij} \times \Delta \bar{u}_j] = \Delta \bar{y}_i, i = 1, \cdots, N_y$$

$$\bar{y}_i = \bar{y}_{i0} + \Delta \bar{y}_i, \bar{y}_{i0} = \frac{\sum_{k=1}^{N_L} y_{ki0}}{N_L}, i = 1, \cdots, N_y$$

$$\bar{u}_j = \bar{u}_{j0} + \Delta \bar{u}_j, \bar{u}_{i0} = \frac{\sum_{k=1}^{N_L} u_{ki0}}{N_L}, j = 1, \cdots, N_u \tag{3}$$

where N_L is the sampled data length, and

$$J_k = \sum_{i=1}^{N_y} \left[b_{ki} \times \bar{y}_i + a_{ki}^2 (\bar{y}_i - y_{dki})^2 \right] + \sum_{j=1}^{N_u} \left[b_{kj} \times \bar{u}_j + a_{kj}^2 (\bar{u}_j - u_{dkj})^2 \right], k = 1, 2, \cdots, N_L$$

2.2 Assessment of optimal yield without tuning control

Consider there is no change in the control tuning, i.e., all variability and constraints remain the same as the base case operation. Compared with base case operation, this scenario considers to move the actual average operating point of y_1 to its optimal operating condition as close as possible without tuning control. This is achieved simply by mean shift, and the distance between average operating point and the optimal operating point could be reduced significantly, meaning increased profit. The inequalities in (2) now become

$$\begin{aligned} Y_{Lki} + 2 \times Y_{stdi0} \le \bar{y}_i \le Y_{Hki} - 2 \times Y_{stdi0} \\ U_{Lkj} + 2 \times U_{qorj0} \le \bar{u}_j \le U_{Hkj} - 2 \times U_{qorj0} \\ i = 1, \cdots, N_y, \quad j = 1, \cdots, N_u \end{aligned}$$

2.3 Assessment of improved yield by reducing variability - relation between economic assessment and MVPA

Under this scenario, we consider the tuning of the control such that the variability of y_1 and/or y_2 can be reduced. Reduction of variability can obviously yield opportunity to push operating point closer to the optimum. The potential of variability reduction can be estimated through performance assessment. In general, however, the reduced variability of one variable can transfer to the increased variability of other variables such as constrained variable y_2. Since y_2 has no direct impact on profit, its variability is not of concern as far as it falls within its constraint. Therefore, the variability of quality variable y_1 may be reduced by transferring the variability to the constrained variable y_2. This type of interacting variance reduction is assessed by MVPA (Huang and Shah, 1999). Here, we introduce two variables, the ratios R_y and R_u, which are defined as the ratio of the targeted variance reduction of input/output variables and the existing variance of input/output variables. Likewise, the inequalities in (2) can now be modified as

$$\begin{aligned} Y_{Lki} + 2 \times Y_{stdi0}(1 + R_{yi}) \le \bar{y}_i \le Y_{Hki} - \\ 2 \times Y_{stdi0}(1 + R_{yi}), \quad i = 1, \cdots, N_y \\ U_{Lkj} + 2 \times U_{qorj0}(1 + R_{uj}) \le \bar{u}_j \le U_{Hkj} - \\ 2 \times U_{qorj0}(1 + R_{uj}), \quad j = 1, \cdots, N_u \end{aligned}$$

R_y and R_u are specified by users but their magnitudes should not be below -1. We call this procedure as benefit potential assessment based on variance reduction. Using MVPA, the potential variability reduction can be readily calculated. However, for MVPA with minimum variance control (MVC) as the benchmark, only variance reduction of output variables can be estimated, and thus it will be limited to the benefit assessment of output variables. For MVPA with LQG as the benchmark (Huang and Shah, 1999), both input and output variables can be considered. As a consequence, we can get an theoretical absolute optimal benefit potential with MVC or LQG as the benchmark.

2.4 Assessment of improved yield by relaxing constraints

If the constraints can be relaxed for all or some variables, the operating condition may be moved further in the direction of increased profit. This move is mainly due to changes of constraints in constraint variables. As a consequence, this will create a new opportunity to transfer more variability from y_1 to y_2 and thus the variability of y_1 could be reduced to increase profit. In this case, we introduce two ratios, S_y and S_u, which are defined as the ratio of targeted (often increased) constraint and the existing constraint. They can be specified by the user, and we call this procedure as benefit potential assessment based on relaxation of constraints. The inequalities in (2) can be modified to accommodate this assessment as

$$\begin{aligned} Y_{Lki} - S_{yi} \times Y_{holki} + 2 \times Y_{stdi0} \le \bar{y}_i \le Y_{Hki} + \\ S_{yi} \times Y_{holki} - 2 \times Y_{stdi0}, \quad i = 1, \cdots, N_y \end{aligned}$$

$$U_{Lkj} - S_{uj} \times U_{holkj} + 2 \times U_{qorj0} \le \bar{u}_j \le U_{Hkj} + S_{uj} \times U_{holkj} - 2 \times U_{qorj0}, \quad j = 1, \cdots, N_u$$

2.5 Assessment of improved yield by reducing variability and relaxing constraint simultaneously

Since benefit potential could be achieved by either reducing variability or relaxing constraint, they can also be considered simultaneously, hoping to achieve higher yield. In this case, the inequalities in (2.3) and (2.4) can be combined as

$$Y_{Lki} - S_{yi} \times Y_{holki} + 2 \times Y_{stdi0}(1 + R_{yi}) \le \bar{y}_i \le Y_{Hki} + S_{yi} \times Y_{holki} - 2 \times Y_{stdi0}(1 + R_{yi}),$$
$$U_{Lkj} - S_{uj} \times U_{holkj} + 2 \times U_{qorj0}(1 + R_{uj}) \le \bar{u}_j \le U_{Hkj} + S_{uj} \times U_{holkj} - 2 \times U_{qorj0}(1 + R_{uj}),$$
$$i = 1, \cdots, N_y, \quad j = 1, \cdots, N_u$$

2.6 Variability tuning for desired yield

In the benefit potential assessment, the ratios R_y and R_u are specified *a priori* by users. If, instead, we use them as decision variables, the optimal R_y and R_u can be found from optimization accordingly. For notation purpose, we use r_y and r_u in place of R_y and R_u , respectively. A targeted ratio, R_V, is defined as the ratio between the targeted benefit and ideal benefit, where R_V should be within 0 and 1. Given R_V, the ratios r_y and r_u may be calculated but the solutions are not unique. However, to minimize tuning effort, we would want r_y and r_u as smaller as possible. The minimum r_y and r_u may be found through the optimization of the following problem:

$$\min_{\bar{y}_i, \bar{u}_j, r_{yi}, r_{uj}, r} -r \tag{4}$$

subject to

$$Y_{Lki} + 2 \times Y_{stdi0}(1 + r_{yi}) \le \bar{y}_i \le Y_{Hki} - 2 \times Y_{stdi0}(1 + r_{yi}), \quad i = 1, \cdots, N_y$$
$$U_{Lkj} + 2 \times U_{qorj0}(1 + r_{uj}) \le \bar{u}_j \le U_{Hkj} - 2 \times U_{qorj0}(1 + r_{uj}), \quad j = 1, \cdots, N_u$$
$$-1 < r_{yi}, -1 < r_{uj}, r_{yi} > r, r_{uj} > r$$
$$J_k = R_V \times J_{k0}, \quad Equalities \quad in \quad (3)$$

2.7 Constraint tuning for desired yield

If the variability could not be reduced further, we may achieve desired benefit potential by tuning the constraints. Similarly, a desired ratio, R_C, is defined as the targeted benefit against that of ideal yield. We would also want the change of the constraints, s_y and s_u (a counterpart of S_y and S_u defined before), to be as small as possible. The minimum s_y and s_u can be solved through

$$\min_{\bar{y}_i, \bar{u}_j, s_{yi}, s_{uj}, s} s \tag{5}$$

subject to

$$Y_{Lki} - s_{yi} \times Y_{holki} + 2 \times Y_{stdi0} \le \bar{y}_i \le Y_{Hki} + s_{yi} \times Y_{holki} - 2 \times Y_{stdi0}, \quad i = 1, \cdots, N_y$$
$$U_{Lkj} - s_{uj} \times U_{holkj} + 2 \times U_{qorj0} \le \bar{u}_j \le U_{Hkj} + s_{uj} \times U_{holkj} - 2 \times U_{qorj0}, \quad j = 1, \cdots, N_u$$
$$s_{yi} < s, s_{uj} < s$$
$$J_k = R_C \times J_{k0}, \quad Equalities \quad in \quad (3)$$

3. SYNTHESIS APPROACH

Economic evaluation of MPC applications includes economic performance assessment, sensitivity analysis and tuning guidelines, which are considered in this section.

3.1 Economic performance assessment

In the assessment of ideal yield, the optimal operating condition is expected to be pushed closest to the constraint for the quality variable without back off. However, in the assessment of optimal yield without tuning control, the benefit potential is obtained by only shifting the mean values of quality variables in the direction of increasing benefit potential without reducing variability and hence the back off depends on the present level of disturbances. By comparing these two scenarios, an economic performance index without tuning can be defined as

$$\eta_E = \frac{\Delta J_E}{\Delta J_I}$$

where ΔJ_E is the optimal yield without tuning control and ΔJ_I is the ideal yield. η_E is the benefit potential ratio that can be realized by just pushing the mean values without reducing the variability, while $1 - \eta_E$ is the benefit potential ratio that is due to no variability. It is noted that $0 \le \eta_E \le 1$. If $\eta_E = 0$, no benefit could be obtained without reducing the variability. If $\eta_E = 1$, there is no disturbance and hence no back off is required under the current control strategy. By introducing MVC or LQG benchmark, MVPA gives a theoretical absolute variance lower bound; thus a theoretical economic performance index can be calculated as

$$\eta_T = \frac{\Delta J_T}{\Delta J_I}$$

where ΔJ_T is the theoretical benefit potential upper bound that could be achieved by MVC or LQG plus steady state optimization. ΔJ_T is in part due to the mean value shift and in part due

to the variability reduction. It can be seen that $0 \leq \eta_T \leq 1$. By comparing with η_E, the following inequality holds,

$$0 \leq \eta_E \leq \eta_T \leq 1$$

Therefore, if no variability can be reduced, η_E (or ΔJ_E) can be adopted in the economic performance assessment. $\eta_E = 0$ (or $\Delta J_E = 0$) shows that no benefit potential can be further obtained without tuning the controller. On the other hand, if the benchmark of MVPA is available, η_T (or ΔJ_T) can be utilized instead, which gives an absolute upper bound on the economic benefit potential that could be realized theoretically. The positive value of η_T (or ΔJ_T) does not mean that this benefit potential is practically achievable since MVC itself is usually not applied in practice, but the benefit potential with positive value of η_E (or ΔJ_E) is really achievable.

3.2 Sensitivity analysis

Sensitivity analysis is applied to investigate the impact of variability change or constraint change on the benefit potential. The result shows the importance of different variables based on their contributions to the benefit potential. The size and direction of the change of variability or constraint can be specified by the user. Some constrained variables may have no impact on the benefit potential, and some variables may have great impact on the benefit potential that should be paid special attention on during the operation. It is thus worthwhile to reduce the variability or relax the constraint of these variables such that more benefit potential can be achieved.

3.3 Tuning guideline

As analyzed in the economic performance, ΔJ_I or ΔJ_T can be regarded as an upper bound on the benefit potential against which other scenarios can be compared. The desired potential benefit can never be greater than this upper bound by tuning variability only. Once the desired variability ratio R_V or the desired constraint ratio R_C is specified, the corresponding optimization problem will result in the tuning guideline for variability reduction or constraint relaxation. The tuning guideline tells directly which variables should be tuned and by how much in order to achieve the desired benefit potential.

4. LMI FORMULATION

The quadratic objective function in (1) alone can be transformed into

$$\underset{\bar{y}_i, \bar{u}_j}{min}\, \gamma \tag{6}$$

subject to

$$\sum_{k=1}^{N_L} \left\{ \sum_{i=1}^{N_y} \left[b_{ki} \times \bar{y}_i + a_{ki}^2 (\bar{y}_i - y_{dki})^2 \right] + \sum_{j=1}^{N_u} \left[b_{kj} \times \bar{u}_j + a_{kj}^2 (\bar{u}_j - u_{dkj})^2 \right] \right\} < \gamma \tag{7}$$

According to *Schur* complement, it can be readily formulated as follows and then solved via LMI technique,

$$\begin{pmatrix} \gamma - Y_{lin} - U_{lin} & X_{lin}^T \\ X_{lin} & I \end{pmatrix} \succ 0 \tag{8}$$

where

$$X_{lin} = \left[\left(\sqrt{\sum_{k=1}^{N_L} a_{k1}^2} \right) \bar{y}_1 \ \cdots \ \left(\sqrt{\sum_{k=1}^{N_L} a_{kN_y}^2} \right) \bar{y}_{N_y} \left(\sqrt{\sum_{k=1}^{N_L} a_{k1}^2} \right) \bar{u}_1 \ \cdots \ \left(\sqrt{\sum_{k=1}^{N_L} a_{kN_u}^2} \right) \bar{u}_{N_u} \right]^T$$

$$Y_{lin} = \sum_{i=1}^{N_y} \left\{ \sum_{k=1}^{N_L} \left(b_{ki} - 2a_{ki}^2 y_{dki} \right) \bar{y}_i \right\} + \sum_{i=1}^{N_y} \left\{ \sum_{k=1}^{N_L} \left(a_{ki}^2 y_{dki}^2 \right) \right\}$$

$$U_{lin} = \sum_{j=1}^{N_u} \left\{ \sum_{k=1}^{N_L} \left(b_{kj} - 2a_{kj}^2 y_{dkj} \right) \bar{u}_j \right\} + \sum_{j=1}^{N_u} \left\{ \sum_{k=1}^{N_L} \left(a_{kj}^2 y_{dkj}^2 \right) \right\}$$

5. CASE STUDY

5.1 Process and controller description

An MPC is applied in the reactor section of the gas oil hydrotreating unit (GOHTU) to maintain gas oil nitrogen/sulphur specifications, maximize catalyst run length, minimize hydrogen/fuel gas consumption and improve operation safety. It has total 41 output variables (y), 15 input variables (u) and 5 disturbance variables (d). The real-time data collected for this analysis include all y, u, d and associated parameters, such as high/low limits, linear coefficients, quadratic coefficients and targeted steady state values. The data collection lasted for approximately 26.5 hours with sampling time 15 second and total 6350 data points were collected.

5.2 Economic performance assessment

The result shows that $\Delta J_I = 196.7428$ and $\Delta J_E = 13.0889$. The performance indices from

MVPA with MVC as the benchmark are given in Figure 2 and accordingly $\Delta J_T = 186.4489$. The economic performance index is calculated as $\eta_E = 6.7\%$, one can thus conclude that the steady-state operation of this MPC has achieved good economic performance given the existing variability within the set of data studied, and the potential for improved benefit is rather small.

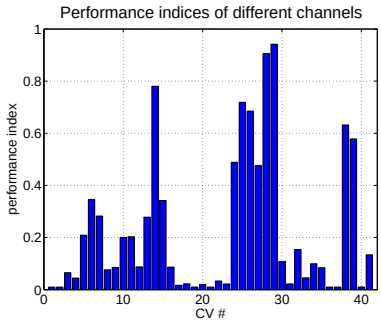


Figure 2. MVPA performance assessment result

5.3 Sensitivity analysis

In the variability sensitivity analysis, the variability of chosen variable was reduced 1% to observe its impact on the benefit potential. The result shows that output variables $(y_1, y_{10}, y_{11}, y_{22}, y_{32}, y_{33}, y_{41})$ and input variables (u_9, u_{10}, y_{15}) have effects on the benefit potential and other variables have no effect at all. The impacts of output variable y_1 and input variable u_{15} are much greater than other sensitive variables, which means output variable y_1 and input variable u_{15} should be the first choice to reduce their variability if it is possible. Similarly, the constraint sensitivity analysis shows that output variables $(y_1, y_{10}, y_{11}, y_{18}, y_{19}, y_{22}, y_{32}, y_{33}, y_{38}, y_{39}, y_{41})$ and input variables $(u_1, u_9, u_{10}, u_{15})$ have effects on the benefit potential while the other variables not at all. The sensitivity analysis shows the importance of different variables in the sense of economic performance. For example, the variability sensitivity analysis for output variables is shown in Figure 3.

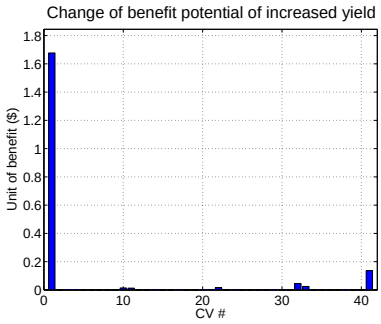


Figure 3. Output variability sensitivity analysis

5.4 Optimal tuning guidelines

The desired variability ratio and constraint ratio are both specified as $R_V = R_C = 80\%$ and the desired benefit potential is equal to 157.3942. This target benefit potential may be achieved by either reducing the variability of output variables $(y_1, y_{10}, y_{11}, y_{22}, y_{32}, y_{33}, y_{38}, y_{39}, y_{41})$ and input variables $(u_1, u_2, u_9, u_{10}, u_{15})$ as suggested by variability tuning guideline, or, relaxing the constraint ranges of output variables $(y_1, y_{10}, y_{11}, y_{18}, y_{19}, y_{22}, y_{32}, y_{33}, y_{38}, y_{39}, y_{41})$ and input variables $(u_1, u_9, u_{10}, u_{15})$ as suggested by constraint tuning guideline. The tuning guideline is shown as percentages. The variability tuning guideline for output variables is shown in Figure 4.

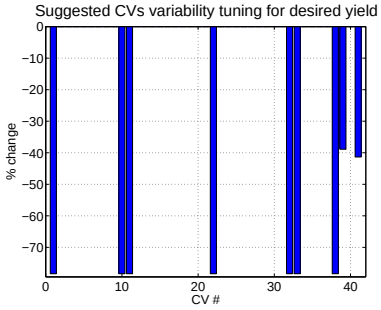


Figure 4. Output variability tuning guideline

6. CONCLUSION

A synthesized approach is proposed for MPC economic performance assessment based on its steady state optimization and variance/constraint tuning. It shows that further benefit potential could be achieved by optimizing its steady state, reducing variability or increasing constraint ranges. The case study demonstrates that it is a powerful tool for the control engineers in the economic performance assessment for existing MPC applications. This tool has been integrated together with MVPA which gives the variability potential improvement. This variability potential could be converted to its economic benefit potential. Synthesis of these two tools will give not only the MPC performance on variance reduction but also its economic benefit potential. They have been integrated into a plant oriented solution for APC performance monitoring.

REFERENCES

Huang, B. and S.L. Shah (1999). *Performance assessment of control loops: theory and applications*. Spring-Verlag, London.

Muske, K.R. (2003). Estimating the economic benefit from improved process control. *Ind. Eng. Chem. Res.* **42**, 4535–4544.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DIAGNOSIS OF FAULTS WITH VARYING INTENSITIES USING POSSIBILISTIC CLUSTERING AND FAULT LINES[1]

Detroja K. P.[1], Gudi R. D.[2*], Patwardhan S. C.[2]

[1] Interdisciplinary Programme on Systems and Control Engineering,
[2] Department of Chemical Engineering
Indian Institute of Technology Bombay, Powai, Mumbai – 400076, India.
* Email: ravigudi@che.iitb.ac.in

Abstract: In this paper, a new approach for fault detection and isolation that is based on the possibilistic clustering algorithm is proposed. Fault detection and isolation (FDI) is shown here to be a pattern classification problem, which can be solved using clustering and classification techniques. The possibilistic clustering approach was proposed to address some of the short comings of the fuzzy c-means (FCM) algorithm. The probabilistic constraint imposed on the membership value in the FCM algorithm is relaxed in the possibilistic clustering algorithm. Because of this relaxation, the possibilistic approach is shown in this paper to give more consistent results in the context of the FDI tasks. The proposed approach addresses the issue of correctly isolating a fault that may occur with varying intensities. The concept of fault lines is introduced, which in conjunction with possibilistic clustering has been effectively used for FDI. Fault signatures that change as a function of the fault intensities are represented as fault lines, which are shown to be useful to classify faults that can manifest with different intensities. The proposed approach has been validated here through simulations involving a co-polymerization reactor simulation.

Keywords: Possibilistic clustering, Fuzzy c-means clustering, Gustafson-Kessel algorithm, Fault Detection (FDD)

1. INTRODUCTION

Online process monitoring for fault detection and diagnosis (FDD) is very important for ensuring plant safety and product quality. The area of FDD has been very active in recent years. Both, model based and process history based methods have been proposed with a fair amount of success.

In a typical process plant, hundreds of variables are measured every few seconds. These measurements bring in useful signatures about the status of the plant. While model-based methods can be used to detect and isolate signals that indicate abnormal operation, such quantitative cause-effect models may be difficult to develop from the first principles. Methods based on historical data attempt to extract maximum information from the archived data and require minimum physical information of the plant. Multivariate statistical monitoring tools such as PCA (Kresta *et al.*, 1991) are developed to extract information from historical process data so as to carry out the task of FDD easily and more efficiently.

For online process monitoring, it is important to not only be able to say whether the plant operation is aberrant but also to be able to isolate the fault. This is typically done through the use of contribution plots, which assesses the relative contribution of each variable to a suitably formulated error criterion.

While multivariate statistical tools can compress data and perform monitoring in the lower dimension space, they are inherently representation-oriented rather than discrimination oriented. They seek to explain or represent the variance in a data set rather than discriminate between dissimilar subsets in the data (Chiang *et al.*, 2000). Thus, there could be overlaps between the clusters representing fault regions and normal operating regions, leading to higher misclassification rates. Overlaps could also still exist, although to a smaller extent, when tools such as multiple discriminant analysis are used. The latter are discrimination oriented and are generally known to yield directions that enhance discrimination between regions. Thus, it is necessary to look at methods that accommodate such overlaps and analyze these regions so as to provide useful indicators to detect and diagnose faults.

[1] A fuller version of this paper has been communicated for journal review

An alternate approach to the task of fault detection and diagnosis is to mine the archived data and examine patterns in the process variables that indicate the occurrence of a fault. Johannesmayer *et al.* (2002) and Singhal and Seborg (2002) proposed pattern matching methods based on similarity factors which attempted to match patterns in the archived plant database. Typically, parametric faults, sensor and actuator biases and disturbances generate different patterns in the process variables. These patterns or signatures can be classified into different clusters that represent normal or aberrant operation. Subsequently, when deployed online, the plant operation can be classified in terms of the belonging or membership of the new data to the known clusters, based on the similarity of the patterns that the data brings. Thus, the problem is often related towards being able to classify plant operation as normal or belonging to one or more of the faults from the available (measurements and manipulated inputs) plant signatures. This can be effectively done by various pattern recognition and clustering techniques.

Clustering techniques, such as fuzzy c-means clustering (FCM) (Duda *et al.*, 2003) and its variants (Fuzzy Gustafson-Kessel (FGK) algorithm for clustering (Gustafson and Kessel, 1979)) have been very popular in image analysis and pattern classification. Since, fault detection and isolation (FDI) is also a pattern classification problem, these clustering techniques can be effectively used for this task. Attempts have also been made to use different clustering algorithms for the task of FDI. The k-means clustering, which is a hard clustering technique, has been used along with principal components analysis (PCA) and Fisher discriminant analysis (FDA) for the task of FDD in a three step procedure proposed by Peter He et al. (2005). Teppola and Minkkinen (1999) have used adaptive FCM for process monitoring of a waste water plant. They also used possibilistic clustering algorithm for fault detection. Choi *et al.* (2003) used credibilistic clustering algorithm, proposed by Chintalpudi and Kam (1998), based approach for process monitoring.

One of the limitations of the proposed clustering based algorithms is that signatures resulting from the same fault but with differing intensities would confound them and may lead to spurious fault isolation. Another important aspect, relevant for the task of FDD, is related to the issue of identifying and classifying novel faults. It is important to recognize that archived data does not necessarily encompass all possible fault scenarios. Therefore, the FDD algorithm also needs to have learning ability, i.e. when deployed online it should be able to identify the occurrence of new faults and establish relevant signatures or patterns that are representative of the novel fault.

In this paper, we propose to overcome the above difficulties, by using the possibilistic clustering algorithm (Krishnapuram and Keller1, 1993) in conjunction with 'fault lines'. Possibilistic clustering algorithm is a powerful technique that is similar to probabilistic clustering methods but differs in the nature of the constraint(s) that bind the objective function. Possibilistic clustering algorithm has a number of advantages when compared with conventional FCM algorithm (Krishnapuram and Keller, 1993). In possibilistic clustering, the number of clusters need not be specified accurately and they can be derived during the classification step. In FCM, however, approximate number of clusters/ classes in the data is determined by various cluster validity measures proposed in the literature (Bezdek, 1981). The formulation of the possibilistic clustering algorithm relaxes some constraints on the nature of the membership functions; here we also show that this relaxation gives (i) more consistency in the classification task and (ii) enables the detection of novel (or not-seen) classes. It is also shown here that that the possibilistic clustering algorithm is relatively insensitive to noise and outliers. These features make the possibilistic clustering algorithm more suited for the FDD task. Here, we also introduce the concept of fault lines for handling faults with varying intensities and detecting novel faults. Fault lines are characterized by cluster centre of the normal operation and cluster centre one of the fault clusters.

We demonstrate the suitability of the proposed approach for the FDD task, through simulation case study involving CSTR simulation for solution co-polymerization of methyl-methacrylate (MMA) and vinyl acetate (VA) (Congalidis *et al.*, 1986).

The paper is organized as follows. In the next section we present a brief review of clustering algorithms, namely the FCM/ FGK and PGK algorithm. In the next section the proposed FDI scheme is described in detail. Finally we present a case study to validate the proposed approach.

2. REVIEW OF CLUSTERING ALGORITHMS

The aim of any clustering analysis is to derive a partition of a set of N data points or objects based on some similarity metric, so that the data points/ objects that get clustered into the same group are similar to one another. Clustering algorithms can be broadly classified into hierarchical and non-hierarchical clustering techniques. Here, due to brevity, we briefly review fuzzy c-means clustering and possibilistic clustering algorithms. In fuzzy c-Means (FCM) algorithms, each data point can be a member of more than one cluster, i.e. the membership of a point can take any value between 0 and 1. The following section briefly describes the FCM clustering algorithm.

2.1. Fuzzy c-means algorithm

In FCM clustering techniques, a data/ feature point can be a member of more than one cluster with different degrees of membership. If the membership value of j^{th} data point to i^{th} cluster is μ_{ij}, we have the condition that $\mu_{ij} \in [0,1]$. For the data set $\{X \mid x_1, x_2, \ldots, x_N \in X\}$, consisting of c clusters, the constraints imposed on the membership value μ_{ij} are given below.

$$\sum_{i=1}^{c} \mu_{ij} = 1 \qquad (1)$$

where, c is the number of clusters and N is total number of data points in the data set. This is also known as the probabilistic constraint. This constraint requires that total membership of a data/ feature point to all the clusters must be unity. Another important constraint on the clusters is that none of the clusters can be empty, and this constraint can be mathematically represented as shown in Equation (2)

$$0 < \sum_{j=1}^{N} \mu_{ij} < N \qquad (2)$$

The FCM algorithm then minimizes the following objective function subjected to constraints in Equation (1) & (2).

$$J = \sum_{i=1}^{c} \sum_{j=1}^{N} \left(\mu_{ij}\right)^m d_{ij}^2 \qquad (3)$$

where, m is the fuzziness exponent, d_{ij} is the distance between a data point x_j and cluster centre v_i.

The algorithm for FCM starts with some initial guess for either the fuzzy partitioning matrix or the cluster centers and iterates till convergence. The convergence of the FCM algorithm is guaranteed (Bezdek, 1981), but it may converge to a local minima.

The FCM membership of a data point to a cluster depends not only on the distance of that point to the cluster centroid, but also on the distance of that point to other cluster centroids. As will be discussed in detail later, this can cause the algorithm to assign very different memberships to points which are similar as measured by their distances from a cluster center, because their distances from other cluster centers can be different. This problem primarily arises due to the probabilistic constraint described by Equation (1).

The FCM algorithm can be modified in several ways depending on the distance measure chosen. The most commonly used distance measures are Euclidean and Mahalanobis distance. Using these distance measures is equivalent to assuming that the data is oriented in each cluster identically. This may not necessarily be true, for example, the orientation of the data could be spherical in the first cluster and elliptical in the second. From an FDD viewpoint, this could mean higher miss-classification rates and hence poorer diagnosis. To overcome this difficulty, FGK algorithm uses an adaptive distance norm which adapts the similarity measure (norm) according to the shape of the cluster. The algorithm is also quite sensitive to the user specified choice of c, the number of clusters in the data set. The results obtained from FCM/ FGK algorithm largely depends on this number of clusters. If c is not specified correctly, the FCM/ FGK algorithm can arbitrarily split or merge the classes in the data to give exactly c clusters. Different cluster validity measures have therefore been proposed to overcome this difficulty. However, Bezdek (1981) pointed out that the concept of cluster validity is open to interpretation and can be formulated in different ways.

2.2. Possibilistic clustering algorithm

As described earlier, the probabilistic constraint (Equation (1)) imposed on the membership assigned by the FCM/ FGK algorithm brings in problems of classification. These can be broadly enumerated as (i) the points equidistant from the centroid may get very different memberships depending upon the placement of the other clusters, although they are similar as measured by the distance metric, and (ii) the points which are equidistant from all the centroids get the same membership irrespective of their relative positions. To overcome these drawbacks, Krishnapuram and Keller (1993) proposed a new clustering technique called possibilistic clustering, in which the probabilistic constraint on the membership is relaxed. We discuss the possibilistic clustering algorithm in the next section.

In possibilistic clustering, the probabilistic constraint on the objective function in equation (3) is relaxed in possibilistic clustering so as to get membership values, which represent the 'degree of typicality' to a cluster. Simply relaxing the probabilistic constraint produces a trivial solution, i.e. the objective function is minimized by assigning all membership values to 0. Therefore the objective function of Equation (3)(3) is modified as

$$J = \sum_{i=1}^{c} \sum_{j=1}^{N} \left(\mu_{ij}\right)^m d_{ij}^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{N} \left(1 - \mu_{ij}\right)^m \qquad (4)$$

The first term in the equation minimizes the distances of data points from the cluster centers, where as the second term forces the membership values to be as large as possible. In this equation as well, the value of m determines the fuzziness of the final possibilistic partition.

The value of parameter η_i determines the distance at which the membership value of a point in a cluster becomes 0.5. Thus, it needs to be chosen depending on the desired bandwidth of the possibility distribution for each cluster. In practice however, the following definition works well (Krishnapuram and Keller, 1993):

$$\eta_i = \frac{\sum_{j=1}^{N} \left(\mu_{ij}\right)^m d_{ij}^2}{\sum_{j=1}^{N} \left(\mu_{ij}\right)^m} \qquad (5)$$

Updating of the membership values depends on the distance measure chosen. Different distance measures lead to different algorithms. If the distance measure chosen is either Euclidean or Mahalanobis distance, the algorithm gives possibilistic c-means (PCM) membership values. However, if the distance measure is chosen based on scaled Mahalanobis distance and fuzzy covariance matrix, the algorithm gives possibilistic Gustafsson-Kessel (PGK) membership values.

The solution to the objective function in equation (4) leads to the values of memberships as,

$$\mu_{ij} = \frac{1}{1+\left(\frac{d_{ij}^2}{\eta_i}\right)^{1/(m-1)}} \quad (6)$$

The iterative part of the algorithm for possibilistic clustering is very much similar to that of the FCM algorithm, except for the additional parameter η_i which should be estimated from the initial partitioning matrix. However, η_i need not be calculated at every iteration.

Since the parameter η_i is independent of the relative location of the clusters, the membership value μ_{ij} depends only on the distance of a point from the cluster centre (centroid). Hence, unlike in the probabilistic case, the membership of a point in a cluster is determined solely by how far a point is from the centroid and is not coupled with its location with respect to other clusters.

The advantages of PCM/ PGK lie in finding meaningful clusters as defined by dense regions. This happens because each cluster is independent of the other cluster in PCM/ PGK algorithm. Hence, the objective function corresponding to cluster *i* can be formulated as in Equation (7) and the overall objective function is collection of c such objective functions.

$$J_i = \sum_{j=1}^{N} \left(\mu_{ij}\right)^m d_{ij}^2 + \eta_i \sum_{j=1}^{N} \left(1-\mu_{ij}\right)^m \quad (7)$$

It has been shown (Krishnapuram and Keller, 1993) that for a given value of η_i, each of the c sub-objective functions is minimized by choosing the centroid location such that the sum of the memberships is maximized. This makes each cluster centroid to converge to a dense region. Thus, even if the true value of the number of clusters is unknown, the outcome of the algorithm will give c 'good' clusters, i.e. dense regions. Thus, PCM/ PGK have self validating capability which can be very useful when c is not known apriori. When the number of clusters is more than the actual number of clusters in the data set, PCM/ PGK give approximately coinciding clusters, indicating that the actual number of clusters is lesser than specified. This could be interpreted accordingly and the clusters could be collapsed into a single cluster for further analysis.

3. PROPOSED SCHEME FOR FDI

Clustering based approaches are aimed at partitioning the historical data into a number of clusters, e.g. normal operation and different fault operations. Depending on the membership value of the data point to different clusters, the plant operation is declared either normal or otherwise. The shift from normal operating cluster to any fault mode cluster is not instantaneous and the transient response depends on the dynamics of the system. The FCM algorithm would assign different membership values as governed by the probabilistic nature (Equation (1)) to the points even during these transients. This may be useful for example, when the dynamics are to be represented (as shown in Venkat and Gudi, 2002) in a composite modeling methodology, where the memberships essentially weigh the model predictions in each cluster. However, for the FDD task, this may yield erroneous results and misleading interpretations. Possibilistic clustering algorithms (PCM/ PGK) appear to be more suited because these points corresponding to the transition region are not governed by the probabilistic constraint and are assigned low memberships to all the clusters. In the following section, we describe the proposed approach for fault detection and isolation.

3.1. Data collection and clustering

The ability of a statistical approach to detect and isolate a fault depends on the availability of rich historical data, containing data corresponding to normal and fault modes of operation. Ideally, the data set used for training the clustering based technique should contain data that represents all possible fault scenarios. In practice however, it may not be possible to have such a data set and the algorithm should have some self-learning abilities.

In general, the historical data consists of measurements of various controlled and manipulated variables at each sampling instant. The clustering approach could either first construct a feature vector from this data or directly work with the measurements. In the former case, the classification is carried out in the space defined by these feature vectors. For example, for incipient fault detection, it may be mandatory to look at the dynamic patterns represented by the feature vector that is constructed from the measurements from the current and past instants. Meel *et al.* (2004) used such an approach to rapidly reject unmeasured disturbances using these pattern recognition techniques, by classifying an appropriately constructed feature vector that was based on apriori knowledge of the dynamics. This approach necessarily requires apriori information of the classification space which is usually difficult to obtain. Alternately, one could directly classify in the space spanned by the measurements (i.e. without constructing feature vectors). This latter approach is taken in this paper.

The clustering algorithm can then be applied on the data. Specifying the exact number of clusters present in the data set is not mandatory for possibilistic clustering approach, as the possibilistic clustering algorithm attempts to search for c good clusters, i.e. dense regions. In the case when the number of clusters specified is more than the actual number of clusters present in the data set, the algorithm will give overlapping clusters indicating that the value of c is over specified. This greatly simplifies the task of clustering of historical data in which the number of clusters present is not known apriori. The outcome of the clustering algorithms would thus yield cluster centroids and fuzzy covariance matrices for each cluster (in case when the GK algorithm is used).

3.2. Generating Fault lines

We next discuss the effect of different fault magnitudes and intensities. As mentioned earlier, different fault intensities of the same fault (for example, sensor bias) can manifest in different data vector signatures / paths and would end up into new cluster. In such cases, a methodology, which is still able to classify the fault as a sensor bias (rather than

as a novel fault), independent of its magnitude is desirable. Towards this end, we propose the concept of fault lines that characterize the movement of the cluster as a function of the intensity of the fault. When the fault intensity increases, the dynamics and the controller effects result in parallel paths that shift to the fault cluster and eventually end up into new clusters. A fault line could therefore be constructed through the centers of the clusters, beginning from the normal cluster to the fault clusters, and would characterize the behavior of the clusters as a function of increasing intensities of that fault. Assuming that during the training step, data corresponding to a particular fault is available; fault lines can then be constructed to characterize the particular fault.

3.3. Online monitoring and fault detection

For online process monitoring and fault detection, the membership value of the data vector, constructed from the measurements at each instant, to each cluster is calculated from Equation (6) in case of possibilistic clustering approach. High membership values to the normal operating cluster imply that the plant is operating normally. When an abnormal event occurs, these reflect in the signatures of the measured variables, which result in changing memberships of the data vector to the known clusters. An analysis of these memberships would help in the interpretation and classification of the fault scenario. The PCM/ PGK membership value for the normal cluster will assume smaller values close to zero, indicating that a fault may have occurred.

3.4. Fault confirmation and isolation

It is important to recognize that the changing memberships due to the occurrence of a fault are influenced by the inherent system dynamics. The membership profiles can also change due to the occurrence of short-term transients (introduced for example by a control loop), measurement noise or outliers. Thus, it is important to confirm the occurrence of a fault after it is detected. For fault confirmation and isolation, we therefore propose to use a window of M sampling instants over which the membership profiles are analyzed. If the memberships to the normal cluster consistently stay below the user specified threshold for a period exceeding M sampling instants, the occurrence of a fault is confirmed. Similarly, if the memberships to a particular fault cluster assume significant values (above a specified threshold) the fault may be isolated as well.

It should be noted that this will happen only if the fault that has occurred is of the same intensity as in the historical data. In case, if the fault has occurred with a different intensity, the membership value to all known fault clusters will remain close to zero, indicating that a new cluster is formed. As pointed out earlier, the objective function for possibilistic clustering can be seen as a set of c objective functions and the membership value in possibilistic clustering is not influenced by how other clusters are placed. Therefore, it is sufficient to find only the new cluster centre from this newly collected data. Once the new cluster centre is computed, its proximity with the fault lines can be examined. If the new cluster centre is close to one of the fault lines, which are generated from the historical data, the fault may be isolated as the fault associated with that fault line.

The specification of the parameter M has to be carefully done to achieve a compromise between false alarms and sensitivity to the fault occurrence. In general, the choice of M can be made from the closed loop process dynamics or plant operator's experience.

3.5. Novel fault detection

This proposed approach also therefore provides a method to flag novel faults. Low membership value to normal operation cluster for M sampling instants confirms the occurrence of a fault. However, if membership to all known fault clusters and proximity to all known fault lines suggest that the fault that has occurred is indeed novel. Thus, proposed approach enables the classification of the plant operation either as (i) normal operation, (ii) belonging to the known fault scenarios, or (iii) novel faults. The approaches based on other clustering approaches can not provide such crisp division of the plant operation. The new cluster information can be merged with the existing knowledge base and used for future fault diagnosis. Thus an added advantage of the proposed scheme is that it reduces the emphasis on exhaustive historical data. In principle, one can start with just the normal operating data and continue building the monitoring scheme as the new fault events occur.

The role of the fuzziness exponent m in the FDD task also merits some important comments. As mentioned earlier, a higher value of m blurs the distinction between the clusters and makes the cluster boundaries to fade. While monitoring a transition from a normal operating region to a fault mode, with higher values of m, the algorithm would confirm the fault early. However, this high value of m would also increase the incidence of false alarms, which would be indicated when the memberships to the normal cluster decrease. Thus, as in the case of window length M, the value of the fuzziness exponent m should also be chosen as a careful compromise between the requirements of early fault detection and confirmation.

Remark: *The above monitoring strategy is restricted to steady state behavior wherein the points belonging to different operating regions cluster together. For the time varying case, for example in a batch process, the method needs further modifications using manifolds that characterize time varying operation. This aspect is currently under investigation.*

4. CASE STUDY

To validate the proposed approach a simulation case study that is based on co-polymerization reactor is presented here. A 4×5 transfer function matrix model (Congalidis *et al.*, 1986) for a CSTR solution co-polymerization of methyl-methacrylate (MMA) and vinyl acetate (VA) was simulated under closed loop conditions. Based on the RGA analysis, the pairings of controllers were chosen and U_1 was kept constant, effectively resulting in a 4×4 system.

To begin with, the historical data set containing data for (i) normal operation and (ii) for the fault case when sensor Y_1 has developed a bias, was collected. This resulted in two clusters F_0 and F_1 and a fault line corresponding to fault F_1 in the knowledge base. When implemented online, the proposed possibilistic clustering algorithm could easily detect and isolate fault F_1.

In the next step, a positive sensor bias in sensor Y_4 was introduced. As this fault is not part of the archived data that was used for training, it was detected as a novel fault after M (20) samples. The new cluster centre for the newly obtained data was computed and it was found that the new cluster centre (say F_2) was not on the fault line corresponding to fault F_1. Distance of F_2 from F_1 fault line was found to be 2.74 units. Hence, the fault was isolated as novel fault and knowledge base was updated with the new cluster centre. The monitoring scheme now had three fault clusters F_0, F_1 and F_2 along with fault lines for F_1 and F_2. The monitoring scheme could now easily detect and isolate fault F_2 (Figure 1).

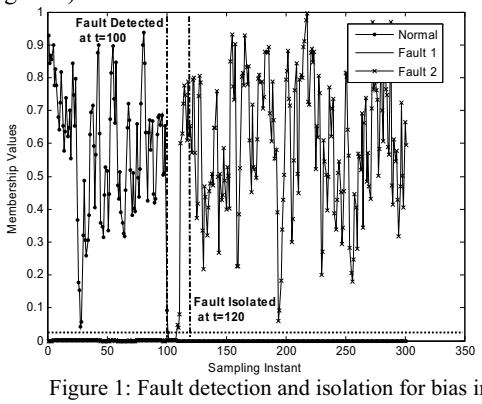


Figure 1: Fault detection and isolation for bias in sensor Y_4

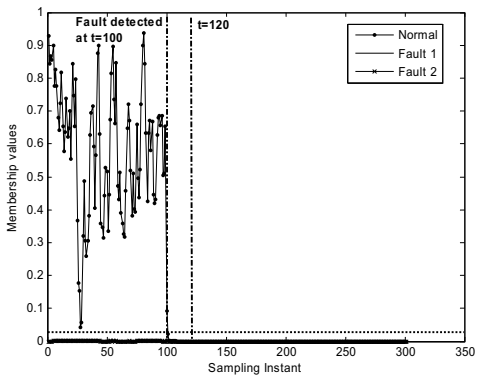


Figure 2: Fault with different intensity than the training set is not be classified as one of the known faults via membership values

As discussed earlier, the same fault can occur with varying intensities during the plant operation. It is therefore important to ascertain that they are not isolated as different faults. To demonstrate the same, a negative bias in sensor Y_4 was introduced. With the proposed approach, it was promptly detected. However, since the intensity of the fault was different than the training set, all the clusters' membership value remained close to zero (Figure 2). Here, fault isolation was performed using the fault lines. The distance of fault lines for F_1 and F_2 were found to be 2.70 and 0.05, indicating that the fault detected is indeed the same fault as F_2 with different intensity.

5. CONCLUSION

A fuzzy clustering and classification based fault detection and diagnosis algorithm was proposed and validated through a simulation case study. The proposed approach is based on the possibilistic clustering methodology of Krishnapuram and Keller (1993) and was found to be vastly superior to other classification methodologies such as the fuzzy c-means and fuzzy credibilistic algorithm. The concept of fault lines was shown to address the difficulty of isolating the same fault with varying intensities. The fault lines were shown to distinguish between scenarios of a novel fault and known fault with different intensities. Thus, the proposed scheme reduced the emphasis on exhaustive historical data and can update the monitoring scheme as the new fault events occur.

REFERENCES

Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function*. Plenum Press, New York.

Chiang L. H., Russell E. L. & Braatz R. D. (2000), Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis, *Chemometrics & Intelligent Laboratory Systems*, **50**, 243-252.

Chintalapudi K. K. & Kam M. (1998), The Credibilistic fuzzy c-means clustering algorithms, *IEEE Int. Conf. Syst., Man, Cybernetics*, **2**, 2034-2039.

Choi S. W., Yoo C. K., & Lee I. (2003), Overall statistical monitoring of static and dynamic patterns, *Industrial and Engineering chemistry research*, **42**, 108-117.

Congalidis J. P., Richards J. R. & Ray W. H. (1986), Modeling and control of copolymerization reactor, *Proceedings of American Control Conference*, **3**, 1779-1793.

Duda R. O., Hart P. E. and Stork D. G (2003), *Pattern Classification*, Wiley-Interscience publication, New York.

Gustafson, D.E. and Kessel W.C. (1979), Fuzzy clustering with a fuzzy covariance matrix. *Proc. IEEE CDC*, San Diego, CA, USA, 761–766.

Johannesmeyer, M.C., Singhal A. and Seborg D.E. (2002), Pattern Matching in Historical Data, *AIChE J.*, **48**, 2022-2038.

Kresta J. V., MacGregor J. F. & Marlin T. E. (1991), Multivariate statistical monitoring of process operating performance. *Canadian Journal of Chemical Engineering*, **69**, 35-47.

Krishnapuram R. and Keller J. M. (1993), A possibilistic approach to clustering, *IEEE transactions on fuzzy systems*, **1**(2), 98-110.

Meel, A., Venkat, A. and Gudi, R.D. (2003), "Disturbance Classification and Rejection using Pattern Recognition methods", *Ind. and Eng. Chem. Res.*, **42**, 3321-3333.

Peter He Q., Wang J. & Joe Qin S. (2005), A new fault diagnosis method using fault directions in Fisher discriminant analysis, *AIChE J.*, **51**(2), 555-571.

Singhal, A., and Seborg D.E. (2002), Pattern Matching in Multivariate Time Series Databases Using A Moving Window Approach, *Ind. & Eng. Chem. Res.*, **41**, 3822-28.

Teppola P. and Minkkinen P. (1999), Possibilistic and fuzzy c-means clustering for process monitoring in an activated sludge waste-water treatment plant, *Journal of Chemometrics*, **13**, 445-459.

Venkat A. and Gudi, R.D. (2002), "Fuzzy Segregation based Identification and Control of Nonlinear Dynamic Systems", *Ind. and Eng. Chem. Res.*, **41**, 538-552.

Keynote 9

The Role of Control in Design: From Fixing Problems to the Design Of Dynamics

A. Banaszuk, P. G. Mehta and G. Hagen
United Technologies

Keynote 10

Distributed Decision Making in Supply Chain Networks

B. E. Ydstie, K. R. Jillson and E. J. Dozal-Mejorada,
Carnegie Mellon University

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

THE ROLE OF CONTROL IN DESIGN: FROM FIXING PROBLEMS TO THE DESIGN OF DYNAMICS

Andrzej Banaszuk * **Prashant G. Mehta** ** **Greg Hagen** ***

* *United Technologies Research Center, East Hartford, CT 06108, USA 860 610 7381,* banasza@utrc.utc.com
** *Department of Mechanical and Industrial Engineering University of Illinois at Urbana-Champaign 1206 W. Green Street Urbana, IL 61801,* mehtapg@uiuc.edu
*** *United Technologies Research Center, East Hartford, CT 06108, USA,* hagengs@utrc.utc.com

Abstract: We will advocate the need to change the role of dynamics and control community from fixing problems related to the detrimental dynamics using active control to the design for beneficial dynamics early in the design cycle. We will summarize lessons learned in industrial research on mitigation of flow and structure oscillations in jet engines. We will show how the decisions on the control system architecture (sensor and actuator location) impact the achievable level of suppression of oscillations (fundamental limitations of performance). Attempts to introduce control late in the design process and without proper attention to control architecture often fail because of high cost to modify the design to add on active control. We will also show how certain aspects of design (symmetry) contribute to the origin of detrimental oscillations and point out how the dynamical systems and control theory methods can guide the design to prevent the oscillations.

Keywords: Fundamental limits, nonlinear control, describing functions, combustion

1. INTRODUCTION

In this paper we will review the typical role of dynamics and control communities in the design cycle of new products and advocate the need to change this role from mainly *reactive* to strongly *proactive*. Case studies in active and passive control of oscillations in jet engines will be used to illustrate both the current and the proposed use of dynamics and control methods as well as the role of dynamics and control experts in developing technologies applicable to jet engines. While author's experience was restricted to jet engines, we hypothesize that the assessment of the current role of the control and dynamics communities and merits of the proposed new role apply broadly across multiple industries.

Dynamical phenomena such as transients and oscillations strongly affects operation of most devices. Active control is often used to modify the dynamics late in the design cycle. However, the dynamics and control communities play relatively insignificant role in the early design cycle for new products. Since not enough attention is paid in early design to the dynamic characteristics of the product, it is often discovered late in the design cycle, when the first prototypes are built and tested, that the dynamic properties of the product are not acceptable. To fix the dynamics problem a costly recovery process is launched. Often only at this point the dynamics and control experts are invited to participate. However, at this stage the design modifications required to modify the dynamics with control are extremely constrained by the hardware already built, cost, and schedule. As a result an active

control solution is rarely accepted and a more practical passive control solution is sought for. Independently of whether an active or passive control solution is selected, its implementation cost at this late stage is much higher than a cost of a similar solution if it were implemented early in the cycle. In this paper we advocate the need for a *Design of Dynamics*, which amounts to an early introduction of dynamics and control methods in the design process to properly address the dynamic characteristics of products.

The paper is organized as follows. We begin with an overview of the current role of dynamics and control communities in the design. We argue that the dynamics and control communities are typically *reactive*, *excluded* from the process of selection of control architecture and a model, biased towards active control solution that is *external* to the product (extra hardware), and with *narrow* focus on the design of control algorithm. We provide case studies in control of flutter and thermoacoustic instabilities that point out the consequences of such behavior. We will show how lack of participation of the control experts in the process of selection of control architecture can lead to an intractable control design problem because of fundamental limitation of performance. We will also show how *engaging* in the process of selecting a control architecture and a model leads to an improved control system with clear understanding of the physical factors that fundamentally limit the control performance. Next, we will show on an example of analysis of wave phenomena in jet engines how manipulation of the natural physical feedback loops in the product can lead to a solution with minimal modifications to the product. We will argues that the dynamics and control experts need to be more *proactive*, *engaged* in the design process, and considering *broad* range of solutions with a preference toward these *internal* to the product. We will conclude by indicating some *technical* and *social* barriers that need to be overcome before the Design of Dynamics is introduced into industrial practice.

2. CURRENT ROLE OF DYNAMICS AND CONTROL IN DESIGN: FIXING PROBLEMS LATE IN DESIGN CYCLE

Despite decades of extensive research detrimental oscillatory wave phenomena still drive jet engines development and maintenance costs and limit their operability. Jet engines are designed for high performance, survivability, operability, and affordability. However, design for increased performance and low observability often leads to excitation of detrimental wave phenomena such as flutter, rotating stall, and thermoacoustic instabilities that reduce engine parts life and limit its operability. In particular, compressor and fan blade failure is still a common problem for all major engine manufacturers in spite of decades of extensive research in the area of design for flutter and High

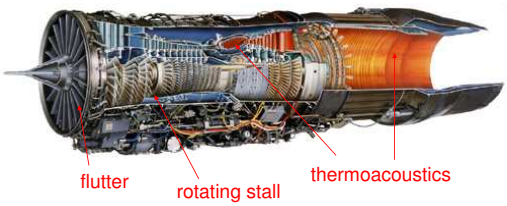


Fig. 1. Detrimental wave phenomena affecting operations of jet engines

Cycle Fatigue mitigation. While first discovered in 1950-ties, augmentor screech and rumble as of today are still common problems for military engines.

Mitigation of detrimental wave phenomena in jet engines is difficult since it involves controlling sensitive and complex dynamics in presence of model uncertainty and severe design constraints. Lightly damped structural, acoustic, and fluid dynamic modes easily become under-damped or unstable because of positive feedback coupling with other flow phenomena and excitation by broad band, and tonal flow disturbances. Physics that causes detrimental oscillations involves complex high Mach and Reynolds number flows, which cannot be reliably and accurately computed with current methods. State of the art computational methods based on CFD are too computationally expensive to be applied. Hence, model analysis is not utilized to exploit the design space and find innovative damping solutions in early design stage. The accuracy of reduced order models utilized is doubtful especially when chemical reaction, flow separation, or shocks are involved. Because of high uncertainty, models are not utilized for a design of robust oscillation mitigation solutions and robustness of chosen solutions to unexpected off-design conditions is not guaranteed. Avoiding oscillations by operating engines at regimes with large stability margins results in unacceptable performance loss.

Passive dampers can be used to control oscillations, but they undesirable additions, since they increase engine weight and complexity. Moreover, the positive effects of the passive damping devices are only utilized at a small portion of flight envelope when instabilities occur, but the negative aspects (like weight) impact engine performance at all operating conditions. Hence, active control is often considered as an alternative to passive dampers. Both passive and active control solutions are typically introduced late in the design cycle as a reaction to dynamics problem discovered when first prototypes are build and tested. The active control community typically becomes engaged late in the design process when the design process owners recognize possibility that active control can solve the dynamics problems. Very often the control experts do not participate in creating the model used for control

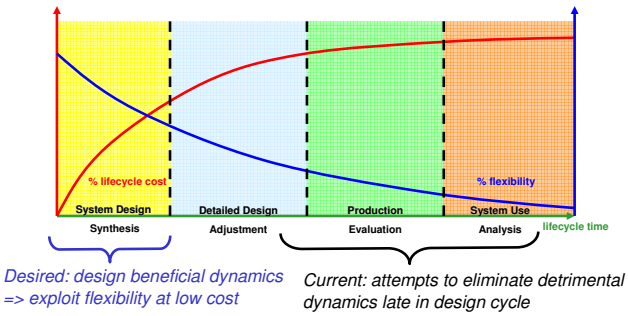


Fig. 2. Impact of dynamics and control methods on the design cycle

design and in the design of the control architecture, limiting its role to designing the control algorithms. In

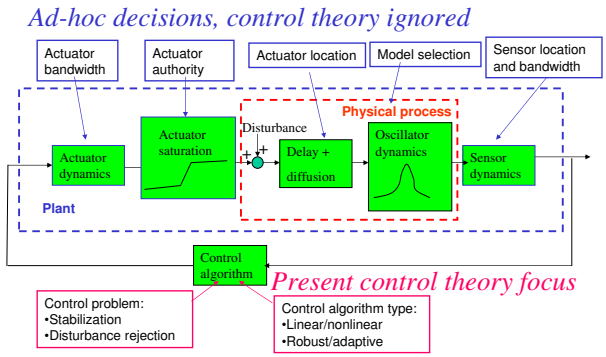


Fig. 3. Decisions that impact control performance

this paper we will show in examples how a selection of a control architecture that ignores the principles of control theory can lead to an intractable control problem in which the desired control performance cannot be achieved because the achievable control performance is *fundamentally limited* by the control architecture and physics of the problem.

The dynamical systems academic community has created tools to *analyze* the dynamics if low dimensional models are available, but not to *design* the beneficial dynamics. This community is typically not engaged with industrial processes missing a significant opportunity for impact. In this paper we will show an example how, by using ideas from the theory of the *dynamical systems with symmetry*, one can identify the root causes of detrimental dynamics and how one can create *beneficial dynamic interactions* that eliminate the detrimental dynamic behavior.

Let us summarize here some attributes of the *current* role that the dynamics and control communities play in the jet engine design process. First, they act *reactively*. They will only act when called upon by the design process owners. This usually means a late entrance into the process, when the acceptable solutions are extremely constrained. Secondly, the control experts will rarely attempt to analyze the natural dynamics of the problem and solve the problem by manipulation of the natural dynamics. Instead, they will typically look for an active control solution which is *external* to the product, i.e., requires adding extra sensors and actuators. The solutions obtained in this way attempt to *override* the natural dynamics and require a nontrivial modifications of the design. In particular, it requires extra hardware, which means extra cost and complexity. This is always the least desirable solution. Third, they will often *accept assumptions* about the control problem definition, proper control architecture, and model of the process made by the process owners without questions. Given the model, the sensors, and the actuators defined by the design process owners, the control engineers will quickly proceed to the design of a control algorithm and its experimental verification. At best such behavior can result in expensive active control solution if the assumptions made by process owners are correct and the extra hardware addition is acceptable. However, since the assumptions that led to the definition of an active control concept were made without involving control experts, they often are incorrect and the active control solution does not satisfy the performance requirements. Even when these assumptions are corrected and the active control performance is acceptable, the active control approach often cannot meet the acceptable criteria in term of cost or complexity and is abandoned in favor of a cheaper and easier to implement passive control solution if the latter is found.

3. THE IMPORTANCE OF PROPER CHOICE OF CONTROL ARCHITECTURE

In this section we will describe an active control project in which initial lack of team play between the design process owners and the control engineers resulted in a failure of the project to achieve an acceptable control performance. After this initial failure, a close collaboration of the design process owners with control engineers was established, which resulted in a discovery of a superior control architecture and demonstration of an excellent control performance.

Despite advances in aeromechanical engineering, fan stall flutter (Forsching, 1984) remains a substantial constraint in jet engine designs. The motivation for the work described in this section was to investigate the extent to which active control of this aeromechanical instability can extend the operability of a given fan design. The control objective was damping augmentation of the flutter modes.

The details of modeling, control design, and experimental demonstration of active flutter control are summarized in the papers (Banaszuk *et al.*, 2002*a*; Banaszuk *et al.*, 2002*b*; Rey *et al.*, 2003).

The experimental setup shown in Figure 4 contains a 17 inch scale fan with flow characteristics and flutter margin comparable to of those found in high by-pass ratio commercial jet engines. The first attempts to provide flutter damping augmentation involved using

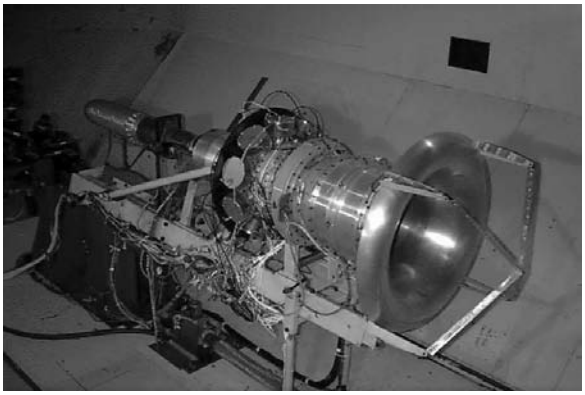


Fig. 4. 17” fan experimental rig

pressure sensors and five circumferentially located valves as actuators. This architecture was chosen by the turbomachinery experts responsible for the the experimental demonstration of a high performance flutter control system.

An active control algorithm was supposed to be designed using a model extracted from a frequency response of the pressure sensors. A typical frequency response with the pressure sensors is shown in Figure 5. Note that the lightly damped flutter pole represented by a spike in the magnitude response around 273Hz is accompanied by a zero only 1Hz apart. Such proxim-

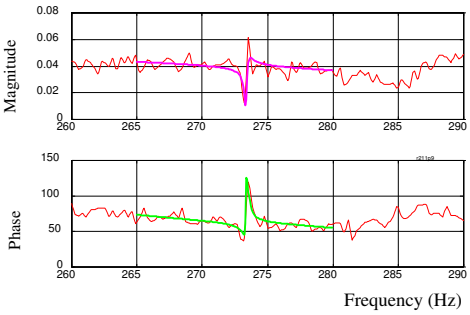


Fig. 5. Flutter frequency response using pressure sensors

ity of zero to the flutter pole resulted in a very difficult control problem. Four control engineers made four separate attempts using different control algorithm design techniques to provide damping augmentation for the 273Hz flutter pole. Since the close proximity of pole and zero indicates severe fundamental limitations of achievable control performance, the damping augmentation achieved was insignificant. Even though several pressure sensor configurations were tested, all sensor configurations resulted in a near pole/zero cancelation.

The failure of the attempts to control flutter using pressure sensors could be explained by an inadequate design of the control architecture (namely selection of sensors and actuators), which resulted in a pole/zero configuration that fundamentally limited achievable control performance. The turbomachinery experts who designed the architecture used their physical intuition, controllability and observability of flutter being the only control aspects that were analyzed properly. They were unaware of an importance of a proper design methodology leading to an architecture that avoids a near pole/zero cancelation. While the control experts understood the detrimental influence of a near pole/zero cancelation on the control performance, they were not involved in the design of the control architecture, simply because they did not insist on a participation in the control architecture design process.

The root cause of a near pole/zero cancelation was discovered during a discussion involving both turbomachinery and control experts that lasted only one and half hour. It was postulated that a strong direct feedthrough from the actuators to the pressure sensors dominating the pressure response is the root cause of a near poles/zero cancelation. When a large direct feedthrough term is added to a smaller transfer function representing flutter, it is easy to show (by combining the terms into one simple fraction) that a near pole/zero cancelation will occur. Linking the origin of a near pole/zero cancelation to the physics of the problem was the key development. The turbomachinery experts started to appreciate the value the control theory methods and became strong promoters of the active control methods.

Another discussion led to an identification of an easily implementable sensing approach that eliminated the direct feed-through. With the new eddy current sensors to measure the blade time arrival the direct feedthough was completely eliminated and zeros close to poles were removed (Rey *et al.*, 2003). Figure 6 shows the comparison of the frequency responses of flutter dynamics using the pressure and eddy current sensors.

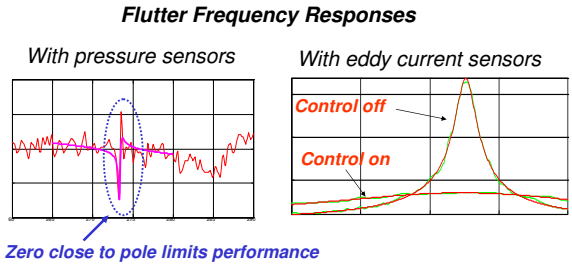


Fig. 6. Flutter magnitude response using pressure and eddy current sensors

The final active control hardware consisted of ten zero mean mass flow actuators equally spaced around the fan case between the blade row and the exit guide vanes. The actuators consisted of regular audio speak-

ers enclosed in a pressure vessel. Flutter was sensed by means of eddy current sensors mounted on the fan casing. As blades speed past these sensors, the sensed signal is used to record the blade arrival time. Early and late arrivals are associated with combinations of forward and backward bending and twisting of the blades from which the flutter modes amplitudes can be derived in real-time.

Figure 7 shows a schematic of the control system. The "Inverse DFT" block in the diagram performs an inverse spatial Fourier Transform which converts each flutter control signal into an actuator command according to the position of the actuator along the circumference. The amplitude of the mass-flow corresponds to that of a sine-wave of the nodal diameter and phase speed of the traveling wave.

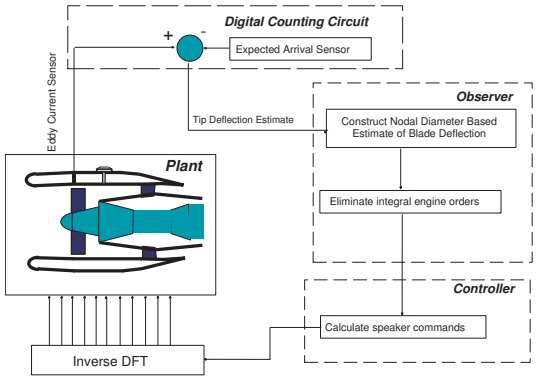


Fig. 7. Flutter control system schematics

The control system was able to add damping to the three critical flutter modes. The damping augmentation achieved was an order of magnitude larger than the intrinsic aeromechanical damping of the flutter modes at the design point. More details on the flutter control algorithms used in this work can be found in (Banaszuk *et al.*, 2002*b*).

Figure 8 shows a summary of the damping augmentation achieved for 0, 1 and 2 nodal diameter flutter of the blade first bending mode. Notice that the range of damping values for the open loop system between the design point (label "A") and the flutter boundary (label "B") is much smaller than the amount of damping added through active control.
Unfortunately, even though feasibility of active control of flutter with off-blade sensors and actuators was demonstrated in a rig, the technology did not make it to the product it was supposed to impact. A fan blade redesign resulted in an elimination of the dynamics problem on the product, and hence active control solutions was no longer needed.

In this section we described an active control project in which an initial lack of team play between the design process owners and the control engineers resulted in a failure of the project to achieve an acceptable control performance. After this initial failure, a close collaboration of the design process owners with control engineers was established, which resulted in a definition

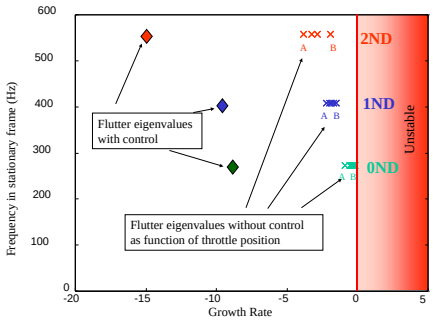


Fig. 8. Summary of flutter control experiments

of a superior control architecture and an experimental demonstration of an excellent control performance. The key factor of the success was an assumption of a *proactive* role by the control experts. Despite this technical success, the active control technology was abandoned in favor of a passive control solutions that was *internal* to the product.

4. THE IMPORTANCE OF CORRECT MODELING ASSUMPTION

In this section we will describe an active control project in which a *wrong modeling assumption* that the dynamics being controlled could be represented as a linearly unstable limit cycling system led to inadequate definition of control objective as a stabilization problem. These wrong assumptions resulted in a failure of the project to explain the poor performance achieved in some of the active control experiments. Eventually, a rigorous analysis revealed the faulty assumptions. A new assumption was stated that the dynamics should be modeled as a noise-driven system with a large control delay. The controlled system can be either stable or unstable depending on the particular values of the parameters. The presence of a large broad-band disturbance driving the system implies that a proper control objective is a disturbance attenuation, rather than stabilization. In addition, the presence of a large delay in the control path causes the achievable control performance to be fundamentally limited. This in turn explains a poor control performance observed in some control experiments. For more details we refer to papers (Banaszuk *et al.*, 1999*a*; Banaszuk *et al.*, 1999*b*; Cohen and Banaszuk, 2003; Mezic and Banaszuk, 2004).

Emphasis on reducing the level of pollutants created by gas turbine combustors has led to the development of premixed combustor designs, especially for industrial applications. Premixing large amounts of air with the fuel prior to its injection into the combustor greatly reduces peak temperatures within the combustor and leads to lower NOx emissions. However, premixed combustors are susceptible to the so-called

thermoacoustic combustion instabilities. These instabilities arises due to a destabilizing feedback coupling between acoustics and combustion (unsteady heat release). It causes large pressure oscillation in the combustor that detrimentally affects the combustor durability and raises environmental noise pollution (Seume *et al.*, 1997).

Active Combustion Instability Control (ACIC) with fuel modulation has appeared an effective approach for reducing pressure oscillations in combustors. Promising experimental results have been reported by researchers at United Technologies Research Center (UTRC) (Cohen *et al.*, 1998; Hibshman *et al.*, 1999), Seimens kWU (Seume *et al.*, 1997; Hoffmann *et al.*, 1998), ABB/Alstom (Paschereit *et al.*, 1999), Honeywell Inc. (Anson *et al.*, 2002), Westinghouse/Georgia Institute of Technology (Sattinger *et al.*, 1998), and the U.S. Department of Energy (Richards *et al.*, 1995). However, the achieved reduction of pressure oscillation varies between these experiments from 6dB to 20dB. In many cases, the attenuation of the oscillation at primary frequency is accompanied by excitation of the oscillation in some other frequency band (Langhorne *et al.*, 1988; Fleifil *et al.*, 1997; Saunders *et al.*, 1999). This phenomenon is commonly referred to as *secondary peaking* or *peak splitting*.

A satisfactory explanation of the different attenuation levels and peak-splitting phenomena has not been presented in the literature. Much of the theoretical attention in the area of ACIC has focused on control design (Bloxsidge *et al.*, 1987; Bloxsidge *et al.*, 1988; Langhorne *et al.*, 1988; Chu *et al.*, 1998; Hathout *et al.*, 2000; Evesque *et al.*, 2000) – that is inherently dependent on the dynamics considered in the model or present in the experiment – and not so much on factors that actually limit the achievable performance. One of the reasons for this is that the thermoacoustic oscillations frequently arise as a limit cycle that requires nonlinear models of combustion dynamics. This limits the mathematical tools available for both control design as well as the analysis of resulting dynamics.

We investigated the factors that determined achievable reduction of the level of pressure oscillation in combustors using fuel control. Our studies have been motivated by experience with ACIC in the experiments conducted at UTRC (Cohen *et al.*, 1998; Hibshman *et al.*, 1999). These experiments were done in sub-scale single nozzle combustors.

An industrial engine is equipped with an annular combustor comprising of several premixing fuel nozzles arranged along the circumference. The ACIC experiments used sector embodiments of the annular combustor. Figure 9 depicts a four megawatt single-nozzle combustor and a three-nozzle sector combustor. In either setup, experiments were carried out at realistic operating conditions and between 10-17% of the net fuel was modulated for control using linear proportional or nonlinear on-off fuel valves. Pressure sensors inside the combustor were used for feedback. Additional details on the experiments appear in (Cohen *et al.*, 1998; Hibshman *et al.*, 1999).

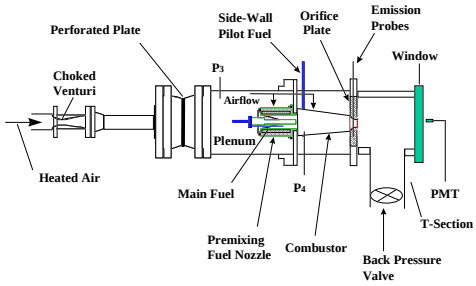


Fig. 9. UTRC single-nozzle 4MW combustor.

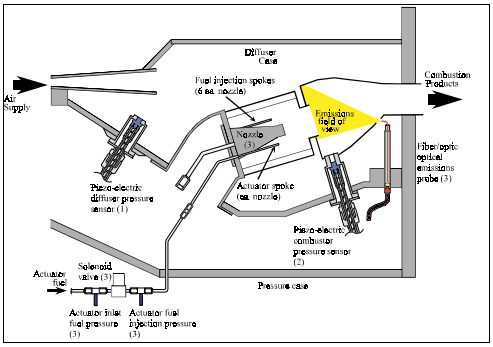


Fig. 10. UTRC three-nozzle sector combustor.

Combustion dynamics arise due to a feedback coupling between the acoustic modes of the combustor cavity and the unsteady heat released due to combustion of fuel-air mixture. The resulting feedback interconnection is typically referred to as a *thermoacoustic loop*. In the simplest setting considered here, the acoustics is modeled by the bulk Hemholtz mode of the combustor cavity. The precise physical mechanisms underlying the unsteady heat release are complex and reduced order models for the same are not well-understood. Here, the unsteady heat release was modeled as a fluctuation in the equivalence ratio (normalized fuel/air ratio) expressed as a nonlinear function of acoustic velocity input. Only the simplest two effects are considered to model the functional relationship. One is the bulk fluid convection effect that is modeled by a time delay and the other is the effect due to time-delay and nonlinearities in the burning rate where the latter that is modeled by a static saturation nonlinearity. The resulting thermoacoustic model equations arise as

$$\frac{d}{dt}\begin{bmatrix} \rho c u_i \\ \rho c u_e \\ p \end{bmatrix} =$$

$$\begin{bmatrix} -M_i c_i/l_i & 0 & c_i/l_i \\ 0 & -M_e c_e/l_e & c_e/l_e \\ -A_i c/V & -A_e c/V & 0 \end{bmatrix} \begin{bmatrix} \rho c u_i \\ \rho c u_e \\ p \end{bmatrix} + \quad (1)$$

$$\begin{bmatrix} 0 \\ 0 \\ H\left(u_i(t-\tau)+u_t(t), w(t-\tau)\right) \end{bmatrix},$$

where p is the combustor chamber (modeled as a capacitance) pressure, $\rho c u_i$ is the upstream nozzle mass velocity, $\rho c u_e$ is the downstream exit mass velocity, $w(t)$ is the fuel mass flow input, and $u_t(t)$ is used to model the stochastic turbulent flow velocity in the nozzle – assumed to be a broad-band white noise. The heat release function $H\left(u_i(t-\tau)+u_t(t), w(t-\tau)\right)$ in (1) captures in a reduced order fashion the nonlinear effects due to combustion. The parameter τ represents the cumulative time delay – primary delay due to convection plus delay because of chemical reaction and fuel-air mixing. For additional details on the model and explicit characteristics of the forcing term, see (Peracchio and Proscia, 1998).

In the existing thermoacoustic literature a commonly accepted assumption was that a presence of peaks in the pressure spectra is an indication of a limit cycle. This assumption was adopted by the UTRC team that included combustion engineers, dynamical system experts, and control engineers. As a consequence of this modeling assumption the adopted control objective was a *stabilization of a linearly unstable limit-cycling plant*. A simple phase-shifting algorithm using pressure sensors was designed to control the fuel valves.

The amount of the pressure amplitude attenuation with control varied between the rigs and between various operating conditions. Figure 11 shows spectra of pressure without and with active fuel control. Note that 5.5x attenuation was achieved in the single-nozzle rig, while in the sector rig the attenuation was only 2x. The attenuation in the sector rig was limited by the peak-splitting phenomenon mentioned above. This result was puzzling, since the peak-splitting could not be easily explained using the limit-cycling plant assumptions.

Eventually a breaktrough was achieved when the limit-cycling model assumption was questioned. Methods presented in the paper (Mezic and Banaszuk, 2004) led to identification of regions of validity of linear and nonlinear models for thermoacoustic oscillations. An alternative hypothesis was formulated that the low amplitude pressure oscillations should be modeled using a *linearly stable, noise-driven model*. To verify this hypothesis, a control-oriented thermoacoustic models were identified by fitting the experimentally obtained frequency response from fuel valve input to the pressure sensor output. The frequency

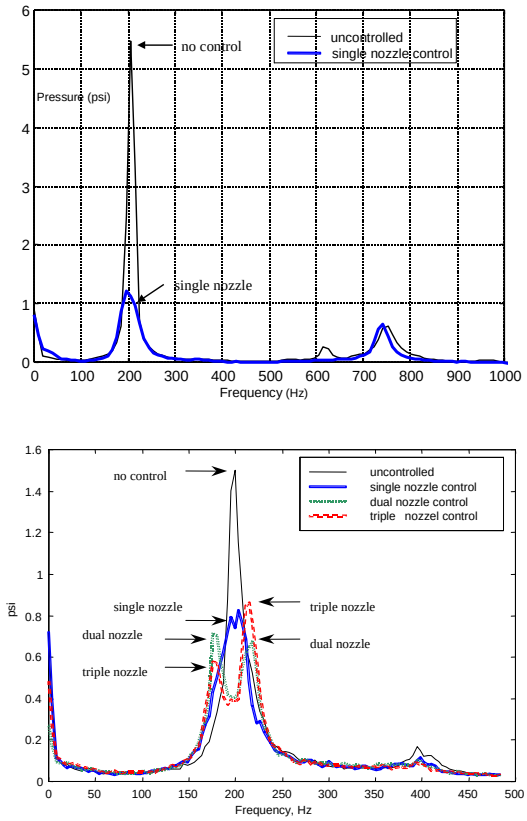


Fig. 11. Open and closed-loop pressure spectra for the single nozzle and sector combustors.

response experiments were carried out a in a range of operating conditions with both the single-nozzle and sector (three-nozzle) combustors. For the single nozzle combustor operating at the high equivalence ratio condition, a linear model consisting of a lightly damped second order system together with a (large) delay was found to fit the data well. In the following, we describe the identification and validation of the linearity hypothesis with this model.

At the high equivalence ratio condition, the pressure oscillations observed in the single nozzle combustor are relatively small and the proportional actuator used for control operates in its linear range. Therefore, it was hypothesized that a linear plant and controller model may be used to analyze the behavior of the controlled system. Figure 12 depicts the structure of the feedback control system. Figure 13 compares the experimentally obtained frequency response to it's model fit. The identified model arises as a second order lightly damped oscillator with a delay of $\tau = 4.4$ ms chosen to match the phase roll-off in the $300 - 400$ Hz frequency range. As the identified model is linear and stable, a model of external noise is needed to account for the pressure oscillations observed in the experiments. We use the model structure for the noise in Eq. (1) together with the identified model to estimate a noise model. In particular, a white noise model is built at the plant input (see Figure 12) to match

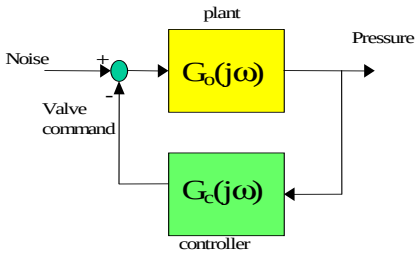


Fig. 12. Feedback control of thermoacoustic plant in the presence of noise: a pressure measurement is used to obtain the fuel valve control input.

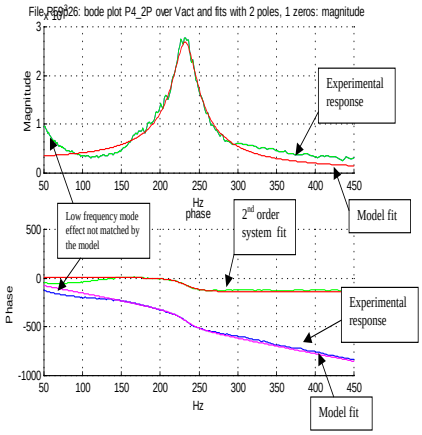


Fig. 13. Results of the linear model identification of the frequency response obtained in the single nozzle rig experiment.

the experimentally obtained PSD of the uncontrolled pressure. Figure 14 shows that the identified noise

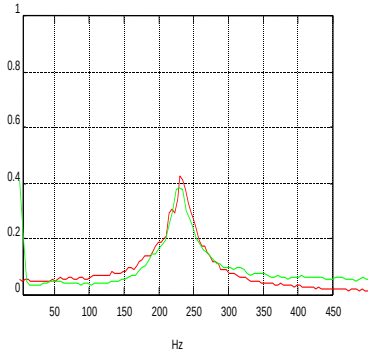


Fig. 14. Square root of the PSD of pressure from experiment and from model simulation.

model allows us to match the experimentally obtained pressure PSD with the results of the model simulations using SIMULINK. In the combustion experiment, a wide band turbulent air velocity fluctuation in the nozzle is one of the sources for the presence of noise. The identified plant model includes the fuel valve actuator dynamics together with the thermoacoustic model dynamics of Eq. (1). The frequency response of the actuator is effectively flat over a wide frequency band about the resonant thermoacoustic frequency ω_r. As a result, additional states are not needed and a second order model with delay consistent with Eq. (1) is sufficient.

Finally, feedback control experiments were used to validate the implicit linearity hypothesis and the noise model. An observer-based phase-shifting controller was used both in the experiment and in the model simulations. Figure 15 compares the experimentally obtained pressure PSD with the PSD obtained from simulations with various phase-shifting controllers. The fact that the two PSDs are nearly identical implies that out plant and noise models are valid and suitable for the control design at the high equivalence ratio condition in the single-nozzle combustor.

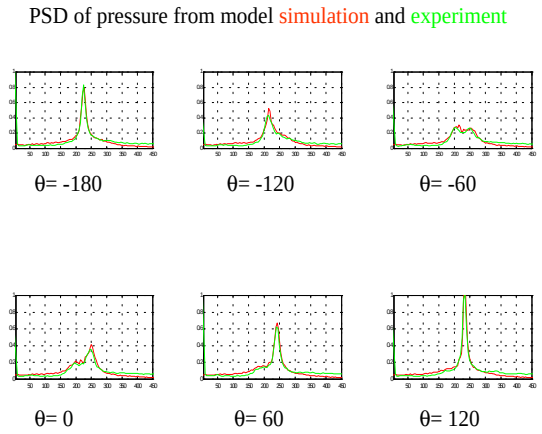


Fig. 15. Results of validation using feedback control: effect of a phase-shifting controller on pressure PSD in experiment and simulation.

The correction of the model assumptions led to an explanation of the peak splitting phenomenon and eventually to understanding of the fundamental limitations of the achievable control performance. This key technical contribution showed the value of the control theory methods to the design process owners (in this case the combustion engineers) and increased *credibility* of the control engineers among the combustion engineers. The process owners became open to learning the basic principles of the feedback control theory. Translation of the control theory principles to the language of physics was most important in breaking the *language barrier* between the dynamics and control group and the process owners.

5. FUNDAMENTAL LIMITATIONS OF PERFORMANCE

In this section we will discuss a relationship between the physics of a problem, the control architecture selection, and the achievable *control performance* using the thermoacoustic problem introduced in the previous

section as an example. The performance limitations of the achievable suppression of oscillations will be described in terms of controller-independent lower bounds on the sensitivity function gain. The lower bounds will depend on the physics of the problem and the selection of control architecture. Since these factors cannot be analyzed independently, it will become mandatory that a high performance control architecture can only be designed by a team including the experts in control and in the physics of the problem. For more details we refer to papers (Banaszuk *et al.*, 1999*a*; Banaszuk *et al.*, 1999*b*; Cohen and Banaszuk, 2003; Mehta *et al.*, 2004).

In this section, the fundamental limitations associated with the feedback control of combustion instabilities are discussed. The theory is applied to obtain bounds on achievable performance in the high equivalence ratio experiments where linear plant and control models are adequate. In particular, the analysis helps explain the peak splitting phenomenon observed in UTRC and other ACIC experiments. In frequency domain, the closed-loop transfer function from the noise model to the pressure measurement is given by

$$\frac{p(j\omega)}{n(j\omega)} = G_0(j\omega)S(j\omega), \tag{2}$$

where

$$S(j\omega) = \frac{1}{(1 + G_0(j\omega)G_c(j\omega))} \tag{3}$$

denotes the sensitivity function. The control objective is to stabilize the closed loop system and shape the sensitivity function with the objective of reducing the noise driven pressure oscillation. In particular, the controller attenuates the noise at frequencies where $|S(j\omega)| < 1$ and amplifies the noise otherwise. Figure 16 depicts the experimentally obtained Nyquist diagram for the controlled single nozzle combustor. The attenuation and excitation frequency bands are also shown. The effect of the phase-shifting controller is to rotate the diagram so that the attenuation is maximized at the resonant frequency ω_r. The presence of a large delay in the loop makes it difficult to achieve broadband attenuation of pressure oscillations – the sidelobes in the diagram are the regions of secondary peaks. This observation in our closed-loop combustion experiments (Cohen *et al.*, 1998; Hibshman *et al.*, 1999) together with a wide range of performance results in the ACIC literature (Seume *et al.*, 1997; Hoffmann *et al.*, 1998; Paschereit *et al.*, 1999; Anson *et al.*, 2002; Sattinger *et al.*, 1998; Richards *et al.*, 1995) motivated us to study the fundamental limitations of ACIC. Our objective was to better understand – in a controller independent fashion – the effect of delay, limited actuator bandwidth and authority and plant dynamics (unstable poles) on the achievable performance and study the resulting trade-offs.

Fundamental limitations in obtainable performance (and robustness) are determined by certain conservation laws that govern the balance of negative and positive areas under the sensitivity (and complementary sensitivity) frequency response (Seron *et al.*, 1997; Freudenberg and Iooze, 1987). These laws are used to obtain controller-independent bounds on performance and robustness with *any* LTI controller. For the sensitivity function, obtainable performance bounds can be derived from the celebrated Bode integral formula

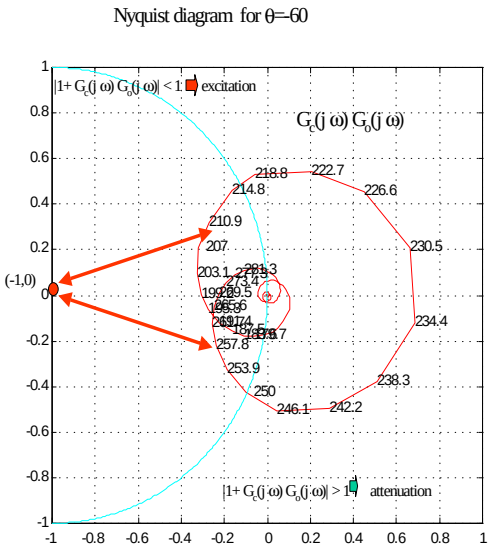


Fig. 16. Nyquist diagram for the phase-shifting controller with optimal phase shift.

$$\int_0^\infty log \mid S(j\omega) \mid d\omega = 2\pi\sigma_r, \tag{4}$$

where σ_r is the real part of the resonant unstable pole-pair; right-hand-side is zero for open-loop stable plant. The integral formula shows that noise attenuation (which requires $\mid S(j\omega) \mid < 1$) over a certain frequency band is always accompanied by noise amplification $\mid S(j\omega) \mid > 1$ over some other frequency band. (This is sometimes referred to as the waterbed effect.) In the presence of unstable poles, a larger penalty is paid in terms of sensitivity amplification. Figure 17 provides a graphical representation of the area formula: sensitivity reduction (negative area in the integral) is always accompanied by sensitivity amplification (positive area).

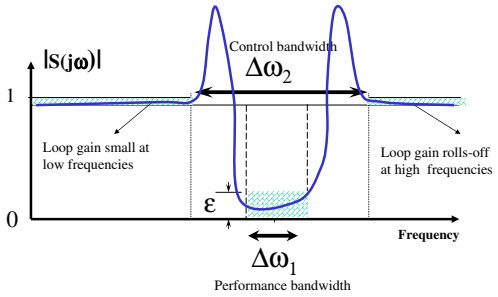


Fig. 17. A typical sensitivity function for control of oscillations

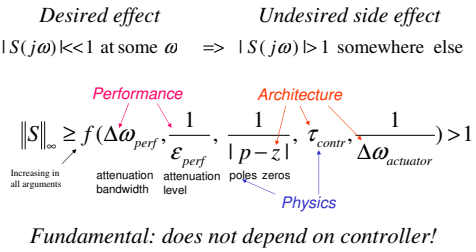


Fig. 18. Fundamental limitations of control performance showed as a lower bound on the sensitivity function gain

The performance objective for ACIC is to shape the sensitivity function so that it is small at and near the resonant frequency ω_r, i.e.,

$$|S(j\omega)| < \epsilon \tag{5}$$

for $\omega \in \Delta\omega_1$ where $\Delta\omega_1$ is the performance bandwith centered at ω_r. Meeting this performance objective creates negative area in the integral and this leads to noise amplification at some other frequencies. If the control bandwidth were infinite, the positive area may be distributed over a wide frequency band so amplification at any given frequency may be designed to be arbitrarily small. However, if the control bandwidth is finite (so the loop rolls off beyond certain low and high frequencies), the positive area would have to be accommodated in a smaller frequency band (where loop gain is high) and this would necessarily result in peaking of the sensitivity function.

In the industrial ACIC settings at UTRC, the linearity hypothesis and subsequent control-oriented analysis of the preceding section applies only to a limited set of operating conditions. For most operating conditions of practical interest, the linearity hypothesis is not applicable because of in the industrial settings, the high power requirements of fuel modulation due to control means that the actuator essentially operates in its saturated nonlinear range. Next, On-Off actuator is a popular and cheap fuel actuator that is widely used for ACIC. ACIC experiments in sector combustor (Hibshman *et al.*, 1999) used On-Off actuators. The resulting closed-loop feedback system was thus nonlinear. Experimental results obtained with a linear controller showed peak splitting for a range of operating conditions. Figure 19 depicts the PSD of the pressure oscillations with one, two, and three fuel nozzles operating.

In order to understand the nonlinear effects because of On-Off actuators, the operating conditions for the uncontrolled case are specifically chosen to verify the linearity hypothesis for the thermoacoustic model. A linear thermoacoustic and noise model are identified from experiments using the procedure described in the previous section The thermoacoustic model now includes a larger time delay of $\tau = 7$ ms and a second order linear system with resonant frequency $f_r = 208.9$ Hz.

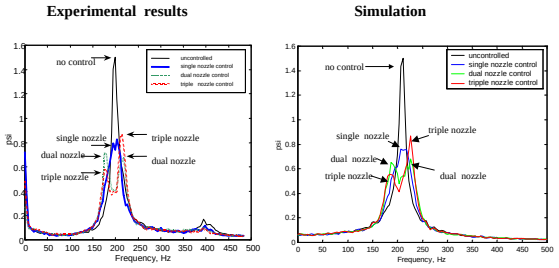


Fig. 19. PSD of pressure signal with on-off control of one, two, or three liquid fuel nozzles showing the peak splitting phenomenon observed in experiment and simulation.

In the models of the ACIC experiments with On-Off actuators, Gaussian balance using Random-Input Describing Functions (Gelb and Velde, 1968) yields an approximation of the feedback loop with respect to the Gaussian noise balance. The analysis in (Banaszuk *et al.*, 1999*b*; Cohen and Banaszuk, 2003) shows that the loop $G_1(j\omega)N_R(A(\sigma),\sigma)$ (where $N_R(A(\sigma),\sigma)$ denotes the Gaussian input describing function) yields a well-posed closed-loop system for all $\sigma \neq \sigma_0$. Note that this is the case independently of the dynamics of the open loop $G_1(j\omega)$, the amplitude of the limit cycle A, and the values of Gaussian process standard deviation σ. The resulting sensitivity function is stable and one can formally write down an area formula which gives peak splitting for the approximation. Under the assumption that the approximation yields a good representation of the nonlinear model, this explains the peak splitting seen in the PSD of the ACIC experiments with On-Off actuators.

The above considerations give a formal framework for extending the fundamental limitations analysis for control of thermoacoustic loops. One considers the *modified sensitivity function* with respect to the noise balance. Peak splitting is a consequence of the area formula as applied to the modified sensitivity function. For the case of On-Off nonlinearity with G_1 stable, we showed that the *modified sensitivity function* is stable and well-posed independent of the dynamics of G_1 and the noise (variance). We expect this to be true for a larger class of nonlinearities.

Analysis provided indicates that the peaking phenomenon observed in ACIC experiments is to a large extent inevitable for combustion systems with large delay controlled with actuators of limited bandwidth. This is reflected in the fact that the sensitivity with the linear actuator case or the modified sensitivity function with the nonlinear On-Off actuator achieves values exceeding 1.

We also used the analysis to explain the difference between the experimental results obtained in single-nozzle and sector combustors (see Fig. 11). Compared to the sector combustor, the single nozzle combustor shows higher open-loop oscillations but in a narrower

band of frequencies. Such is the case because of lower damping of the thermoacoustics in the single nozzle combustor. Next, the plant delay identified from the frequency responses is higher in the sector combustor than in the single-nozzle combustor. As a result, limitations and peaking in the sector combustor – with its broadband performance objective for a thermoacoustic plant with large delay – is more severe than in the single nozzle combustor.

In this section we discussed a relationship between the physics of a problem, the control architecture selection, and the achievable *control performance*. Since these factors cannot be analyzed independently, a high performance control architecture can only be designed by a *team* including the experts in control and the physics of the problem. Such teamwork helps articulate the value of analytic methods of control theory and greatly facilitates breaking the *language* barrier between the control theory expert and the design process owners. The best possible outcome of this process is convincing the design process owners that the analysis methods of dynamical systems and control theory are an efficient way of exploiting the physics of the problem.

6. DESIGN OF BENEFICIAL DYNAMIC INTERACTIONS

In this section we will discuss how using methods of dynamics and control to analyze the natural dynamics of the product can lead to a solution of a dynamics problem that is *internal* to the product, and hence easily implementable without necessity of adding extra hardware and complexity. Such solution can only be found if the experts in the physics of the problem and the dynamics and control experts work together as a *team*. Since such team work is not a natural act, it requires prior establishment of credibility and breaking of the language barrier. Working on the active control projects such as flutter and thermoacoustics described in the previous sections can greatly facilitate creation of such a team, even if the active control technology is not implemented on a product.

In (Hagen and Banaszuk, 2004) we examined how spatial variations of the system parameters can affect the system stability properties. Recent work has focussed on analysis of heterogeneous distributed systems (Dullerud and D'Andrea, 1999; Hagen, 2004; Jovanovic *et al.*, 2003). Symmetry-breaking is commonly referred to as mistuning in the literature regarding the dynamics of arrays of turbine blades on a disk. Studies of stability properties of turbine blade flutter through the introduction of spatial nonuniformities has appeared in (Bendiksen, 2000; Rivas-Guerra and Mignolet, 2003). Optimal mistuning in arrays of bladed disks has appeared in (Petrov *et al.*, 2000; Shapiro, 1998). A study of the effects of asymmetry on compressor stall inception has appeared in (Graf *et al.*, 1998).

As in the case of mistuning in arrays of bladed disks in turbines, this form of passive control is often more feasible than implementing an active control scheme. This may also be true for the case in combustion chambers, where high temperatures prohibit adequate sensing and may damage the actuators required for active control. Furthermore, symmetry-breaking can be a more cost-effective means of stability enhancement.

Within recent years at UTRC the analysis of the role of jet engine design symmetry in the dynamics of detrimental rotating waves led to explanation of the origin of the waves and practical means of their passive control demonstrated in an engine test. It is worth pointing out that these developments were inspired by Igor Mezic analysis of the impact of the symmetry structure of DNA molecules on DNA dynamic behavior (Mezic, 2005) that provided an inspiration to the authors of the current paper that led to a discovery of the beneficial and detrimental symmetry patterns in jet engines. This key inspiration ultimately led to the concept of the Design of Dynamics for the wave phenomena in jet engines described in this paper.

Oscillatory phenomena such as thermoacoustic instabilities and turbomachinery fan blade flutter could be modeled using wave equation with a nonlinear dynamic feedback representing coupling of lightly damped acoustic or structural waves with flow or combustion. An elegant explanation of the role of jet engine design symmetry in the inception and suppression of instabilities such as thermoacoustics and flutter was provided. The explanation does not require any particular physics-based model for combustion and flow phenomena, because it only utilizes its symmetry properties. In particular, it was shown that the so-called skew-symmetric feedback is always detrimental, while breaking the symmetry of the circumferential wave speed pattern is always beneficial. The research led to a methodology for designing engines with greater dynamic stability margins that was transitioned to an engine company. The effectiveness of symmetry breaking in quenching detrimental rotating wave oscillations was demonstrated in a full-scale engine test.

To derive results on stability, it was shown that under the assumption of identical feedback elements (identical combustion flameholders, identical fan blades, etc.), any feedback model can be decomposed as a sum of symmetric and a skew-symmetric feedback. Conceptually, the symmetric feedback corresponds to dynamics that have reflection (about centerline) symmetry while the skew-symmetry is a result of local asymmetry in feedback. The symmetric feedback causes the two eigenvalues to move as a pair in the same directions. It can either stabilize or de-stabilize depending upon the feedback model. The skew-symmetric feedback, on the other hand, is always detrimental

regardless of the feedback model. It splits the eigenvalues, causing one rotating mode to gain damping while causing the other rotating mode to lose the same amount of damping. Using only the time-series data from experiments, the instability such as flutter and screech seen in experiments was explained as a consequence of the skew-symmetric feedback. The presence of a skew-symmetric feedback also explains why rotating wave instabilities in jet engines have preferential direction of rotation.

The second idea was to modify the structural aspects of the model in order to control the instability. This was accomplished by introducing precise spatial variations (mistuning) in the "mean properties" such as wave speed of the wave equation. While the skew-symmetric feedback causes the two eigenvalues to move apart, mistuning causes the eigenvalues to move closer. In either case, the net amount of damping in the system remains the same. This net damping depends upon the net symmetric feedback due to the presence of liner etc. and is not affected by spatial variation in mean. In effect, the mistuning utilizes the more heavily damped system modes to augment the damping of the lightly damped modes.

For a given skew-symmetric feedback (split of eigenvalues), there is an optimal amount of mean variations that reverses the detrimental effect of skew-symmetric feedback. This optimal amount corresponds to the eigenvalue diagram where the nominally double eigenvalues are the closest. Decreasing the amount of mistuning from the optimal amount causes one of the modes to become more damped at the expense of the other mode, which becomes less damped. On the other hand, increasing the mistuning beyond the optimal amount causes the frequencies of the two counter-rotating modes to shift without any additional damping augmentation. Figures 20 and 21 illustrate the concepts described above.

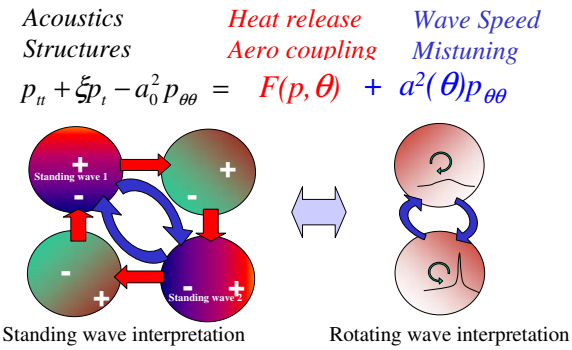


Fig. 20. Harmful and beneficial energy exchange between the rotating waves in engines: the effect of skew-symmetric feedback and symmetry breaking (wave-speed mistuning)

Finally, we comment on the robustness of the method that makes the symmetry breaking feasible for practical applications. The method exploits the dynamics of the problem for the purpose of creating beneficial

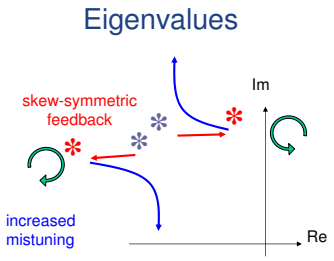


Fig. 21. Harmful and beneficial energy exchange between the rotating waves in engines: the effect of skew-symmetric feedback and symmetry breaking (wave-speed mistuning) on the model eigenvalues and the amplitudes of the rotating waves

energy exchange between traveling waves and thus enjoys several advantages. In particular, the method works by using the heavily damped rotating wave to provide damping augmentation for the lightly damped (or unstable) one, resulting in an overall decrease in the oscillation amplitude. In case of thermoacoustic waves, the method is applicable to general combustion schemes including swirl and bluff-body stabilized combustors. The approach does not require very accurate physics-based dynamic models for unsteady combustion or aero coupling and is robust to many un-modeled physical effects, such as changes in frequency, as long as the modal structure of the problem is approximately preserved.

Let us summarize what the Design of Dynamics means. If natural dynamics of a product needs modification, active or passive control using *external* devices is just one possible solution and often the least desirable one. However, the principles of dynamics and control can be utilized to find a solution that is *internal* to the product. The idea is to find a decomposition of a model of the dynamics into a *system of interacting components* and use the dynamics and control methods to create *beneficial dynamic interactions* between the component. For instance, the control of oscillations using external devices can be realized by interconnection of the lightly damped or unstable mode of the system with a heavily damped external device by choosing appropriate gain and phase of the closed-loop system. The same principle can be used to interconnect a lightly damped or unstable mode of the system with a heavily damped natural mode of the system. For instance, the wave speed mistuning interconnects underdamped traveling waves with heavily damped waves traveling in the opposite direction using the wave-speed perturbation as an interconnecting feedback. Figure 22 illustrates this idea.

7. BARRIERS IN DESIGN OF DYNAMICS

Several barriers are present in a way of introduction of Design of Dynamics into the industrial practice.

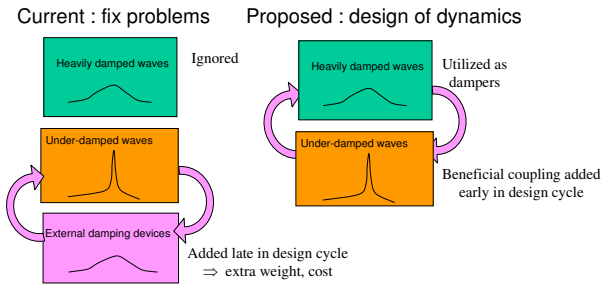


Fig. 22. The idea behind the Design of Dynamics

For the purpose of this paper we will group them into two areas: the first one technical related to an intrinsic difficulty of analysis of the dynamic phenomena and the other mostly social related to the perception of the role of the dynamics and control community.

The technical barriers in mitigation of dynamic problems in products affected by unsteady flow phenomena (such as aerospace and chemical industry) become clearly visible when one realizes that the technical problem amounts to controlling sensitive and complex dynamics in presence of model uncertainty and severe design constraints. Lightly damped structural, acoustic, and fluid dynamic modes easily become underdamped or unstable because of positive feedback coupling with other flow phenomena and excitation by broad band, and tonal flow disturbances. Physics that causes detrimental oscillations involves complex high Mach and Reynolds number flows, which cannot be reliably and accurately computed with current methods. State of the art computational methods based on CFD are too expensive to be applied. Hence model analysis is not utilized to exploit the design space and find innovative dynamics mitigation solutions in early design stage. Accuracy of reduced order models utilized is doubtful especially when chemical reaction, flow separation, and shocks are involved. Because of high uncertainty, models are not utilized for design of robust oscillation mitigation solutions and robustness of chosen solutions to unexpected off-design conditions is not guaranteed. In this paper we partially addressed mitigation of the technical difficulties in modeling and analysis of dynamics by utilizing the symmetry properties of the product.

The social barriers that prevent the Design of Dynamics from becoming industrial practice are related to an underestimation of the full potential of the dynamics and control methods by the design process owners and the dynamics and control experts alike. In this paper we presented three case studies of dynamics and control analysis applied to jet engines technologies and discussed how the social barriers influenced the impact of the technologies on the product.

Let's reverse the negative aspects of the current role of the dynamics and control community outlined in Section 2 in the design process and postulate a new role for these communities. First, they should act *proactively* and play a critical role in early design process when a design flexibility is high and a cost of introduction of solutions to mitigate dynamics problems is the lowest. Second, they should focus on finding a solution that is *internal* to the product, i.e., does not involve extra hardware. This can be accomplished by analysis of the natural dynamics of the problem. The solution of the problem should *exploit* the natural dynamics by creation of beneficial dynamic interactions within the products with minimum external intervention. Third, rather than accepting assumptions from the design process owners, the dynamics and control community should *accept responsibility* for defining the best control architecture in terms of performance and cost. This direction will involve control and dynamics experts learning more of the physics of the problem than they typically accept as necessary. In particular, they will have to drive the modeling and experimental activities rather than be just the users of the results.

We hope that the three case studies presented in this paper will serve its intended purpose of convincing the control and dynamics experts about the benefits of the Design of Dynamics and the attributes they need to have to succeed. However, a much harder problem is how to articulate the benefits of the Design of Dynamics to the current design process owners so they will embrace the expanded role of the dynamics control methods and experts in the early design stages. The social barriers that need to be overcome here should not be underestimated. In fact, in author's experience, these are the most difficult barriers to overcome.

First, the design process owners have a *perception of limited applicability* of dynamics and control methods. Active control solutions are perceived as external to the product, involving extra cost and complexity, and are rarely implementable. As such they are treated as last resources. Moreover, because of the perception that active control methods are only useful for design of an active control system, the idea to use the dynamics and control methods in passive control design does not naturally occur to the design process owners.

Second, a *language barrier* between the dynamics and control community and the design process owners limits ability of the process owners to fully appreciate the potential of the dynamics and control methods to be used in a non-standard way postulated in this paper. The design process owners use the language of the most relevant discipline to the process (physics, chemistry, biology, etc.) and they require that all potential dynamics mitigation technologies be explained in this language. On the other hand, the dynamics and control experts use the language of dynamical systems or control theory often biased towards mathematical rigor and formalism. Inability to adequately articulate the control concepts in the native tongue of the design

process owners will often result in the control concept being rejected.

Third, consider a *lack of credibility* of dynamics and control experts among the early design process owners combined with *territorial behavior* of the latter group. This combination will likely result in a control solution proposed by the dynamics and control experts to be treated with suspicion and possibly rejected. After all, the dynamics and control experts typically are not experts in any of the disciplines considered the most relevant to the product design. To contrary, the design process owners typically are experts in the disciplines most relevant to the product being designed. Why would they even consider solutions proposed by non-experts, when the experts struggle to exploit as much of the domain expertise to solve their problems? Use of a different type of models by the process owners and dynamics and control experts exacerbates the situation further. Low order models typically required by the dynamics and control experts are considered simplistic and inadequate by the design process owners. Using a suspect model as a basis of proposed design modification is a likely reason for a rejection of the proposed modification.

The last issue is a danger of a *competition* between the established design process owners and the dynamics and control experts. When the established process owners insist on solving the problem themselves without external help, they are likely to treat any attempt to bring outside expertise as an unnecessary distraction and a competition for limited resources available to solve the problem. This type of competition or perception of such is never healthy and should be avoided at all cost.

The social barriers mentioned above can be overcome, but the process of doing so is lengthy, difficult, frustrating, and fragile. In fact, it often fails. Here are some necessary conditions for increasing the role of dynamics and design communities in the early design process.

First, it is necessary to reach the state when a *perception of inadequacy* of the current design process is widely accepted among the process owners and their management. Such a perception is typically a result of a major *crisis* in a product design process. When the current design process is widely acknowledged to be faulty, the technical design process owners and even more so their managers become more open to a control solution. This is a best point of entry for the dynamics and control experts to get involved.

At this point it is important that the control and dynamics experts learn as much as possible about the scientific disciplines most relevant to the problem and their relationship with the internal dynamics of the product. They also have to show *commitment to work towards solving the problem* using the simplest possible means, including passive methods and exploitation of the natural dynamics, and avoid the trap of pushing for an active control solution at all cost. In this way they will present themselves as *team players*.

Along the way the dynamics and control experts need to show some partial *successes*. This can be accomplished utilizing traditional strengths that the dynamics and control communities exhibit, such as an ability to extract a low order model of a process directly form experimental data or assess the validity of a given physics-based model. Utilizing the rigor of mathematics is a great value, but it has to be balanced with translation to the language of physics tom be convincing.

Last but not least, just demonstrating the technical progress is not sufficient. It is extremely important that the dynamics and control experts use every opportunity to *educate* the process owners on all levels about the basic principles of dynamics and control theories and show why these principles are relevant to solving the problem at hand. Ability to translate the physics of the problem the the language of dynamics, finding solution in the dynamics domain, and translating the dynamic solution back to the language of physics will go particularly long ways towards eliminating the language barrier. The best possible outcome of this educational process is convincing the design process owners that the analysis methods of dynamical systems and control theory are just alternative efficient ways of exploiting the physics of the problem. This behavior will help establish the *credibility* with the design process owners and their management.

8. CONCLUSION

We advocated the need to change the role of dynamics and control community from fixing problems related to the detrimental dynamics using active control to the design for beneficial dynamics early in the design cycle. The paper summarized lessons learned in industrial research on mitigation of flow and structure oscillations in jet engines (thermoacoustic instabilities and turbomachinery flutter). We showed how the decisions on the control system architecture (sensor and actuator location) impacted the achievable level of suppression of oscillations (fundamental limitations of performance). Attempts to introduce control late in the design process and without proper attention to control architecture often fail because of high cost to modify the design to add on active control. We also showed how certain aspects of design (symmetry) contribute to the origin of detrimental oscillations and point out how the dynamical systems and control theory methods can guide the design to prevent the oscillations. The control and dynamics methods used early in design allow one to manipulate the physical feedback loops in the system to create beneficial dynamics and exploit design flexibility at low cost. To increase impact of experts in control and dynamics on the design process, the experts need to establish

credibility in the technical community that owns the design process. In industrial environment this can be accomplished by playing a key role in a response to a crisis, and following up with teaching of basic principles of dynamics and control to the design community and their management.

ACKNOWLEDGEMENT

This work was partly supported by AFOSR grants F49620-01-C-0021 and FA9550-04-C-0042, which is gratefully acknowledged. We also acknowledge help and support of a turbomachinery expert Dan Gysling and combustion engineers Jeff Cohen and William Proscia. The control engineers who contributed significantly to the flutter and combustion control projects are Clas Jacobson and Gonzalo Rey. The test engineer Karen Teerlinck contributed to the flutter experiments. Collaboration with Igor Mezic from UCSB influenced various elements of the projects. More specifically, methods developed by Igor Mezic (Mezic and Banaszuk, 2004) led to identification of regions of validity of linear and nonlinear models for thermoacoustic oscillations. Moreover, Igor Mezic analysis of the impact of the symmetry structure of DNA molecules on DNA dynamic behavior (Mezic, 2005) provided an inspiration to the authors of the current paper that led to a discovery of the beneficial and detrimental symmetry patterns in jet engines. This key inspiration ultimately led to the concept of the Design of Dynamics for the wave phenomena in jet engines described in this paper. Mihai Huzmezan provided valuable suggestions how to improve the readability of this paper.

REFERENCES

Anson, B., I. Critchley, J. Schumacher and M. Scott (2002). Acyive control of combustion dynamics for lean premixed gas fired systems. In: *ASME Paper GT-2002-30068*. American Society of Mechanical Engineers.

Banaszuk, Andrzej, Clas A. Jacobson, Alex I. Khibnik and Prashant G. Mehta (1999*a*). Linear and nonlinear analysis of controlled combustion processes. part i: Linear analysis. In: *1999 Conference on Control Applications*. Hawaii.

Banaszuk, Andrzej, Clas A. Jacobson, Alex I. Khibnik and Prashant G. Mehta (1999*b*). Linear and nonlinear analysis of controlled combustion processes. part ii: Nonlinear analysis. In: *1999 Conference on Control Applications*. Hawaii.

Banaszuk, Andrzej, Gonzalo Rey and Daniel Gysling (2002*a*). Active control of flutter in turbomachinery using off blade actuators and sensors. part i: Modeling for control. In: *Proceedings of 15th Triennial World Congress of IFAC*. Barcelona, Spain.

Banaszuk, Andrzej, Gonzalo Rey and Daniel Gysling (2002*b*). Active control of flutter in turbomachinery using off blade actuators and sensors. part ii: Control algorithm. In: *IEEE Conference on Decision and Control*. Las Vegas, NV.

Bendiksen, O. O. (2000). Localization phenomena in structural dynamics. *Chaos, Solitons, and Fractals* **11**, 1621–1660.

Bloxsidge, G.J., A.P. Dowling, N. Hooper and P.J. Langhorne (1987). Active control of acoustically driven combustion instability. *Journal of Theoretical and Applied Mechanics* **6**, 161–175. special issue, supplement to Vol. 6.

Bloxsidge, G.J., A.P. Dowling, N. Hooper and P.J. Langhorne (1988). Active control of reheat buzz. *AIAA Journal* **26**, 783–790.

Chu, Y. C., A. P. Dowling and A. P. Glover (1998). Robust control of combustion oscillations. In: *Procs. of the IEEE Conference on Control Applications*. Trieste, Italy. pp. 1165–1169.

Cohen, J. M., N. M. Rey, C. A. Jacobson and T. J. Anderson (1998). Active control of combustion instability in a liquid fueled low NOx combustor. In: *1998 ASME Gas Turbine and Aerospace Congress*. ASME.

Cohen, J.M. and A. Banaszuk (2003). Factors affecting the control of unstable combustors. *Journal of Propulsion and Power* **19**, 811–821.

Dullerud, G. and R. D'Andrea (1999). Distributed control of inhomogeneous systems with boundary conditions. *Proceedings of the 38th IEEE Control Conference on Decision and Control* pp. 186–190.

Evesque, S., A. P. Dowling and A. M. Annaswamy (2000). Adaptive algorithms for control of combustion. In: *Procs. of the NATO/RTO Active Control Symposium*. Braunscweig, Germany.

Fleifil, M., A.M. Annaswamy, J.P. Hathout and A.F. Ghoniem (1997). The origin of secondary peaks with active control of thermoacoustic instability. In: *Proceedings of the AIAA Joint Propulsion Conference, Seattle 1997*.

Forsching, H. (1984). Aeroelastic stability of cascades in turbomachinery. *Prog. Aerospace Sci.* **30**, 213–266.

Freudenberg, J.S. and D.P. Iooze (1987). A sensitivity tradeoff for plants with time delay. *IEEE Trans. Automatic Control* **AC-32**, 99–104.

Gelb, A. and W.E. Vender Velde (1968). *Multiple–Input Describing Functions and Nonlinear System Design*. McGraw–Hill.

Graf, M. B., T. S. Wong, E. M. Greitzer, F. E. Marble, C. S. Tan, H. W. Shin and D. C. Wisler (1998). Effects of nonaxisymmetric tip clearance on axial compressor performance and stability. *Transactions ASME Journal of Turbomachinery*.

Hagen, G. (2004). Absolute stability of a heterogeneous semilinear dissipative parabolic pde. *Submitted to the 43rd IEEE Conference on Decision and Control*.

Hagen, G. and A. Banaszuk (2004). Symmetry-breaking and uncertainty propagation in a reduced order thermo-acoustic model. In: *2004 Conference on Decision and Control*. Bahamas.

Hathout, J. P., A. M. Annaswamy and A. F. Ghoniem (2000). Modeling and control of combustion instability using fuel injection. In: *Procs. of the NATO/RTO Active Control Symposium*. Braunscweig, Germany.

Hibshman, J.R., J.M. Cohen, A. Banaszuk T.J. Anderson and H.A. Alholm (1999). Active control of combustion instability in a liquid-fueled sector combustor. In: *1998 ASME Turbo Expo*. ASME.

Hoffmann, S., G. Weber, H. Judith, J. Hermann and A. Orthmann (1998). Application of active combustion instability control to siemens heavy duty gas turbines. In: *RTO-MP-14, Symposium of the AVT panel on gas turbine engine combustion, Emission and alternative fuels*. AGARD, RTA. Neuilly-Sur-Seine.

Jovanovic, M. R., B. Bamieh and M. Grebeck (2003). Parametric resonance in spatially distributed systems. *Proceedings of the 2003 American Control Conference, Denver, CO.* pp. 119–124.

Langhorne, P.J., A.P. Dowling and N. Hooper (1988). Practical active control system of combustion ocsillations. *Journal of Propulsion* **6**, 324–333.

Mehta, P., A. Banaszuk and M. Soteriou (2004). Impact of convection and diffusion processes in fundamental limitations of combustion control. In: *2nd AIAA Flow Control Conference*. Portland, OR.

Mezic, I. (2005). Dynamics and control of large-scale molecular motion. In: *Proceedings of 2005 IFAC World Congress*.

Mezic, I. and A. Banaszuk (2004). Comparison of systems with complex behavior. *Physica D, Nonlinear Phenomena* **197**, 101–133.

Paschereit, C., E. Gutmar and W. Weisenstein (1999). Control of combustion driven oscillations by equivalence ratio modulations. In: *ASME Paper 99-GT-118*. American Society of Mechanical Engineers.

Peracchio, A.A. and W. Proscia (1998). Nonlinear heat release/acoustic model for thermoacoustic instability in lean premixed combustors. In: *1998 ASME Gas Turbine and Aerospace Congress.* ASME.

Petrov, E. P., R. Vitali and R. T. Haftka (2000). Optimization of mistuned bladed discs using gradient-based response surface approximations. *AIAA-2000-1522*.

Rey, Gonzalo, Andrzej Banaszuk and Daniel Gysling (2003). Active control of flutter in turbomachinery using off blade actuators and sensors. part iii: Experimental demonstration. In: *Proceedings of AIAA Conference*. Reno, NV.

Richards, G., M. Yip, E. Robey, L. Cowell and D. Rawlins (1995). Combustion oscillation control by cyclic fuel injection. In: *ASME Paper 95-GT-224*. American Society of Mechanical Engineers.

Rivas-Guerra, A. J. and M. P. Mignolet (2003). Local/global effects of mistuning on the forced response of bladed disks. *Journal of Engineering for Gas Turbines and Power* **125**, 1–11.

Sattinger, S., Y. Neumeier, A. Nabi, B. Zinn, D. Amos and D. Darling (1998). Sub-scale demonstration of the active feedback control of gas-turbine combustion instabilities. In: *ASME Paper 98-GT-258*. American Society of Mechanical Engineers.

Saunders, W.R., M.A. Vaudrey, B.A. Eisenhauer, U. Vandsburger and C.A. Fannin (1999). Perspectives on linear compensator designs for active combustion control. In: *AIAA paper 2000-0717, 37th AIAA Aerospace Sciences Meeting, Reno, January 1999*. AIAA.

Seron, M.M., J.H. Braslavsky and G.C. Goodwin (1997). *Fundamental Limitations in Filtering and Control*. Springer. New York.

Seume, J.R., N. Vortmeyer, W. Krause, J. Hermann, C.-C. Hantschk, P. Zangl, S. Gleis, D. Vortmeyer and A. Orthmann (1997). Application of active combustion instability control to a heavy duty gas turbine. In: *ASME Paper 97-AA-119, Proc. of ASME Asia '97 Congress and Exhibition, Singapore, October 1997*. ASME.

Shapiro, B. (1998). A symmetry approach to extension of flutter boundaries via mistuning. *Journal of Propulsion and Power* **14**(3), 354–366.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DISTRIBUTED DECISION MAKING IN SUPPLY CHAIN NETWORKS

B. Erik Ydstie, Kendell R. Jillson, and Eduardo J. Dozal-Mejorada

Carnegie Mellon Department of Chemical Engineering
Pittsburgh, PA 15213

Abstract: The supply chain system is modeled as a "Value Added Network" (VAN) which performs the following tasks: assembly, storage, routing, processing and transportation. Many activities interact and complexity increases as the number of business activities and links in the VAN increase. In order to develop a model which can deal with changing market conditions and evolving technology it is necessary to adapt as the supplier and demand structures change. Recent developments focus on decentralized business structures and software solutions to reduce complexity and maintain scalability. It has been claimed that decentralized decisions lead to sub-optimal solutions. We show that this is not necessarily so. We present novel abstraction of an integrated system of decision makers, software and physical devices which allows for optimal decentralized decision making. The objective function captures the idea that investment and resource use decisions in a VAN (capacity expansion expansion, how much inventory to carry, which markets to address and which technology to use) carries value. The decentralized decision making processes we cover may be quite complex and may include local feedback corrections as well as decentralized, optimal (model predictive) strategies.

Keywords: Distributed control, supply chain management, self-optimization, optimal control, inventory control, flow control.

1. INTRODUCTION

Information Technology (IT) tools can be used to improve the resource allocation, flow of materials and diffusion of knowledge within companies and entire enterprises. Enterprise Resource Planning (ERP) systems supplied by companies like SAP, i2, Oracle, J.D. Edwards and others integrate business processes, streamline production systems and provide company wide access to information related to critical work processes [13]. Such systems can also be used to track inventory levels, identify bottle necks, smooth flows and evaluate performance. Impressive gains have been reported in a great variety of industries, including the computer industry (hardware and software), discrete parts manufacture and commodity chemicals [1,4,12,14,11].

The application of ERP tools has made it apparent that it does not suffice to focus on the internal processes alone. Upsets are often created by factors beyond the control of a single company. This led to the development of Advanced Planning Systems (APS) that link the database capabilities of the ERP system to market forecasts and process models. Such tools enable a company to evaluate scenarios and respond to changes in the market

place by applying feedforward and planning. However, it is clear that improved agility and better performance can be achieved by application of active feedback and tools from process systems theory, like distributed control and real time optimization [8].

In our context a supply chain is thought of as a "network of organizations that are involved, through upstream and downstream linkages, in different processes and products [3]." The objective of the supply chain is to *create value* through a sequence of operations which we refer to as activities. Such activities include assembly, storage, routing, processing and transportation. In this context Stadtler and Kilger [13] define Supply Chain Management (SCM) as "the task of integrating organizational units along a supply chain and coordinating materials, information, and financial flows in order to fulfill (ultimate) customer demands with the aim of improving competitiveness of a supply chain as a whole." The SCM perspective therefore includes the idea of two or more legally separate partners working together towards a common goal within a business sector.

A number of models of supply chain systems have been developed. Recently, control theory methods have been introduced to manage and adapt flows within the supply chain so that it remains competitive in the market place. For example, a centralized Model Predictive Control (MPC) framework for optimization of supply chains was developed in [9,8]. These models are suitable in static systems where the models and boundary conditions do not change.

In the current paper we are interested in models that are flexible, adaptive and self-optimizing. This approach leads to the study of structural properties, stability and optimality using distributed feedback in lieu of centralized planning. The study of industrial dynamics and feedback control was advanced further by Forrester [5] who elucidated an instability in supply chains referred to as demand amplification. His ideas on feedback loops and systems theory formed the basis for very fruitful developments that continue to have a significant impact to this day [10].

2. SUPPLY CHAIN NETWORKS

A supply chain system (SCS) is an integrated network of activities which transports and transforms assets so that their intrinsic values [2] are maximized. An asset may be a tangible product like a gallon of oil, a piece needed in an assembly line or an intangible item like an order, information or intellectual property. An increase in value may be the result of an asset been transported to a location closer to the customer, a transformation (e.g. chemical reaction or assembly) or because time progresses and market parameters change. The objective of this section is to describe the conservation laws that constrain the dynamic behavior of assets in the supply chain. In the next section we introduce the value function.

Consider an SCS with n distinct assets a_i. The index i identifies an asset by its name, SKU-number, chemical composition or some other index which should be unique. The asset space $A = \{a_i\}$ defines the nature of the business. The amount (inventory) of each asset is given by a non-negative real number $v_i(x,t)$, where x denotes the location and t denotes the time. The vector of inventories is represented by the vector $v^T = (v_1, ..., v_n)$.

The topology of the SCS is represented by the graph $\mathcal{G} = \{\mathcal{H}, \mathcal{A}\}$. $\mathcal{H}$ represents the set of edges, along which we allow assets to flow, while $\mathcal{A}$ represents vertices where assets are stored, transformed, shipped or routed. A non-empty collection of edges and vertices is called an *activity*.

We find it sometimes useful to introduce a little more structure and distinguish among four different classes of activities. These include *transportation, manufacture, storage, terminals (shipping/receiving)* and *routing*. This additional structure allows us to define the Supply Chain Graph [1]:

$$\mathcal{G} = \{\mathcal{H}, \underbrace{\mathcal{M}, \mathcal{S}, \mathcal{T}, \mathcal{R}}_{\mathcal{A}}\}$$

Elements, $h_i \in \mathcal{H}, i = 1, ..., n_h$ denote the transportation of assets. Elements $m_i \in \mathcal{M}, i = 1, ..., n_p$ represent the manufacturing with assembly or disassembly of chemical constituents or parts into pre-cursors and products. Elements $s_i \in \mathcal{S}, i = 1, ..., n_s$ denote the storage facilities. Elements $t_i \in \mathcal{T}, i = 1, ..., n_t$ denote terminals for receiving and shipping. Elements $r_i \in \mathcal{R}, i = 1, ..., n_r$ represent points where material, energy, money and data can be routed in different directions.

Example 1. Consider the production facility shown in Figure 1. There is a terminal where materials are received from the supplier. There are storage locations for raw materials and products next to the terminal, an assembly plant, storage for finished products and a shipping terminal. All nodes are connected by edges representing flow of assets. More vertices and edges can be added to represent flow of services, orders, information, capital and energy. There are two routing points in this figure. At routing point 1 decisions are made

[1] The notation and order has been chosen in memory of our beloved hamster TicTac

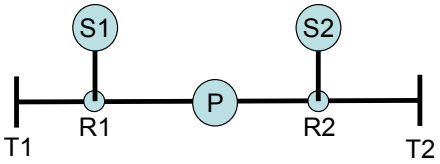


Fig. 1. Graph of an activity in a supply chain system consisting of two terminals, two storage locations, one production facility and six transportation links.

about sending raw materials to storage (Storage 1) or production. At routing point 2 decisions are made about sending finished products to the plant warehouse (Storage 2) or shipping.

We now develop the conservation laws that govern the transformation and flow of assets.

(1) *Transportation:* Asset flow is represented using the hyperbolic, partial differential equation
$$\frac{\partial \rho}{\partial t} + \frac{\partial f}{\partial x} = 0$$
This equation model describes a pipeline where $\rho(x,t)$ is the local "density of an asset" at the point x and time t while $f(x,t)$ is the local flow rate.

(2) *Manufacture:* Manufacture is represented as a source or sink so that
$$f_i = p$$
where p is rate of production/consumption of an asset. This notation allows us to model transformation of assets via assembly or disassembly.

(3) *Storage:* The rate of change in storage is given by the differential equation
$$f_i = \frac{dv}{dt}, \qquad v(0) = v_0$$
where v_0 is the amount stored initially. Negative f_i denotes flow out of storage whereas positive f_i denotes flow into storage. This type of storage, referred to as "tanque pulmon", represents a capacitor in an electrical network.

(4) *Terminals:* Applying the conservation law to the terminal gives
$$0 = f_i + f_T \tag{1}$$
where f_T is the shipping/receiving rate. Receiving is positive and shipping is negative.

(5) *Routing points:* Asset flow through routing points, like terminals, is conserved. We therefore have
$$0 = \sum_{\text{Connections}} f_i \tag{2}$$
The summation is carried out so that the index i ranges over all edges connected to the corresponding routing point in the network.

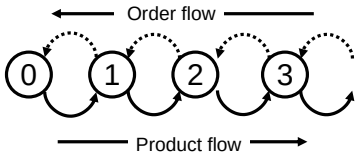


Fig. 2. Three echelon supply chain network, representing retailer, distribution center, warehouse, and production center.

An activity is an arbitrary collection of the basic building blocks. By combining building blocks and eliminating the internal flows we see that the dynamics of activities are represented by the inventory balance
$$\frac{dv}{dt} = \sum_{\text{Terminals}} f_{T,i} + p \tag{3}$$
where
$$p = \sum_{\text{Production sites}} p_i$$
It is often convenient to use projected and transformed variables so that $\bar{v} = Tv$ where T is a linear operator. T is often non-square and projects the high dimensional asset space into a lower dimensional space. It is possible to let T be a differential operator (to allow prediction) and/or an integral operator (the Fourier-Laplace transforms for example).

Example 2. In the last decade the world changed from a marketplace with several large independent markets to a highly integrated global market that demands a large variety of products and services complying with high quality, reliability and environmental standards. Furthermore, the fast development of new products as well as customer focus and increasing competitiveness pose new challenges in the area of modeling and control of global supply chain systems. Here we will develop a model control technique carried out in cooperation with Unilever.

The problem we consider is illustrated in Figure 2. This system has three echelons corresponding to the retailer, distribution center and plant warehouse. There are two classes of flows, one corresponding to the flow of orders and the other the flow of goods in response to the demand. There is only one product is this example.

For the flow and storage of goods we have
$$\frac{dI_i}{dt} = f_{i-1,i} - f_{i,i+1}, \qquad i = 1,2,3$$
where I_1 is the inventory in the plant warehouse, I_2 is the inventory in the distribution center and I_3 is the inventory at the retailer. We develop a similar equation for the order flow so that
$$\frac{dO_i}{dt} = f^o_{i-1,i} - f^o_{i,i+1}, \qquad i = 1,2,3,4$$

Where O_0 is the backlog of orders in the plant, O_1 the backlog for the plant warehouse, O_2 is the backlog in the distribution center and O_3 is the back at the retail level.

The objective is to ensure a high level of service at the retail level. In this example we will work on the basis that we should have $f_{3,4} = f^o_{3,4}$ indicating that the demand is satisfied exactly. We furthermore want to achieve this objective without carrying too high inventory anywhere in the supply chain system.

There are 4 inventory flows and 3 inventories, 4 order flows and 4 order levels in this problem. So there are 15 variables. It follows that we have $15-7 = 8$ degrees of freedom. These correspond to the flows that must be managed. According to the objectives we would like to manage these flows so that inventories and back-orders follow setpoints so that

$$\begin{aligned} O_i &= O_i^*, \qquad i = 1,2,3 \\ I_i &= I_i^*, \qquad i = 1,2,3 \end{aligned}$$

Ideally we would like to use $O_i^* = I_i^* = 0$ indicating the the inventory and order levels are equal to zero [2].

Starting at the retailer level we see that we should use the following inventory controller for the order buffer,

$$f_{3,4} = f^o_{3,4} + K_O(O_3 - O_3^*)$$

This means that the we deliver product at the rate of incoming orders plus a proportionality constant times the size of the back-orders. This distributed control policy quickly converges so that $f_{3,4} = f^o_{3,4}$ indicating that we deliver at the same rate as the orders come. If we want to track a specific order then the average delay in fulfilling the order is given by the number

$$d = O/f$$

The policy can only be implemented if there is sufficient material in storage to fulfill the orders at the rate given by the inventory controller. In order to ensure that the inventory is also controlled we need to use another inventory controller for the retail storage. Ideally we would like to set

$$f_{2,3} = f_{3,4} - K_I(I_3 - I_3^*)$$

This means that we use a combination of feedforward and feedback control to manage the inventory.

However, this method cannot be used exactly as indicated since the retailer does not control the rate of arrivals directly. The retailer has to send an order to the distribution center and wait until the order is fulfilled. The controller for inventory therefore becomes

$$f^o_{2,3} = f_{3,4} - K_I(I_3 - I_3^*)$$

Indicating that orders are sent to the distribution center at the same rate that material is shipped plus a term which is proportional to current inventory level.

The distribution centers and plant warehouse use a similar policy. We will assume for now that the production plant is very responsive so that we can set

$$f^o_{0,1} = f_{0,1}$$

Indicating that the plant warehouse is re-stocked as soon as an order is sent.

Applying these idea to the entire supply chain gives

$$\begin{aligned} f_{i,i+1} &= f^0_{i+1,i} + K_O(O_i - O_i^*), \qquad i = 0,1,2 \\ f_{i-1,i} &= f^0_{i-1,i} - K_I(I_i - I_i^*), \qquad i = 0,1,2 \end{aligned}$$

There are seven of these controllers so there is now one degree of freedom, corresponding to the demand rate at the retail level, which acts as a disturbance. This effect can be seen by developing the closed loop expression for the supply chain. The inventories are seen to satisfy the expression

$$\frac{dI_i}{dt} = -K_I(I_i - I_i^*) + \Delta_i(t)$$

where

$$\Delta_i(t) = f^0_{i-1,i} - f_{i-1,i}$$

represents the discrepancy between the order rate and supply rate to node i. If this is equal to zero then the supply chain system is stable. If this number is not equal to zero then the supply chain dynamics may exhibit instabilities, and disturbances may even be amplified causing bullwhip.

The problem we consider is how to manage the flow through, routing points, terminals, storage and production sites so that assets flow through the system and are distributed in the best manner. In order to solve this problem we must assign values to the assets as functions of time and location in the SCS.

3. VALUE ADDED NETWORKS

The instantaneous profit is the difference between the revenues from sales and the activity costs:

$$P = R - C \tag{4}$$

This measure is also called the *rate of accounting earnings*. Integrated and discounted over time into

[2] Walmart is a company that has been able to move in this direction by elimination of distribution centers.

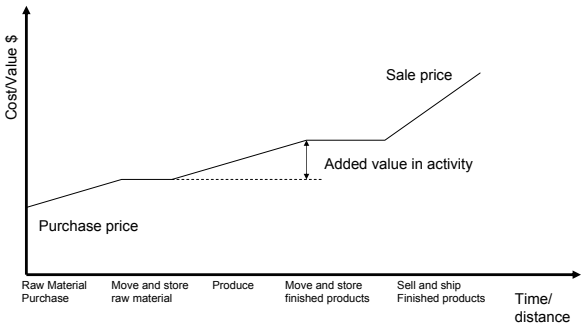


Fig. 3. The generation of value through the production process.

the future the expected accounting earnings gives an indication of the performance of the SCS.

Approaching the supply chain management problem from the point of view of maximizing the discounted income in a distributed network of activities in this way results in a type of analysis, called activity based analysis [6].

We now need to make some basic assumptions about the supply chain system.

Assumption 1. *Consider a supply chain system.*

(1) The inventory of assets represents the state of the system.
(2) There a exists a positively homogeneous degree one function $A(v)$ which defines the value of the assets.
(3) Any activity cost is positive.

The first assumption provides the concept of state. The basic idea here is that the state of a company can be defined by determining the magnitude of its assets. The second assumption implies that the value of the company (for example the discounted cash value) can be expressed in terms of its current state and that it is a homogeneous function. The third assumption states that all activities cost something. The cornerstone for our developments is then given by the Legendre-Fenchel dual

$$A^*(c) = \max_v (A(v) - c^T v) \tag{5}$$

The vector c represents the value of adding one unit of the corresponding asset to the inventory at location x. We see that c represents the Lagrange multiplier corresponding to the inventory vector.

Euler's theorem for homogeneous functions gives

$$A = v^T c = \sum_{\text{Assets}} v_i^T c_i \tag{6}$$

We connect this equation with the inventory balances (3) which define the dynamics of the SCS process. First we note the following orthogonality relationship

$$v^T dc = 0 \tag{7}$$

which is referred to as the Gibbs-Duhem equation. By differentiating $A(v)$ we have

$$\frac{dA}{dt} = c^T \frac{dv}{dt}$$

We now use equation (3) with equation (7) to give

$$\frac{dA}{dt} = c_r^T r - c_p^T s + p_A \tag{8}$$

where c_r and c_p represent the per unit value of the resource and the product and the variable

$$p_A = \underbrace{(c - c_r)^T r - (c - c_p)^T s}_{\text{Transportation}} + \underbrace{c^T p}_{\text{Production}} \tag{9}$$

represents the activity cost. The activity cost is positive (in accordance with Assumption A3. This leads to the following important result.

Lemma 1. *The value $A(v)$ is concave.*

Proof. Follows from the homogeneous degree one property and positivity of p_A. □

The result is important since it shows that value based analysis in supply chain systems can be approached using convex optimization.

The cost is the sum of the cost of resources and activities so that

$$C = p_A + c_r^T r$$

Combining this expression with equations (4) and (8) gives

$$\frac{dP}{dt} = R - \left(\frac{dA}{dt} + c_p^T s \right) \tag{10}$$

By using the definition for the income from sales we get

$$\frac{dP}{dt} = \sum_{\text{Sales}} (c_{s,i} - c_{p,i}) s_i - \frac{dA}{dt} \tag{11}$$

In this expression we let $c_{s,i}$ denote the ith component of the vector c_s and $c_{p,i}$ denote the ith component of the vector p_s. We note that $c_{s,i}$ denotes the sales price whereas as $c_{p,i}$ denotes the price "at cost" for item i. We therefore have

$$c_{s,i} - c_{p,i} = \begin{cases} > 0, & \text{sell @ profit} \\ = 0, & \text{sell @ cost} \\ < 0, & \text{sell @ loss} \end{cases}$$

In the case $c_s - c_p = 0$ there is no mark-up. This is often the case for internal customers and the cost c_p is then referred to as a "transfer price".

Expressions (9) and (11) highlight the main issues in supply-chain management [3]:

(1) The profit increases at a faster rate when the markup $c_s - c_p > 0$ is large and sales volume high. Larger markup can be achieved by raising the per unit price of the item sold. But

[3] There can be a considerable phase shift between the movement of goods and the associated financial transaction. Ignoring this phase shift is referred to as accrual.

higher prices also tend to give reduced sales. This expression emphasize the importance of marketing and sales.

(2) The profit can be increased by reducing current inventory and fixed assets since we have $dA/dt < 0$. This expression emphasizes the importance of being "lean" [14].

(3) The transportation and production costs as defined in (9) should be minimized. This expression emphasizes the importance of planning, scheduling and process control and new process technology.

The added value of a path consisting of several activities is given by the formula

$$w = \sum_{\text{Segments}} w_i \tag{12}$$

where w_i represents the added value of each subactivity. This number may be positive, zero or negative and it does not depend on the path taken since the function $A(v)$ is unique. It follows that for a cyclical activity we have

$$0 = \sum_{\text{Loop}} w_i \tag{13}$$

This expression conveys the idea that there is no value in a cyclical activity. However there is cost associated with every activity, and cyclical activities therefore add cost but no value.

Just like for the conservation laws (3), it is convenient to project or transform the activity costs $\bar{w} = Lw$, and $\bar{c} = Lc$. Equation (13) still holds if these transformations are linear.

We now have the following extremely important result for transportation, storage and production in an SCS.

Theorem 1. *Consider an supply chain network with with linear network operators T and L. We have*

$$\sum_{storage} \frac{d\bar{v}}{dt}^T \bar{c} = \sum_{transportation} \bar{f}^T \bar{w} + \sum_{production} \bar{p}^T \bar{c} + \sum_{terminals} \bar{c}^T \bar{f}$$

Proof. See [7] □

This result expresses the interesting fact that the spaces of inventories and are cost variables are orthogonal.

4. OPTIMALITY OF DECENTRALIZED DECISION MAKING

The problem we want to solve is how to stabilize the dynamics and balance the load in the supply chain while maximizing the intrinsic value. The discussion given above shows that we can formulate this problem so that

$$\min_{f_i, p_i} \sum_{i=1}^{M} A(v_i)$$

subject to equations (3) and (5). In centralized decision making all information is collected and the problem is solved using all available information. In decentralized decision making the problem is solved by distributing computational effort amongst the node points. In either case we want to implement the strategies using feedback laws of the type

$$f = \hat{f}(w), p = \hat{p}(c)$$

where $\hat{f}, \hat{p}$ determine the transportation and production rates as functions of the cost.

In order to develop production schedules that balance system load, we need to evaluate the activity costs and their sensitivity with respect to changes in the activity rate. In the simplest case this may be a linear function with a downward trend to reflect discounts for larger volumes. Let Δ be the difference operator so that for any variable z we have $\Delta z = z_2 - z_1$.

Definition 1. *An activity is said to be* positive *if for any f_1 and f_2*

$$\Delta f \Delta w \geq 0$$

and for any p_1 and p_2

$$\Delta p \Delta c \geq 0$$

Positive rate ensures that the cost of a given activity does not increase with increasing traffic. Examples include the barrier function, which describes capacity constraints, gradient directions that result from optimization of convex cost functions and more generally any cost which is monotonic in the sense that higher added value gives incentive to larger shipments. We may for example have

$$f = 0, \text{ if } w < w_{\min} \text{ and } f = f_{\max} \text{ otherwise}$$

and

$$p = 0, \text{ if } c < c_{\min} \text{ and } p = p_{\max} \text{ otherwise}$$

In this case there is no activity if the value added is below a certain threshold and we operate at maximum capacity otherwise.

The activity costs are used to solve load balancing and resource allocation problems since they show how the cost varies with respect to production volume. Without such costs load balancing is not a well posed problem.

We now show that the decentralized policy solves the optimal control problem. We proceed in two steps. We first show that the decentralized control

system is stable and converges to a unique solution provided the boundary conditions are fixed. We then show that the obtained stationary point is optimal.

Theorem 2. *Consider an enterprise network with fixed boundary costs and positive feedback controls. The inventories are then stable and converge to stationary values.*

Proof. Details given in full length paper available from the authors. □

Theorem 3. *Consider an enterprise network with fixed boundary costs and positive feedback controls. The total activity cost is then minimized.*

Proof. Details given in full length paper available from the authors. □

These two theorems show that there exists a unique, stationary solution to the enterprise network. This solution, furthermore is stable and optimized under decentralized control policies. The concavity result given in the previous section shows that optimum is global due to the concavity of A.

5. SUMMARY AND CONCLUSIONS

Distributed decision making in supply chain systems arises naturally in several ways: The systems we want to model are distributed since process segments, business units and enterprizes are integrated into a complex, diverse and highly dynamic global market. Information, physical infrastructure and human resources are distributed across the globe and the computer networks we use for information exchange are also distributed. It is often thought that decentralized decision making is sub-optimal. In this paper we show that this not necessarily the case. Optimal and stable decision making processes can be constructed when we modeled the SCN as a VAN with assembly, storage, routing, processing and transportation. The decentralized decision making processes may be quite complex and may include local feedback corrections as well as decentralized, optimal (MPC) strategies. The use of distributed decision making allows the topology of the network to change and adapt as new needs arise. Old subsystems can be exchanged with newer ones, new products and processes can be brought on line and new businesses can be added or old ones closed without changing the overall management strategy.

REFERENCES

[1] Backx, T., O. Bosgra and W. Marquardt (1998), "Towards intentional dynamics in suplly chain conscious process operations", *Proc. 3rd Int. Conf. on Foundations of Computer Aided Process Operations, AIChE* , p. 5

[2] Buffett, W. (2002), "Berkshire Hathaway, Shareholder Letters", `www.berkshirehathaway.com/letters.html`

[3] Christopher, M. (1998), *Logistics and Supply Chain Management - strategies for reducing cost and improving service, 2nd Ed.*, London et al. UK.

[4] Ettl, M., G.E. Feigin, G. Y. Lin and D.D. Yao (1996), " A supply network model with base stock control and service requirements ", *IBM Research Report* , RC 20473 (06/04/96) Computer Science/Mathematics

[5] Forrester, J. W. . (1979), *Industrial Dynamics*, MIT Press, Cambridge MA

[6] Helfert, E. A. (2001), *Financial Analysis Tools and Techniques*, McGraw Hill, New York

[7] Jillson, K. R. and Ydstie, B. E.(2005), Complex Process Networks: Passivity and Optimality, *IFAC Triennal World Congress*, Prague, July 2005.

[8] Perea Lopez, E. (2002), *Centralized and decentralized approaches for Supply Chain Management based on Dynamic Optimization and Control*, PhD Thesis, Carnegie Mellon

[9] Perea Lopez, E., I, E. Grossmann and B.E. Ydstie (2000), "Towards the integration of dynamics and control for supply chain management ", *Computers and Chemical Engineering*, Vol 24, pp 1143-1150.

[10] Sterman, J. D. (2000), *Business Dynamics: Systems Thinking and Modeling for Complex World*, McGraw Hill, NY.

[11] Sery, S., V. Presti and D. E. Shobrys (2001), "Optimization models for restructuring BASF North America's distribution system ", *Interfaces*, Vol 31, pp 55-65.

[12] Selcuk Erenguc, S, N.C. Simpson and A .J. Vakharia(1999), "Integrated production/distribution planning in supply chains: an invited review", *European Journal of Operations Research* Vol 115, pp. 219- 236

[13] Stadtler, H. and C. Kilger (2000), *Supply Chain Management and Advanced Planning* , Springer Verlag, Berlin.

[14] Taylor, D. and D. Brunt (2000), *Manufacturing operations and supply chain management: The lean approach*, Thomson Learning, London.

Session 7.1

Optimization and Design Applications

Scheduled Optimization of an MMA Polymerization Process

R. Lepore, A. Vande Wouwer, M. Remy, R. Findeisen, Z. Nagy, and F. Allgöwer,
Faculté Polytechnique de Mons, University of Stuttgart

Opportunity for Real-Time Optimization In A Newsprint Mill: A Simulation Case Study

A. Berton, M. Perrier, and P. Stuart
École Polytechnique de Montréal

Product Design via PLS Modeling: Stepping Out of Historical Data into Unknown Operating Space

N. Lu, Y. Yao, and F. Gao, *Hong Kong University of Science and Technology*

Adaptive Control of Bromelain Precipitation in a Fed-Batch Stirred Tank

F. V. da Silva, R. L. A. dos Santos and A. M. F. Fileti
University of Campinas

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

SCHEDULED OPTIMIZATION OF AN MMA POLYMERIZATION PROCESS

R. Lepore [1] A. Vande Wouwer M. Remy
R. Findeisen Z. Nagy F. Allgöwer

Laboratoire d'Automatique
Faculté Polytechnique de Mons
Boulevard Dolez 31, B-7000 Mons, Belgium
Institute for Systems Theory in Engineering
University of Stuttgart
Pfa enwaldring 9, D-70550 Stuttgart, Germany
Control Engineering Department
University of Loughborough
LE11 3TU Loughborough, United Kingdom

Abstract: In this study, attention is focused on the design of a scheduled-optimization strategy for a batch MMA polymerization process. The objective of this strategy is to track an optimal temperature, despite uncertainties in the heat transfer and the gel e ect. This strategy makes use of an (uncertain) physical model and on-line temperature measurements. The uncertain parameters are re-estimated on line, so as the optimal temperature trajectory. The good decoupling (in time) between the two major disturbances allows good performance to be achieved.

Keywords: Polymerization, batch control, process control, optimization.

1. INTRODUCTION

From an industrial viewpoint, the methyl methacrylate (or *MMA*, in the abbreviated form) polymer compounds hold an important place in the production of plastics. In this area as in other sectors of the chemical industry, batch processes have gained much interest essentially thanks to better production exibility, easier scale-up from laboratory setup and increased safety (reduced dimensions).
From a scienti c viewpoint, polymerization processes are relevant, essentially because complex temperature-dependent chain reactions and heat transfers lead to highly nonlinear algebraic and di erential equations. In addition, control design for batch processes is a challenging task since 1) in pure batch, no in uential input (e.g., feed) allows to alter the reactor contents, only the reaction rates can be modi ed by adjustment of the reactor temperature 2) only a few variables (temperatures) can be measured on line, the polymer properties being usually measured at the end of the batch only. Batch processes are often run in open loop using a prede ned (often heuristic) trajectory. Several variants of this strategy are proposed in the literature, such as classical feedback allowing an optimal trajectory to be tracked, repeated estimation-optimization during the batch, batchwise enhancement of the trajectories (run-to-run optimization). Whatever the technique, the calculation of a trajectory satisfying well-de ned

[1] Author to whom correspondence should be addressed:
e-mail: Renato.Lepore@fpms.ac.be
phone: +32 (0)65374140 fax: +32 (0)65374136

end-of-batch properties requires that a physical, rst-principles model of the process be developed. However, due to time and cost constraints, the model is often of limited accuracy, so that the control strategy has to take the model-plant mismatch into account and to enhance robustness, eventually in detriment of nominal performance. These industrial and scienti c aspects are abundantly covered in previous works, such as (Kiparissides, 1996) and (Terwiesch *et al.*, 1994; Bonvin, 1998) and the references therein.

An MMA batch polymerization process is under study, which is a laboratory scale plant at the Aristotle University (Thessaloniki, Greece), and a nonlinear state space model is used, which describes the gel e ect using a deterministic law (Kiparissides *et al.*, 2002; Mourikas, 1998). In contrast with G. Mourikas' approach, attention is focused on the more common situation where only temperature measurements are available on line and the model is subject to two sources of uncertainty: one a ects a gel e ect parameter (e.g., due to an inaccurate identi cation), the other in uences the heat exchange coe cient between the reactor wall and the solution (e.g., due to fouling). When a robust worst-case approach is used, improved performance is achieved in the pure open-loop variant (Lepore *et al.*, 2004; Nagy and Braatz, 2004), i.e., the best input pro le is calculated so as to minimize the worst criterion value obtained when the gel parameter ranges within speci ed limits (min-max problem). However, in the less conservative variant, which uses a feedback controller for immediate disturbance rejection, the feedback required for the gel effect must be positive and cannot deal with the heat exchange disturbance. For these reasons, we have selected a scheduled- optimization approach, which uses on-line measurements in order to estimate both the time-varying parameters (gel e ect and heat exchange) and to update the model, so that a new optimal trajectory is calculated. If the heat exchange disturbance occurs batchwise, decoupled e ects of the disturbances on the solution temperature are used to obtain accurate, reliable estimates of the coe cients.

The paper is organized as follows. In section 2, the process is described and a nonlinear state space model is derived from mass and energy balances. Section 3 is devoted to the de nition of the control objectives. Section 4 describes and assesses the worst-case strategy used to incorporate the model uncertainties. In section 5, the principle of the scheduled-optimization strategy is presented and some results are discussed. Finally, conclusions are drawn in section 6.

2. PROCESS DESCRIPTION AND MODELLING

The reactor depicted in gure 1 contains the reactant (monomer) and the product (polymer) just mixed with water. The solution is continuously stirred, and its temperature is adjusted by feeding the jacket with hot and cold water. Two independent valves are regulated by a split range controller (low-level control of the jacket temperature T_J). The main process characteristics are:

- the process is of the bulk type, i.e., the solution consists of pure monomer and water, no other agent or solvent is added,
- a homogeneous mixture is considered, i.e., the monomer is miscible with its polymer,
- the reaction kinetics is based on a free-radical mechanism, i.e., intermediate, active radicals, generated from a monomer unit and a catalyst (initiator), grow or propagate by addition of monomer units, then terminate into polymer chains.

In this process, the viscosity of the mix causes poor heat transfer characteristics within the solution and with the jacket (due to polymer deposits on the reactor wall).
A particular phenomenon, well known as the *gel e ect* may cause, if ignored, poor properties of the nal product or very low conversion. In fact, when the monomer conversion is su cient, the termination reactions become *di usion-controlled*, i.e., large free-radical chains terminate hardly, whereas the propagation phenomenon accelerates. A natural counteraction consists in heating the solution.

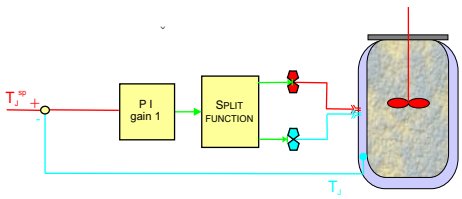


Fig. 1. Polymerization reactor: equipment and low-level control.

The process objectives are mainly concerned with nal-product quality (speci c physical properties), process performance (conversion percentage and/or batch time), safety (temperature limitations). The disturbances (model uncertainties) affecting the process behaviour are of several types: wrong initial conditions (e.g., initiator), varying coe cients due to impurities or polymer deposits (heat exchange between the metal wall and the solution, initiator e ciency), poor accuracy when identifying the gel e ect. Only temperature measurements are available at su cient rate and reliable, whereas the polymer properties are accurately measured at the end of the batch only.

De ning R_i and P_i the molar concentration of the free-radical and the polymer species, respectively, the general k^{th} noncentered moments are:

$$_k = \sum_{i=1}^{\infty} i^k R_i \tag{1a}$$

$$_k = \sum_{i=1}^{\infty} i^k P_i. \tag{1b}$$

With $\boldsymbol{\lambda} = [\ _0 \quad _1 \quad _2]^T$ and $\boldsymbol{\mu} = [\ _0 \quad _1 \quad _2]^T$, mass balances expressed in terms of the rst three noncentered moments lead to the following set of di erential-algebraic equations (DAEs):

$$\frac{d\boldsymbol{\xi}_1}{dt} = \boldsymbol{f}_1(\boldsymbol{\xi}_1; \boldsymbol{\theta}_1) \tag{2a}$$

$$\boldsymbol{g}_1(\boldsymbol{\lambda}, \boldsymbol{\xi}_1; \boldsymbol{\theta}_1) = 0 \tag{2b}$$

$$\boldsymbol{\xi}_1(0) = \boldsymbol{\xi}_{1;0}, \tag{2c}$$

where:

- $\boldsymbol{\xi}_1 = \left[c_M\ I\ (\boldsymbol{\mu}^T\ V)\right]^T$, c_M is the monomer conversion factor, I is the initiator molar concentration, V is the sum of the monomer and polymer volumes,
- $\boldsymbol{\xi}_{1;0} = [0\ I_0\ 0\ 0\ 0]^T$, I_0 is the initial molar concentration of the initiator,
- $\boldsymbol{\theta}_1$ is the parameter vector, related to the reaction rates.

Due to the faster dynamics of the free-radical species, the quasi-steady state assumption (QSSA) holds for the free-radical chains, leading to the purely algebraic equations (2b).

Another set of equations derives from thermodynamic balances between the physical components of the reactor. According to (Mourikas, 1998) and the references therein, one can assume that 1) the temperatures of the metal wall and of the solution are uniform (e cient stirring); the sensor for the solution temperature is modelled by a rst-order system 2) as the heat exchange coe cient between the jacket and the metal wall highly depends on the jacket temperature distribution, the jacket is discretized into four zones where the heat exchange parameters are lumped. Expressing the variations of the internal energy as the net amount of heat transfer leads to ordinary di erential equations (ODEs) for variables T_R (reacting solution), T_M (metal wall), $T_{J;k},\ k = 1..4$ (jacket zones) and T_S (sensor).

$$\frac{d\boldsymbol{\xi}_2}{dt} = \boldsymbol{f}_2(\boldsymbol{\xi}_2, \boldsymbol{F}_w; \boldsymbol{\theta}_2) \tag{3a}$$

$$\boldsymbol{\xi}_2(0) = \boldsymbol{\xi}_{2;0}, \tag{3b}$$

where:

- $\boldsymbol{\xi}_2 = [T_R\ T_{J;1}\ T_{J;2}\ T_{J;3}\ T_{J;4}\ T_S\ T_M]^T$,
- $\boldsymbol{\xi}_{2;0;i} = 300\,\text{K}, i = 1..7$,
- $\boldsymbol{F}_w$ is the two-component vector of hot and cold water ow rates,
- $\boldsymbol{\theta}_2$, the parameter vector, contains the heat exchange and speci c heat coe cients, which are complex functions of the temperature.

One element of $\boldsymbol{\theta}_2$, the heat exchange coe cient between the metal wall and the solution, noted h_{ms}, can vary during the batch or batchwise due to accumulation of impurities (fouling).

A deterministic law, which describes the termination rate coe cient, is de ned as follows (Mourikas, 1998):

$$k_t = k_{t0}\ g_t, \tag{4a}$$

$$g_t = f_{gel}\left(T_R,\ _0, c_M; A\right), \tag{4b}$$

where k_t and k_{t0} are the real and low-conversion termination rate coe cients, respectively, $_0$, according to (1a), is the total concentration in the free-radical species, A is a scalar parameter which accounts here for the inaccuracy when identifying the gel e ect. A can vary between two bounds (lower and upper).

Assembling equations (2) and (3), and augmenting them with one variable, named η, and one equation accounting for the low-level control allows to rede ne a new, complete system:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x}, u; \boldsymbol{p}), \tag{5a}$$

$$\boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{5b}$$

where:

- $\boldsymbol{x}^T = \left[\boldsymbol{\xi}_1^T\ \boldsymbol{\xi}_2^T\ \eta\right]$, $\boldsymbol{x}_0^T = \left[\boldsymbol{\xi}_1^T\ \boldsymbol{\xi}_2^T\ 0\right]$,
- u is the jacket temperature setpoint T_J^{sp},
- $\boldsymbol{p} = [p_h\ p_g]^T$ accounts for the two aforementioned model uncertainties, i.e., p_h is such that the heat exchange coe cient between the metal wall and the solution $h_{ms}^{real} = h_{ms}^{nom}(1 + p_h)$, p_g a ects the model of the gel e ect through parameter A in relation (4b) as $A^{real} = A^{nom}(1 + p_g)$.

3. CONTROL OBJECTIVES

The control objective consists of a trade-o between several end-of-batch properties related to product quality and quantity. As in (Thomas and Kiparissides, 1984) and (Mourikas, 1998; Kiparissides *et al.*, 2002), a terminal cost $\Phi(\boldsymbol{x}(t_f))$ is de ned as follows:

$$\Phi\left(\boldsymbol{x}(t_f)\right) = \epsilon_{c_M}^2 + \epsilon_{Mn}^2 + \epsilon_{Mw}^2 \tag{6a}$$

$$\epsilon_{c_M} = \left(1 \quad \frac{c_M(t_f)}{c_{Md}}\right) \tag{6b}$$

$$\epsilon_{Mn} = \left(1 \quad \frac{Mn(t_f)}{Mn_d}\right) \tag{6c}$$

$$\epsilon_{Mw} = \left(1 \quad \frac{Mw(t_f)}{Mw_d}\right). \tag{6d}$$

In expressions (6), $c_M(t)$ is the conversion factor, $Mn(t)$ and $Mw(t)$ are the *number average molecular weight* and the *weight average molecular weight*, respectively:

$$Mn(t) = MW\ \frac{_1(t)}{_0(t)} \tag{7a}$$

$$Mw(t) = MW\ \frac{_2(t)}{_1(t)}, \tag{7b}$$

where MW is the molar weight of the monomer. $Mn(t)$ is exactly the mean length of the polymer chains, whereas $Mw(t)$ encompasses the length and the dispersion of the polymer chains. In expressions (6), c_{Md}, Mn_d and Mw_d are the target (or desired) values.

Generally, a dynamic optimization problem is stated as follows: given the desired values c_{Md}, Mn_d and Mw_d, nd the input pro le $u(t)$ which minimizes $\Phi(\boldsymbol{x}(t_f))$, while satisfying the system constraints (5), input constraints, path and terminal constraints. In the following, the dynamic optimization problem is solved using a direct single-shooting method, i.e., the input $u(t)$ is parameterized with N linear segments and box constraints apply on the input only. On the other hand, it is considered that the constant input of 337 K is optimal in the nominal case ($\boldsymbol{p} = \boldsymbol{0}$) for a batch of 120 min.

4. ROBUST WORST-CASE STRATEGIES

In a standard worst-case strategy, one solves an optimization problem, which accounts for the model uncertainty. In our case, it is assumed that the gel e ect is unknown (with no loss of generality, p_g is between 0.0 and 0.05). Due to the structure of the cost function (a sum of square deviations from target values), the optimization problem is of type min-max:

$$\min_{u(t)} \max_{p_g} \Phi(\boldsymbol{x}(t_f)), \tag{8}$$

subject to the system constraints (5) and subject to input and disturbance bounds, $u^{min} \leq u \leq u^{max}$ and $0.0 \leq p_g \leq 0.05$ respectively.

In the pure open-loop variant, the minimization is performed using the input pro le only. Figure 2 shows the evolution of the terminal cost, as a function of the gel e ect parameter p_g, either in the nominal design (i.e., no uncertainty is accounted for, which leads to a classical open-loop optimization) or in the worst-case design. It is noted that 1) the decrease in the terminal cost is signi cant when applying the worst-case input for higher values of p_g 2) the worst-case input yields some degradation for lower values of p_g, however not very signi cant. Robustness of the strategy is also exhibited with respect to perturbations in the heat transfer. In regards of this conservative variant, another variant is of major interest, which uses an internal feedback controller for immediate disturbance rejection (whose parameters may also be optimized). However, the rejection of the (residual) gel e ect requires a positive reaction, which is not compatible when dealing with the heat exchange disturbance.

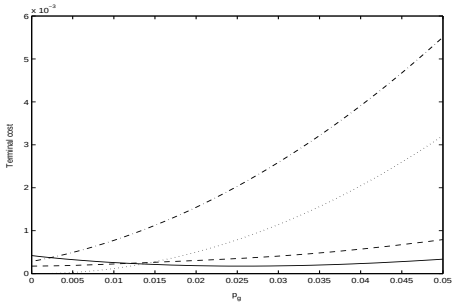


Fig. 2. Terminal cost corresponding to optimal open-loop input. No heat exchange disturbance: nominal (dotted) and worst-case (solid) design ; heat exchange disturbance: nominal (dash-dotted) and worst-case (dashed) design.

5. SCHEDULED-OPTIMIZATION STRATEGY

This strategy lies on the principle of estimation-optimization, i.e., the uncertain parameters are estimated on line and a new trajectory is calculated using the adapted model. In our investigation, we consider that the batch preparation is ideal, i.e., the initial conditions are known. The algorithm is designed to deal e ciently with two disturbances in the same batch, which are in the heat exchange and in the gel e ect and both vary batchwise only. In fact, the gel e ect exhibits only when the monomer conversion is su cient whereas the heat exchange has an impact during the whole batch, especially at the beginning where the input contains su cient excitation (strong heating). Therefore, decoupled, reliable estimation can be achieved, according to the following output least-square error problem (9).

$$\min_{p \in \mathcal{P}} \int_{t_0}^{t^{EOT}} \left(y(\tau) \quad y_m(\tau)\right)^2 d\tau, \tag{9}$$

subject to system constraints:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u; p), \tag{10a}$$
$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{10b}$$
$$y = h(\boldsymbol{x}, u), \tag{10c}$$

where:

- t_0 is the initial time of the measurement sequence and t^{EOT} is the estimation-optimization time (last time in the measurement sequence),
- $y(t)$ and $y_m(t)$ are the model and measured solution temperatures at time t, respectively.

Additionally, a lower bound on the parameter variance $\overset{2}{p}$ is given by the inverse of the (scalar) Fisher information matrix and is approximated as follows (Walter and Pronzato, 1997):

$$\overset{2}{p} = \frac{\overset{2}{y}}{\int_{t_0}^{t^{EOT}} \left(\left(\frac{\partial h}{\partial p} \right)(\tau) \right)^2 d\tau} \tag{11}$$

where:

- $\overset{2}{y}$ is the variance of the temperature measurements,
- $\frac{\partial h}{\partial p}$ is the rst-order sensitivity function of the temperature variable with respect to the parameter.

The heat exchange coe cient is estimated once at time noted t^{he} ($t_0 = 0$ and $t^{EOT} = t^{he}$). The gel e ect is estimated at any time t^{gel} where the solution temperature deviates su ciently from the most recently-calculated optimal trajectory ($t_0 = t^{he}$ and $t^{EOT} = t^{gel}$). After each estimation (at time t^{EOT}), a new trajectory is calculated by solving problem (12).

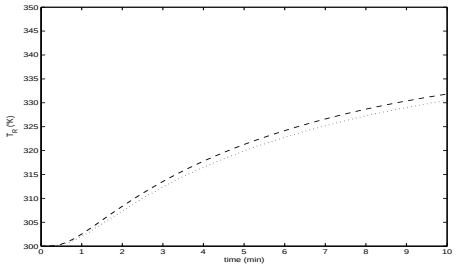
$$\min_{u(t)} \Phi(\boldsymbol{x}(t_f)), \tag{12}$$

subject to system and input constraints :

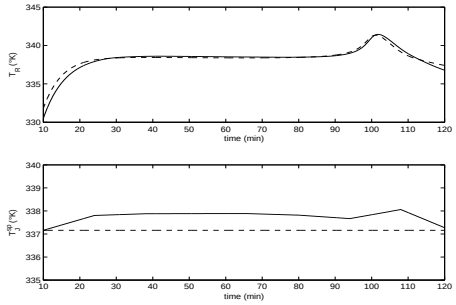
$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u; \boldsymbol{p}^{est}), \tag{13a}$$
$$\boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{13b}$$
$$u(t) \in [u^{min}, u^{max}]. \tag{13c}$$

An illustrative experiment is performed, under the following operating conditions:

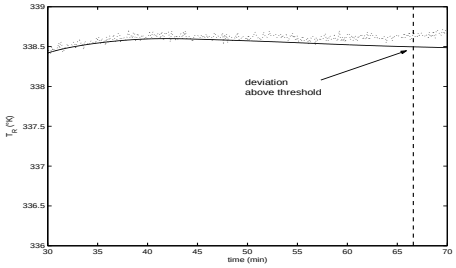
- model uncertainties: $p_h = \ \ 0.5$, $p_g = 0.05$,
- measurements of the solution temperature are available every 0.1 min and are a ected by white Gaussian noise, with zero mean and 0.033 K standard deviation (maximum error: 0.1 K),
- the estimation-optimization task is performed at time $t^{he} = 10$ min, accounting for the variation of the heat exchange,



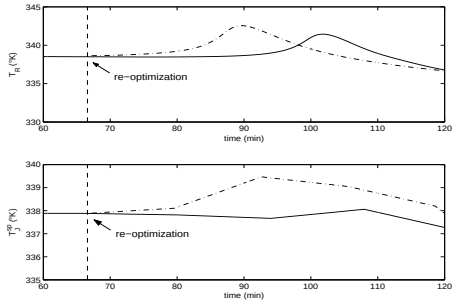
(a) Heat exchange mismatch: deviation between optimal (dashed) and real (dotted) trajectories ($t \leq 10$)



(b) Trajectories re-optimized at time 10 (solid) compared to off line-calculated trajectories (dashed)



(c) Gel effect mismatch: deviation between optimal (solid) and real (dotted) trajectories ($30 \leq t \leq 70$)



(d) Trajectories re-optimized at time 66.6 (dash-dotted) compared trajectories re-optimized at time 10 (solid)

Fig. 3. Scheduled-optimization strategy: illustration

- the estimation-optimization task is performed at time t^{gel} where a deviation of 2 times the maximum error is detected, i.e., 0.2 K.

Figure 3 illustrates the experiment and calls for the following comments:

- in 10 min, su cient information is available from the real temperature which deviates

sensitively from the optimal temperature trajectory calculated o line (3(a)),
- at time $t = 10$ min, the heat exchange coe cient is estimated and the state vector is obtained by simulation from the known initial conditions, then new optimal trajectories are calculated based on the adapted model (3(b)),
- at time $t = 66.6$ min, the solution temperature deviates sensitively from the optimal trajectory (3(c)), which is attributed to the gel e ect mismatch,
- the gel e ect coe cient is estimated and, again, the state vector is obtained by simulation, then the new optimal trajectories are calculated based on the adapted model (3(d)),
- from time $t = 66.6$ min on, no more deviations above the threshold are detected.

In this simple experiment, satisfactory end-of-batch performance is achieved (the terminal cost is 0.33 10 ^{3}), as well as accurate estimation results, such as:

- estimation of $p_h = 0.5$, with a standard deviation of 7 10 ^{4},
- estimation of $p_g = 0.049$, with a standard deviation of 9 10 ^{4}.

If the accuracy of the temperature sensor is lower (higher measurement error) whereas the threshold factor is kept unchanged (for example, equal to 2), a degradation (increase) of the terminal cost can be expected due to the lag in the detection/re-optimization. Figure 4 illustrates the evolution of the detection time and of the terminal cost for various values of the maximum absolute measurement error (from 0.1 to 2.0 K). Clearly, the terminal cost is kept at reasonable values even when the detection is very late (time 85 corresponds to the gel e ect phenomenon).

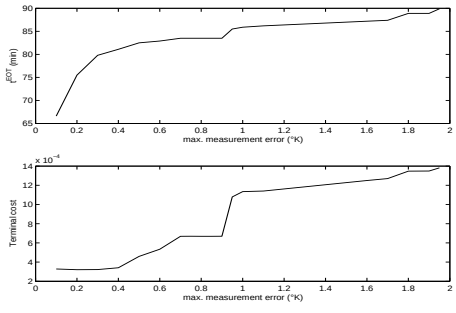


Fig. 4. Scheduled optimization-based strategy: e ect of noise on temperature measurements.

6. CONCLUSION

This paper reports work on the design of controllers for an MMA polymerization reactor, based on temperature measurements only. The approach, based on a worst-case analysis (min-max problem), gives acceptable results in open loop. However, it is very conservative unless combined with a feedback controller, which cannot be achieved here since only a positive feedback is suitable for rejection of the gel e ect disturbance. Better results are obtained with a scheduled-optimization strategy, i.e., the model parameters are adapted on line, and optimal trajectories are re-evaluated. Provided the reasonable assumption that the parameters vary batchwise only, a decoupled, reliable estimation of these is achieved and this latter approach gives very satisfactory results.

REFERENCES

Bonvin, D. (1998). Optimal operation of batch reactors: a personal view. *Journal of Process Control* **8**(5-6), 355–368.

Kiparissides, C. (1996). Polymerization reactor modeling: a review of recent developments and future directions. *Chemical Engineering Science* **51**(10), 1637–1659.

Kiparissides, C., P. Seferlis, G. Mourikas and A.J. Morris (2002). On-line optimizing control of molecular weight properties in batch-free polymerization reactors. *Industrial and Engineering Chemistry Research* **41**, 6120–6131.

Lepore, R., R. Findeisen, Z.K. Nagy, F. Allgöwer and A. Vande Wouwer (2004). Optimal open- and closed-loop control for disturbance rejection in batch process control: a MMA polymerization example. In: *Symposium on Knowledge-driven Batch Processes (BATCH-PRO)*. pp. 235–241. Poros, Greece.

Mourikas, G. (1998). Modelling, estimation and optimisation of polymerisation processes. PhD thesis. The University of Newcastle. Newcastle upon Tyne, United Kingdom.

Nagy, Z. K. and R.D. Braatz (2004). Open-loop and closed-loop robust optimal control of batch processes using distributional and worst-case analysis. *Journal of Process Control* **14**, 411–422.

Terwiesch, P., M. Agarwal and D.W.T. Rippin (1994). Batch unit optimization with imperfect modelling: a survey. *Journal of Process Control* **4**(4), 238–258.

Thomas, I.M. and C. Kiparissides (1984). Computation of the near-optimal temperature and initiator policies for a batch polymerization reactor. *The Canadian Journal of Chemical Engineering* **62**, 284–291.

Walter, E. and L. Pronzato (1997). *Identi cation of parametric models from experimental data*. Springer-Verlag. London.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

OPPORTUNITY FOR REAL-TIME OPTIMIZATION IN A NEWSPRINT MILL: A SIMULATION CASE STUDY

Antoine Berton, Michel Perrier and Paul Stuart

NSERC Chair in Environmental Design Engineering
Chemical Engineering Department
École Polytechnique, Montréal, Canada
email: paul.stuart@polymtl.ca

Abstract: Real-time optimization (RTO) is applied for the broke recirculation in the stock approach system of an integrated newsprint mill simulation. New optimal broke ratio profiles are set for the paper machine recipes each time the process is confronted with an important disturbance. The variability in the four paper machine headboxes is minimized while the inventory management is handled. The proposed approach is compared to the actual mill operation, making use of a detailed plant simulation. While maintaining broke tank level within an acceptable range, RTO brings the pulp variability down about an order of magnitude. It is expected that it could significantly reduce the sheet break occurrence in a real plant and therefore enhance its economical efficiency.

Keywords: Real time optimization, newsprint mill, sheet break, paper machine.

1. INTRODUCTION

A typical state-of-the-art integrated newsprint mill involves the production of pulp by mechanical pulping and deink pulping, which are combined with broke pulp and delivered to a high-speed newsprint paper machine, typically running at speeds over 1 000 m/min. The broke pulp consists of recycled dry paper and wet pulp coming from the paper machine. The machine uptime is mainly limited by breaking of the paper sheet, causing stops several times per day, for periods of up to one and a half hour. In addition to the loss of production, these stops have drawbacks on the process by creating new instability resulting in future breaks on the sheet of paper. In fact, sheet breaks (or rupture) are major perturbations for the paper making process and accounts for a net loss of 10 to 15% of production (Bonissone *et al.*, 2002) resulting in revenue loss of billions every year for the global pulp and paper industry.
Paper breaks generally occur due to local weaknesses in the mechanical resistance of the sheet or when the sheet is subjected to sudden load increase. A number of studies have been carried out on characterizing breaks (Khanbaghi *et al.*, 1997; Akrour *et al.*, 2006) and in the identification of its causes (Linstrom *et al.*, 1994).
During a break, the paper machine is in continuous operation while action is taken to restore the continuity of the paper sheet. Until this is achieved, pulp is still fed to the paper machine but is directly collected in pits located below the machine and subsequently diluted and reutilized as it contains valuable raw materials (broke pulp). The broke pulp and fresh pulp have different properties and must be mixed in appropriate proportions depending on the type of paper produced. Recirculation of broke pulp is a major disturbance that increases the variability of the pulp feeding the paper machine affecting its performance. The variability is shown to increase the sheet break occurrence, bringing the process into a vicious circle, as a high variability in the headbox is a source of paper sheet breaks. Various studies on broke recirculation and its management have been published (Bonhivers *et al.*, 2002; Croteau and Roche, 1987; Orccotoma *et al.*, 1997; Ogawa *et*

al., 2004).
The present work is part of a vast research project on simulation, control and optimization of newsprint mills. The final objective is to design new control and optimization strategies on a plant-wide scale.
In the present paper, a Real-time optimization (RTO) strategy (Forbes and Marlin, 1996) is tested considering a key area of the plant: the stock approach system, the broke system and the four paper machines. Very few RTO applications exist in the pulp and paper industry (Flisberg and Rönnqvist, 2002).
First, the process is presented in detail and the actual control strategy for broke recirculation is shown. Then, the metric used for comparing operating strategies performance with regards to the headbox variability is presented. Using mill databases, an informal validation of this metric is established, showing a clear link between sheet break occurrence and headbox variability. In section 4, the proposed RTO control strategy is presented. Section 5 finally presents the results comparing the RTO control strategy to the usual plant operation. Through the use of simulation, the control approaches are compared considering the headbox variability and the behavior of the broke tank level. This is illustrated for a typical scenario of one day of operation.

2. PROCESS DESCRIPTION AND ACTUAL PRACTICES

This study has been performed in collaboration with an existing newsprint mill in Canada. However, the proposed approach is not specific to that mill and can be applied to other newsprint mills.

2.1 Mill description

The mill considered in this study produces newsprint at a rate of approximately 1000 t/d. It has four different paper machines that are fed with thermomechanical pulp (TMP) and deinked pulp (DIP). In addition, broke pulp, is also recycled to the headbox, at a rate appropriate to maintain reasonable inventory level in the broke storage tank. The broke and white water system is shared by the four machines. Figure 1 shows a simplified view of the process with the three principal storage tanks and the four paper machines. For the sake of simplicity, the white water system, as well as many intermediate storage tanks considered in the study are not illustrated. The mixed pulp composition is adjusted several times a day depending on pulp inventory, type of paper produced, presence of breaks on a paper sheet, etc. Since the three different sources of pulp are characterized by different properties, a change in the recipe introduces variability in the properties of the mixed pulp feeding the headbox. Table 1 shows how five pulp properties considered in this study vary depending on the type of pulp (Dabros

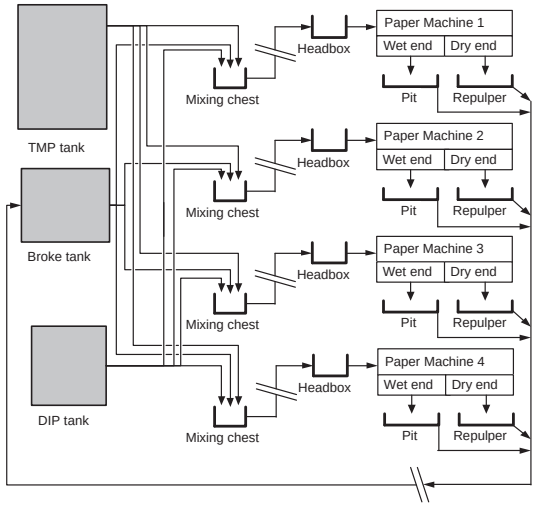


Fig. 1. Simplified flowsheet of the mill stock approach system and paper machines.

et al., 2004). The problem consists of adjusting the different pulp ratio of the mixed pulp, taking into consideration the inventory levels for the different break scenarios. When a break occurs, broke pulp is feeding the broke tank at a higher rate creating a disturbance in the control of its level. In the present study, only the management of the broke pulp is considered, having a constant DIP feed. The remainder of the pulp is TMP.

Table 1. Typical characteristic values of TMP, DIP and broke pulp at the mill.

Property	TMP	DIP	Broke
Consistency (%)	3.4	4.0	3.2
Temperature (°C)	65	52	58
Fibre content (%)	79	79	49
Fine content (%)	21	21	51
Dissolved solids (%)	0.4	0.4	0.08

2.2 Mill operating strategy

The current method for adjusting the broke ratio in the mill is manual. Operators supervise the inventory level of the broke tank and when it reaches too high a level or when a break occurs, they increase the broke ratio usually by increments of 5 %. This method for adjusting the broke ratio causes sudden changes in the mixed pulp properties, highly affecting the headbox stability.

2.3 Process simulation

A dynamic simulation of the papermaking section of the mill was developed using the WinGEMS 5.3 sequential simulation software, which is commonly used for simulation of pulp and paper processes (PacSim, 2005). The model includes the four different paper machines, the white water and the broke systems. An informal validation of the simulation was performed using mill data (Dabros *et al.*, 2004). The dynamic input variables include, for each paper machine, the mixed pulp flowrate,

the proportion of TMP, DIP and broke pulp as well as the break status (if there is currently a break on the paper machines). The simulator then provides the tank levels, the flowrates and all the properties listed in Table 1.

3. HEADBOX VARIABILITY

In this study, the headbox variability, referring to the variability of several physical properties of the pulp at the headbox is used for two objectives. First, it is included in the optimization objective function and then, it is used in order to compare operating strategies.

3.1 Definition of headbox variability

In previous work (Bonissone *et al.*, 2002), five pulp properties were selected for characterizing the headbox pulp variability (HV): consistency (%), temperature (°C), flowrate (l/min), fines content (%) and dissolved solids (%). The metric that will be used to compare the performance of different operating policies is a summation of the squared normalized incremental changes of those five properties:

$$HV(t) = \sum_{k=1}^{5} (c_k(P_k(t) - P_k(t-1))^2 \qquad (1)$$

where P_k denotes a specific property. Normalizing coefficients (c_k) are included as weighting factors for the five properties. The c_k coefficient are chosen to be the inverse of the average incremental change over a standard day of operation. In that way, each property should contribute equally to the headbox variability metric (HV). The simulation sampling time is one minute, making a new metric evaluation available every minute.

3.2 Informal validation of the objective function

Recently, extensive statistical studies of mill data have been carried out that proved that there is a clear correlation between the headbox variability and sheet break occurrence (Akrour *et al.*, 2006). Figure 2 offers a qualitative relation between the probability of break and the properties involved in the HV as defined in Equation 1. Note that the actual values of HV cannot be determined since there is no available information due to lack of sensors in the plant to measure all the properties in Equation 1. In Figure 2 a surrogate of HV is composed with two of the measurements of Equation 1 (consistency, flowrate) and with a surrogate of the other variables (white water temperature, broke ratio). The broke ratio has been included as its change is strongly correlated with changes in dissolved solids and fines content. Figure 2 is made from data over a 10 months period. For the right half of the histogram, very few observations were available making the corresponding probabilities

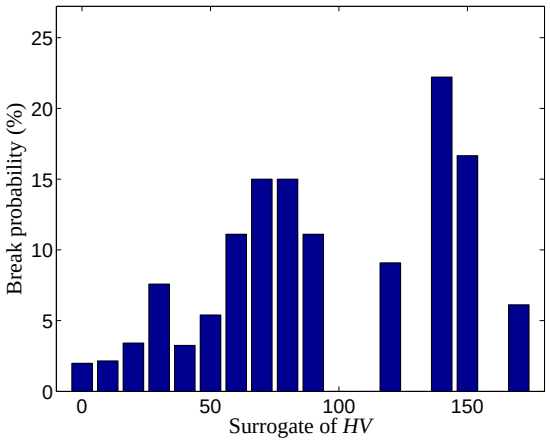


Fig. 2. Observed sheet break probability in the mill with respect to a surrogate of HV.

inaccurate. For some classes, there was even no observation (for instance, between 90 and 115 in the y axis). The actual value of HV (x axis) has no physical meaning but it is clear that the probability of break occurrence is enhanced when its value is higher. By keeping the HV function below a value of around 10, the break occurrence should be significantly minimized. Even if it is not exactly the same metric, the clear tendency observed in the break probability validates the hypothesis that high headbox variability increases break occurrence and is therefore a suitable metric to evaluate the performance of different operating strategies.

4. PROPOSED APPROACH FOR BROKE MANAGEMENT

The proposed strategy for broke management is basically inspired by mill operator's methods, but in an automated and smoother way. In the mill, the operators adjust the broke ratio so that the broke inventory remains within an acceptable range. The proposed strategy is simultaneously adjusting the four broke ratios, taking into consideration the broke tank level.
The time to perform RTO could be set periodically or when the process is facing a significant disturbance. In the present study, RTO occurs every time the break status changes. It means that every time a break occurs or ceases, a new optimization is performed. In addition, when the break status is not changing for a long time, a new optimization is performed for safety.

4.1 Process models

The RTO strategy uses a WinGEMS dynamic model which is similar to the simulation model that is to be used as a test mill. The WinGEMS model used for RTO purposes is simpler in order to enable the fast simulation which is critical for practical implementation. The main simplification is that no convergence of the steady state simulation is obtained, signifying that the starting point

of the optimization model does not agree strictly to material conservation.

4.2 Optimization window

Each time RTO is performed, a snapshot of the current states of the plant (in this study the plant simulator) is taken. This snapshot includes all variables necessary to initiate a new simulation. It also includes the current break status, i.e. which machine is affected by breaks. From that point, the broke ratio profiles must be optimized for a certain time window in the future. This time window was set to 30 minutes, thus 30 sampling period. For this chosen time window, the break status is assumed to be constant.

4.3 Optimization parameters

Ideally, the optimization parameters would be the four broke ratios at each sampling periods. However, it would make the optimization problem too big and practically impossible to implement.
Choosing a fixed value for the broke ratios in the optimization window would bring the number of parameters to only four. In that way, the broke ratio profiles would be a succession of steps of different amplitude. These sudden changes would significantly increase the probability of breaks.
In a recent study, Dabros et al. (2004) show that the best way to minimize headbox variability when performing a modification of the pulp proportions was to implement gradual changes spread over time. The same idea is used here by setting each broke ratio profile to be gradually increasing or decreasing. In that way, the headbox variability is minimized and the optimization problem remains simple, having only four optimization parameters. Those parameters are the incremental rates of change, constant at each sampling period.

4.4 Objective function

In order to satisfy the dual objective of minimizing HV over the optimization window and keeping the broke tank level within an acceptable range, the broke tank level ($L(t)$) is included in the objective function. Also, the actual break status is included as it has a direct impact on the broke tank level. When a break occurs, the pulp is recycled and thus requires more of the storage space in the broke tank. For a given optimization window, the objective function corresponds to the sum over time of the headbox variability function ($HV(t)$) for the four paper machines added to a squared sum of the broke tank level deviation from its setpoint:

$$J = \sum_{4}\sum_{t} HV(t) + \sum_{t}(\kappa(1+n_b)(L(t)-L_{sp}))^2 + \lambda\sum_{4}(poslin(BR_{max}-BR_{lim}))^2 \quad (2)$$

The inclusion of the number of breaks (n_b) in the objective function allows automatically more emphasis on the tank level control when there is break(s). The broke tank level setpoint (L_{sp}) was set to 30%. The objective function also includes a term limiting the range of the broke ratio. The third term of Equation 2 is the squared difference of the broke ratio's maximum value over the optimization window (BR_{max}) and its soft limit (BR_{lim}). The term *poslin* denotes a transfer function equal to one when the inner term is positive and 0 when it is negative. The coefficients κ and λ are normalizing parameters in order to put a similar weight on the three terms of J in Equation 2. The broke ratio limit (BR_{lim}) can be different on each machine depending on the type of paper produced. In the present study, it was set to 35% for the four machines. Although none were used, the actual architecture also allows implemention of hard constraints which might be necessary for some operating conditions.
The RTO strategy is depicted in Figure 3. For each block of the flowchart, the box format indicates if the operation is made in Excel, Matlab (Mathworks, 2005) or WinGEMS. The three softwares are necessary as no communication protocol is available between Matlab and WinGEMS. A first guess is made for the incremental changes and is sent along with simulation parameters to the simplified 30 minutes dynamic simulation. After it is completed, the objective function J is evaluated. The Optimization toolbox from Matlab is used to solve the optimization problem. Once the algorithm has converged, the recommended incremental changes are sent to the WinGEMS mill simulation. This dynamic model simulates the mill from the present time until the next change in the break status, at t_{newD} (for *new disturbance*). From that point, a snapshot of the mill is taken and sent back for a new optimization. For the

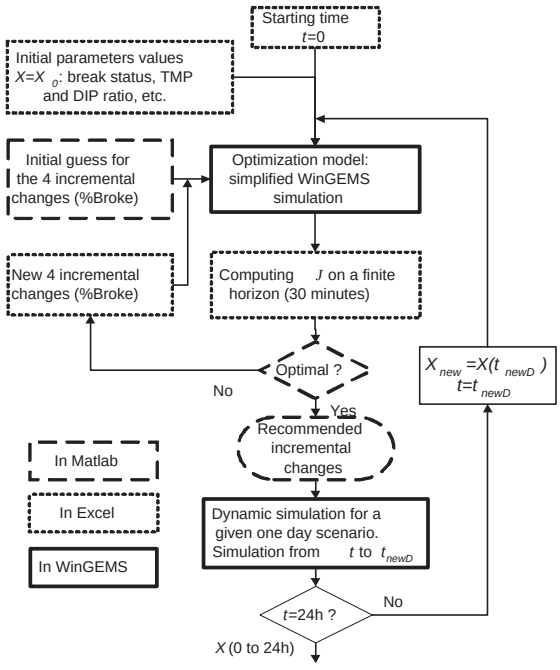


Fig. 3. Flowsheet of the RTO algorithm.

purpose of the study, this routine is repeated until the one day scenario is completed.

5. CASE STUDY AND RESULTS

The proposed strategy for broke management is compared with the actual mill operation in the simulated newsprint mill for a typical one day scenario. Breaks occurrence and length correspond to a real day of operation in May 2003. For that day, a total of 24 breaks occurred among the four machines. Figure 4 illustrates the occurrence and the length of the breaks. The scenario chosen is not a particularly good day for the mill as the uptime of the four machines was low (88 %). This is therefore a good candidate to test if the proposed procedure can handle difficult situations. For that day, the broke ratio set by the operator was also recorded and applied to the newsprint mill simulation to be compared to the proposed RTO strategy.

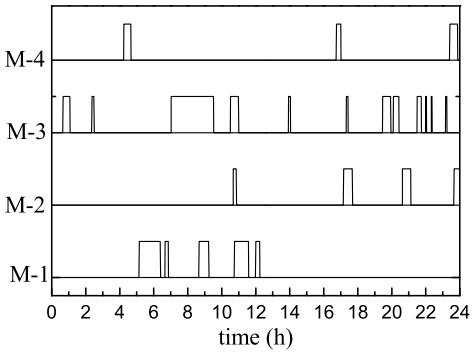


Fig. 4. Occurrence and length of breaks during the one-day scenario of study for the four paper machines (M-1 to M-4).

5.1 Impact on the headbox variability

As defined earlier, the headbox variability (HV) is used as a comparison metric to evaluate the operating strategies. Figure 5 shows the broke ratio profiles for the whole day of operation as performed by mill operators (left axes). On the right axes, the headbox variability function is shown on the bottom part of the graphs. When the operators change the broke ratio in the plant, it increases dramatically the variability of the headbox pulp properties, resulting in very high values of HV.
The peaks are observed for two main reasons: change in broke ratio and sheet break. Referring to Figure 2, the high peaks after a broke ratio change can highly increase the probability of breaks. This can be seen in Figure 5 where for machine 1, breaks occurred just after the first and second changes of the broke ratio. This observation agrees with the results observed over a 10 months period shown in Figure 2.

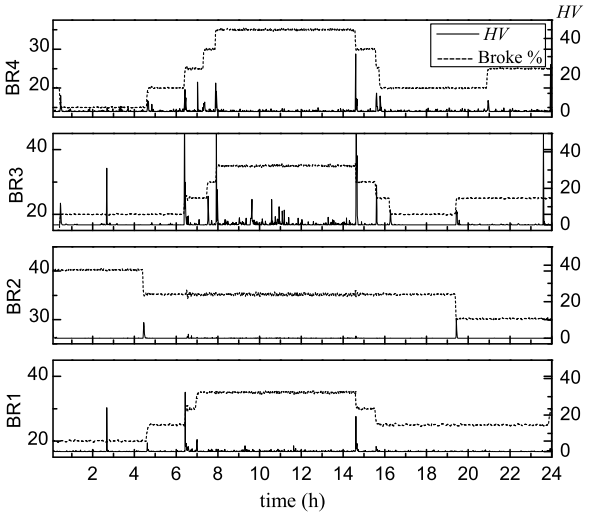


Fig. 5. Broke ratio and headbox variability profiles of the four paper machines.

By implementing the broke ratio changes more gradually while considering the safety aspects of broke tank level and broke ratio limitations, the RTO strategy (illustrated in Figure 6) gives better results than the operator's strategy. From the results of Figure 6, the improvement on the variability is obvious, the peaks for HV being decreased by about an order of magnitude compared to Figure 5. The broke ratio is adjusted more smoothly in the proposed approach, resulting in a higher stability of the headbox pulp properties.
For the mill operating strategy, peaks in HV are the results of both breaks (disturbance) and radical changes in the broke ratio. For RTO, the remaining peaks, while being attenuated, are almost only due to breaks (see Figure 4 for the break occurrence). The operator disturbance, which is indeed the most important one, is eliminated. The proposed approach shows the advantage of minimizing one type of disturbances while not creating others.
Referring to historical databases, this improvement can be translated into a decrease of more

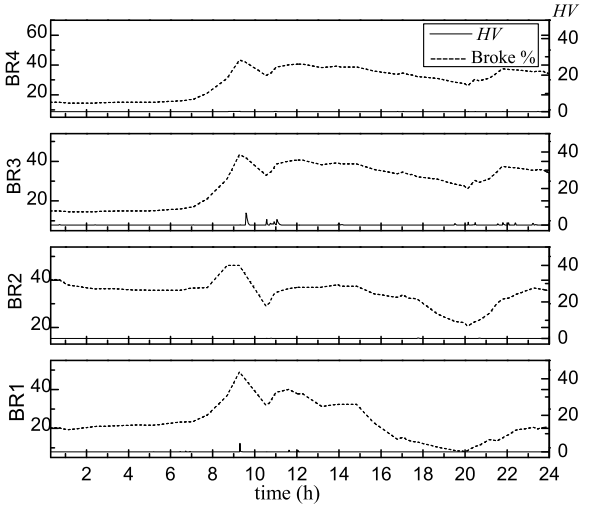


Fig. 6. Broke ratio and headbox variability profiles with RTO.

than 40% of breaks potentially caused by headbox instability. According to data available in Akrour et al. (2005), this improvement could result in a decrease of around 20% of the total mill loss due to breaks.

5.2 Impact on the broke tank level

A major concern of the plant personnel when a smoother broke ratio management approach is proposed is that during breaks, the effect on the broke tank level is too important to be able to slow down the broke ratio changes. This is an incorrect belief as the tank is able to absorb a large number of breaks. Figure 7 shows the broke tank level behavior for the two operating strategies. The total number of breaks is also shown in order to identify the most demanding periods for broke inventory.
Although RTO could seem very different than regular mill operation, the broke tank levels remain in a similar range. Therefore, the RTO strategy brings the headbox variability down without having any negative effect on inventory. Furthermore, RTO even brings the broke tank level lower by 20% at the end of the day, which represents an improvement.

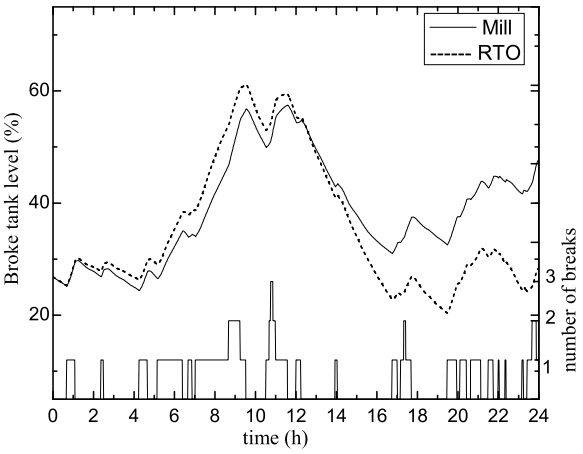


Fig. 7. Broke tank profile for manual and RTO strategies along with the number of breaks.

6. CONCLUSION

The proposed RTO strategy shows promising results for the control of stock approach systems in newsprint mills. The approach, based on the online optimization of a WinGEMS mill simulation, can bring tangible improvement in the mill performance. By keeping the headbox variability down, the probability of break occurrence of the mill operation can be lowered significantly, having a direct impact on its profitability. In future work, it is planned to refine and develop the RTO strategy to take into account the whole mill including the DIP and TMP plants.

ACKNOWLEDGEMENTS

Support for this project was provided by *NSERC* and *FQRNT*. We also thank Madhukar Gundappa, Michel Ruel (TopControl), Sylvain Gendron (Paprican) as well as Martin Fairbank and Pascal Hébert (Abitibi Consolidated).

REFERENCES

Akrour, F., A. Berton, M. Fairbank, M. Perrier and P.R. Stuart (2006). Probabilistic Index for Paper Machine Web Break Prediction due to Headbox Instabilities. *submitted to TAPPI journal.*

Bonhivers, J. C., Michel Perrier and Jean Paris (2002). Pulp properties: management of broke recirculation in an integrated paper mill. *Pulp and Paper Canada* **103**(2), 44–49.

Bonissone, P., K. Goebel and Y. Chen (2002). When will it break? A Hybrid Soft Computing Model to Predict Time-to-break Margins in Paper Machines. In: *Proceedings of SPIE 47th Annual Meeting.*

Croteau, A. P. and A. Roche (1987). Study of broke handling and white water management using a dynamic simulation. *Pulp and Paper Canada* **88**(11), 420–423.

Dabros, M., M. Perrier, F. Forbes, M. Fairbank and P.R. Stuart (2004). Improving the broke recirculation strategy in a newsprint mill. *Pulp and Paper Canada* **105**(11), 45–48.

Flisberg, P. and M. Rönnqvist (2002). Optimized control of the bleaching process at pulp mills. In: *Preprints of Control Systems.* pp. 210–214.

Forbes, J.F. and T.E. Marlin (1996). Design cost: a systematic approach to technology selection for model-based real-time optimization systems. *Computers & Chemical Engineering* **20**, 717–734.

Khanbaghi, M., R. Malhame, M. Perrier and A. Roche (1997). A statistical model of paper breaks in an integrated TMP-newsprint mill. *Journal of Pulp and Paper Science* **23**(6), 282–288.

Linstrom, R., W. H. Manfield, A. F. Tracz and J. Mardon (1994). Coping with an avalanche of breaks. *Appita Journal* **47**(2), 163–172.

Mathworks (2005). *Matlab 6.5: Optimization toolbox.*

Ogawa, S., B. Allison, G. Dumont and M. Davies (2004). Automatic control of broke storage tanks. In: *Preprints of Control Systems.* pp. 203–206.

Orccotoma, J. A., J. Paris, M. Perrier and A. Roche (1997). Dynamics of whitewater networks during web breaks. *Tappi Journal* **80**(12), 101–110.

PacSim (2005). *WinGEMS 5.3: Pulp and paper simulation software.*

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PRODUCT DESIGN VIA PLS MODELING: STEPPING OUT OF HISTORICAL DATA INTO UNKNOWN OPERATING SPACE

Ningyun LU, Yuan YAO and Furong GAO*

Dept. of Chemical Engineering
Hong Kong University of Science & Technology
Clear Water Bay, Hong Kong, China

Abstract: A product design and analysis method is given in this paper to step out of the historical data space to search for operating conditions meeting new quality specifications. Iterative piecewise PLS modelling is adopted as the implementation framework for this purpose. The historical linear PLS model is extended systemically and iteratively to track the likely nonlinear property in the newly discovered operating space. Application to an injection molding process shows the good feasibility of the proposed method.

Keywords: Product design, Multivariate quality control, Modelling, Piecewise analysis, Iterative method.

1. INTRODUCTION

Modern industrial processes can be operated over a range of operating conditions to produce a wide variety of products to meet the rapidly changing market. To respond to such frequent product changeover, it is necessary to develop methods that can quickly and economically find new operating conditions to achieve the desired product qualities. Existing solutions to this subject can be grouped into three categories: theoretical model based, design of experiment (DoE) based, and experience based. Although theoretical model can cover complete operating space, such a model is rarely available due to complicated process physical and chemical behaviours. Factorial design of experiment can provide balanced and representative data covering the design space. The number of experiment can be still large for a process with large number of variables. Experience based methods are highly dependent on the knowledge of the experts, and they are applicable only to specific processes.

* To whom correspondence should be addressed:
Tel: (852) -2358-7139, Fax: (852)-2358-0054, Email: kefgao@ust.hk

For a modern industrial process, there exists many historical data that can be explored to reveal the relationship between the existing product and its corresponding operating condition. For a new quality specification, it may be useful to start with the analysis of the historical data using multivariate statistical methods to extract information guiding experiments for searching for operating conditions to meet the new product requirement. This procedure is referred as product design via multivariate analysis. The first work was reported by Moteki & Arai (1986), where Principal Component Analysis (PCA) is combined with a theoretical method for operation planning and quality design. More recently, Jaeckle & MacGregor (1998) developed a methodology based on latent variable techniques using historical data to determine a window of process operating conditions for new quality specifications. This method has been successfully applied to a semi-batch emulsion polymerization process and a batch solution polymerization process (Jaeckle and MacGregor, 2000). Industrial case study has been reported by Chen & Wang (2000) and Sebzalli & Wang (2001), using PCA and clustering method to identify operating spaces and operating strategies for desired products. A product design method combining PCA with genetic programming has also been reported by

Lakshminarayanan et al. (2000) to determine new operating conditions.

The above methods are all data-based with a common implicit assumption that new product quality specifications and operating conditions are within the range and structure of historical data. This assumption, stated in the work of Jaeckle & MacGregor (1998), will limit the applications to a certain degree. For many industrial processes, it may be common that the process has been only operated in the past under certain specific operating conditions, which span only a narrow subspace of the entire feasible operating space. The operating condition for a new quality specification may be highly likely to be outside, rather than within, the range of historical data. It is thus necessary to develop methods to search for operating conditions outside the historical envelope. This paper attempts to do so. Section 2 analyzes possible conditions when stepping out of the historical data into new operating space. An iterative piecewise PLS modelling method is proposed in section 3 as a possible solution to the problem. The results are illustrated on an injection molding process in section 4. Finally, conclusions are drawn in section 5.

2. PROBLEM ANALYSIS

Let's first have a brief review on the key ideas of the existing methods. Based on the available historical data, the existing methods attempt to build an empirical model (M_{method}) between operating conditions (X) and product quality (Y). Under the assumption that the relationship between new product quality $\mathbf{y}_{des}$ and new operating condition $\mathbf{x}_{new}$ still obey the model (M_{method}), the new operating condition is obtained by $\mathbf{x}_{new}^T = \mathbf{y}_{des}^T \cdot M_{method}^T$ (Jaeckle & MacGregor, 1998).

For industrial processes, if the historical database indeed covers the entire operating space, the existing methods can be directly and successfully applied. This, however, is a quite ideal case. It is more likely that the historical data is only a small sub-set of the full scope of products. For this case, we should consider the following two scenarios: Scenario A, where the historical model (M_{method}) can be applicable to new operating space, and Scenario B, where the historical model can no longer be accurately applied in the new operating space. As the behaviour over the entire product range is typically nonlinear for many industrial processes, Scenario B can widely exist, which is the focus of this paper.

One-dimensional examples are given in Fig. 1 to illustrate the two scenarios, where Fig.1(a) is for Scenario A and Fig.1(b) is for Scenario B. The rectangle represents the historical operating region; solid line is the true model (f_*) over the entire operating space; the dashed line is the model (f_0) derived from the limited historical data. Generally, the true model, i.e., the global model, is nonlinear; while the historical model, i.e., the local model, may be linear for most industrial processes over a narrow operating space, as shown in Fig. 1(b).

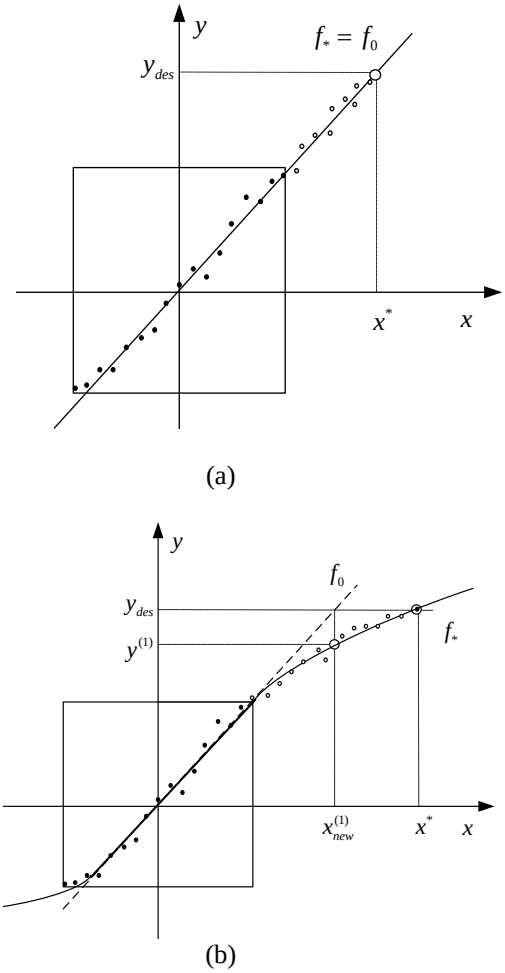


Fig. 1. Illustration of two scenarios in discovering new operating space
(a) Scenario A; (b) Scenario B.

For Scenario A, the local model can be directly applied to new operating space. But for Scenario B, the operating condition ($x_{new}^{(1)}$) obtained from the historical model (f_0) result in the actual quality ($y^{(1)}$), rather than the desired quality (y_{des}). To find the desired operating condition (x^*), relationship between product qualities and operating conditions in the new operating space is necessary. Product design in this case can be viewed as a coupled procedure of model updating. The challenge in such model updating lies in that, there is no data available in the unknown operating space. New experiments need to be designed in searching for the desired operating condition quickly and economically. New model in the desired local operating space can be typically represented by a linear model; a procedure needs to be developed for obtaining such a model to track the globally nonlinear process behaviour.

Based on the above analysis, a strategy is proposed in the next section to migrate the historical model to new operating space in the frame of PLS modelling for the aforementioned product design issue.

3. METHODOLOGY

In the following, we assume that, (1) a set of historical data is available, consisting of existing product quality (Y) and the corresponding settings on all manipulated process variables (X); (2) X and Y have been mean-centred and scaled; (3) a PLS model, introduced in section 3.1, has been built on X and Y, which can represent the relationship between X and Y in the historical data space; (4) new quality specification is feasible for the process, satisfying all physical constraints; (5) In the entire operating space of the process, the relationship between operating conditions and qualities changes mildly and continuously.

3.1 PLS

Partial Least Squares (PLS) is a popular regression method that can project high dimensional correlated process data down to a few number of latent variables and then model the latent variables by one-dimensional linear regression. It had many successful applications in process monitoring, fault detection and diagnosis, quality prediction, product design, etc. Mathematically, PLS is formulated by an outer relationship in X and Y (Eq.1) and an inner relationship between X and Y (Eq.2),

$$\begin{aligned} X &= T \cdot P^T + E = \sum_a \mathbf{t}_a \mathbf{p}_a^T + E \\ Y &= U \cdot Q^T + F = \sum_a \mathbf{u}_a \mathbf{q}_a^T + F \end{aligned} \tag{1}$$

$$\begin{aligned} \mathbf{u}_a &= \mathbf{t}_a \cdot b_a + r \\ b_a &= \mathbf{u}_a^T \mathbf{t}_a / \mathbf{t}_a^T \mathbf{t}_a \\ (a &= 1, \cdots, A) \end{aligned} \tag{2}$$

where $\{\mathbf{t}_a, \mathbf{u}_a\}$ is a pair of latent variables in X and Y spaces; $\{\mathbf{p}_a, \mathbf{q}_a\}$ are the corresponding loading matrices; T, U, P and Q are in the matrix form; E, F and r are model residuals; a is the index of latent variable; and A is the number of latent variables retained. The detailed PLS algorithm can be found in literature (Geladi and Kowalski, 1986; Höskuldsson, 1988).

3.2 Piecewise regression

Piecewise regression is a popular nonlinear regression method, where linear regression models over different regions are lumped together to approximate a globally nonlinear model. The simplest piecewise-regression model that joins two straight lines sharply at the changepoint is formulated as,

$$y = \begin{cases} f_1(x, \theta_1) & x \le \alpha \\ f_2(x, \theta_2) & x > \alpha \end{cases} \tag{3}$$

where the model parameters θ and the changepoint α are typically unknown and must be estimated. For details on piecewise regression, one can refer to the literature (Seber and Wild, 1989).

3.3 Iterative piecewise PLS method

For easy understanding, we shall first illustrate the iterative piecewise regression method using the one-dimensional example of Fig. 1(b). The first thing is to check the validity of the old model in a new operating space, where new experiment is unavoidably needed. With the assumptions, the operating condition of the first new experimental trial can be calculated by the historical model (f_0), e.g., $x_{new}^{(1)} = f_0^{-1}(y_{des})$, as illustrated in Fig. 2, where the desired quality data and all new experimental data in future are mean-centred and scaled by the same normalization parameters obtained from historical data. The actual quality value will be $y^{(1)} = f_*(x_{new}^{(1)})$, rather than y_{des}. If the difference $\left\| y^{(1)} - y_{des} \right\|$ is beyond the tolerance limit, that is, the old model is no longer valid for new data $(x_{new}^{(1)}, y^{(1)})$, piecewise regression can be performed here for modelling the relationship between operating conditions and qualities in the expanded operating space.

To do so, the changepoint and parameters of the new linear model have to be determined. With the assumption that the historical local model has reasonable performance for the historical data, and the newly obtained data that does not obey the historical model, the "boundary" point $(x_{b+} = f_0^{-1}(y_{b+}), y_{b+})$ can be chosen as the changepoint between the historical model $f_0(x)$ and a new linear model $f_1(x)$, where y_{b+} stands for the closest quality data in the historical envelope to the desired quality. The piecewise model that joins the above two models is formulated as,

$$y = \begin{cases} f_0(x) & x \in R^0 \\ f_1(x) & x \in R^1 \end{cases} \tag{4}$$

where R^0 is the historical data space; and R^1 is the new data space, e.g. $R^0 = [x_{b-}, x_{b+}]$ and $R^1 = [x_{b+}, \infty]$ for the example of Fig. 2. The linear model $f_1(x)$ can be roughly determined by the changepoint (x_{b+}, y_{b+}) and new experimental data $(x_{new}^{(1)}, y^{(1)})$. Although it is impossible to ensure the confidence of the new model derived only from the two data points from the statistical point of view, it can still be used to provide the right direction in searching for the desired operating condition from the application point of view, supposed that the new experimental data is clean and informative.

After the first experimental trial, there are two possible results in terms of the newly obtained quality data $y^{(1)}$. For case I, as illustrated in Fig. 2, $y^{(1)}$ is still smaller than the desired value, indicating that the optimal operating condition is still outside the newly explored space. Further experimental trial $(x_{new}^{(2)}, y_{new}^{(2)})$ is needed in the unknown space based on

the local model f_1, and a three-piece regression model will be derived with the new changepoint $(x_{new}^{(1)}, y^{(1)})$ as,

$$y = \begin{cases} f_0(x) & x \in R^0 \\ f_1(x) & x \in R^1 \\ f_2(x) & x \in R^2 \end{cases}, \quad (5)$$

where $R^0 = [x_{b-}, x_{b+}]$, $R^1 = [x_{b+}, x_{new}^{(1)}]$ and $R^2 = [x_{new}^{(1)}, \infty]$ for the example of Fig. 2.

For case II, as illustrated in Fig. 3, $y^{(1)}$ is larger than the desired value. In this case, we should search for the desired operating condition within the newly explored operating space. Similarly, a new experimental data $(x_{new}^{(2)}, y_{new}^{(2)})$ is obtained based on the model f_1. Then, the local model $f_1(x)$ will be replaced by two sub local models as,

$$y = \begin{cases} f_0(x) & x \in R^0 \\ f_{02}(x) & x \in R^{02} \\ f_{21}(x) & x \in R^{21} \end{cases}, \quad (6)$$

where $R^0 = [x_{b-}, x_{b+}]$, $R^{02} = [x_{b+}, x_{new}^{(2)}]$ and $R^{21} = [x_{new}^{(2)}, x_{new}^{(1)}]$ for the example of Fig. 3. The new linear models f_{02} and f_{21} are determined by the data $\{(x_{b+}, y_{b+}), (x_{new}^{(2)}, y_{new}^{(2)})\}$ and $\{(x_{new}^{(1)}, y_{new}^{(1)}), (x_{new}^{(2)}, y_{new}^{(2)})\}$, respectively.

The above piecewise regression modelling can be repeated further until the updated process model can approach the true relationship around the desired operating condition (x^*).

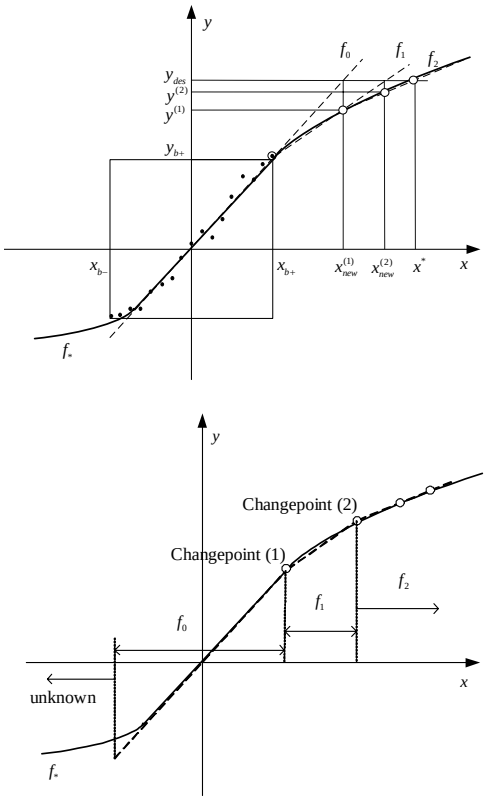


Fig. 2. Illustration of iterative piecewise regression modelling (Case I).

The above iterative piecewise modelling is simple and easy to be implemented in the one-dimensional case. For high dimensional and highly correlated industrial data, the above procedures can also be implemented with the aid of PLS modelling. In the implementation, the outer relationships of the historical PLS model will be kept unchanged, only the inner relationship is updated to track the nonlinearity in the expanded operating space, where piecewise regression is performed in each pair of latent variables $\{\mathbf{t}_a, \mathbf{u}_a\}$ $(a = 1, \cdots, A)$. This is feasible as discussed by Qin & McAvoy (1992) in their nonlinear PLS method, where neural network is used to approximate the nonlinear inner relationship and the outer relationships are kept in linear structure to attain the robust generalization property.

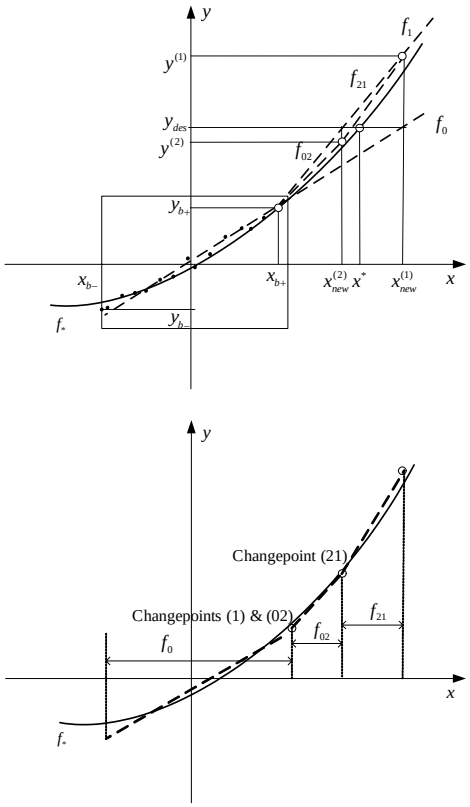


Fig.3. Illustration of iterative piecewise regression modelling (Case II).

The proposed iterative piecewise PLS modelling can be summarized as below.

Step I: Use the historical PLS model to get a new operating condition ($\mathbf{x}_{new}^{(1)}$), the corresponding quality data ($\mathbf{y}^{(1)}$), and the latent variable scores $\mathbf{t}_{new}^{(1)}$ and $\mathbf{u}_{new}^{(1)}$, as shown in Table 1.

Step II: Keep the outer relationship unchanged, piecewise updating the inner relationship to track the nonlinearity in the expanded operating space similar to Eq.(5) or Eq.(6).

Step III: Repeat step I & II using the historical PLS outer model and the updated inner model until achieving the desired operating conditions and quality.

Based on the updated nonlinear PLS model over the expanded operating space, the existing product

design methods can be applied to find the feasible operating conditions for new quality specifications, as illustrated in the next section.

Table 1. Procedures in the first step of iterative piecewise PLS modelling method

Historical PLS model parameters:
Outer model: P_0 and W_0 for X;
Q_0 and C_0 for Y
Inner model: B_0 between T and U

Step I of recursive piecewise PLS method:

(1) $\mathbf{u}_{des} = \mathbf{y}_{des} C_0$;

(2) $\mathbf{t}^{(1)} = \mathbf{u}_{des} (B_0)^{-1}$;

(3) $\mathbf{x}_{new}^{(1)} = \mathbf{t}^{(1)} P_0^T$;

(4) Modify $\mathbf{x}_{new}^{(1)}$ to satisfy physical constraints;

(5) Do experiment under $\mathbf{x}_{new}^{(1)}$ to get quality data $\mathbf{y}^{(1)}$;

(6) $\mathbf{t}_{new}^{(1)} = \mathbf{x}_{new}^{(1)} W_0$ and $\mathbf{u}_{new}^{(1)} = \mathbf{y}^{(1)} C_0$, $\{ \mathbf{t}_{new}^{(1)}, \mathbf{u}_{new}^{(1)} \}$ for step II.

4. ILLUSTRATION

The proposed iterative piecewise PLS modelling method for product design is applied to an injection molding process to demonstrate its feasibility and effectiveness.

Injection molding process can be operated over a wide range of operating conditions. For the machine in our lab, when processing high-density polyethylene (HDPE), the normal settings of Packing Pressure (P.P.), Barrel Temperature (B.T.) and Mold Temperatures (M.T.) can be within the ranges of 150bar ~ 450bar, 180 ºC ~ 220 ºC, and 15 ºC ~ 55 ºC, respectively. The relationship between these settings and dimensional qualities such as part weight and length can be accurately described by the first pair of PLS latent variables, as detailed in the authors' previous work (Lu & Gao, 2005). It is clearly nonlinear over the entire operating space, as shown in Fig. 4.

To illustrate the proposed method, data from 9 different operating conditions are collected to form the "historical data", as shown in Table 2, where the ranges of product weight and length are 23.36g ~ 27.41g and 116.67cm ~ 117.27cm, respectively. The linear PLS model derived from these historical data have good performance in quality prediction, as shown in Fig. 5. A new product quality specification, weight= $27.86g$ and length= $117.52cm$, is required now, which is beyond the range of historical products, but achievable on the machine. The results of the proposed iterative piecewise PLS method are shown in Fig. 6 and Table 2, where the method of Jaeckle & MacGregor (1998) is adopted to invert the PLS model to find the corresponding operating conditions.

From Fig. 6, the final PLS model has three-piece inner relationships (B_0 , B_1 , and B_2), and the desired operating conditions for the new quality setting can be achieved by the third inner model (B_2). Only two trials are conducted by the proposed method in searching for the right operating condition to achieve the desired product qualities. This obviously reduces the effort and time in designing new products.

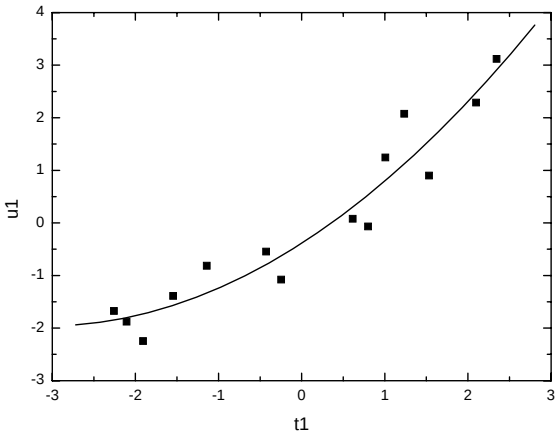
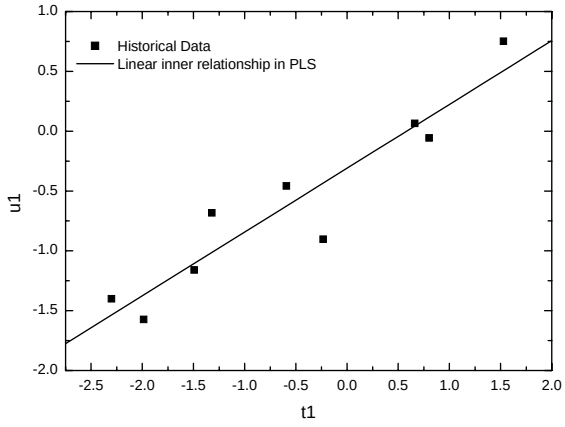
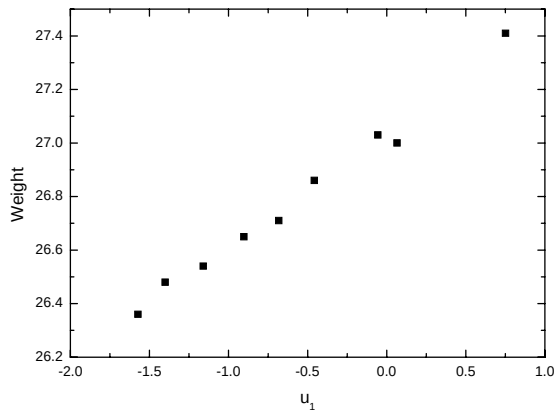


Fig.4. Nonlinear relationship illustration by the first pair of PLS latent variables over the entire operating space.



(a)



(b)

Fig. 5. Illustration of goodness of the historical linear PLS model in the historical operating space.
(a) Linear inner relationship;
(b) Linear outer relationship in Y.

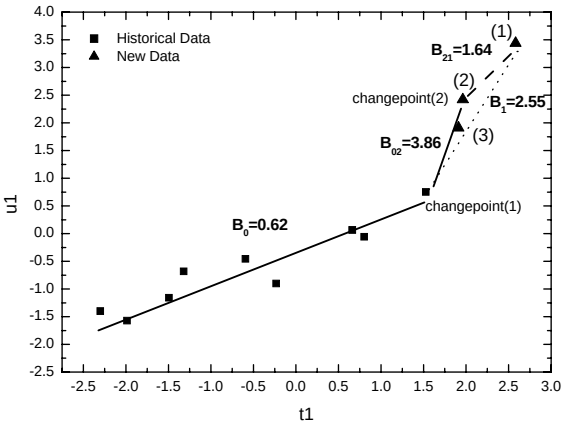


Fig. 6. Iterative piecewise PLS modelling of the inner relationship on the first pair of latent variables.

5. CONCLUSION

Analysis on the product design into unknown operating space has been given in the paper. An iterative piecewise PLS modelling method has been adopted for the above purpose. The application on an injection molding process has demonstrated good feasibility and effectiveness of the proposed method.

ACKNOWLEDGEMENT

This work is supported in part by Hong Kong Research Grant Council, project number 601104.

REFERENCE

Chen, F.Z. and Wang, X.Z. (2000). Discovery of operational spaces from process data for production of multiple grades of products. *Ind. Eng. Chem. Res.*, 39, 2378

Geladi, P. and Kowalshi, B.R. (1986). Partial least squares regression: a tutorial. *Analytica Chimica Acta*, **185**, 1.

Hoskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, **2**, 211.

Jaeckle, C.M. and MacGregor, J.F. (1998). Product design through multivariate statistical analysis of process data. *AIChE J.*, **44**, 1105.

Jaeckle, C.M. and MacGregor, J.F. (2000). Industrial applications of product design through the inversion of latent variable models. *Chemometrics and Intelligent Laboratory Systems*, **50**, 199.

Lakshminarayanan, S., Fujii, H., Grosman, B., Dassau, E. and Lewin, D.R. (2000). New product design via analysis of historical databases. *Computers and Chemical Engineering*, **24**, 671.

Lu, N.Y. and Gao, F.R. (2005). Stage-based process analysis and quality prediction for batch processes. *Ind. Eng. Chem. Res.*, **44**, 3547.

Moteki, Y. and Arai, Y. (1986). Operation planning and quality design of a polymer process. *IFAC Control of Distillation Columns and Chemical Reactors (DYCORD), Bournemouth, UK*, 159.

Qin, S.J. and McAvoy, T.J. (1992). Nonlinear PLS modeling using neural networks. *Computers chem. Engng.*, **16**, 379.

Seber, G.A.F. and Wild, C.J. (1989). Nonlinear regression. *John Wiley & Sons*.

Sebzalli, Y.M. and Wang, X.Z. (2001). Knowledge discovery from process operating data using PCA and fuzzy clustering. *Engineering Applications of Artificial Intelligence*, **14**, 607.

Table 2. Operating conditions, scores of latent variables and quality measurements for historical and new experimental data.

	Operating conditions (P.P., B.T., M.T.)			PLS scores (t1, u1)		quality measurements (Weight, Length)	
				Historical Data			
	150	180	15	-0.59	-0.46	26.86	116.96
	150	180	35	-1.32	-0.68	26.71	116.93
	150	180	55	-2.30	-1.40	26.48	116.69
	300	200	35	0.66	0.06	27.00	117.15
	300	200	55	-0.23	-0.90	26.65	116.85
	120	220	15	-1.49	-1.16	26.54	116.78
	450	220	15	1.53	0.75	27.41	117.27
	150	220	35	-1.99	-1.57	26.36	116.67
	450	220	35	0.80	-0.06	27.03	117.07
				New Experimental Data			
(1)	450	189	15	2.58	3.44	28.17	117.70
(2)	450	193	25	1.96	2.42	27.94	117.56
(3)	440	196	32	1.91	1.91	27.86	117.52

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ADAPTIVE CONTROL OF BROMELAIN PRECIPITATION IN A FED-BATCH STIRRED TANK

Flávio Vasconcelos da Silva
Regina Lúcia de Andrade dos Santos
Ana Maria Frattini Fileti

State University of Campinas (UNICAMP), School of Chemical Engineering, Department of Chemical Systems Engineering, CP 6066, CEP13083-970, Campinas, SP, Brazil
e-mail: flavio@feq.unicamp.br

Abstract: In this work, bromelain is recovered from triturated pineapple stem and rind (usually kitchen waste) through the precipitation process with alcohol at low temperature. The temperature control is crucial to avoid the irreversible protein denaturation and consequently improving the precipitation yield. The process is carried out in a fed-batch system, so that its dynamic nature poses challenging control system design. Conventional and adaptive controllers are properly designed and on-line implemented through a fieldbus digital control system. Closed loop performance was improved under adaptive PID controller. Overshoot and response time decreased and no control action saturation occurred. *Copyright © 2002 IFAC.*

Keywords: Adaptive control, bromelain, enzyme precipitation, PID Controller, fieldbus.

1. INTRODUCTION

In many biotechnological industries, including food and pharmaceutical ones, the selective separation of a protein out of fermentation broths or vegetable sources has been a primary research interest for downstream processing operations. It is difficult and expensive to selectively recover a targeted protein from a broth due to the low protein concentration and the similarity of the physical properties between proteins present in the same solution.

Among the practical methods being applied to the large-scale recovery and purification of proteins from dilute solution, protein precipitation is regarded as a key operational process, which is used during the early stages of the downstream processing. Protein precipitation is frequently featured by the spontaneous fractionation and concentration as well as a low additive consumption and protein non-denaturation (Kim *et al*., 2002; Chen and Berg, 1993; Clark and Clatz, 1987). Protein precipitation usually produces insoluble protein by contacting precipitants, such as neutral salts, acids, organics solvents, or metallic ions, with the desired protein in a stirred tank.

The need for rapid monitoring in biotechnology has been highlighted by several authors (Paliwal *et al*., 1993; Ransohoff *et al*., 1990). The high proportion of the process cost attributed to the downstream processing steps for many products, means that ensuring that the purification sequence is performed within specified limits at high yield, and knowing rapidly when such limits are crossed is an extremely important consideration (Foster *et al*., 1986). Particularly, for enzyme precipitate products, quality control is mandatory to ensure structural authenticity. If, in addition, prior process knowledge allows feedback control using on-line information, then costly run-stoppage times or disposal of non-specification material may be avoided (Holwill *et al*., 1996).

While fermentation control problem is well addressed, the practical application of control in downstream processing has not been properly studied. Most of bioprocesses are carried out in batch or semi-batch systems, so that their dynamic nature poses challenging control system design. Non-linearity is usually found as well, therefore conventional feedback controllers are not supposed to be able to follow set point specifications.

The present work is concerned about the experimental control system development for fruit bromelain precipitation. Bromelain is the name of a group of powerful protein-digesting, or proteolytic, enzymes that are found in the pineapple plant (Ananas comosus). Discovered in 1957, and widely studied since then, bromelain is particularly useful for reducing muscle and tissue inflammation and as a digestive aid. Besides the pharmacological effects, bromelain is also employed in food industries, such as breweries and meat processing.

It is a fact that the majority of processes in the chemical industries can be satisfactorily controlled using conventional controllers. However, the conventional PID control method is inadequate for bioprocesses control in which processes dynamics will change in known ways during operation (time

delays, process non–linearities and interactions). For these difficult problems, it is important to generate the initial settings with a model-based tuning strategy. Controller tuning involves the selection of the best values of Kc, Ti and Td. This is often a subjective procedure and is certainly process dependent. A number of methods have been proposed in the literature over the last 50 years. However, the most well-known tuning technique is the method of Ziegler and Nichols.

In this work, bromelain is recovered from triturated pineapple stem and rind (usually kitchen waste) through the precipitation process with alcohol at low temperature. The process temperature control is crucial to avoid the irreversible protein denaturation and consequently improving the precipitation yield. Conventional and adaptive controllers are properly designed and on-line implemented through a fieldbus digital control system.

2. THE PRECIPITATION PROCESS

The precipitation process aims to achieve separation by conversion of solutes to solids. Precipitation can result in both concentration and purification methods. The advantages of using precipitation for concentration and purification are: easy scale-up, involves simple equipments and can be based on a large number of alternative precipitants, some of them inexpensive or used in very low concentration. Precipitants can be chosen which do not denature biological products, and the precipitate form is often more stable than the solute material.

Solubility of Proteins

A larger number of water-miscible organic solvents like ethanol or methanol can be used to precipitate proteins. A typical globular protein presents to the solvent a surface consisting of regions of positive and negative charge, along with polar, but uncharged, hydrophilic regions and nonpolar, hydrophobic regions. The complex interactions between the protein surface and surrounding solvent determine the solubility. A protein is made insoluble by changing either the surface characteristics of the protein itself or by changing the solvent. The change in solubility that results is sufficiently great to be viewed as step change between soluble and insoluble. The resulting high level of supersaturation leads to rapid formation of an amorphous solid.

As in the case of salting-out, this phenomenon has been described in terms of removal of water from the hydration spheres of the protein allowing electrostatic forces to bring oppositely charged regions of the protein together. Water is removed both by bulk replacement by the organic solvent and by structuring of the water around the organic molecules. The solvent property affected is the dielectric constant. The hydrophobic area of the protein would tend to become more soluble, but the net result is decrease in solubility.

Fed-Batch Precipitation Tank

The bromelain precipitation process is carried out in a fed-batch stirred tank (500mL), according to Figure 1. Inside the jacked tank, the protein will be exposed to a range of operating conditions during the period in which the precipitant agent (ethanol 99.5%) is added. The first material precipitated will form at other than the final conditions. Overprecipitation is less likely to occur by adding precipitant slowly and well dispersing it. A micropump (pump 1) is employed to continuously feed the alcohol, at environmental temperature (about 25ºC), to the tank.

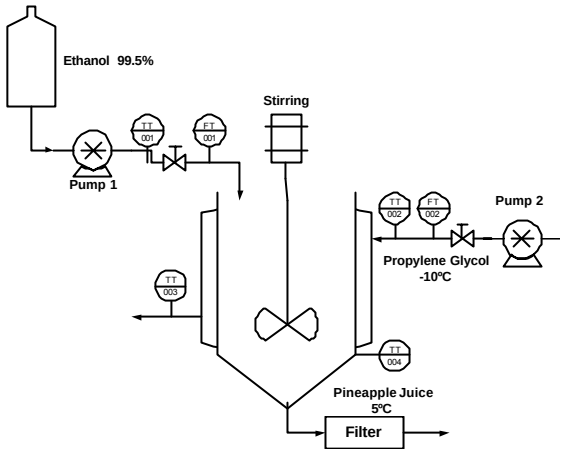


Fig 1. Fed-batch stirred precipitation tank.

Since the upper bound for bulk temperature is 10ºC, in order to avoid protein denaturation process, the cooling flow rate is manipulated through a variable speed pump (pump 2). Calibrated J-type thermocouples, located at precipitation mixture bulk and at the inlet and outlet of cooling fluid (propylene glycol), provide temperature measurements. The set point of bulk temperature is 5ºC.

3. THE DIGITAL CONTROL SYSTEM

The management of the digital control system is performed through a Foundation Fieldbus communication system, according to Figure 2. Four field devices compose the fieldbus network used to monitor and control the precipitation tank:

- Distributed Fieldbus Interface (DFI302): bridge to link different speed networks. It manages the communication between the Local Area Network (High Speed Ethernet) and the Fieldbus network (H1);

- Temperature Transmitters (TT302-1 and TT302-2): perform temperature data acquisitions and transmit them to the interface (DFI302);

- Fieldbus/Electric Current Converter (FI302): receives digital signal from DFI302 and converts to 4-20mA to operate the variable speed pump.

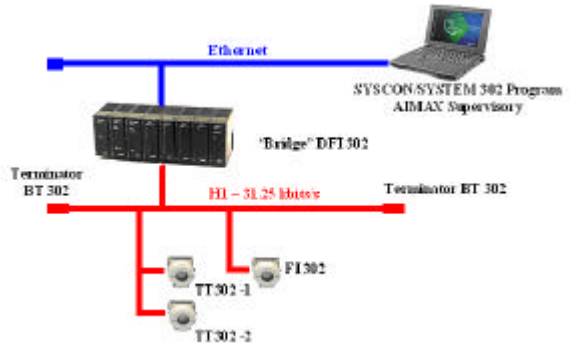


Fig. 2. Fieldbus network for bromelain precipitation tank monitoring and control.

The digital signal sent to the pump is computed by controllers (conventional or adaptive ones), which were implemented through the software Syscon (function blocks) and supervisory software Aimax (adaptive tuning equations).

4. MATERIAL AND METHODS

Pineapple juice preparation

Stem and rind of the pineapple fruit (species "Perola") are triturated and filtered. Distilled water is employed at dilution rate 1:1. The filtrate, called pineapple juice, contains the enzyme bromelain. Samples containing 100mL of pineapple juice are frozen at -18ºC (Cesar *et al*., 1999).

Controller tuning

Tuning PID systems is both a science and an art form. Proper parameters depend entirely on the system being tuned. Scaling conventions will differ depending on a particular PID implementation. When tuning a PID algorithm, generally the aim is to match some preconceived ideal response profile for the closed loop system.
It has long been recognized that first-order-plus-dead (FOPDT) approximations can adequately explain the behavior of a range of processes:

$$G(s) = \frac{K_p e^{-tds}}{t_p s + 1} \qquad (1)$$

Where Kp is the process gain, tp the process time constant and the td is the process time delay.
The Zeigler Nichols Open-Loop Tuning Method is a way of relating the process parameters to the controller parameters.
The process reaction procedure was carried out at five different tank levels. Pseudo-steady state was reached before step disturbances were implemented in cooling flow rate. Process parameters (gain, time constant and time delay) are then graphically obtained from the monitored tank temperature response (TT-004 – Figure 1). Due to the transient nature and process non-linearity, these parameters are not constant. In order to improve the design of the PID linear controller, Ziegler-Nichols tuning equations (Ogata, 1997) are applied for every pseudo-steady state studied.

Accurate tuning is hardly possible using trial and error method. The classical methods for controller settings can not be precise enough and it is expected the controller function can be improved considerably if they are better tuned.
From this methodology, an adaptive PID controller is obtained, implemented and compared to a well-tuned conventional PID.
A comprehensive summary of adaptive control was recently published by Sastry and Bodson (1994).

5. RESULTS

From the process reaction curve procedure, implemented at five different mixture volumes (100, 200, 300, 400 and 500 mL), and Ziegler-Nichols tuning equations, the scheduling relationships for PID controller parameters - Kc, Ti and Td - were determined (Figure 3).

The tuned and well tuned parameters obtained from Ziegler and Nichols method and fine tuned method (trial and error), respectively, are shown in Table 1.

Table 1. Tuned and well-tuned parameters (Kc, Ti and Td) obtained at different tank volume levels using Zigler and Nichols method and fine tuning method (trial and error).

Volume (mL)	Tuned			Well-tuned		
	Kc	Ti	Td	Kc	Ti	Td
100.0	433.1	26.0	6.5	101.6	156.0	3.3
200.0	223.1	44.0	11.0	52.8	264.0	5.5
300.0	159.9	74.0	18.5	37.4	444.0	9.3
400.0	144.8	78.0	19.5	34.2	468.0	9.8
500.0	110.9	112.0	28.0	26.7	672.0	14.0

From Figure 3, it is observed that the controller gain decreases as soon as tank volume increases. Indeed the process becomes more sensitive since the heat transfer area increases and consequently a small control action is required to regulate bulk temperature. Still according to PID Ziegler-Nichols tuning, the integral action must decrease and the effect of derivative term must increase as volume increases.

Experimental runs were carried out by loading the stirred tank with 100mL of pineapple juice at 5ºC and continuously adding ethanol 99.5%, at environmental temperature, until the tank volume reaches 500mL.

Since ethanol pump operates at fixed flow rate (0.18mL/s) and the batch time is known, the tank volume is computed and the obtained tuning relationships (Figure 3) are on-line applied in the adaptive PID controller.

Although similar behaviour is obtained from conventional PID controller application (Figure 5), undesirable control action saturation occurred. Fixed tuning parameters obtained at the early stage of the batch (volume: 100mL) were employed.

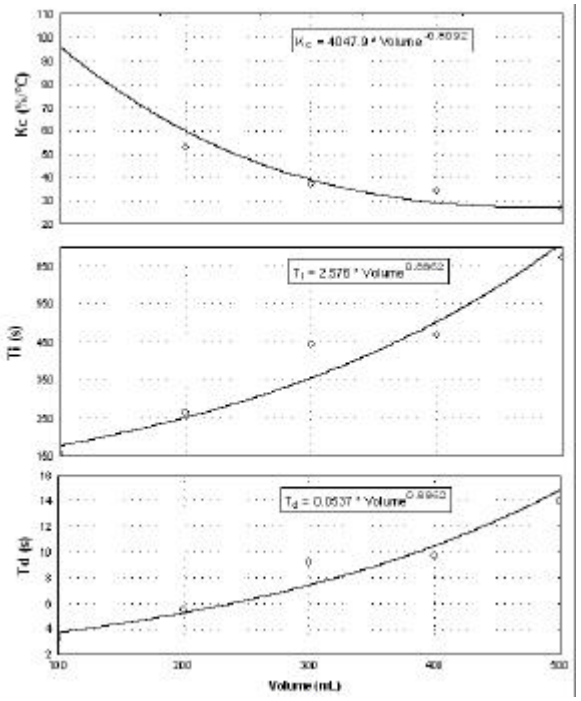


Fig. 3. Experimental values of controller gain (Kc), integral time (Ti) and derivative time (Td) obtained at different tank volume levels.

Figure 4 presents the results from the adaptive controller implementation. The behaviour of the controlled variable (bulk temperature deviation) and the manipulated variable (pump speed) is shown. An open-loop run was also carried out (pump speed equals to 50%). In the early moments of ethanol addition, there is a heat generation that immediately cause an increase on pump speed in the closed-loop system. Soon afterwards, the temperature decreases and pump speed leads to a minimum cooling flow rate. Besides precipitation heat generation, the ethanol inlet at environmental temperature adds heat continuously to the tank bulk mixture.

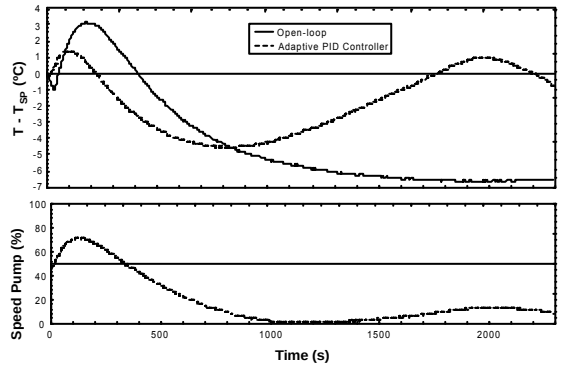


Fig. 4. Bulk temperature deviation (T-Tsp) and manipulated variable behavior obtained from experimental runs: open-loop and adaptive PID implementation.

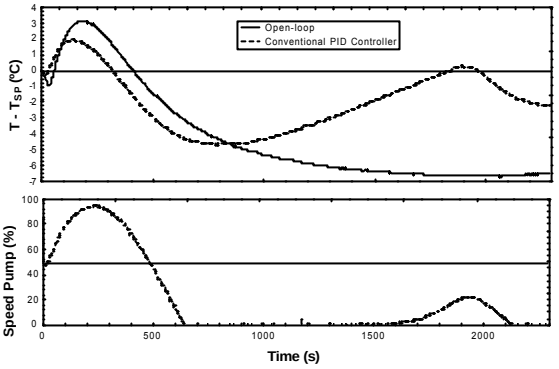


Fig. 5. Bulk temperature deviation (T-Tsp) and manipulated variable behavior obtained from experimental runs: open-loop and conventional PID implementation (Kc= 101.6 %/ºC, Ti= 156.0s, Td= 3.2s).

In order to compare conventional and adaptive PID controllers, performance criteria are computed from the experimental runs of Figures 4 and 5 and summarized on Table 2.

Table 2. PID and PID adaptive performance criteria from experimental runs.

Performance criteria	Controller	
	Adaptive PID	Conventional PID
ITAE	4,736,074	5,346,902
Overshoot (ºC)	1.3	1.9
Rise time (s)	220	320

From the results, it could be concluded that both controllers are suitable for the precipitation tank control. However, from Table 1, the adaptive controller showed better global performance criterion (ITAE) and reduced overshoot and rise time.

6. CONCLUSION

In the present work, the pineapple juice (raw material) is obtained from the stem and rind of the fruit, both considered as kitchen waste in most brazilian restaurants and industries. The precipitation process is usually the first step of the downstream processing, and this was the practical method adopted here in order to recover bromelain enzyme from the pineapple juice.

Looking forward to avoiding bromelain denaturation during the precipitation process, the temperature must be monitored and controlled. Due to the transient nature and non-linear features of bromelain precipitation fed-batch process, designing a temperature controller for the experimental apparatus is not a trivial task.

Conventional and adaptive controllers were properly designed and on-line implemented through a fieldbus digital control system. From the results, both controllers were considered suitable to control the temperature of the fed-batch tank. However, the performance was improved when adaptive controller

was employed, since no saturation on control action occurred and the overshoot and rise time decreased.

REFERENCES

Chen, W., and Berg, J. C. (1993). The effect of polyelectrolyte dosage on floc formation in protein precipitation by polyelectrolytes. *Chemical Engineering Science*, 48, 1775-1784

Clark, K. M. and Clatz, C. E. (1987). Polymer dosage considerations in polyelectrolyte precipitation of protein. *Biotechnology Progress*, 3, 241-247

Cesar, A. C., Silva, R. and Lucarini, A. C. (1999). Recuperação das enzimas proteolíticas presentes na casca e no talo do abacaxi. *Revista de Iniciação Científica – São Carlos/SP*.(1), 47-53.

Foster, P. R., Dickson, A. J., Stenhouse, A. and Walker, E. P.,(1986). A process control system for the fractional precipitation of human plasma proteins. *J. Chem. Tech. Biotechnol*. 36: 461-466

Holwill, I., Gill, A., Harrison, J., Hoare, M.and Lowe, P.A. (1996). Rapid analysis of biosensor data using initial rate determination and its application to bioprocess monitoring. *Process Control and Quality* 8 (4): 133-145.

Kim, W.-S., Hirasawa, I., Kim, W.-S. (2002). Aging characteristics of protein precipitates by polyelectrolyte precipitation in turbulently agitated reactor. *Chem. Eng. Science* 57, 4077-4085

Kim, W.-S., Hirasawa, I., Kim, W.-S. (2001). Eeffects of experimental conditions on the mechanism of particle aggregation in protein precipitation by polyelectrolytes with a high molecular weight. *Chem. Eng. Science*, 56, 6525-6534.

Locher, G, Sonnleitner, B., Fiechter, A. (1992). On-line measurement in biotechnology: Techniques. *J. Biotech*. 25: 23-53

Ogata, K. Modern Control Engineering, 3rd edition, 1997 (Prentice-Hall, Upper Saddle River, New Jersey).

Paliwal, S. K. Nadler, T. K., Regnier, F. E.(1993) Rapid process monitoring in biotechnology. *Tibtech* 11: 95-101

Ransohoff, T C., Murphy, M K., Levine, H. L. 1990. Automation of biopharmaceutical process. Biopharm.(March):20-25

Sastry, S., Bodson, M. Adaptive Control: Stability, Convergence, and Robustness Prentice-Hall Advanced Reference Series (Engineering), 1994.

Session 7.2

Control of Complex Systems

Distributed Model Predictive Control of a Four-Tank System

M. Mercangöz and F. J. Doyle III
University of California, Santa Barbara

Coordinated Decentralized MPC for Plant-Wide Control of a Pulp Mill Benchmark Problem

R. Cheng, J. F. Forbes, and W. S. Yip
University of Alberta

Optimizing Hybrid Dynamic Processes by Embedding Genetic Algorithms into MPC

T. Tometzki, O. Stursberg, C. Sonntag, and S. Engell
Dortmund University

Optimal Control of Multivariable Block Structured Models

G. Harnischmacher and W. Marquardt,
RWTH Aachen University

Operability of Multivariable Non-Square Systems

F. Lima and C. Georgakis,
Tufts University

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DISTRIBUTED MODEL PREDICTIVE CONTROL OF A FOUR-TANK SYSTEM

Mehmet Mercangöz
Francis J. Doyle III

Department of Chemical Engineering,
University of California, Santa Barbara, CA 93106, USA

Abstract: A distributed model predictive control (DMPC) framework is proposed. The physical plant structure and the plant mathematical model are used to partition the control duties over self-sufficient estimation and control nodes. Linear models and local measurements at the nodes are used to estimate the relevant plant states. This information is then used in the model predictive control calculations. Communication among relevant nodes during estimation and control calculations provides improvement over the performance of completely decentralized controllers. The DMPC framework is demonstrated for the level control of an interacting four-tank system. The performance of the DMPC system for disturbance rejection is compared with other control configurations. The results indicate that the performance of the proposed framework provides significant improvement over completely decentralized MPC controllers, and approaches the performance of a fully centralized design.

Keywords: Distributed decentralized estimation and control, model based control, network control, plantwide control.

1. INTRODUCTION

Efficient plantwide control of chemical processing plants provides a significant economic advantage by enabling closer operation to optimization constraints, decreasing the number of shut-downs and by reducing the amount of off-specification products. Efficient control systems can also provide environmental and operational safety for these chemical plants.

Control systems in typical modern chemical plants are built in a hierarchical structure, where a large number of digital PI, PID and other simple controllers enable stable operation of most unit operations. These controllers are then connected to multivariate systems spanning several unit operations to control the important quality variables or to achieve more sophisticated tasks such as waste minimization, economic optimization or production scheduling as shown in figure 1 (Skogestad, 2004). The information flow in these hierarchical structures is in a vertical direction and the systems at the same level are not aware of the existence of their neighbors even though they may be interacting.

The objective of this paper is to develop a framework for a horizontal connection among the different control systems in a chemical plant at the multivariate control level. The industry standard for multivariate control is model predictive control (MPC) and the proposed framework provides a communication structure for estimation and control among different distributed MPC applications.

Decentralized estimation and control problems have attracted attention from several different fields. The Control of vehicle formations or a group of robots in mechanical or aerospace engineering, control of power grids in electrical engineering and coordination of wireless systems in computer science are examples of these problems. Achievements, especially in the field of multivariate estimation and control, include the development of parallel partially decentralized controllers (Siljak, 1991). Multi-level hierarchical control systems have been designed based on decomposition and coordination strategies (Findeisen *et al*., 1980). A fully distributed decentralized estimation and control structure (DDEC) to achieve the same performance of a centralized algorithm under certain conditions have also been developed (Mutambara, 1998). This method has been successfully applied to a chemical engineering plantwide control problem for a state-feedback control law (Vadigepalli and Doyle III, 2003). Distributed approaches to MPC applications have also been investigated (Jia and Krogh, 2001). However, these approaches involved assumptions on the worst-case interactions to design a stabilizing hierarchical control strategy.

In the present study, a nodal estimation network is designed as an extension on the scalable DDEC methodology of Mutambara (1998), however the simple state feedback based control structure of the DDEC is replaced with an MPC algorithm. The nodal communication of state information in the original DDEC methodology is preserved and extended to include the communication and renewal of MPC results among relevant controllers before the implementation of control action. Additional requirements regarding model decomposition are also considered for the design of a DMPC network.

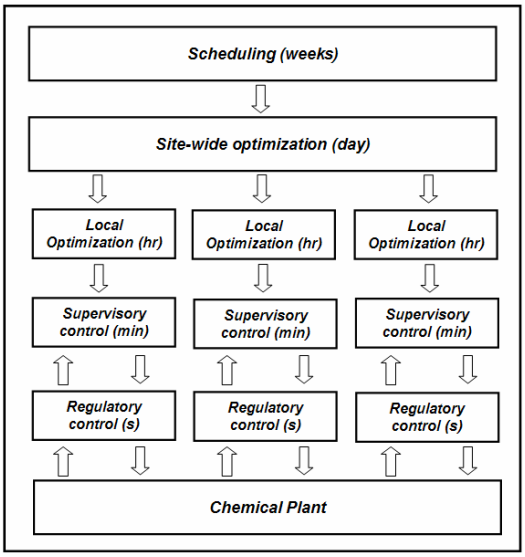


Fig.1. Hierarchical organization of process control systems in chemical plants.

In the following sections the DDEC methodology is revisited and the DMPC framework is outlined. The application of the proposed framework to an interacting four-tank system is provided. The simulation results to compare the efficiency of the DMPC algorithm to conventional centralized and completely decentralized MPC formulations precede a discussion of these results and a conclusion section.

2. DDEC METHODOLOGY

The DDEC methodology is based on a linear discrete-time plant with n_u inputs and n_y outputs of the following state–space form:

$$x(k)=\Phi x(k-1)+Bu(k-1)+w(k-1) \quad (1)$$

$$y(k)=Hx(k)+v(k) \quad (2)$$

where $x(k)\in\Re^n$ is the n-dimensional state of interest at time k; $\Phi:\Re^n\to\Re^n$ is the state transition matrix from time $(k-1)$ to k; $u(k)\in\Re^{n_u}$ and $B:\Re^n\to\Re^{n_u}$ are the input vector and matrix, respectively; $y(k)\in\Re^{n_y}$ is the measurements vector at time k; $H:\Re^n\to\Re^{n_y}$ is the observation matrix; and, $w(k)\sim N(0,Q)$ and $v(k)\sim N(0,R)$ are the associated process and measurement noise vectors, respectively, and are modelled as uncorrelated, zero mean sequences with covariance matrices Q and R, respectively.

2.1 Model Decomposition

The plant model given in (1) and (2) is partitioned according to either physical structure or based on an analysis of the mathematical model. At this stage the number of DDEC nodes is determined along with the allocation of measurements and control inputs among different nodes. Even though two nodes can share measurements, a control input cannot be assigned to more than a single node. The number of nodes should also be chosen carefully to evenly distribute the computational requirements and the communication load due to overlapping states. Linear transformations T_i for each node i are then designed to obtain the local state transition and observations given by:

$$x_i(k)=\Phi_i x_i(k-1)+B_i u_i(k-1)+w_i(k-1) \quad (3)$$

$$y_i(k)=C_i x_i(k)+v_i(k) \quad (4)$$

where $u_i(k)$ are the inputs affecting the local states. Φ_i is related to the global state transition matrix $\Phi(k)$ as $\Phi_i=T_i\Phi(k)T_i^{\dagger}$, where $T_i^{\dagger}$ is the generalized inverse of T_i. The local state vector at node i, $x_i(k)$, is related to the global state vector $x(k)$ by $x_i(k)=T_i x(k)$.

2.2 Distributed Prediction and Estimation

The prediction and estimation calculations at every time step are done according to the distributed and decentralized Kalman filter (DDKF). The state $x_i(k)$ at node i, is predicted according to:

$$\hat{x}_i(k|k-1)=\Phi_i\hat{x}_i(k-1|k-1)+B_i u_i(k-1) \quad (5)$$

$$P_i(k|k-1)=\Phi_i P_i(k-1|k-1)\Phi_i^{\mathrm{T}}+Q_i \quad (6)$$

where Q_i and R_i represent the local covariance matrices of process and measurement noise, respectively. The estimation step follows in three stages: (i) local estimation, (ii) internodal communication and (iii) assimilation to produce a global estimate. The Local covariance and state estimates are computed from local measurements as follows:

$$P_i(k|y_i(k))=[C_i^T R_i^{\dagger}C_i]^{\dagger} \quad (7)$$

$$\hat{x}_i(k|y_i(k))=P_i(k|y_i(k))[C_i^T R_i^{\dagger}]y_i(k) \quad (8)$$

The relevant subset of local estimates of the state and prediction error covariances are communicated to relevant nodes and the information at each node is transformed into the local state subspace. The transformed state and covariance estimates are given by

$$P_i^{\dagger}(k|y_j(k))=T_i[T_j^T P_j^{\dagger}(k|y_j(k))T_j]^{\dagger}T_i^T \quad (9)$$

$$\hat{x}_i(k|y_j(k))=T_i T_j^{\dagger}\hat{x}_j(k|y_j(k)) \quad (10)$$

The transformed states are assimilated locally to produce state and covariance estimates according to:

$$\hat{x}_i(k|k) = P_i(k|k)[P_i^{-1}(k|k-1)\hat{x}_i(k|k-1) + \sum_{j=1}^{N} P_i^{\dagger}(k|y_j(k))\,\hat{x}_i(k|y_j(k))] \quad (11)$$

$$P_i(k|k) = [P_i^{-1}(k|k-1) + \sum_{j=1}^{N} P_i^{\dagger}(k|y_j(k))]^{\dagger} \quad (12)$$

This combined process of local prediction, inter-nodal communication and assimilation among N nodes produces estimates identical to those obtained from an equivalent centralized Kalman filter algorithm.

2.3 Distributed Control

A nodal control law obtained as a cost minimizing control function is given by

$$u_i(k) = K_{ci}[x_{ri}(k) - \hat{x}_i(k|k)] \quad (13)$$

where $x_{ri}(k)$ is the local state reference, $\hat{x}_i(k|k)$ is the local optimal state estimate, and K_{ci} is the optimal control gain computed from the solution to a distributed and decentralized backward Riccati recursion. The prediction, estimation and control stages of the DDEC algorithm are shown in figure 2.

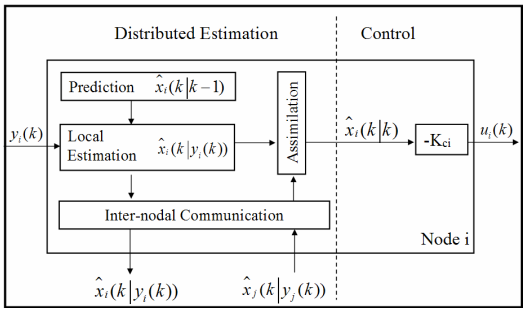


Fig.2. Organization of the distributed estimation and state-feedback based control in the DDEC algorithm.

3. DMPC FRAMEWORK

The DMPC framework relies on a similar model decomposition structure as given in (1) to (4). However, local models are developed by explicitly indicating interactions due to control moves from *n* neighboring nodes given as

$$x_i(k) = \Phi_i x_i(k\text{-}1) + B_i u_i(k\text{-}1) + \sum_{j=1}^{n} B_j u_j(k\text{-}1) + w_i(k\text{-}1) \quad (14)$$

$$y_i(k) = C_i x_i(k) + v_i(k) \quad (15)$$

The DMPC framework also requires "self-sufficiency" of local subsystems, meaning that every subsystem will be able to estimate the local states and achieve the local control objectives with the measurements and control inputs allocated to that node. This translates into the requirement that every subsystem will be observable with the allocated measurements and controllable with the assigned control inputs. "Self-sufficiency" will enable successful operation of a node in the case of a failure in other nodes or during an intermission in the communication structure. In the present study, it is assumed that an off-line Kalman filter with innovation gain M_i is available at every node prior to the start of the algorithm based on the local nodal models (14) and (15), and also on the expected disturbances. There are no requirements on the connectivity of the nodal DMPC network. The states of the local models x_i are separated into two sets, x_i^o and x_i^{no}, denoting the overlapping and non-overlapping components respectively. Moreover, every node *i* has reliability factors $r_{i(l)}^j$ for any overlapping state *l* of x_i^o coming from a relevant node *j* as determined during the design of the local Kalman filters.

3.1 DMPC estimation

The DMPC algorithm is initiated when a node *i* obtains its corresponding measurements $y_i(k)$. These local measurements are used with the predictions from the previous time step $\hat{x}_i(k|k-1)$ to produce local state estimates according to

$$\hat{x}_i(k|y_i(k)) = \hat{x}_i(k|k-1) + M_i[y_i(k) - C_i\,\hat{x}_i(k|k-1)] \quad (16)$$

Subsets of these local state estimates among x_i^o are broadcasted to relevant nodes and, in return, external state information is received back. The estimate for state *l* sent by a node *j* to another node *i* is denoted by $\hat{x}_{i(l)}^j(k|y_i(k))$, where *l* denotes a certain state in x_i^o. A given node *i* can interact with multiple other nodes and the shared states can be completely different or have common elements between different interconnections.

After the communication step, the received state estimates are weighted and fused together with the local estimates at node *i,* according to the pre-assigned reliability factors as:

$$\hat{x}_{i(l)}(k|k) = \sum_{j=1}^{q} \hat{x}_{i(l)}^j(k|y_j(k))\,r_{i(l)}^j \quad (17)$$

where *q* is the number of overlapping nodes for the state $x_{i(l)}$, including node *i* itself and the corresponding reliability factors for the different estimates are related by

$$\sum_{j=1}^{q} r_{i(l)}^{j} = 1 \qquad (18)$$

This step ensures that the information about the shared states x_i^o is distributed throughout the network.

In the DDEC scheme, the assimilation step uses the inverses of the estimation error covariances to weight the information coming from all other nodes. Since the DMPC framework is based on an off-line suboptimal Kalman filter, inverses of the steady-state error covariances can be used as reliability factors for the corresponding local estimators. However, because of the generality of the DMPC framework in terms of the interconnection structure and the number of overlapping states, there is no restriction on the choice of $r_{i(l)}^{j}$ besides (18). One potential negative aspect of the DMPC estimation scheme is the loss of equivalence to an optimal centralized estimator, as is the case with the DDEC methodology.

3.2 DMPC prediction and control

State estimates obtained in the previous section are used to initialize the local models (14) and (15). At each time step the nodes also receive information about the control moves of the neighboring nodes in the previous time step. This information is then used in the state transition equations (14) to include the effects of input interactions by assuming constant values throughout the prediction horizon. The prediction and control stages are conducted locally at each node *i*, according to the MPC algorithm by solving a numerical optimization problem given as

$$\begin{aligned} \min_{u_i(k|k),\ldots,u_i(m-1+k|k)} & \sum_{t=0}^{p-1}\Big(\sum_{s=1}^{n_{y_i}}\big[w_{t+1,s}^{y}\big(y_{it}(k+t+1|k) \\ & -y_{is}^{ref}(k+t+1)\big)\big]^2+\sum_{s=1}^{n_{u_i}}\big[w_{t,s}^{u_i}u_{is}(k+t|k)\big]^2\Big) \\ s.t. \quad & u_i^{low} \le u_i \le u_i^{high} \\ & \Delta u_i^{low} \le \Delta u_i \le \Delta u_i^{high} \\ & u_i(k+h|k)=0 \quad for \quad h=m,m+1,\ldots,p-1 \end{aligned} \qquad (19)$$

where s denotes the s^{th} component of a vector, $(k+t|k)$ denotes the t steps ahead prediction using the local models, based on information available at time k and finally, y_i^{ref} denotes the output reference for sample time k.

When a node obtains the solution of the local MPC problem, it sends the control moves for the next time step to its interacting neighbors before implementing it in the actual system. The nodes then use this information to update the input interactions in the prediction models and repeat the MPC calculations with the same initial states. The MPC calculations can be repeated for a certain number of iterations or based on a convergence criterion. The convergence properties of the DMPC algorithm will be reported in a subsequent publication. After a satisfactory solution is obtained from the MPC calculations, local nodes implement the control moves for the current time step and the predictions for state information is send to the next estimation stage which starts again as new measurements are received.

The prediction and control structure of the DMPC framework introduces cooperation for control calculations in the nodal network, whereas in the DDEC scheme the network only cooperates for state estimation and the control calculations are performed locally. Even though repeated MPC calculations are computationally more cumbersome compared to LQG based local controllers, the ability to include constraints, and the flexibility in the design of controllers considerably improve the performance of a DMPC network over a DDEC counterpart. The prediction estimation and control stages of the DMPC algorithm are depicted in figure 3.

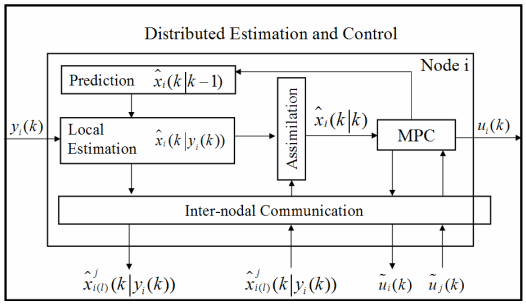


Fig.3. Organization of the distributed estimation, MPC based control and the inter-nodal communication stages in the DMPC framework.

4. CASE STUDY

The DMPC framework is applied in a case study involving level controls in an interacting four-tank system. This problem provides a simple, illustrative example for system decomposition and yet has challenging dynamic behaviour that can distinguish the performance of different control strategies.

4.1 System Description

The simulated four-tank problem considered here is a variant of the experimental system described by Gatzke *et al*. (2000). A schematic of this process is shown in figure 4. The system has two inputs (pump speeds) which can be manipulated to control the two outputs (levels in tanks 3 and 4). The multivariate dynamics is created by the cross-recycle streams feeding the two different overhead tanks 1 and 2. In this case study the dynamics are enriched by adding first order lags between the control signals and the pump throughput, and the system is simulated based on a full nonlinear mass balance model given in 20. Bernoulli's law is used for the flows out of the tanks, A_i stands for the cross sectional area, h_i for the liquid level and k_i for the flow factors from tank *i*. F_{in} stands for the input flows to the overhead tanks, d_i for the flow disturbances, γ_i for the recycle flow ratios, v_i for actual pump throughput and finally u_i for the controller output.

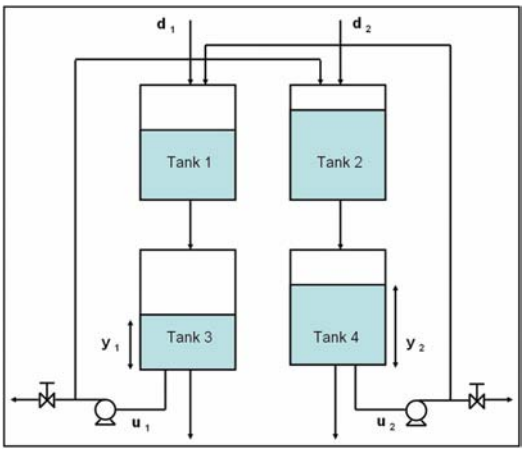


Fig.4. Schematic description of the four-tank system.

$$
\begin{aligned}
A_1 \frac{dh_1}{dt} &= F_{in} + d_1 + \gamma_1 v_2 - k_1\sqrt{h_1} \\
A_2 \frac{dh_2}{dt} &= F_{in} + d_2 + \gamma_2 v_1 - k_2\sqrt{h_2} \\
A_4 \frac{dh_4}{dt} &= k_2\sqrt{h_2} - v_2 - k_4\sqrt{h_4} \\
A_3 \frac{dh_3}{dt} &= k_1\sqrt{h_1} - v_1 - k_3\sqrt{h_3} \\
\tau_1 \frac{dv_1}{dt} &= -v_1 + u_1 \\
\tau_2 \frac{dv_2}{dt} &= -v_2 + u_2
\end{aligned}
\tag{20}
$$

By omitting all information about the flow disturbances, the nonlinear model equations with specified parameter values can be linearized and rearranged to give the following discrete-time system with a sampling frequency of 1 s^{-1}. This system will serve as the starting centralized model in the DMPC framework, corresponding to equations (1) and (2).

$$
x(k+1) = \begin{bmatrix}
0.89 & 0.26 & -0.2 & 0.03 & 0 & 0 \\
0 & 0.34 & 0 & 0.08 & 0 & 0 \\
0 & 0 & 0.37 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.45 & 0 & 0 \\
0 & 0 & 0.03 & -0.1 & 0.82 & 0.27 \\
0 & 0 & 0.09 & 0 & 0 & 0.25
\end{bmatrix} x(k) + \begin{bmatrix}
-0.1 & 0.01 \\
0 & 0.06 \\
0.63 & 0 \\
0 & 0.69 \\
0.01 & -0.1 \\
0.07 & 0
\end{bmatrix} u(k)
\tag{21}
$$

$$
y(k) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} x(k)
$$

4.2 Nodal Decomposition

The control objective in the four-tank system is to keep the levels at tanks 3 and 4 at the specified reference values in the face of flow disturbances. This objective and an examination of the rest of the process lead to a clear division of the system into two physical subsystems with each subsystem containing a controlled tank 3 or 4 and their matching overhead tanks 1 or 2. The physical separation of the process into two subsystems also allocates the measurements (tank levels 3 or 4) and the manipulated variables corresponding to each controlled tank as well (pump speeds 1 or 2).

The states of the system described in (21) are pre-arranged for the demonstration of mathematical decomposition. The measured states 1 and 5 are the focal points for the distribution of states among the two subsystems and they are positioned diagonally on the two sides of the state transition matrix. This arrangement forms a disconnected system, except for states 3 and 4, which are shared between the two subsystems leading to nodal models given below. These models correspond to equations (14) and (15) of the DMPC framework. In this distribution node 1 contains states 1 to 4, output 1 and input 1, whereas node 2 contains states 3 to 6, output 2 and input 2. This decomposition also satisfies the "self-sufficiency" requirement for the nodes in a DMPC network.

$$
x_1(k+1) = \begin{bmatrix}
0.89 & 0.26 & -0.2 & 0.03 \\
0 & 0.34 & 0 & 0.08 \\
0 & 0 & 0.37 & 0 \\
0 & 0 & 0 & 0.45
\end{bmatrix} x_1(k) + \begin{bmatrix} -0.1 \\ 0 \\ 0.63 \\ 0 \end{bmatrix} u_1(k) + \begin{bmatrix} 0.01 \\ 0.06 \\ 0 \\ 0.69 \end{bmatrix} u_2(k)
\tag{22}
$$

$$
y_1(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} x_1(k)
$$

$$
x_2(k+1) = \begin{bmatrix}
0.37 & 0 & 0 & 0 \\
0 & 0.45 & 0 & 0 \\
0.03 & -0.1 & 0.82 & 0.27 \\
0.09 & 0 & 0 & 0.25
\end{bmatrix} x_2(k) + \begin{bmatrix} 0 \\ 0.69 \\ -0.1 \\ 0 \end{bmatrix} u_2(k) + \begin{bmatrix} 0.63 \\ 0 \\ 0.01 \\ 0.07 \end{bmatrix} u_1(k)
\tag{23}
$$

$$
y_2(k) = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} x_2(k)
$$

4.3 DMPC design

The next stage in the construction of a DMPC network is to add the anticipated disturbance and noise models on the different subsystems. In this case, a single unmeasured disturbance and output noise channels are added to both nodes. Kalman filters, based on these models are then designed with expected covariance values for the process and measurement noises.

The overlapping states between the subsystems were determined during the nodal decomposition step, however for the DMPC design, the reliability factors for these states at each node has to be specified. For simplicity, in this case study the reliability of state estimation at both nodes are assumed to be the same and factors $r^2_{1(3)}$, $r^2_{1(4)}$, $r^1_{2(3)}$, $r^1_{2(4)}$ are all taken as 0.5 in accordance with (18).

For the MPC design, upper and lower limits for pump speeds are taken as 5 and 0, and only output weights of 100 are used in both subsystems. The prediction horizons are specified as 8 and the move horizons are specified as 3. The DMPC framework does not require symmetry for the design of the nodal MPC controllers, however similar computational loads will create a balanced network and nodal communication will proceed without long delays. In this case study, the performance of the DMPC network is compared to two other control strategies. The first one employs a centralized MPC controller for the whole system and the second one has two

completely decentralized MPC controllers. Both of these strategies have the same input constraints, output weights, prediction and control horizons as in the DMPC design. As a final design parameter, the MPC solutions are repeated only once by the DMPC nodes.

4.4 Performance Comparison

The performance of the DMPC network is compared with the centralized and completely decentralized MPC controllers in a simulation study using the nonlinear four-tank system. Concurrent feed flow disturbances were considered in the simulation scenarios in the form of 1 and 1.5 m^3/s steps for disturbances 1 and 2 respectively. Different control strategies were compared based on the maximum deviations from the set-points, the integral absolute errors and the settling times for the two outputs.

The simulation results are shown in figure 5 and the performance measures are listed in table 1. The results show that in the present configuration the concurrent disturbances effect output 1 more profoundly and in return the controllers have more difficulty managing this output. According to the results, the DMPC formulation outperforms the fully decentralized MPC controllers by a large margin. For both outputs all three performance measures are in favour of the DMPC but for output 1 the difference is more pronounced. Comparison of the DMPC formulation with the centralized MPC controller reveals that the DMPC comes fairly close to the performance of the centralized MPC. Moreover, even though the overall performance of the centralized MPC is better than the DMPC scheme, the DMPC has better results for the settling time and IAE for output 2.

The short oscillations in output 1 after the disturbance are caused by the distributed state estimation due to the initial differences in the state estimates for the shared states. The oscillations disappear as both estimators converge to the correct state estimates. This behaviour is not observed in output 2.

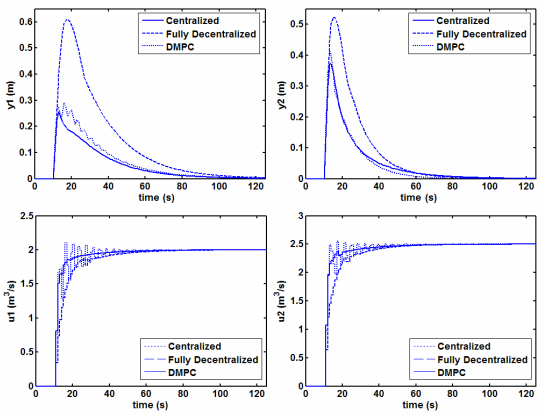


Fig.5. Dynamic response of the four-tank system to concurrent feed flow disturbances.

Table 1 Performance measures for different control formulations

	Settling time (s)	Maximum Deviation (m)	IAE (m·s)
Centralized MPC (y1)	90	0.24	6.1
Centralized MPC (y2)	90	0.37	5.7
DMPC (y1)	90	0.28	7.5
DMPC (y2)	70	0.42	5.2
Decentralized MPC (y1)	110	0.61	16.3
Decentralized MPC (y2)	90	0.52	9.2

5. CONCLUSION

In the present study, a distributed model predictive control framework is presented. The methodology is demonstrated on a four-tank level control problem. The results show that the new methodology performs significantly better than a completely decentralized set of controllers. In terms of computational parallelization, the size of the individual problems is reduced by more than 33 % compared to a completely centralized formulation. Iterative solutions of the DMPC provide feed-forward anticipation of the interactions due to control inputs, bringing the performance of the DMPC closer to that of a completely centralized MPC. In addition, the "self-sufficiency" criterion required for the DMPC enables the subsystems to stay functional in case of failures in other subsystems. With these properties, the DMPC framework can be a viable candidate to provide a connecting link between existing MPC applications in chemical engineering systems.

REFERENCES

Findeisen, W., Brdys, M., Malinowski, K., Tatjewski, P. and Wozniak, A. (1980). *Control and coordination in hierarchical systems*, Wiley Inc., New York, NY.

Gatzke, E.P., Meadows, E.S., Wang, C. and Doyle III, F. J. (2000). Model based control of a four-tank system. *Comput. Chem. Eng.,* **24**, 1503-1509.

Jia, D., & Krogh, B. H. (2001). Distributed model predictive control. In: *Proceedings of the American Control Conference*, pp. 2767–2772. Arlington, VA.

Mutambara, A.G.O. (1998). *Decentralized estimation and control for multisensor systems*, CRC Press, Boca Raton, FL.

Siljak, D.D. (1991). Decentralized control of complex systems. *Mathematics in Series and Engineering,* **184**, Academic Press, San Diego, CA.

Skogestad, S. (2004). Control structure design for complete chemical plants. *Comput. Chem. Eng.,* **28**, 219-234.

Vadigepalli, R. and Doyle III, F. J. (2003). A distributed state estimation and control algorithm for plantwide processes. *IEEE Transactions Contr. Syst. Technol.*, **11**, 119-127.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

COORDINATED DECENTRALIZED MPC FOR PLANT-WIDE CONTROL OF A PULP MILL BENCHMARK PROBLEM

Ruoyu Cheng * **J. Fraser Forbes** *,[1] **W. San Yip** **

* *Department of Chemical and Materials Engineering*
University of Alberta, Edmonton, T6G 2G6, Canada
** *Suncor Energy Inc., Ft. McMurray, T9H 3E3, Canada*

Abstract: In large-scale model predictive control (MPC) applications, such as plant-wide control, the coordination of unit-based MPC controllers has been identified as both an opportunity and a challenge in enhancing the plant-wide control performance. This work discusses an efficient strategy for the coordination of decentralized MPC systems and illustrates the approach with an application to the pulp mill benchmark problem proposed by Castro and Doyle III (2004*a*). The decentralized unit-based MPC controllers are coordinated at the MPC steady-state target calculation stage by employing decentralized optimization techniques. The off-diagonal element abstraction technique and the price-driven coordination algorithm are used in the development of a coordination mechanism. The pulp mill case study shows that this coordinated, decentralized MPC framework is an effective approach to plant-wide MPC applications, which has high reliability, accuracy and efficiency.

Keywords: Decentralized MPC, Target calculation, Price-driven coordination, Decentralized optimization

1. INTRODUCTION

In many plant-wide control and optimization applications, a large-scale process model is decomposed into several smaller subsystems and a controller is developed for each subsystem. This may lead to a decentralized unit-based MPC framework. The coordination of the unit-based controllers has been identified as having significant potential benefit (Havlena and Lu, 2005).

The decomposition and coordination approaches to solving complex large-scale control problems attracted attention in 1970's and 1980's (Wismer, 1971; Titli, 1978; Jamshidi, 1983); but the interest diminished thereafter for a number of reasons (Havlena and Lu, 2005) including: limited implementation opportunities; inherent complexity and difficulty of the problem; and computational issues. The industrial success in applying control schemes with a decentralized architecture has stimulated increasing interest in coordination of decentralized MPC (Lu, 2003); however, the need for more research in this area has been well recognized (Havlena and Lu, 2005; Isaksson *et al.*, 2005).

Most commercial MPC products employ two-stages: a steady-state target calculation and a dynamic control calculation (Qin and Badgewell, 2003; Ying and Joseph, 1999; Rao and Rawlings, 1999). In the case of decentralized unit-based MPC, without plant-wide coordination,

[1] Email : fraser.forbes@ualberta.ca, phone : 780 492-0873, fax : 780 492-2881.

the optimum operations achieved by each unit-based MPC may provide significantly worse performance than the plant-wide optimum solution (Havlena and Lu, 2005).

The potential benefit of coordinating decentralized control schemes has garnered increasing interest by both researchers and practitioners. Camponogara *et al.* (2002) have proposed a distributed MPC scheme, where local control agents broadcast their states and optimization results to every other agent under pre-specified rules to help reach a plant-wide optimum. Decentralized optimization via Nash bargaining has been applied for solving multi-player coordination problem by Waslander *et al.* (2004). Venkat *et al.* (2004) have used augmented states to model interactions and their scheme involves iterative negotiations among decentralized MPC systems. One common feature of the above schemes is that the decentralized MPC controllers exchange information directly and thus stand at an equal status within their negotiation hierarchy.

In process industries, however, a wide-spread belief among practitioners is that the trend toward decentralization will continue until the control system consists of seamlessly collabrating autonomous and intelligent nodes with a supervisory coordinator overseeing the whole process (Scheiber, 2004). One approach to coordinating decentralized MPC is to employ a centralized optimization layer to perform a plant-wide target calculation (e.g., Honeywell's *ad hoc* technology); while an alternative approach is to take advantage of decentralized optimization with an additional coordination system (Havlena and Lu, 2005). Our previous work (Cheng *et al.*, 2004; Cheng *et al.*, 2005*b*) adopts this approach, where the Dantzig-Wolfe decomposition and price-driven coordination strategies are tailored to yield a coordination system for decentralized MPC.

This work discusses the development of a coordination system for decentralized MPC that employs the price-driven coordination algorithm and off-diagonal element abstraction technique (Cheng *et al.*, 2005*b*). The case study based on the pulp mill benchmark problem shows that the proposed coordinated, decentralized MPC framework can be a viable approach to solving plant-wide MPC problems.

2. PLANT-WIDE MPC

In the process industry, centralized or monolithic MPC schemes are considered to be not viable for complex process control and optimization problems (Lu, 2003; Havlena and Lu, 2005). Consequently, industrial practice has tended toward a decentralized MPC architecture.

Usually, any limited cooperation between decentralized MPC controllers is through an upper level optimization, such as real-time optimization (RTO), at a sampling time of hours; however, disturbances or setpoint changes in the interval between two RTO executions may drive the optimum operations away from the targets given by the RTO system; thus, it is necessary to perform re-optimization at a higher frequency to maintain optimum operations. This section focuses on the coordination strategies for decentralized MPC at the target calculation level, and as a result, at a sampling time comparable with that of the MPC control calculation.

2.1 Unit-based MPC

In this work, unit-based MPC refers to the decentralized MPC subsystems developed for individual operating units as shown in Figure 1.

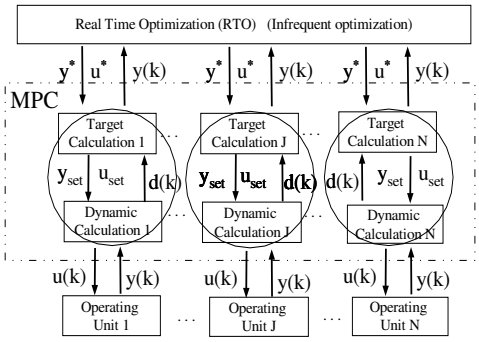


Fig. 1. Two-stage unit-based MPC system

Consider the following constrained quadratic programming (QP) formulation of MPC target calculation for an individual operating unit (Ying and Joseph, 1999):

$$\min_{\mathbf{y}_{set},\mathbf{u}_{set}} \quad z = (\mathbf{y}_{set}(k) - \mathbf{y}^*)^T \mathbf{Q}_y (\mathbf{y}_{set}(k) - \mathbf{y}^*) + (\mathbf{u}_{set}(k) - \mathbf{u}^*)^T \mathbf{Q}_u (\mathbf{u}_{set}(k) - \mathbf{u}^*) + \mathbf{c}_y (\mathbf{y}_{set}(k) - \mathbf{y}^*) + \mathbf{c}_u (\mathbf{u}_{set}(k) - \mathbf{u}^*) + \epsilon^T \mathbf{c}_\epsilon^T \mathbf{c}_\epsilon^T \epsilon \quad (1)$$

s. t.

$$\begin{aligned} \mathbf{y}_{set}(k) &= \mathbf{K}\mathbf{u}_{set}(k) + \mathbf{d}(k) \\ \mathbf{d}(k) &= \mathbf{d}(k-1) + \delta(k) \\ \mathbf{y}_{min} - \epsilon \le \mathbf{y}_{set}(k) &\le \mathbf{y}_{max} + \epsilon \\ \mathbf{u}_{min} \le \mathbf{u}_{set}(k) &\le \mathbf{u}_{max} \\ \epsilon &\ge 0 \end{aligned} \quad (2)$$

where $\mathbf{y}^*$ and $\mathbf{u}^*$ are the optimal nominal "targets" computed by upper level optimizers, $\mathbf{y}_{set}(k)$ and $\mathbf{u}_{set}(k)$ are the achievable targets to be optimized, while $\mathbf{d}(k)$ is the estimated disturbance updated by:

$$\delta(k) = \mathbf{y}_m(k) - \mathbf{y}_{set}(k|k-1), \quad (3)$$

where $\mathbf{y}_m(k)$ are the measured outputs at time k and $\mathbf{y}_{set}(k|k-1)$ is the prediction of outputs in the previous control execution. ϵ may be defined as a violation tolerance of the output constraints that ensures a feasible solution to the QP. The steady-state gain matrix $\mathbf{K}$ can be calculated via linearization of the nonlinear model used in an upper optimizing layer or abstracted from the linear model used by lower level MPC dynamic control. Note that the above formulation considers only the local unit.

2.2 Coordination of Decentralized MPC

Centralized Optimization A centralized optimization approach for coordinating decentralized MPC systems has been discussed in Havlena and Lu (2005). In that framework, as is shown in Figure 2, a monolithic optimization problem is formulated

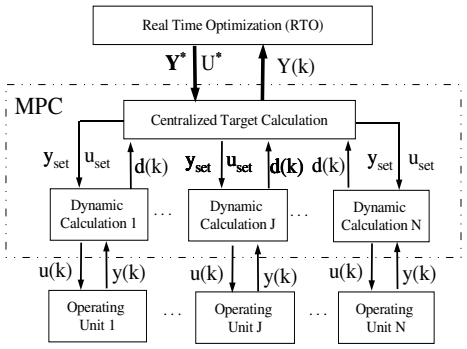


Fig. 2. Centralized MPC target calculation

at the target calculation stage for the entire plant. A plant-wide gain matrix is used, which relates the manipulated variables (MVs) and controlled variables (CVs) of all decentralized MPC subsystems. Although a centralized steady-state optimization approach may accurately track optimal plant operations, it can lack the reliability of the decentralized control structure.

Decentralized Optimization Depicted in Figure 3, a coordinator is designed to deal with the interactions among decentralized MPC controllers and makes use of the price-driven coordination method (Cheng *et al.*, 2005*a*). The task of the coordinator is to ensure that the coordinated system finds the optimal plant operations. Note that in the figure "S. I." denotes the term *sensitivity information*, which is the Lagrangian-like information flow used in the coordination mechanism.

A key step in coordinator design is to identify appropriate interactions for linking constraint formulation. The linking constraints contain process variables from multiple operating units (or unit-based MPC controllers). These linking constraints are used in the coordinator's optimizing scheme.

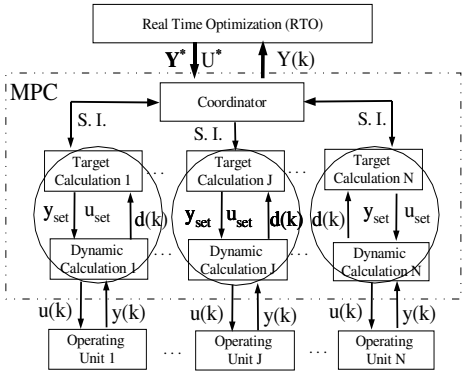


Fig. 3. coordinated, decentralized MPC target calculation

Several methods can be used to model the interactions, such as the interstream consistency (Cheng *et al.*, 2005*a*), off-diagonal elements abstraction, and ratio control constraint augmentation (Havlena and Lu, 2005).

Off-diagonal Elements Abstraction Here we briefly discuss the off-diagonal elements abstraction method for constructing linking constraints. Quite often, advanced control strategies are designed and implemented at different times for different operating units. In this situation, the CVs and MVs have been specified and grouped in a unit-based sense. Assume that we have a full gain matrix for a plant with N operating units:

$$\mathbb{A} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \dots & \mathbf{K}_{1N} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \dots & \mathbf{K}_{2N} \\ \dots & \dots & \dots & \dots \\ \mathbf{K}_{N1} & \mathbf{K}_{N2} & \dots & \mathbf{K}_{NN} \end{bmatrix} \tag{4}$$

A unit-based implementation of MPC in (2) uses the block-diagonal information $\mathbf{K}_{ii}$ of the plant model in their calculations, while the off-diagonal blocks may be treated as disturbances in their models. This way of dealing with the off-diagonal information can introduce undesirable uncertainty when the interactions are significant. Note that the plant-wide model:

$$Y(k) = \mathbb{A}U(k) + D(k) \tag{5}$$

where $Y(k)$ and $U(k)$ are vectors containing the CVs and MVs of N local operating units, respectively, and is equivalent to:

$$\mathbf{y}_i(k) = \mathbf{K}_{ii}\mathbf{u}_i(k) + \mathbf{e}_i(k) + \mathbf{d}_i(k) \tag{6}$$

$$\mathbf{e}_i(k) - \sum_{j=1}^{N} \mathbf{K}_{ij}\mathbf{u}_j(k) = \mathbf{0} \quad j \neq i \tag{7}$$

The auxiliary variable $\mathbf{e}_i$, which is an abstraction of the off-diagonal elements, represents the influence of the inputs of other operating units on the local system. Without the equations in (7), the auxiliary vector $\mathbf{e}_i$ can be regarded as local variables only involved with the i^{th} operating unit, and they can play a role as decision variables in the optimization.

The abstracted equality constraints in (7) are the linking constraints to be incorporated into the coordinator's optimization problem. Different decentralized optimization strategies have different usage of the linking constraints, but all of them aim to find a set of $[\mathbf{y}_i, \mathbf{u}_i, \mathbf{e}_i]$ so that the optimum plant operations are achieved.

Price-driven Coordination Method In our previous work (Cheng *et al.*, 2005*a*), the price-driven coordination method was developed to efficiently solve large-scale QP problems with equality linking constraints in a decentralized optimization manner. The work was based on ideas presented in Jose and Ungar (1998*a*) and (1998*b*).

Using the price-driven coordination method, the MPC target calcultion for a local operating unit can be modified as:

$$\min_{\mathbf{y}_{set},\mathbf{u}_{set}} \quad z = (\mathbf{y}_{set}(k) - \mathbf{y}^*)^T \mathbf{Q}_y (\mathbf{y}_{set}(k) - \mathbf{y}^*) + (\mathbf{u}_{set}(k) - \mathbf{u}^*)^T \mathbf{Q}_u (\mathbf{u}_{set}(k) - \mathbf{u}^*) + \mathbf{c}_y (\mathbf{y}_{set}(k) - \mathbf{y}^*) + \mathbf{c}_u (\mathbf{u}_{set}(k) - \mathbf{u}^*) + \epsilon^T \mathbf{c}_\epsilon^T \mathbf{c}_\epsilon^T \epsilon - \mathbf{p}^T \mathbf{e}(k) \quad (8)$$

s. t.

$$\begin{aligned} \mathbf{y}_{set}(k) - \mathbf{K}\mathbf{u}_{set}(k) &= \mathbf{e}(k) + \mathbf{d}(k) \\ \mathbf{d}(k) &= \mathbf{d}(k-1) + \delta(k) \\ \mathbf{y}_{min} - \epsilon \le \mathbf{y}_{set}(k) &\le \mathbf{y}_{max} + \epsilon \\ \mathbf{u}_{min} \le \mathbf{u}_{set}(k) &\le \mathbf{u}_{max} \\ \epsilon &\ge 0 \end{aligned} \quad (9)$$

where we omitted the subscript i for simplicity. Note that there is a minor modification to the objective function and unit model.

Note that a price vector $\mathbf{p}$ is introduced in (8). It has been proved that there exists an equilibrium price vector $\mathbf{p}^*$ that optimally coordinates the independently solved subproblems (unit-based optimization problems). To find the equilibrium price vector, the generalized Newton's method with stepsize determination is used to solve the following system of equations:

$$\mathbf{p}^T \Delta(\mathbf{p}) = 0 \quad (10)$$

$$\Delta(\mathbf{p}) = \mathbf{e}_i - \sum_{j=1}^{N} \mathbf{K}_{ij} \mathbf{u}_j \qquad i = 1...N \quad j \neq i \quad (11)$$

for updating the price vector $\mathbf{p}$, and the equilibrium price vector $\mathbf{p}^*$ satisfies the above two equations. One may also notice that Δ in (11) is an implicit function of the price vector $\mathbf{p}$. When the price vector is appropriately updated, the composition of unit-based MPC solutions will converge to the plant-wide optimum.

3. PLANT-WIDE CONTROL OF A PULP MILL PROCESS

The pulp mill model given in Castro and Doyle III (2004*a*) is a newly published industrial benchmark problem, which may be suitable for the study of process modeling and estimation, process control and optimization, and fault detection and diagnosis. This pulp mill model includes the fiber-line and the chemical recovery loop. The primary goal of the pulp mill is to produce wood pulp of a given Kappa number or brightness while minimizing energy costs, utilities and chemical make-up streams. The control objectives, modes of operation, process constraints and measurements are all defined in Castro and Doyle III (2004*a*).

3.1 Existing Unit-based MPC Schemes

In Castro and Doyle III (2004*b*), a decentralized control system has been proposed. At the unit level, it involves two control layers: unit-based MPC and decentralized regulatory control loops. This work focuses on the MPC layer.

The existing MPC consists of four separate controllers, one each for the digester and oxygen reactor, the bleach plant, the evaporators, and the lime kiln/recaust areas, respectively. In their configuration, the MPC layer only contains the dynamic control calculation stage and involves totally 21 CVs and 20 MVs. The MPC is designed to track the set-point trajectories given by an upper level optimization.

3.2 Modeling for Target Calculation

Since we focus on MPC target calculation, the plant-wide linear steady-state model matrix $\mathbb{A}$ in (4), from the MVs to CVs, is obtained via step response tests to ensure that the steady-state gains are consistent with the dynamic simulation.

In this work, the effect of disturbances are compensated via the bias update strategy in Ying and Joseph (1999).

3.3 Unit-based MPC Target Calculation

This study uses the decentralized, two-stage MPC system discussed in Ying and Joseph (1999) and takes the formulation given by (1) and (2) for target calculation. The control calculations in this paper use the configuration of Castro and Doyle III (2004*b*).

In the unit-based MPC target calculation, the interactions between units were ignored. Thus, the gain matrix $\mathbf{K}$ in (2) is actually $\mathbf{K}_{ii}$, the block-diagonal elements of the overall-plant gain matrix $\mathbb{A}$. The effect of off-diagonal elements was treated as disturbances, through $\mathbf{d}(k)$. The bounds for variables are the same as in the dynamic control calculation, and the weightings $\mathbf{Q}_y$ and $\mathbf{c}_y$ are given in Table 1.

3.4 Closed-loop Performance

This section compares three control schemes: the centralized, the decentralized, and the coordinated, decentralized MPC target calculation. The centralized optimization scheme uses the entire plant-wide gain matrix and is used to define the performance benchmark for our study.

It is desired to closely track the setpoints given by an upper level optimization at the same time maximize production rate and minimize oxygen reactor coolant flow and kiln fuel flow. In this case study, the plant-wide objective function is defined as a combination of those objectives with weightings given in Table 1. The optimization problems in all of the schemes are formulated as minimization problems. The weightings for the MVs used in all MPC control schemes are adopted from the work by Castro and Doyle III (2004*b*).

Table 1. Important CV Weightings

Controlled variables	$\mathbf{Q}_y/100$	$\mathbf{c}_y$
production rate	1.5	-80
digester kappa No.	1.5	0
oxygen reactor kappa No.	1.0	0
oxygen reactor caustic flow	1.0	0
oxygen reactor steam flow	0.5	0
oxygen reactor coolant flow	0.75	30
E kappa No.	1.0	0
D_2 brightness	1.0	0
slaker temperature	1.0	0
kiln $\mathbf{O}_2$ excess %	1.0	0
kiln fuel flow	0.5	30

Using the coordination strategy given in Section 2.2, the plant operation can be driven to the optimum operation. This usually takes a few communication cycles between the coordinator and subsystems.

Results based on a 8000-minute (about 140 hours) closed-loop simulation are reported. The disturbance set imposed on the process was adopted from Castro and Doyle III (2004*b*). Because the coordinated scheme provides identical performance to that of the centralized scheme, the following figures only show the closed-loop responses for the coordinated scheme and the original decentralized scheme.

The responses of some key process variables are reported in Figure 4. Note that, if the existing decentralized MPC behaves satisfactorily (i.e., is stabilizing and robust) under certain disturbances, the proposed coordination mechanism will not impact these characteristics. Moreover, the proposed control scheme can provide the optimum plant operations as given by the centralized scheme. In this study, the decentralized scheme exhibits significant offset from the optimum production rate and use of raw materials and energy.

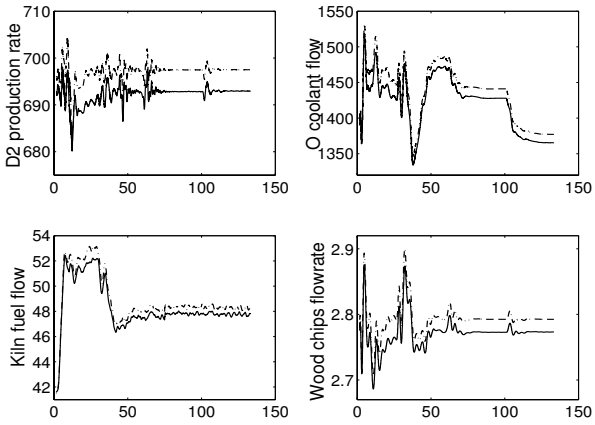


Fig. 4. Pulp mill closed-loop responses: solid line (coordinated); dash line (decentralized)

Table 2 reports the profit/cost function values and computational times for all three control schemes. Note that the accumulated value function is a time-integration of the objective function evaluated using the measured process variables. The optimization time is an average value based on the observed computational times. As we have defined the value function of the centralized scheme as a benchmark, we can see that the coordinated decentralized MPC provides the same plant-wide operations, which produces an 8.2% improvement on that of the decentralized scheme. In the case study, the optimization problems involve dozens of decision variables and hundreds of constraints. The coordinated MPC scheme provides solutions at a reasonable computational speed and as a result, exhibits a good trade-off between accuracy, reliability and computational load.

Table 2. Performance comparisons

control schemes	value function	optimization time* (per MPC execution)
centralized	1.22×10^5	0.06 s
unit-based	1.32×10^5	0.04 s
coordinated	1.22×10^5	0.14 s

* Simulations performed in Matlab 6.1, AMD Athlon 1.4G Hz, 1024M RAM machine.

3.5 Remarks: Some Implementation Issues

In the case study, the sampling time for target calculation is chosen as 10 minutes, which is the least common multiple of the sampling times of MPC subsystems. Based on our observations from simulations, the selection of sampling time for coordination should also depend on the frequencies of the disturbances that the control system must deal with.

In general, good initial points can substantially enhance the efficiency of optimization. A practical target calculation formulation should not provide

too aggressive changes in the reference trajectories, so the optimal solutions from two consecutive target calculation executions should not differ too much. Therefore, the equilibrium price vector from the previous execution works very well as an initial guess for the current target calculation.

4. CONCLUSIONS AND FUTURE WORK

In the MPC applications for plant-wide control, the coordination of unit-based MPC controllers can substantially improve plant operations. With the price-driven coordination algorithm and the off-diagonal element abstraction technique, a coordination mechanism was developed for a coordinated, decentralized MPC framework. The proposed control approach is applied to a pulp mill benchmark problem, and shows a significant improvement in performance in comparison to the existing decentralized MPC systems. Thus, the case study shows that this coordinated, decentralized MPC framework may be a viable technology for plant-wide MPC applications.

A number of issues still need to be further investigated. One of these is an understanding of the complexity and scaling behavior of the price-driven coordination algorithm. In addition, it is vital to understand the relationship between the structure of the decentralized MPC system and the performance of the coordinated system.

5. ACKNOWLEDGMENTS

This study is part of a project on decomposition and coordination approaches to large-scale operations optimization. The authors would like to express their gratitude to NSERC and the department of Chemical and Materials Engineering at the University of Alberta for financial support.

REFERENCES

Camponogara, E., D. Jia, B. H. Krogh and S. Talukdar (2002). Distributed model predictive control. *IEEE Control Systems Magazine* **0272-1708/02**, 44–52.

Castro, J. J. and F. J. Doyle III (2004*a*). A pulp mill benchmark problem for control: Application of plantwide control design. *Journal of Process Control* **3**, 329–347.

Castro, J. J. and F. J. Doyle III (2004*b*). A pulp mill benchmark problem for control: Problem description. *Journal of Process Control* **14**, 17–29.

Cheng, R., J. F. Forbes and W. S. Yip (2004). Dantzig-Wolfe decomposition and large-scale constrained MPC problems. In: *DYCOPS 7*. Boston, USA.

Cheng, R., J. F. Forbes and W. S. Yip (2005*a*). Price-driven coordination for solving plant-wide MPC problems. In: *16th IFAC World Congress 2005*. Prague, Czech.

Cheng, R., J. F. Forbes, W. S. Yip and J. V. Cresta (2005*b*). Plant-wide MPC: A cooperative decentralized approach. In: *2005 IEEE-IAS APC*. Vancouver, Canada.

Havlena, V. and J. Lu (2005). A distributed automation framework for plant-wide control, optimization, scheduling and planning. In: *16th IFAC World Congress 2005*. Prague, Czech.

Isaksson, A., B. J. Cott, K. Klatt, J. A. Mandler and P. Daoutidis (2005). Panel discussion: Industrial perspectives on process control. In: *16th IFAC World Congress 2005*. Prague, Czech.

Jamshidi, M. (1983). *Large-Scale Systems Modeling and Control*. Elsevier Science Publishing Co.,Inc.

Jose, R. A. and L. H. Ungar (1998*a*). Auction-driven coordination for plantwide optimization. *Foundations of Computer-aided Process Operation FOCAPO*.

Jose, R. A. and L. H. Ungar (1998*b*). Pricing inter-process streams using slack auctions. *AIChE Journal* **46**, 575–587.

Lu, J. Z. (2003). Challenging control problems and emerging technologies in enterprise optimization. *Control Engineering Practice* **11**, 847–858.

Qin, S. J. and T. A. Badgewell (2003). A survey of industrial model predictive control technology. *Control Engineering Practice* **11**, 733–764.

Rao, C. V. and J. B. Rawlings (1999). Steady states and constraints in model predictive control. *AIChE Journal* **45**, 1266–1278.

Scheiber, S. (2004). Decentralized control. *Control Engineering* pp. 44–47.

Titli, M.G. Singhand A. (1978). *Systems: Decomposition, Optimization and Control*. 1 ed.. Pergamon Press.

Venkat, A. N., J. B. Rawlings and S. J. Wright (2004). Plant-wide optimal control with decentralized MPC. In: *DYCOPS 7*. Boston, USA.

Waslander, S. L., G. Inalhan and C. J. Tomlin (2004). Decentralized optimization via nash bargaining. In: *Thoery and Algorithms for Cooperative Systems*. pp. 565–585.

Wismer, D. A. (1971). *Optimization Methods for Large-Scale Systems – with Applications*. McGraw-Hill.

Ying, C. and B. Joseph (1999). Performance and stability analysis of LP-MPC and QP-MPC cascade control systems. *AIChE Journal* **45**, 1521–1534.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

OPTIMIZING HYBRID DYNAMIC PROCESSES BY EMBEDDING GENETIC ALGORITHMS INTO MPC

Thomas Tometzki * **Olaf Stursberg** *,1
Christian Sonntag * **Sebastian Engell** *

* *Process Control Laboratory (BCI-AST)*
University of Dortmund, 44221 Dortmund, Germany.

Abstract: Optimizing the operation of processes with hybrid dynamics is challenging since the discrete dynamics (i.e. abrupt changes of states or inputs) introduces discontinuities with which gradient-based solvers often cannot cope very well. This contribution suggests a scheme that combines model predictive control (MPC) with genetic algorithms and embedded simulation of the hybrid dynamics. As demonstrated for the example of a chemical reactor, the genetic algorithm provides good results even if the prediction horizon includes points of discontinuities of the continuous dynamics.

Keywords: Automata, genetic algorithms, hybrid systems, optimal control, predictive control.

1. INTRODUCTION

Automated industrial processes are suitably modeled by hybrid dynamic systems if the evolution of continuous quantities (like levels, temperatures, or concentrations) is superposed by switching behavior. The latter arises e.g. from discretely operated actuators or autonomous abrupt changes between qualitatively different dynamics. The consideration of hybrid dynamics is particularly important if the process performs transitions between significantly different operating points as they occur for shutdown, start-up, or product changeover. This contribution proposes a technique to algorithmically compute (near-) optimal control strategies to realize such transition procedures. The starting point is a hybrid model of the plant given as a hybrid automaton with nonlinear continuous dynamics and discrete as well as continuous inputs. An optimization problem is formulated to determine the optimal input trajectories to drive the hybrid automaton from an initial state into a state within a target region such that an appropriate cost criterion is minimized, the hybrid dynamics is considered as a constraint, and unsafe state sets are not reached during the transition. The cost criterion represents the combination of the transition time and costs formulated over the state and the input trajectories.

Our previous work on this task revealed the following problems: If the optimization constraints are approximated by algebraic discrete-time linear models and if the optimization problem is solved iteratively using a model predictive control (MPC) scheme and mixed-integer linear programming (Stursberg and Engell, 2002), the solution performance suffers from large numbers of auxiliary variables required to encode the switching dynamics (Till *et al.*, 2004). If alternatively a graph search algorithm with embedded nonlinear programming (NLP) is used (Stursberg, 2004*a*; Stursberg, 2004*b*), the non-smoothness arising from the switching can lead to a lack of convergence in

[1] Corresponding author: olaf.stursberg@uni-dortmund.de

the NLP step. This paper introduces a method that combines the advantages of the two previous approaches in the following sense: The search for (near-) optimal state and input trajectories is carried out using a moving horizon scheme to benefit from the advantage of linearly increasing complexity with the overall time period required for the transition (if a fixed prediction horizon is used). The optimization in any iteration of the MPC scheme is carried out using a genetic algorithm (GA) with embedded hybrid simulation. Since this approach does not employ the gradients of the cost function and of the constraints, the convergence problems mentioned above for NLP do not occur here.

Apart from the work referenced above, a number of alternative approaches to optimal control of hybrid systems were published in recent years: While e.g. (Branicky *et al.*, 1998; Sussmann, 1999; Shaikh and Caines, 2003) aim at extending the maximum principle and calculus of variations to discontinuities, the approaches in (among others) (Shah and Pantelides, 1996; Buss *et al.*, 2000; Zhang and Cassandras, 2001; Bemporad *et al.*, 2002; Lee and Barton, 2003; Stein *et al.*, 2004) address different issues for efficiently computing optimal controllers for certain subclasses of hybrid systems (mostly piecewise affine systems). To the knowledge of the authors of this paper, only the publications (Wegele *et al.*, 2002) and (Olaru *et al.*, 2004) consider the use of GA for hybrid system optimization. While the first does not propose a specific solution algorithm, the second describes a method that solves the optimization problem by GA within an MPC scheme. In contrast to the approach proposed in this paper (which considers continuous-time and nonlinear continuous dynamics), the method in (Olaru *et al.*, 2004) is restricted, however, to MLD-systems, i.e. discrete-time piecewise affine systems.

2. PREDICTIVE CONTROL OF HYBRID AUTOMATA

The model considered in this contribution is formulated as a *hybrid automaton* according to (Stursberg, 2004*a*). Such a model is suitable to represent the transition procedure mentioned above, as it includes continuous and discrete inputs, and it can express the state-dependent switching between different (possibly unstable) continuous dynamics:

Definition 1. Hybrid Automaton
A hybrid automaton with mixed inputs $HA = (X, U, V, Z, inv, \Theta, g, r, f)$ consists of the following elements:

- the *state vector* x defined on the continuous state space $X \subseteq \mathbb{R}^{n_x}$;
- the *continuous inputs* u defined on the continuous input space $U = [u_1^-, u_1^+] \times \ldots \times [u_{n_u}^-, u_{n_u}^+]$, $u_j^-, u_j^+ \in \mathbb{R}$;
- a finite number n_d of *discrete inputs* $v_j \in \mathbb{R}^{n_v}$ defined on the discrete input space $V = \{v_1, \ldots, v_{n_d}\}$;
- a finite set of *locations* $Z = \{z_1, \ldots, z_{n_z}\}$;
- an *invariant* mapping $inv : Z \rightarrow 2^X$ which assigns a polyhedral set $inv(z_j) = \{x \mid \exists\, n_{p_j} \in \mathbb{N}, C_j \in \mathbb{R}^{n_{p_j} \times n_x}, d_j \in \mathbb{R}^{n_{p_j}}, x \in X : C_j \cdot x \leq d_j\}$ to each location $z_j \in Z$;
- the set $\Theta \subseteq Z \times Z$ of *transitions*, denoted by (z_1, z_2) for a transition from $z_1 \in Z$ into $z_2 \in Z$;
- a *guard* mapping $g : \Theta \rightarrow 2^X$ that associates a polyhedral set $g((z_1, z_2)) \subseteq X$ with each $(z_1, z_2) \in \Theta$. For each location $z \in Z$, it is required for all pairs of transitions originating from z that the corresponding guard sets are disjoint;
- a *reset function* $r : \Theta \times X \rightarrow X$ assigning an updated state $x' \in X$ to each $(z_1, z_2) \in \Theta$ and $x \in g((z_1, z_2))$;
- a *flow function* $f : Z \times X \times U \times V \rightarrow \mathbb{R}^{n_x}$ defining a continuous vector field $\dot{x} = f(z, x, u, v)$ for each pair $z \in Z$, $v \in V$.

Let $T = \{t_0, t_1, t_2, \ldots\}$ be an ordered set of time points, such that T contains the initial time t_0 and all points of time at which an input change or a transition occurs. The continuous states x_k, the locations z_k, and the inputs u_k and v_k are defined for the points $t_k \in T$. The values of u_k and v_k are piecewise constant on each interval $[t_k, t_{k+1}[$, $k \in \mathbb{N} \cup \{0\}$. Σ denotes the set of all valid hybrid states $\sigma_k := (z_k, x_k)$ with $z_k \in Z$ and $x_k \in inv(z_k)$. For given sequences of values u_k and v_k, a *feasible run* ϕ_σ of HA is then given as a sequence $\phi_\sigma = (\sigma_0, \sigma_1, \sigma_2, \ldots)$ with $\sigma_k \in \Sigma$ such that:

- σ_0 is initialized to a given $z_0 \in Z$ and $x_0 \in inv(z_0)$ (but x_0 not contained in any guard set).
- σ_{k+1} results from σ_k by: (1) continuous evolution: $\chi : [0, \tau] \rightarrow X$ with $\tau = t_{k+1} - t_k$ and $\chi(0) = x_k$, $\chi(\tau) = \int_0^\tau f(z_k, \chi(t), u_k, v_k)dt$ with $\chi(t) \in inv(z_k)$ and for all $t \in [0, \tau[$: $\chi(t) \notin g((z_k, \bullet))$ for any guard set associated with transitions leaving z_k; and (2) a transition $(z_k, z_{k+1}) \in \Theta$ if $\chi(\tau) \in g(z_k, z_{k+1})$, such that $x_{k+1} = r((z_k, z_{k+1}), \chi(\tau)) \in inv(z_{k+1})$, else $z_{k+1} = z_k$, $x_{k+1} = \chi(\tau)$. $\diamondsuit$

The optimal control problem considered in this paper is posed as follows: Assume that an initial state $\sigma_0 \in \Sigma$, a target set $\Sigma_t \subset \Sigma$ with $\Sigma_t = \{(z_t, x) | \exists_1 z_t \in Z : x \in X_t \subset inv(z_t)\}$, as well as a forbidden state set $F = \bigcup_{j=1}^{n_f} F_j \subset \Sigma$ with $F_j = \{(z_{f,j}, x) | \exists_1 z_{f,j} \in Z : x \in X_{f,j} \subset inv(z_{f,j})\}$ are given. ($X_{f,j}$ is polyhedral, and $X_{f,j} \cap X_t =$

$\emptyset$). Assume furthermore that the ordered set of time points $T = \{t_0, t_1, t_2, \dots, t_f\}$ is finite, and that the inputs can only be changed at $t_k \in T_s \subset T$. Let $\Phi_{u,s}$ contain all possible continuous input trajectories $\phi_u = (u_0, u_1, u_2, \dots)$ defined on T_s, and $\Phi_{v,s}$ correspondingly all discrete input trajectories $\phi_v = (v_0, v_1, v_2, \dots)$. Note that ϕ_σ remains defined on T, and that a run of HA according to Def. 1 is deterministic for any choice of ϕ_u and ϕ_v.

The control task is to determine input trajectories ϕ_u^* and ϕ_v^* that lead to a run ϕ_σ^* of HA from σ_0 into Σ_t such that no hybrid state in F is encountered and a cost function Ω is minimized:

$$\min_{\phi_u \in \Phi_{u,s}, \phi_v \in \Phi_{v,s}} \Omega\left(t_f, \phi_\sigma\right) \tag{1}$$

$$\begin{aligned} s.t.\ & \phi_\sigma = (\sigma_0, \dots, \sigma_f) \text{ with } \sigma_0 = (z_0, x_0), \\ & \sigma_f := (z(t_f), x(t_f)) \in \Sigma_t, \text{ and for } \phi_\sigma \text{ applies} \\ & \text{in each phase of continuous evolution acc. to} \\ & \text{Def. 1: } (z_k, \chi(t)) \notin F_j \ \forall\ F_j \in F,\ \forall\ t \in [0, \tau]. \end{aligned}$$

σ_f denotes the first hybrid state along ϕ_σ which is contained in Σ_t. Since computing the solution of the problem in Eq. 1 requires that $|T_s|$ choices are made for the values of v and u, the solution space of the optimization problem grows exponentially with $|T_s|$. In order to allow for a computationally tractable solution also for larger sets T_s, the approach presented here approximates the problem in Eq. 1 by employing a moving horizon scheme. The optimization problem is divided into $|T_s| - 1$ subproblems which are solved iteratively and yield a solution $\hat{\phi}_u \in \Phi_{u,s}$, $\hat{\phi}_v \in \Phi_{v,s}$ of the control task, which leads to a run $\hat{\phi}_\sigma$ of HA. Starting from t_0, a subproblem is solved for every $t_k \in T_s$ as follows: for a given $h \in \mathbb{N}^{>0}$, an ordered time set $T_k = \{t_k, \dots, t_{k+h-1}\}$ denotes the time horizon for which the problem:

$$\min_{\overline{\phi}_u \in \Phi_{u,h}, \overline{\phi}_v \in \Phi_{v,h}} \Omega\left(t_k, \dots, t_{k+h}, \overline{\phi}_\sigma\right), \tag{2}$$

is solved. The input trajectories are defined as $\overline{\phi}_u = \{u_k, \dots, u_{k+h-1}\}$, $\overline{\phi}_v = \{v_k, \dots, v_{k+h-1}\}$ and are taken from the sets $\Phi_{u,h}$ and $\Phi_{v,h}$ of all input trajectories of length h. The value of h is either equal to a user-specified parameter, or, if Σ_t is reached within the horizon, equal to the number of time steps required to reach Σ_t. $\overline{\phi}_\sigma = (\sigma_k, \dots, \sigma_{k+h})$ denotes the corresponding feasible run of HA, and in each iteration k, σ_k is set equal to the hybrid state σ_{k+1} of the run obtained from the optimization in the preceding iteration. The optimization according to Eq. 2 is subject to the same constraint for the forbidden sets as the optimization according to Eq. 1. The cost function Ω depends on the time points $t_k \in T_k$, since the fraction of each time interval $[t_k, t_{k+1}]$, for which the evolution of HA is outside of Σ_t, leads to a cost contribution. A second contribution is determined as the distance between each $\sigma_k \in \overline{\phi}_\sigma$ and Σ_t. (All state variables are normalized to the interval $[0, 1]$ for the distance computation.)

When all subproblems are solved, the complete continuous input trajectory $\hat{\phi}_u$ is obtained from concatenating the first elements u_k of every trajectory $\overline{\phi}_u$. $\hat{\phi}_v$ is constructed accordingly.

3. A GA-BASED OPTIMIZATION ALGORITHM

The subproblems in Eq. 2 are solved employing genetic algorithms (GAs), which were first introduced in (Holland, 1975). GAs provide an efficient means to solve optimization problems, even if discontinuities in the cost function or the constraints are present. Fig. 1 shows the scheme of the algorithm consisting of three modules: the genetic algorithm, an evaluation module, and a module that performs the simulation of the hybrid dynamics of HA. In an iteration k of the MPC scheme, the GA is initialized with the initial hybrid state σ_k and the horizon T_k (see Sec. 2). It performs n_{max} iterations to update the *population* $P_n = \{s_{1,n}, \dots, s_{\mu,n}\}$, which is the set of *individuals* considered in iteration n, and $\mu \in \mathbb{N}^{>0}$ is an even number which represents the population size. n_{max} is determined as the number of generations after which no increase in the best fitness value (see below) is observed. Each individual s represents a combination of continuous and discrete input trajectories. While continuous inputs are represented using real-valued numbers, the discrete inputs are encoded as bit sequences. These elements, the *genes*, lead to individuals defined by:

$$s_{i,n} := (s_{i,n,1}, s_{i,n,2}, \dots, s_{i,n,h-1}), \tag{3}$$

$$\text{with} \quad s_{i,n,j} := (a_1, \dots, a_{n_u}, b_1, \dots, b_{n,v}). \tag{4}$$

The parameter $a_l \in \mathbb{R}$ with $l \in \{1, \dots, n_u\}$ represents the value of the continuous input u_l, and a bit sequence $b_p \in \mathbb{B}^{1 \times \lceil log(c_p) \rceil}$, $p \in \{1, \dots, n_v\}$ encodes the value of the discrete input v_p (where c_p is the number of possible discrete values of v_p). The individuals of the initial population P_1 are generated randomly.

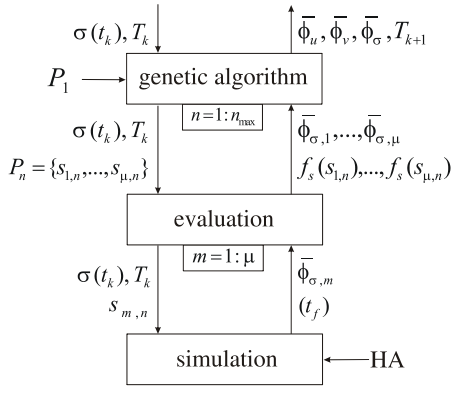


Fig. 1. Scheme of the algorithm.

In every iteration n, the algorithm carries out the following steps: For the input trajectories $\overline{\phi}_{u,i}$, $\overline{\phi}_{v,i}$ encoded by an individual $s_{i,n}$ in P_n, the corresponding state trajectory $\overline{\phi}_{\sigma,i,n}$ is computed. This is accomplished by numerical simulation which approximates the run of HA for the time horizon T_k according to the semantics in Def. 1. The *fitness* of $\overline{\phi}_{\sigma,i,n}$ is then determined by evaluating the function:

$$f_s(s_{i,n}) = \frac{1}{\Omega'(t_k, \ldots, t_{k+h}, \overline{\phi}_{\sigma,i,n})}, \quad i \in \{1, \ldots \mu\},$$

in which Ω' is the cost function from Eq. 2, but with two extensions: (a) an additional term is added which penalizes state trajectories that enter the forbidden hybrid state sets F; (b) the hybrid states in $\overline{\phi}_{\sigma,i,n}$ are weighted over the time horizon T_k. The weighting factors w for the hybrid states are determined according to

$$w(t_{k+o}) = \begin{cases} 1 & , o = h \\ f_w \cdot w(t_{k+o+1}) & , o = 1, \ldots, h-1 \end{cases},$$

with $f_w \in \mathbb{R}^{>0}$.

After the fitness evaluation, the population P_{n+1} for the next iteration of the GA is determined. Fig. 2 shows the algorithm used for generating P_{n+1} from P_n.

$P_{n+1} := \emptyset$;
`WHILE` $(|P_{n+1}| < |P_n|)$ `DO` {
  $p_1 := select(P_n)$;
  $p_2 := select(P_n)$;
  $(c_1, c_2) := crossover(p_1, p_2, w_c)$;
  $c_1 := mutate(c_1)$;
  $c_2 := mutate(c_2)$;
  $P_{n+1} := P_{n+1} \cup \{c_1, c_2\}$;
}
$P_{n+1} := elitist(P_n, P_{n+1})$

Fig. 2. Algorithm for generating P_{n+1} from P_n.

First, two individuals p_1 and p_2 (the *parents*) are selected from P_n by the function *select*. This function stochastically chooses an individual using a rank-based selection mechanism. In a large number of experiments in which several selection mechanisms were tested, it was found that this mechanism allows for good optimization results for the application example described in the following section. Assuming that P_n is sorted in ascending order with respect to the fitness of its individuals, the probability that an individual $s_{i,n}$, $i \in \{1, \ldots, \mu\}$, is selected from P_n is computed as:

$$w_s(s_{i,n}) = \frac{pos(s_{i,n})}{\sum_i pos(s_{i,n})},$$

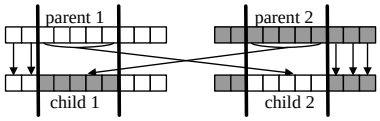


Fig. 3. Two-point crossover.

where the operator *pos* returns the position of $s_{i,n}$ within P_n.

The function *crossover* either performs a two-point crossover of p_1 and p_2 (with probability $w_c \in [0, 1]$) or leaves p_1 and p_2 unchanged (with probability $1 - w_c$). The main idea of the two-point crossover is to generate individuals, the *children*, which combine the advantages of both parents. Two cut-off points of the parents are determined randomly, and the children c_1 and c_2 are constructed by exchanging the segment between the cut-off points (Fig. 3), which may involve continuous and discrete variables. If no crossover is performed, p_1 and p_2 are just copied to c_1 and c_2.

The function *mutate* is used to randomly change genes of the two children. Each gene of a child c_i is changed with probability $w_m = 1/|c_i|$, with the number $|c_i|$ of genes of c_i. Modified real-valued genes a'_j are created from the original gene a_j according to $a'_j = a_j + m_n$; m_n is a realization of a normally distributed random variable with expected value 0 and decreasing variance over the iterations n. Binary-valued genes are changed according to $b'_j = 1 - b_j$.

The procedure of selection, crossover, and mutation is repeated until the number of individuals in P_{n+1} and P_n is equal. Finally, the function *elitist* replaces an arbitrary individual in P_{n+1} with the individual from P_n with the highest fitness value. This function is used to avoid that a good solution for the optimization according to Eq. 2 is lost by applying crossover and mutation.

After n_{max} iterations of the GA, the input trajectories encoded by the individual of $P_{n_{max}}$ with the highest fitness value determine the solution of the optimization in iteration k of the MPC scheme.

The set of time points T_{k+1} for the next iteration of the scheme is computed according to $T_{k+1} = (t_{k+1}, \ldots, t_{k+h-1}, t_{new})$. The last time t_{new} is obtained from $t_{new} = t_{k+h-1} + \Delta t$ with

$$\Delta t = \frac{f_{\Delta t}}{(\|f(z_{k+h}, x_{k+h}, u_{k+h-1}, v_{k-h-1})\|_2)}, \quad (5)$$

where $f_{\Delta t}$ is a tuning factor, $(z_{k+h}, x_{k+h}) = \sigma_{k+h}$ is the last hybrid state of $\overline{\phi}_\sigma$, and u_{k+h-1} and v_{k+h-1} are the last entries of $\overline{\phi}_u$ and $\overline{\phi}_v$. This procedure adjusts the time interval Δt to the dynamics of HA for the current value of the state variables and the inputs.

4. APPLICATION EXAMPLE

The approach is illustrated for the start-up procedure of a chemical reactor as described in (Stursberg, 2004*a*). It consists of a tank equipped with two inlets, a heater, a cooling device, and one outlet.

The state variables considered for optimization are the volume of liquid V_R, the temperature T_R, and the concentrations c_A and c_B of two substances A and B, which react exothermically to form a product. The system has five input variables: the inlet flows F_1, F_2 as well as a variable s_H (representing the status of the heater: *on* or *off*) can be switched between two values. In this example, however, the discrete input set V is restricted to four combinations of values by assuming that the inlet flows F_1 and F_2 are always equal. Hence, the discrete input vector is given by $v := (F_1, F_2, s_H)^\mathrm{T}$, $F_1 = F_2$. The continuous inputs of the system comprise the outlet flow F_3 and a cooling flow F_C, thus $u := (F_3, F_C)^\mathrm{T}$. According to Def. 1, the state vector is defined as $x := (V_R, T_R, c_A, c_B)^\mathrm{T} \in X$. The vector of locations is given by $Z = \{z_1, z_2\}$, with $inv(z_1) = \{x \mid x \in X : V_R \in [0.1, 0.8[\}$ and $inv(z_2) = \{x \mid x \in X : V_R \geq 0.8\}$ to account for the fact that the heating is only effective for a certain range of V_R (i.e. in z_2). The set of transitions is $\Theta = \{(z_1, z_2), (z_2, z_1)\}$ with $g((z_1, z_2)) = g((z_2, z_1)) = \{x \mid x \in X : V_R = 0.8\}$. The flow functions for the two locations are specified as:

$$f^I := f(z_1, x, u, v) = \begin{pmatrix} F_1 + F_2 - F_3 \\ (F_1(T_1 - T_R) + F_2(T_2 - T_R))/V_R \\ \quad + F_C k_1 (T_C - T_R)(k_2/V_R + k_3) - k_4 q \\ (F_1 c_{A,1} - c_A(F_1 + F_2))/V_R + k_9 q \\ (F_2 c_{B,2} - c_B(F_1 + F_2))/V_R + k_{10} q \end{pmatrix}, \tag{6}$$

$$f^{II} := f(z_2, x, u, v) = \left(f_1^I,\ f_2^I + s_H k_6 (T_H - T_R)(k_7 - \frac{k_8}{V_R}),\ f_3^I,\ f_4^I \right)^\mathrm{T}, \tag{7}$$

with $q = c_A c_B^2 \exp(-k_5/T_R)$ and appropriate constants T_1, T_2, T_C, $c_{A,1}$, $c_{B,2}$, T_H and k_1 to k_{10}. State resets do not occur, and the model comprises three forbidden regions $F_1 = (z_1, \{x \mid x \in X : T_R \geq 360\})$, $F_2 = (z_2, \{x \mid x \in X : T_R \geq 360\})$, and $F_3 = (z_2, \{x \mid x \in X : V_R \geq 1.8\})$.

The optimization task is to drive the system from an initial state $\sigma_0 = (z_1, x_0)$ with $x_0 = (0.1, 300, 0, 0)^\mathrm{T}$, which corresponds to an almost empty reactor, into an operating region in which the reactor is filled, the temperature has a desired level, and the production rate is sufficiently high, i.e. the concentrations of A and B have low values.

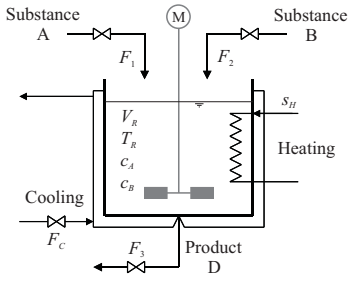


Fig. 4. Scheme of the CSTR.

The target location is z_2, and $X_t \subset inv(z_2)$ is a hyper-box given by $X_t = [1.49, 1.51] \times [343, 347] \times [0.46, 0.5] \times [0.18, 0.22]$.

Several parameters of the algorithm presented in Sec. 3 can be adjusted to tune the performance for the given example. These parameters include the optimization horizon h, the factor determining the weighting of the hybrid states in the fitness computation f_w, the population size μ, the crossover probability w_c, and the tuning factor for the computation of the time steps $f_{\Delta t}$.

In a large number of simulation studies with systematic variation of the values of these parameters, the following effects were observed: The cost function value obtained for the complete transition has a minimum for an optimization horizon of length $h = 4$. The fact that the costs increase for higher values of h may be attributed to the exponential growth of the search space with an increasing value of h, and thus a relatively sparse sampling of the search space for fixed numbers of generations and individuals in the population. Increasing μ or decreasing $f_{\Delta t}$ increases, not surprisingly, the optimization performance (i.e. leads to lower transition costs) but also the computation time. However, since the gain in optimization performance is negligible for higher values of these parameters, it is reasonable to limit them to relatively low values (see Tab. 1) for the sake of a small computation time. Maybe surprisingly, it was found that the crossover probability has only a very small influence on the optimization

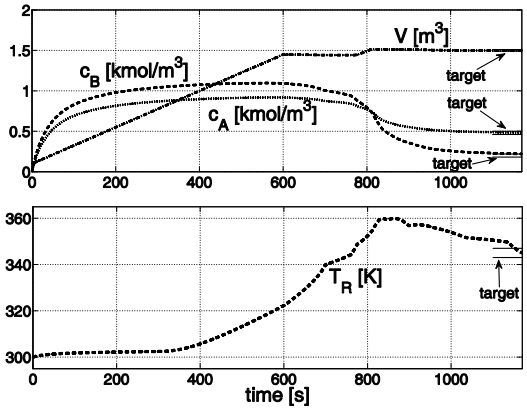


Fig. 5. Optimized state trajectories.

Table 1. An efficient parameterization.

h	f_w	μ	w_c	$f_{\Delta t}$
4	0.2	60	0.4	0.5

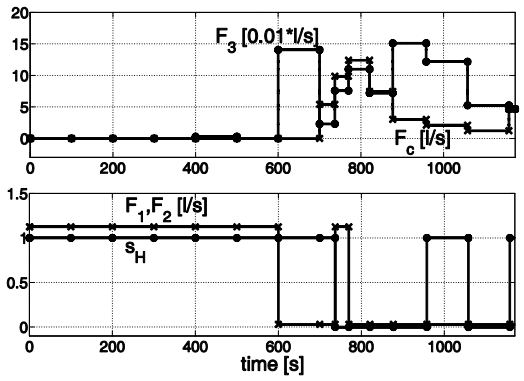


Fig. 6. Optimized input trajectories. F_1 and F_2 are switched between the values 0.03 $\frac{l}{s}$ and 1.125 $\frac{l}{s}$.

performance. The numbers shown in Tab. 1 represent the parameterization that was identified as a suitable compromise between the optimization performance and the computation effort. The figures 5 and 6 depict the trajectories of the state and input variables determined as the optimization result for the parameterization in Tab. 1.

5. CONCLUSION

This paper describes an approach for using GA embedded into an MPC scheme for the optimization of hybrid dynamic models. The solution is qualitatively and quantitatively comparable to the one obtained with the approach in (Stursberg, 2004*a*). In contrast to the latter, the method proposed here has no difficulties to cope with the discrete dynamics of HA. For an extended version of the example, this result has been obtained also for the case that reset functions introduce discontinuities into the state trajectory of the hybrid model (rather than only in the flow functions). By defining an absolute upper limit for the number of generations n_{max} and choosing suitable parameters of the algorithm, the required optimization time per MPC iteration can be held sufficiently small for online application. In this case, the outputs u_k and v_k are applied to the plant in order to obtain the real state (z_{k+1}, x_{k+1}) for the next iteration.

REFERENCES

Bemporad, A., A. Giua and C. Seatzu (2002). A master-slave algorithm for the optimal control switched affine systems. In: 41^{st} *IEEE Conf. on Dec. and Control*. pp. 1976–1981.

Branicky, M. S., V. S. Borkar and S. K. Mitter (1998). A unified framework for hybrid control: Model and optimal control theory. *IEEE Trans. Automatic Control* **43**(1), 31–45.

Buss, M., O. von Stryk, R. Bulirsch and G. Schmidt (2000). Towards hybrid optimal control. *Automatisierungstechnik* **9**, 448–459.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. MIT Press.

Lee, C. K. and P. I. Barton (2003). Determining the optimal mode sequence. In: *IFAC Conf. on Analysis and Design of Hybrid Systems*. pp. 153–158.

Olaru, S., J. Thomas, D. Dumur and J. Buisson (2004). Genetic algorithm based model predictive control for hybrid systems under a modified mld form. *Int. Journal of Hybrid Systems* **4**(1-2), 113–132.

Shah, V.D. Dimitriadis N. and C.C. Pantelides (1996). Optimal design of hybrid controllers for hybrid process systems. In: *Hybrid Systems III*. Vol. 1066 of *LNCS*. Springer. pp. 224–257.

Shaikh, M.S. and P.E. Caines (2003). On the optimal control of hybrid systems. In: *Hybrid Systems: Comp. and Control*. Vol. 2623 of *LNCS*. Springer. pp. 466–481.

Stein, O., J. Oldenburg and W. Marquardt (2004). Continuous reformulations of discrete - continuous optimization problems. *Comp. and Chemical Eng.* **28**(10), 1951–1966.

Stursberg, O. (2004*a*). Dynamic optimization of processing systems with mixed degrees of freedom. In: 7^{th} *Int. IFAC Symp. Dyn. and Control of Process Systems*. number 164.

Stursberg, O. (2004*b*). A graph-search algorithm for optimal control of hybrid systems. In: 43^{rd} *IEEE Conf. on Decision and Control*. pp. 1412–1417.

Stursberg, O. and S. Engell (2002). Optimal control of switched continuous systems using mixed-integer programming. In: 15^{th} *IFAC Congr. on Automatic Control*. Vol. Th-A06-4.

Sussmann, H.J. (1999). A maximum principle for hybrid optimal control problems. In: 38^{th} *IEEE Conf. Dec. and Control*. pp. 425–430.

Till, J., S. Engell, S. Panek and O. Stursberg (2004). Applied hybrid system optimization - an empirical investigation of complexity. *Control Eng. Practice* **12**(10), 1269–1278.

Wegele, S., E. Schnieder and M. Chouikha (2002). Automatic design of controllers for hybrid systems using genetic algorithms. In: *Mod., Analysis, and Design of Hybrid Systems*. Vol. 279 of *LNCIS*. Springer. pp. 285–294.

Zhang, P. and C. Cassandras (2001). An improved forward algorithm for optimal control of a class of hybrid systems. In: 40^{th} *IEEE Conf. Decision and Control*. pp. 1235–1236.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

OPTIMAL CONTROL OF MULTIVARIABLE PROCESSES USING BLOCK STRUCTURED MODELS

G. Harnischmacher, **W. Marquardt** [1]

Lehrstuhl für Prozesstechnik, RWTH Aachen University
D–52056 Aachen, Germany

Abstract: Block structured models have been used in nonlinear model predictive control to reduce computational cost. The solution of the nonlinear dynamic optimization problem has been evaded by inverting the nonlinear element and solving the resulting linear problem in the past. However, by exploiting the block structure for sensitivity calculation, the original nonlinear problem can also be solved at low computational cost, and at the same time this o ers much greater modeling exibility. This paper deals with dynamic optimization and, in particular, the e cient calculation of rst order sensitivity information for the case of multivariable Hammerstein and Uryson systems. In a simulation example the method is shown to combine low computational cost with the possibility to signi cantly reduce the losses of optimality compared to the previous methods.

Keywords: Hammerstein model, sensitivity system, nonlinear model predictive control, dynamic optimization, multivariable block structured model

1. INTRODUCTION

Nonlinear model predictive control (NMPC) poses challenging problems both in modeling and computation. Obtaining nonlinear, dynamic process models either requires large amounts of identi - cation data or deep physical insight for rigorous modeling. Afterwards, the optimization problem has to be solved within short sampling times required in closed loop NMPC. Numerous model reduction techniques have been explored to reduce the original process model (Marquardt, 2002), or to totally avoid online optimization (Kadam *et al.*, 2005).

Block structured models consisting of nonlinear static and linear dynamic elements have been used to reduce both the modeling and computation e orts. Structuring the model in this way leads to an approximate model, which is inferior in prediction quality to a rigorous nonlinear model, but provides a viable compromise between the low predictive capabilities of a linear model and the costly development of a non-structured nonlinear dynamic model. Applications range from such di erent elds as neuroprothesis, where a rigorous nonlinear model could not be obtained (Hunt *et al.*, 1998), to the control of an industrial C2-splitter (Norquay *et al.*, 1999). For Wiener (Norquay *et al.*, 1999) and Hammerstein (Zhu and Seborg, 1994) models tailored solution algorithms have been developed. They are based on the inversion of the nonlinear element to reduce the original nonlinear dynamic optimization problem to a linear one. We will refer to this method as the "inversion based method" in the sequel. To obtain a unique solution with the inversion based method, the nonlinearity of the model needs to be bijective, which is generally not the case. Especially for the multi-input multi-output (MIMO) case, this poses restrictions on the model structures. In particular, the MIMO model structure suggested

[1] Corresponding author: marquardt@lpt.rwth-aachen.de

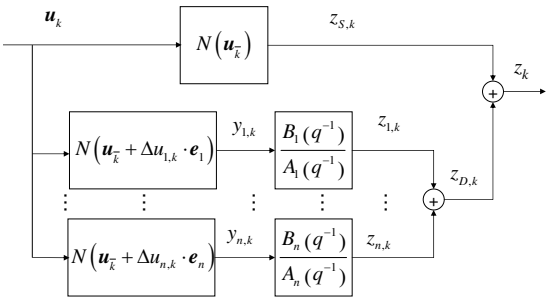


Fig. 1. Block diagram of the HM model.

by Kortmann and Unbehauen (Kortmann and Unbehauen, 1987) has been used previously and we will refer to it as the KU model in the sequel.

In contrast to the inversion based method, we are directly solving the nonlinear dynamic optimization problem constrained by block structured models. Therefore, rst order derivatives of the objective and constraints with respect to the degrees of freedom of the dynamic optimization problem are required. For rigorous dynamic models the calculation of this sensitivity information oftentimes dominates the computational cost of the solution process. We aim at reducing the computational cost by exploiting the block structure for e cient calculation of sensitivity information. Our method covers all MIMO Hammerstein as well as Uryson (Gallman, 1975) models. It allows the solution of the offline optimal control problem. State estimation for such models, required for closed loop control implementation, is the focus of current research.

2. PROBLEM STATEMENT

The constrained, discrete time optimal control problem

$$\min_{\{\boldsymbol{u}_k\}} \ \Phi(\{\boldsymbol{x}_k\},\{\boldsymbol{u}_k\}) \tag{1a}$$
$$s.t. \ \ \boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{(k\ \ 1)}, \boldsymbol{u}_{(k\ \ 1)}) \tag{1b}$$
$$\mathbf{0} \geq \boldsymbol{g}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k) \tag{1c}$$
$$\boldsymbol{x}_0, \boldsymbol{u}_0 \tag{1d}$$
$$k = 1 \ldots K \tag{1e}$$

is given with the objective function $\Phi(\ )$, the manipulated variables $\{\boldsymbol{u}_k\}$, partly measurable state variables $\{\boldsymbol{x}_k\}$, inequality constraints $\boldsymbol{g}(\ )$, process model $\boldsymbol{f}(\ )$, and initial conditions $\boldsymbol{x}_0$, $\boldsymbol{u}_0$. By $\{\ \}$ we denote discrete time sequences of variables, while bold symbols denote vector variables. A function $h(\{\boldsymbol{x}_k\})$ denotes $h(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_K)$. Given the limited computation time available for NMPC, some form of model reduction is required for large process models $\boldsymbol{f}(\ )$. In this paper we assume, that $\boldsymbol{f}(\ )$ can be approximated by a discrete time Hammerstein or Uryson model (Pearson, 1999). Gradient based solution methods require at least rst order derivatives of the objective and constraints with respect to the degrees of freedom, for which sensitivity equations are developed in this paper.

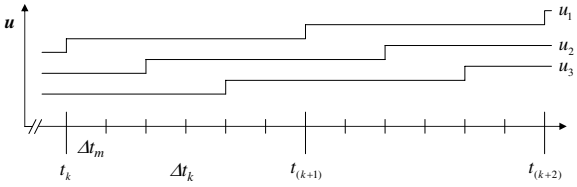


Fig. 2. Oversampling example.

To derive the sensitivity equations we rst treat the single-input single-output (SISO) case for the sake of simplicity. For this case, we approximate problem (1) by

$$\min_{\{u_k\}} \ \Phi(\{z_k\},\{u_k\}) \tag{2a}$$
$$s.t. \ \ \sum_{i=0}^{dim(\boldsymbol{a})} a_i \ \ z_{(k\ \ i)} = \sum_{i=0}^{dim(\boldsymbol{b})} b_i \ \ y_{(k\ \ i)} \tag{2b}$$
$$y_k = N(u_k) \tag{2c}$$
$$0 \geq \boldsymbol{g}(z_k, u_k, t_k) \tag{2d}$$
$$\{z_0\} = \boldsymbol{z}_0, \ \{u_0\} = \boldsymbol{u}_0 \tag{2e}$$
$$k = 1 \ldots K\,. \tag{2f}$$

In problem (2) the reduced process model is de ned by the nonlinear static map $N(\ ) : \mathbb{R}^1 \rightarrow \mathbb{R}^1$, and a linear dynamic process model with gain normalized to one, which is de ned by $\boldsymbol{a}$ and $\boldsymbol{b}$. $\{u_k\}$ and $\{z_k\}$ are the measurable input and output variables and $\{y_k\}$ is the nonmeasurable intermediate variable. $\Phi(\ )$ is the objective function, $\boldsymbol{g}(\ )$ are inequality constraints, and $\{u_0\}$ and $\{z_0\}$ are sequences of delayed inputs and outputs at t_0 de ning the initial condition of the system. Note that the di erence between problems (1) and (2) is a replacement of the original process model by a reduced model. The objective and inequality constraints in (2) only contain the measurable variable $\{z_k\}$ instead of the full state vector $\{\boldsymbol{x}_k\}$.

We extend the method to the more relevant MIMO case in Section 3.2. For the MIMO case several Hammerstein structures have been developed. In this paper we will use the Hammerstein model based on deviation dynamics, which is discussed in detail and compared to the other structures by Harnischmacher and Marquardt (2005). This model is the only one to consistently extend the concept of the Hammerstein model comprising a nonlinear static map followed by an independent linear process model to the multi-input single-output (MISO) case. We will term it HM model in the sequel. The model consists of a static channel and $n = dim(\boldsymbol{u})$ dynamic channels j as depicted in Figure 1. As this model is similar to Uryson models (Gallman, 1975), the results for the MISO case straightforwardly extend to this model class as well. For the MIMO case, problem (1) is approximated using the HM model by

$$\min_{\{\boldsymbol{u}_k\}} \Phi(\{\boldsymbol{z}_k\},\{\boldsymbol{u}_k\}) \tag{3a}$$

$$s.t.\ z_{l,k} = N_l(\boldsymbol{u}_k) + \sum_{j=1}^{dim(\boldsymbol{u})} z_{l,j,k} \tag{3b}$$

$$\sum_{i=0}^{dim(\boldsymbol{a}_{l,j})} a_{l,j,i}\ \ z_{l,j,(k\ \ i)} = \sum_{i=0}^{dim(\boldsymbol{b}_{l,j})} b_{l,j,i}\ \ y_{l,j,(k\ \ i)} \tag{3c}$$

$$y_{l,j,k} = N_l(\boldsymbol{u}_{\overline{k}} + \ \ u_{j,k}\ \boldsymbol{e}_j) \tag{3d}$$

$$0 \geq \boldsymbol{g}(\boldsymbol{z}_k, \boldsymbol{u}_k, t_k) \tag{3e}$$

$$\{\boldsymbol{z}_{l,0}\} = Z_{l,0}, \{\boldsymbol{u}_0\} = U_0 \tag{3f}$$

$$k = 1...K,\ l = 1...dim(\boldsymbol{z}),\ j = 1...dim(\boldsymbol{u}). \tag{3g}$$

In this case, each element of the input and output sequences $\{\boldsymbol{u}_k\}$ and $\{\boldsymbol{z}_k\}$ is of dimension $dim(\boldsymbol{u})$ and $dim(\boldsymbol{z})$ respectively. $\boldsymbol{N}(\ ) : \mathbb{R}^{dim(\boldsymbol{u})} \rightarrow \mathbb{R}^{dim(\boldsymbol{z})}$ is a nonlinear static map of the process and $N_l(\ ) : \mathbb{R}^{dim(\boldsymbol{u})} \rightarrow \mathbb{R}^1$ denotes the l^{th} component of $\boldsymbol{N}(\ )$. $\boldsymbol{u}_{\overline{k}}$ is a reference value for $\boldsymbol{u}$, which is updated at every t_k, and $u_{j,k} = u_{j,k} \ \ u_{j,\overline{k}}$ is the deviation thereof in the direction of the unit vector $\boldsymbol{e}_j$. In this structure the nonlinear element in each channel j represents the local gain of the nonlinear map $N_l(\ )$ in the direction of $\boldsymbol{e}_j$ at $\boldsymbol{u}_k$. To derive the linear elements, linear SISO systems $G_{l,j} : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ are identi ed for all $l = 1...dim(\boldsymbol{u})$ and $j = 1...dim(\boldsymbol{z})$. The parameters $\boldsymbol{a}_{l,j}$ and $\boldsymbol{b}_{l,j}$ are then derived analytically after normalizing the gain to one just as in the SISO case. $\boldsymbol{g}(\ )$ are inequality constraints, $\Phi(\ )$ the objective, and U_0, $Z_{l,0}$ the initial conditions as before.

Model (3b-d) decouples the static response of the system with respect to its inputs to maintain the independence of the nonlinear and linear elements. This decoupling is based on the decomposition of the Taylor expansion of $N_l(\ )$. It is exact, i the second and higher order terms of the Taylor expansion, e.g. $u_{j_1,k}\ \ u_{j_2,k} \frac{\partial^2 N_l(\cdot)}{\partial u_{j_1} \partial u_{j_2}}\Big|_{\boldsymbol{u}=\boldsymbol{u}_{\overline{k}}}$ are equal to zero, which is generally not the case. To meet this condition, we ensure that the input $u_{j,k}$ is di erent from zero for at most one j for $dim(\boldsymbol{b}_{j^*})$ 1 intervals by oversampling the model. The model is sampled at an internal sampling interval of t_m, such that $t_k = t_m \sum_{j=1}^{dim(\boldsymbol{u})} (dim(\boldsymbol{b}_j)$ 1). The response of the system to the input $\boldsymbol{u}_k$ is then calculated by sequentially processing the inputs u_j. We de ne

$$\boldsymbol{u}_{k_n} = [u_{1,k}, ..., u_{n,k}, u_{(n+1),(k\ \ 1)}, ..., u_{dim(\boldsymbol{u}),(k\ \ 1)}]^T \tag{4}$$

for $n = 1...dim(\boldsymbol{u})$. The sequential processing is depicted in Fig. 2 for an example with $dim(\boldsymbol{u}) = 3$, $dim(\boldsymbol{z}) = 1$, and $dim(\boldsymbol{b}_j) = 3\ \forall\ j$. At time t_k the input $u_{1,k}$ is processed and the input is held constant for the following interval t_m. Hence, for the oversampled model, the input is $\boldsymbol{u}_{k_1}$ for $dim(\boldsymbol{b}_1)$ 1 intervals. $u_{2,k}$ is processed at $t_k + 2\ t_m$ and again the input unchanged in the following interval t_m ensuring a constant input $\boldsymbol{u}_{k_2}$ for $dim(\boldsymbol{b}_2)$ 1 intervals and so on. By oversampling, the input $\boldsymbol{u}_k$ is turned into a sequence of inputs $\boldsymbol{u}_m$ for the oversampled model, which will be of importance for the sensitivity calculation. The input $\boldsymbol{u}_m$ to the oversampled model is given by

$$u_{j,m} = \begin{cases} u_{j,k\ \ 1} \ \forall\ t_k \leq t_m < t_k + \sum_{i=1}^{j} (dim(\boldsymbol{b}_i)\ \ 1)\ \ t_m \\ u_{j,k} \ \ \forall\ t_k + \sum_{i=1}^{j} (dim(\boldsymbol{b}_i)\ \ 1)\ \ t_m \leq t_m < t_{k+1}. \end{cases} \tag{5}$$

3. SENSITIVITY EQUATIONS FOR HAMMERSTEIN SYSTEMS

3.1 SISO Case

For the SISO case the sensitivity of z_k with respect to an input u_{k^*} is straightforwardly calculated using the chain rule of di erentiation from

$$\frac{\partial z_k}{\partial u_{k^*}} = \frac{\partial z_k}{\partial y_{k^*}} \ \ \frac{\partial y_{k^*}}{\partial u_{k^*}}. \tag{6}$$

As Eq. (2b) is linear in y_k, solving the recursion for z_k yields

$$z_k = \xi_{k,k^*}(\boldsymbol{a},\boldsymbol{b})\ y_{k^*} + \ \ (\boldsymbol{a},\boldsymbol{b},\{y_{k \neq k^*}\},\boldsymbol{u}_0,\boldsymbol{z}_0), \tag{7}$$

where ( ) is a polynomial containing all elements of $\{y_k\}$ but y_{k^*} and ξ_{k,k^*} is a constant polynomial of $\boldsymbol{a}$ and $\boldsymbol{b}$. The rst term of Eq. (6) is therefore

$$\frac{\partial z_k}{\partial y_{k^*}} = \xi_{k,k^*}(\boldsymbol{a},\boldsymbol{b}) := const. \tag{8}$$

The second term of Eq. (6)

$$\frac{\partial y_{k^*}}{\partial u_{k^*}} = \frac{\partial N(u)}{\partial u}\bigg|_{u=u_{k^*}} \tag{9}$$

is just the rst order derivative of the nonlinear static element $N(u)$ at $u = u_{k^*}$.

Due to the structure of the Hammerstein model, the sensitivity calculation can thus be reduced to the calculation of one rst order derivative of $N(\ )$ and one vector multiplication

$$\frac{\partial \{z_k\}}{\partial u_{k^*}} = \boldsymbol{\xi}_{k^*} \ \ \frac{\partial N(u)}{\partial u}\bigg|_{u=u_{k^*}} \tag{10}$$

with $\boldsymbol{\xi}_{k^*} = [\xi_{1,k^*}, \ldots, \xi_{K,k^*}]$.

3.2 MIMO Case

MIMO Hammerstein and Uryson structures generally consist of parallel branches of MISO or SISO Hammerstein models. Hence, the sensitivity calculation is a straight forward extension of the SISO case. The computational e ort varies with the respective Hammerstein structure. For the KU model (Kortmann and Unbehauen, 1987) only the derivatives of $dim(\boldsymbol{u})$ scalar functions are required, while the model based on combined nonlinearities (Eskinat *et al.*, 1991) requires $dim(\boldsymbol{u})$

gradients of the respective nonlinear models. However, to our knowledge no control application based on the solution of the nonlinear dynamic optimization problem has been reported.

Because of the oversampling the sensitivity calculation for the HM model is a little more complex, but since it also consists of parallel Hammerstein channels, the structure of the solution remains the same. As Eq. (3) contains $dim(\boldsymbol{z})$ parallel MISO models, we will only treat the MISO case in this section and therefore drop the index l of Eq. (3) for the remainder of this section to ease the notation. Since Eq. (3) consists of parallel branches of Hammerstein systems, the sensitivity equations developed in this section are structurally equivalent to the SISO case. In particular Eq. (8) holds for each of the dynamic channels of Eq. (3c). We therefore use the following notation for the remainder of this section:

$$\xi_{j,k,k^*} := \frac{\partial z_{j,k}}{\partial y_{j,k^*}}. \tag{11}$$

The sensitivity of z_k with respect to $\boldsymbol{u}_{k^*}$ is given by

$$\frac{\partial z_k}{\partial \boldsymbol{u}_{k^*}} = \frac{\partial z_{S,k}}{\partial \boldsymbol{u}_{k^*}} + \sum_{j=1}^{dim(\boldsymbol{u})} \frac{\partial z_{j,k}}{\partial \boldsymbol{u}_{k^*}}. \tag{12}$$

The first term in Eq. (12) contains the sensitivity of the static channel S of the model, which is simply

$$\frac{\partial z_{S,k}}{\partial \boldsymbol{u}_{k^*}} = \left.\frac{\partial N(\boldsymbol{u})}{\partial \boldsymbol{u}}\right|_{\boldsymbol{u}=\boldsymbol{u}_{k^*}} \tag{13}$$

and zero for all $k \neq k$

The sensitivity calculation for the dynamic channels follows the same concept and the same simplification as in the SISO case. However as depicted in Fig. 2 the input $\boldsymbol{u}_{k^*}$ is in fact an input sequence to the oversampled model. $\left.\frac{\partial N(\boldsymbol{u})}{\partial \boldsymbol{u}_{k^*}}\right|_{\boldsymbol{u}=\boldsymbol{u}_{k_n}}$ is nonzero for the sequence $\{\boldsymbol{u}_{k_n^*}, ..., \boldsymbol{u}_{(k^*+1)_{n\ \ 1}}\}$. $\frac{\partial z_{j,k}}{\partial u_{j,k^*}}$ for the dynamic channels is then given by

$$\frac{\partial z_{n,k}}{\partial u_{j,k^*}} = \xi_{n,k,k^*} \left.\frac{\partial N(\boldsymbol{u})}{\partial u_j}\right|_{\boldsymbol{u}_{k_n^*}} \xi_{n,k,(k^*+1)} \left.\frac{\partial N(\boldsymbol{u})}{\partial u_j}\right|_{\boldsymbol{u}_{(k^*+1)_n}} \tag{14}$$

for channel $n = j$, by

$$\frac{\partial z_{n,k}}{\partial u_{j,k^*}} = \xi_{n,k,(k^*+1)} \left(\left.\frac{\partial N(\boldsymbol{u})}{\partial u_j}\right|_{\boldsymbol{u}_{(k^*+1)_n}} \left.\frac{\partial N(\boldsymbol{u})}{\partial u_j}\right|_{\boldsymbol{u}_{(k^*+1)_{(n\ \ 1)}}} \right) \tag{15}$$

for all channels $n = 1 \ldots j \quad 1$, and analogously

$$\frac{\partial z_{n,k}}{\partial u_{j,k^*}} = \xi_{n,k,k^*} \left(\left.\frac{\partial N(\boldsymbol{u})}{\partial u_j}\right|_{\boldsymbol{u}_{k_n^*}} \left.\frac{\partial N(\boldsymbol{u})}{\partial u_j}\right|_{\boldsymbol{u}_{k^*_{(n\ \ 1)}}} \right) \tag{16}$$

for all channels $n = j + 1 \ldots dim(\boldsymbol{u})$.

As in the SISO case, the integration of the sensitivity system for the MISO case can therefore be reduced to calculation of the 2 $dim(\boldsymbol{u})$ gradients of $N(\boldsymbol{u})$ at $\boldsymbol{u}_{k_1^*} \ldots \boldsymbol{u}_{(k^*+1)_{dim(\boldsymbol{u})}}$ and a set of matrix multiplications

$$\begin{aligned} \frac{\partial\{z_k\}}{\partial \boldsymbol{u}_{k^*}} = & \sum_{j=1}^{dim(\boldsymbol{u})} \left.\frac{\partial N(\boldsymbol{u})}{\partial \boldsymbol{u}}\right|_{u_{k_j^*}} \Xi_{k^*,j} \ + \\ & \sum_{j=1}^{dim(\boldsymbol{u})} \left.\frac{\partial N(\boldsymbol{u})}{\partial \boldsymbol{u}}\right|_{u_{(k^*+1)_j}} \Xi_{(k^*+1),j} \ , \end{aligned} \tag{17}$$

where $\Xi_{k^*,j}$ and $\Xi_{(k^*+1),j}$ contain the respective vectors $\boldsymbol{\xi}_{j,k^*}$ and $\boldsymbol{\xi}_{j,(k^*+1)}$ analogously to Eq. (10).

4. COMPARISON WITH COMPETING METHODS

Directly competing are the inversion based methods using Wiener or Hammerstein models (Zhu and Seborg (1994), Norquay *et al.* (1999)). They o er slight advantages in computational cost, but are known to possibly su er from non-uniqueness, when the nonlinear map is not bijective over the input space. This severely limits the nonlinear maps as well as the multivariable structures that can be used. Further, the objective function of the linear optimization problem contains the intermediate variable of the model as a proxy variable for either the output or the input to the system. As these are nonlinearly linked, the solution of the linear problem generally does not minimize the original objective. Finally, the inversion based solution of nonlinear dynamic optimization problems constrained by Uryson models is not possible, because intermediate variables y_κ of the different channels κ of the Uryson model, which are independent variables in the linear optimization problem, are in fact nonlinearly coupled.

The e ciency of the sensitivity calculation for Hammerstein systems is greatly increased by making use of Eq. (10), which does not hold for Wiener systems. The sensitivity of z_k with respect to u_{k^*} for a SISO Wiener system can be calculated from

$$\frac{\partial z_k}{\partial u_{k^*}} = \frac{\partial z_k}{\partial y_k} \ \frac{\partial y_k}{\partial u_{k^*}}. \tag{18}$$

In this case $\frac{\partial z_k}{\partial y_k} = \left.\frac{\partial N(\cdot)}{\partial y}\right|_{y=y_k}$ needs to be evaluated at every t_k. Thus, for Wiener systems the solution of the nonlinear dynamic optimization problem is computationally much more demanding, because the derivative of the nonlinear map has to be evaluated on the discretization of the output instead of the discretization of the input. When nonlinear maps other than polynomials are used, the evaluation of the nonlinear map dominates the computational cost (Harnischmacher *et al.*, 2006).

5. SIMULATION EXAMPLE

As a simulation example we choose the industrially relevant uid catalytic cracking (FCC) unit, for which several models exist in the open literature. We use the model originally developed by Kurihara and comprehensively discussed by Denn (1986). This model has been validated and used for control by Ansari and Tadé (2000). We will not restate the equations here due to space limitations. The nomenclature and units used in the sequel are the same as those of Denn (1986), where the complete model may be found. Ansari and Tadé (2000) also state the complete model, but with some typographical error and a slightly di erent notation. Detailed process descriptions can be found in both references. The example shows, that the solution of the nonlinear dynamic optimization problem can be performed in very short time and the increased modeling exibility leads to signi cant improvements in performance.

5.1 Simulated FCC Unit

The main manipulated variables of the process are the air owrate R_{ai} and the catalyst circulation rate R_{rc}, while the feed rate R_{tf} and feed temperature T_{fp} are treated as disturbances. To control the main quality variable, the cracking severity, several controlled variables have been explored due to the complex dynamics of the system. However the riser outlet temperature T_{ra} is directly related to the cracking severity and has recently been used for control (Jia *et al.*, 2003). The control problem is therefore non-square with manipulated variables R_{ai} and R_{rc} and controlled variable T_{ra}.

5.2 Identi cation

The simulated FCC unit is identi ed using two di erent Hammerstein model structures. For the inversion based method we use the KU model (Kortmann and Unbehauen, 1987). Quadratic functions are used in each of the two channels of the model. For the proposed method, the HM model (Harnischmacher and Marquardt, 2005) is used. Here, the nonlinear map is an arti cial neural network (ANN) identi ed from steady state data. For both models fourth order linear elements are identi ed from step response data.

The FCC process is known to exhibit a two timescale behavior (Christo des and Daoutidis, 1997). The models identi ed above give a poor description of the short time scale behavior of the process and a Uryson model, containing two dynamic channels for each input, is much more suitable (Gallman, 1975). As the response on the fast time scale is close to linear, constant gains are used in these two channels, while the same ANN as in the HM model is used in the two long time scale channels. The long time scale dynamic behavior of the system is described by rst order models, while models of third order are identi ed for the fast time scale channels.

5.3 Open-Loop Optimal Control

The control objective

$$\Phi = (\boldsymbol{T}_{ra} \quad \boldsymbol{T}_{set})^T(\boldsymbol{T}_{ra} \quad \boldsymbol{T}_{set}) + \quad \sum R_{rc,i} \quad (19)$$

is to be minimized. The time horizon is 1000 intervals t_k corresponding to two hours simulation time. The inputs $R_{ai} \in [390; 420]\frac{Mlb}{hr}$ and $R_{rc} \in [40; 42]\frac{ton}{min}$ are piecewise constant for 100 intervals t_k. $\boldsymbol{T}_{ra} = [z_{50}, z_{100}, ..., z_{1000}]^T$ contains the model output sampled every 50 intervals. The set point $\boldsymbol{T}_{set}$ changes from 950°F to 960°F at $k = 201$. $\boldsymbol{R}_{rc}$ contains the absolute values of R_{rc} as a proxy for process cost. $= 10^{-4}$ is a weighting parameter.

For the inversion based method $\{\boldsymbol{u}_k\}$ is given by the roots of two independent quadratic functions, i.e. the nonlinear maps of the model. This leads to four possible solutions. In our case, however, the nonlinear functions are monotonous on the respective input spaces. While this leads to a poor description of the process nonlinearity in a certain section of the input space with steady state errors of up to 13°F, it follows that only one of the four solutions lies in the input space and the solution of the optimization problem is therefore unique. Such behavior of the nonlinear map cannot be expected in general and would pose severe restrictions on the nonlinear map.

5.4 Discussion

The nonlinear optimization problems with both the Hammerstein and Uryson models are solved in less than 1 second using MATLAB on a 1.5 GHz PC. Such computation times are well acceptable for NMPC applications in the process industry.

Simulation results for the manipulated variable trajectories obtained by using the di erent models are depicted in Fig. 3, which as a reference also contains the result obtained by solving the original dynamic optimization problem with the original model. This solution clearly outperforms all approximate solutions. It should be noted though, that for this simulation example, there is absolutely no plant model mismatch when the original model is used. The inversion based method, in contrast, performs worst. We compare the performance by the objective values obtained by simulating the original model with the inputs $\{\boldsymbol{u}_k\}$ calculated with the four di erent models. Using the HM model leads to a slight improvement of 15% in the original objective compared to the inversion based method. The weak performance of

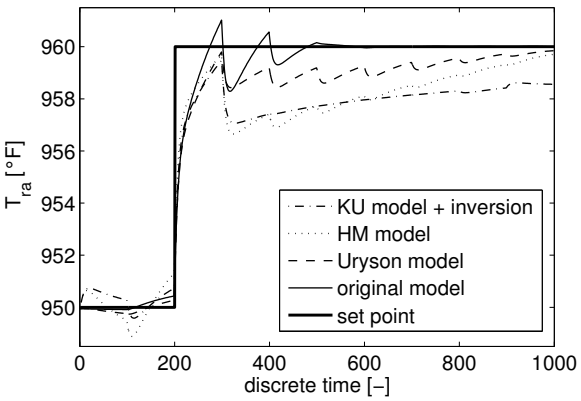


Fig. 3. Trajectory optimization results.

both methods is mainly due to insu cient modeling of the process dynamics.

Solving the nonlinear dynamic optimization problem constrained by a Uryson model leads to a reduction of over 80% in the original objective compared to the inversion based method. Further performance increases can be achieved by using a rigorous steady state model instead of the ANN. This leads to a reduction of 85% in the original objective. However this slight additional improvement comes at a cost of 160 seconds of computation time making this model computationally unattractive. For comparison the improvement in objective for the original model is 94% after 270 seconds of computation time.

6. CONCLUSIONS

Block structured models are well suited for nonlinear model predictive control because of the simple identi cation and low computational cost. Previous approaches aimed at reducing the computational cost by the inversion of the nonlinear element. This requires the nonlinear map to be bijective, excludes the use of Uryson models, and leads to a loss in optimality because of the nonlinear coupling between the proxy variable used in the objective of the linear optimization problem and its counterpart in the original objective. Sensitivity equations have been derived for multivariable Hammerstein and Uryson models to allow the solution of nonlinear optimization problems constrained by these models at low computational cost. An example problem with a non-square controller with two inputs parameterized on 10 intervals each was solved in less than 1 second and at the same time reduced the optimality loss by over 80% compared to previous methods, because of the increased modeling exibility. Future research will be directed at developing a tailored state estimation method for multivariable Hammerstein models to solve the closed loop NMPC problem. Further, online updating methods for the linear elements will be investigated to increase model accuracy.

REFERENCES

Ansari, R.M. and M.O. Tadé (2000). Constrained nonlinear multivariable control of a uid catalytic cracking process. *J. Proc. Contr.* **10**(6), 539–555.

Christo des, P.D. and P. Daoutidis (1997). Robust control of multivariable two-time-scale nonlinear systems. *J. Proc. Contr.* **7**(5), 313–328.

Denn, M. M. (1986). *Process Modeling.* Pitman Publishing. Marsh eld, MA.

Eskinat, E., S.H. Johnson and W.L. Luyben (1991). Use of Hammerstein models in identi cation of nonlinear systems. *AIChE J.* **37**(2), 255–268.

Gallman, P.G. (1975). An iterative method for the identi cation of nonlinear systems using a Uryson model. *IEEE Trans. Automat. Contr.* **20**(6), 771–775.

Harnischmacher, G. and W. Marquardt (2005). A multi-variable Hammerstein model for input-directional dynamics. *Submitted to IEEE Trans. Automat. Contr.*

Harnischmacher, G., O. Kahrs and W. Marquardt (2006). Identi cation of a uid catalytic cracking unit by means of a new multivariable Hammerstein model. *Accepted for: IFAC Symposium on System Identi cation, Newcastle, 2006.*

Hunt, K.J., M. Munih, N. Donaldson and M.D. Barr (1998). Optimal control of ankle joint moment: Toward unsupported standing in paraplegia. *IEEE Trans. Automat. Contr.* **43**(6), 819–832.

Jia, C., S. Rohani and A. Jutan (2003). FCC unit modeling, identi cation and model predictive control, a simulation study. *Chem. Eng. Process.* **42**(4), 311–325.

Kadam, J., B. Srinivasan, D. Bonvin and W. Marquardt (2005). Optimal grade transition in industrial polymerization processes via NCO tracking. *Submitted to AIChE J.*

Kortmann, M. and H. Unbehauen (1987). Identi - cation methods for nonlinear MISO systems. In: *Proceedings of the IFAC World Congress.* Munich. pp. 225–230.

Marquardt, W. (2002). Nonlinear model reduction for optimization based control of transient chemical processes. In: *AIChE Symp. Ser. 326 Vol. 98.* pp. 12–42.

Norquay, S.J., A. Palazoglu and J.A. Romagnoli (1999). Application of Wiener model predictive control (WMPC) to an industrial C2-splitter. *J. Proc. Contr.* **9**(6), 461–473.

Pearson, R.K. (1999). *Discrete-Time Dynamic Models.* Oxford University Press. Oxford.

Zhu, X. and D.E. Seborg (1994). Nonlinear predictive control based on Hammerstein models. In: *Proceedings of PSE 1994.* pp. 995–1000.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

OPERABILITY OF MULTIVARIABLE NON-SQUARE SYSTEMS

Fernando Lima and Christos Georgakis

Department of Chemical and Biological Engineering &
Systems Research Institute, Tufts University

Abstract: Non-square process control systems, with fewer inputs than the controlled outputs, are quite common in chemical processes. In these systems, it is impossible to control all measured variables at specific set points and many of the outputs are controlled within an interval. The objective of this paper is to introduce a multivariable Operability methodology for such non-square systems to be used in the design of non-square constrained controllers. In order to motivate the new concepts, we examine some simple non-square systems obtained from the control system of a Steam Methane Reformer process. *Copyright © 2006 IFAC*

Keywords: Operability, Intervals, Output Variables, Polygons, MIMO, Model Based Control

1. INTRODUCTION

In this section, the problem definition is described and a brief introduction to prior work, concerning non-square systems and operability issues, is presented.

1.1 *Problem Definition*

In recent years, in the face of increasing complexity of chemical processes due to the integration of units, process optimization and strict environmental regulations the use of tools to evaluate the performance of a control structure has become very important. This has to be done in a more systematic manner than by trial and error closed-loop simulations and before the final controller step. Georgakis et al (2003) mentioned that it has been since the last decade that the integration of process and control design has received considerable attention. Skogestad (2004) emphasized that the field of control structure design in plant-wide control problems, which includes the selection of manipulated and control variables, is underdeveloped. Moreover the majority of the controllability methods developed address the design of multiple input – multiple output (MIMO) systems with respect to interactions and loop pairings, and often apply only to unconstrained systems. Few methods take into account the limited range available for the control inputs during the design phase. The Operability framework developed by Vinson and Georgakis (2000) was a contribution in this direction. The Operability methodology is an effort to integrate the process design and control objectives, helping to cope with the complexity of chemical processes. Essentially, the Operability

measure can quantify the ability of a process to change from one steady-state to another and reject expected disturbances utilizing a limited control action available. This measure is important because once the design is fixed, no control methodology can overcome limitations on operability. It is only with a tool to evaluate the operability of a chemical process that one could analyze appropriately the economic aspects of the process. A review of the literature on integration of process and control design related to square systems can be found in Vinson (2000). A survey concerning mathematical and process-oriented approaches in plant-wide control was presented by Larsson and Skogestad (2000).

Based on the linear and steady-state Operability framework initially developed by Vinson and Georgakis (2000), the objective of this paper is the development of a multivariable *non-square* Operability methodology for linear systems. This would help in the design of non-square Model Predictive Controllers (MPC), with more outputs than inputs, commonly encountered in industrial chemical processes. Basically, MPC controllers are model-based controllers which account for process constraints. Based on the input constraints, generally specified a priori due to physical limitations of the process, an important task is to define the output ranges or constraints within which we want to control the process. The problem is that very tight constraints make the control design difficult, with the possibility of not being able to find the appropriate input variables to achieve the control objective. On the other hand, if the constraints are not tight enough, output specifications, such as desired product quality or environmental regulations, cannot be achieved. Therefore the Operability methodology can serve an important role in solving this problem. Through this framework it is possible to verify achievability of control objectives before implementing the MPC controller. In addition, according to Vinson (2000) some of the main features of the MPC strategy, such as being predominantly linear and using constraints for each manipulated and controlled variable, are directly associated with the developed operability framework. This problem functions as motivation for the current effort. The outline of this paper is the following: first a summary of the prior work concerning non-square systems and operability issues is given. Then, the basic theory used in the development of this paper is explained. The results from the analysis of some simple systems are presented next, closing with conclusions.

1.2 *Summary of prior work*

Non-square systems with more outputs than inputs are quite common in chemical processes. Apart from the common non-square nature of some chemical processes, a system with more outputs than inputs may occur if one of the actuators of an original square system is operating at constantly saturated levels. Several studies that analyze aspects of non-square systems can be found in the literature. Reeves and Arkun (1989) developed a block relative gain array (RGA) measure for non-square linear systems as a tool to analyze and evaluate control structures in steady-state before the controller design in order to specify the appropriate control structure. Similarly, Chang and Yu (1990) extended RGA for non-square multivariable systems, defining the non-square relative gain array (NRG). Both studies suggested that for non-square systems with more outputs than inputs, the outputs have to be controlled in the least square sense, minimizing offsets. One very important contribution that examines the design of non-square systems is the concept of Partial Control introduced by Shinnar (1981), and mathematically analyzed by Kothare et al. (2000). This methodology helps the control engineer to choose the appropriate set of measured variables to be controlled at the set-point, in a system having limited degrees of freedom. This choice must be made so that the other outputs can still be controlled at specified ranges while satisfying all the input and performance variable constraints and rejecting all the expected disturbances. This methodology would be useful in selecting the variables in the control design stage after the process operability quantification proposed here has been evaluated.

The Operability Index (OI) was introduced by Vinson and Georgakis (2000) and Vinson (2000) as a measure to access the input-output open-loop controllability of a multivariable square chemical process. The concept of operability given by Vinson (2000) is the following: A process is operable if the available set of inputs is capable of satisfying the desired steady-state and dynamic performance requirements defined at the design stage, in the presence of the set of anticipated disturbances, without violating any process constraints. Vinson and Georgakis (2002) have demonstrated that the Operability Index is independent of the inventory control structure. This property allows one to compare the operability of competing designs before the process control structure is selected or implemented, i.e., during the process synthesis stage. Vinson and Georgakis (2000) have also shown that this measure can be applied to SISO and MIMO

systems and is more appropriate than other design tools such as RGA or minimum singular values. Concerning non-square systems, Vinson (2000) analyzed the ability of the OI to enhance the performance of a non-square MPC controller, specifically DMCplus™ (AspenTech). Finally, an overview of all Operability definitions and concepts has been done by Georgakis et al (2003). In the next section, a brief explanation of the concepts and definitions of the Operability framework is given.

2. PROPOSED APPROACH FOR PROBLEM SOLUTION

The Operability Index (OI) was introduced by Vinson and Georgakis (2000) for analyzing square systems. It provides a quantitative result for multivariable systems and a graphical representation for systems less than 3-D, permitting the design to be modified in order to improve process operability before the control structure selection.

2.1 *Operability of Square Systems: Servo Operability Measure*

To make the idea of the Operability measure clear, it is necessary to define some useful spaces. The Available Input Space (*AIS*) is the set of values that the process input variables can take based on the design of process, limited by process constraints. Mathematically for an *n x n* square system:

$$AIS = \left\{ \underset{\sim}{u} \mid u_{A,i}^{\min} \le u_{A,i} \le u_{A,i}^{\max}; 1 \le i \le n \right\}.$$

Moreover, the Desired Output Space (*DOS*) is given by the desired values of the outputs of the process and is represented by:

$$DOS = \left\{ \underset{\sim}{y} \mid y_{D,i}^{\min} \le y_{D,i} \le y_{D,i}^{\max}; 1 \le i \le n \right\}.$$

Based on the steady-state model of the process, expressed by the process gain matrix (*G*), the Achievable Output Space (*AOS*) is defined by the output values that can be achieved using the inputs inside the *AIS*. We will use the notation $AOS_u(d^N)$ for the *AOS* calculated considering all points inside *AIS* (subscript u) when the disturbances lie in their nominal values (d^N), i.e., for the servo problem. For this problem, the outputs in the $AOS_u(d^N)$ are calculated by: $y=G(u)$, where $u \in AIS$. Based on those definitions, the Servo Operability Index with respect to the outputs is the following:

$$s - OI_y = \frac{\mu(AOS_u(d^N) \cap DOS)}{\mu(DOS)} \quad (1)$$

Where μ represents a function that calculates the size of the space, for example for 3-D it is the volume and in 2-D the area. This index quantifies how much of the region of desired outputs can be achieved using the available inputs in the absence of disturbances and is useful in analyzing changes in the existing plant design to enlarge *AOS*. This Index has a value between 0 and 1. A process is considered completely operable if the index is equal to 1. If the OI is less than one, some regions in the *DOS* are not achievable. It is worth emphasizing that to calculate the OI, mathematical operations involving intersections of polytopes have to be performed. In this work, the MATLAB (Mathworks™, Inc) Geometric Boundary Toolbox (GBT) has been used. It was developed by Veres et al (1996) to evaluate those intersections. Details concerning the servo Operability Index with respect to the inputs and the Desired Input Space (*DIS*) can be found in Georgakis et al (2003), as well as in Vinson and Georgakis (2000).

2.2 *Operability of Non-Square Systems*

In order to quantify the operability of non-square linear systems, some modifications to the definitions initially proposed by Vinson and Georgakis (2000) are required. First, it is worth classifying the process outputs according to the way that they have to be controlled into two categories: *set-point controlled*: variables that are controlled in their exact set-point (for instance, production rates and product qualities); *set-interval controlled*: variables that are controlled at specified ranges (pressure, temperature and level); the operability in the latter case is defined as *interval operability*. The set-point and range variables can be chosen according to an economic objective and given by a supervisory strategy. The idea of the new definition of operability is to fix critical outputs at their set-points, allowing the others to vary within their maximum and minimum limits defined a priori. This definition should also recognize the necessity to control some outputs at ranges rather than require that all points of the *DOS* be reached. In interval operability, process outputs must have at least one feasible operating point within the desired interval for the process to be considered operable. Using the same *AOS* definition as in the square case will lead us to a poor definition of operability. Therefore, it is necessary to redefine the Achievable Output Space in

order to analyze the non-square operability issue properly. At this point, it is necessary to define the Expected Disturbance Space (*EDS*). This space represents all the steady-state disturbances that affect the process which can also be used to reflect uncertainties in model parameters employed in the design. Finally, the goal is to formulate a multivariable steady-state methodology to obtain the *AOS,* given *EDS* and *AIS*, to quantify the operability of any non-square linear system. As a starting point in this development, we will examine 2 simple cases in the next section which involve sub-systems of the Steam Methane Reformer (SMR) process that has 4 manipulated, 1 disturbance and 9 controlled variables.

3. RESULTS

A 2 x 1 sub-system of the SMR model cited in the previous section will be used as a starting point to demonstrate the importance of the proposed problem. It is worth mentioning that the SMR model has only non-integrating outputs, and the process dynamics are neglected here since we are studying steady-state Operability. Using the same notation as above and considering the effect a disturbance has on the process, we write:

$$y = G u_1 + G_d w_1 \quad \Rightarrow \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} u_1 + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} w_1 \qquad (2)$$

Where w_1 represents the Expected Disturbance Space ($EDS = \{w_1 \mid -1 \le w_1 \le 1\}$) and G_d is the disturbance gain matrix. $AIS = \{u_1 \mid -1 \le u_1 \le 1\}$ and $DOS = \left\{ y \in \Re^2 \mid \|y\|_\infty \le 1 \right\}$.

Rearranging equation (2); we have:

$$y_1 = a_{11}u_1 + d_1 w_1 \Rightarrow u_1 = \frac{y_1 - d_1 w_1}{a_{11}} \qquad (3)$$

$$y_2 = a_{21}u_1 + d_2 w_1 \Rightarrow y_2 = a_{21}\frac{y_1 - d_1 w_1}{a_{11}} + d_2 w_1 \qquad (4)$$

Based on the system of equations above, two cases result.

Case 1: Consider the following scaled steady-state gain matrices: $G = [1.41, 0.66]^T$; $G_d = [1, 0]^T$; In this particular case, since $d_2=0$, equation (4) can be rewritten as:

$$y_2 = a_{21}u_1 \Rightarrow y_2 = a_{21}\frac{y_1 - d_1 w_1}{a_{11}} \qquad (5)$$

Thus, the base case servo *AOS* (*AOS* when $w_1 = 0$) is given by a straight line. In this particular case, the disturbance gains shift the servo *AOS* horizontally. This can be observed in Figure 1, where we have also sketched the *DOS*.

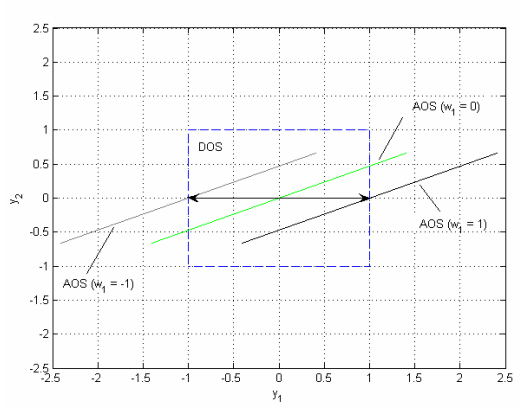


Figure 1: servo *AOS*, shifted *AOS* and *DOS*

The movement of the *AOS* with different disturbance values depends on the values of G_d. If $G_d = [0, 1]^T$, the servo *AOS* would be moving vertically rather than horizontally as in the previous case. For every value of the disturbance, a different straight line is obtained. The union of all the Achievable Output Spaces that correspond to all the expected disturbance values will be called *AOS(d)*. This space and the *DOS* are plotted in Figure 2, along with the Achievable Output Interval Space (*AOIS*). This new space, *AOIS* represents the rectangle that touches, but does not cross, the lines associated with the minimum and maximum disturbance values of *AOS(d)*. This means that if we control the two outputs within some constraints that are not larger than the *AOIS*, the process will not be interval operable with the available input ranges and in the presence of the expected disturbances. In other words, the system will be interval operable if *DOS* covers *AOIS* completely. Therefore, the Interval Operability Index with respect to the outputs (IOI_y) can be now defined as:

$$IOI_y = \frac{\mu(DOS \cap AOIS)}{\mu(AOIS)} \qquad (6)$$

Where μ is the area of the corresponding polygon in this example, and the volume for the 3-D case. It is worth mentioning that we are assuming a rectangular

DOS. If the *DOS* is not rectangular, the solution has to be modified appropriately.

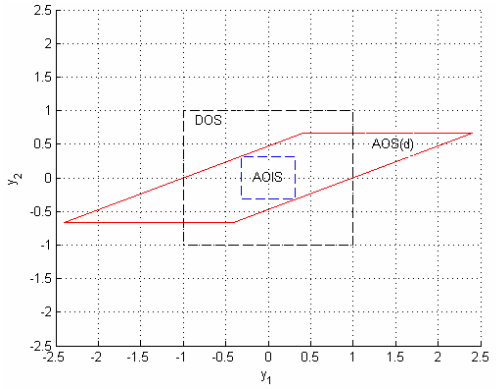


Figure 2: *AOS(d)*, *DOS* and *AOIS* for case 1

Case 2: Disturbance inserted in both output variables: Assume the same *AIS*, *DOS* and *EDS* from the previous case and the following gain matrices: $G = [1.41,\ 0.66]^T$; $G_d = [-0.6,\ 0.4]^T$;

In this case, the system of equations (3) and (4) holds. The servo *AOS* is now shifted along a diagonal, as shown in Figure 3. In this figure we have also plotted an example of a *DOS* and *AOIS* calculated for this case.

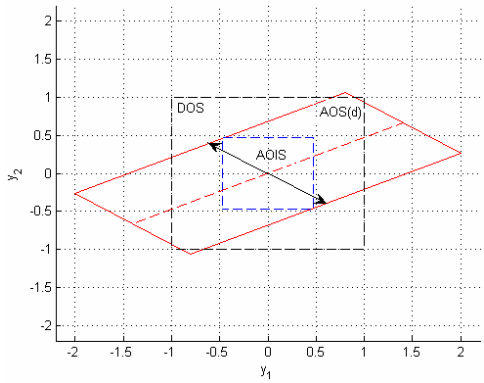


Figure 3: *AOS(d)*, *DOS* and *AOIS* for case 2

For both 2x1 cases above, the system is fully interval operable, since *DOS* is large enough to cover *AOIS*. We can actually see that if we make the *DOS* smaller than the *AOIS*, the process will not be interval operable.

An inoperable case would happen if the disturbances affecting the process were increased in absolute value: $EDS = \{w_1 \mid -3 \le w_1 \le 3\}$. Thus, the *AOIS* would also be enlarged. Figure 4 shows *AOS(d)*, *DOS* and *AOIS* for this scenario. In this case, we notice that the IOI_y would be smaller than 1. This means that the system is only interval operable for some disturbance values considered. In order to make it fully operable, the process constraints should be relaxed, which means the *DOS* should be enlarged to cover the *AOIS* entirely.

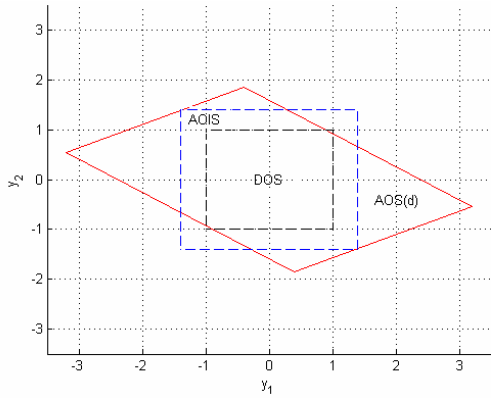


Figure 4: *AOS(d)*, *DOS* and *AOIS* for inoperable case

Now, the problem of controlling 3 outputs at ranges, when having only 2 inputs, will be presented here. The system of equations considered is:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} w_1 \qquad (7)$$

The gain matrices from the SMR process are the following:

$$G = \begin{bmatrix} 1.41 & 0.27 \\ -0.39 & -0.20 \\ 0.66 & 0.49 \end{bmatrix};\ G_d = \begin{bmatrix} 0.2 \\ 0.4 \\ 0.4 \end{bmatrix};$$

Also: $DOS = \{y \in \Re^3 \mid \|y\|_\infty \le 1\}$, $AIS = \{u \in \Re^2 \mid \|u\|_\infty \le 1\}$ and *EDS* is assumed the same as before.

The *AOS(d)* is now the union of all planes, instead of straight lines, corresponding to different values of disturbances affecting the process. Thus, *AOS(d)* will be an oblique parallelepiped, and *DOS* and *AOIS* are, in general, orthogonal parallelepipeds and, in this case, cubes. *AOS(d)* and *DOS* are displayed on Figure 5. Figure 6 shows *AOS(d)* and *AOIS*, and Figure 7 shows *DOS* and *AOIS*.

Observing Figure 6, one notices that *AOIS* touches both extreme planes of *AOS(d)*. As drawn in Figure 7, the *DOS* is quite larger than the calculated *AOIS*. This implies that the *DOS* can be reduced in size and the process will continue to be output operable as long as *AOIS* continues to be a subset of *DOS*.

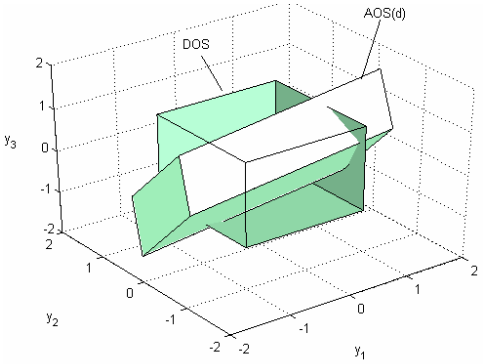


Figure 5: *AOS(d)* and *DOS* - 3x2 problem

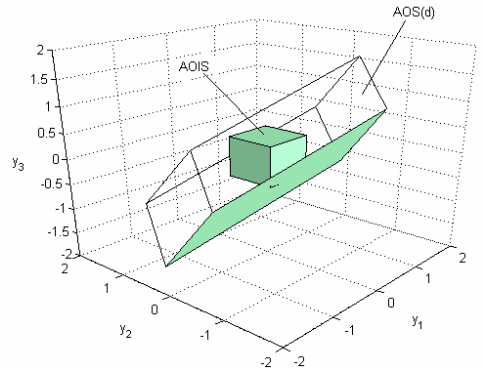


Figure 6: *AOS(d)* and *AOIS* - 3x2 problem

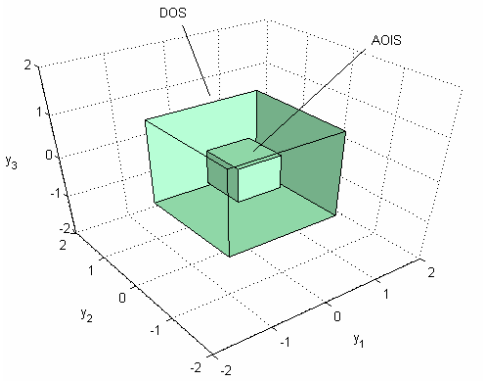


Figure 7: *DOS* and *AOIS* - 3x2 problem

4. CONCLUSIONS

In this paper we presented an extension of the previously defined concept of operability to the case of non-square systems, where some of the output variables need to be controlled within intervals rather than a set-point. Through the detailed examination of 2 case studies we have demonstrated the motivation for calculation of the Achievable Output Interval Space (*AOIS*) as the smallest possible interval constraints for the outputs that can be achieved with the available range of the manipulated variables and when the disturbances remain within their expected values.

REFERENCES

Chang, J. W., Yu, C. C. (1990). The relative gain for Nonsquare multivariable systems. *Chemical Engineering Science*, *45*, 1309-1323.

Georgakis, C., Uzturk, D., Subramanian, S., Vinson, D. R. (2003). On the operability of continuous processes. *Control Engineering Practice, 11*, 859-869.

Kothare, M. V., Shinnar, R., Rinard, I., Morari, M. (2000). On defining the partial control problem: Concepts and examples. *AIChE Journal*, *46*, 2456-2474.

Larsson, T., Skogestad, S. (2000). Plant-wide control - A review and a new design procedure. *Modeling, Identification and Control, 21*, 209-240.

Reeves, D. E., Arkun, Y. (1989). Interaction measures for Nonsquare decentralized control structures. *AIChE Journal*, *35*, 603-613.

Shinnar, R. (1981). Chemical Reactor Modeling for Purposes of Controller Design. *Chemical Engineering Communications*, *9*, 73-99.

Skogestad, S. (2004). Control structure design for complete chemical plants. *Computers & Chemical Engineering, 28*, 219-234.

Veres, S. M., Kuntsevich, A. V., Valyi, I., Wall, D. S., Hermsmeyer, S. (1996). Geometric bounding toolbox for MATLAB^{TM}. Edgbaston, UK: The University of Birmingham.

Vinson, D. R., Georgakis, C. (2000). A new measure of process output controllability. *Journal of Process Control, 10*, 185–194.

Vinson, D. R. (2000). *A new measure of process operability for the improved steady-state design of chemical processes*. Ph.D. thesis, Lehigh University, USA.

Vinson, D. R., Georgakis, C. (2002). Inventory Control Structure Independence of the Process Operability Index. *Industrial and Engineering Chemistry Research, 41*, 3970-3983.

Session 7.3

Process Control

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

EXPERIMENTAL VALIDATION OF MODEL-BASED CONTROL STRATEGIES FOR MULTICOMPONENT AZEOTROPIC DISTILLATION

Lina Rueda, Thomas F. Edgar, R. Bruce Eldridge

Department of Chemical Engineering, University of Texas at Austin

Abstract: This work presents the results from dynamic modeling and control of an azeotropic distillation system. The model was validated with experimental data from a packed distillation unit. The physically-based process dynamic model, developed in HYSYS, was linked online with the control software used in the process. Model parameters were modified online using a feedback configuration to eliminate the difference between the process and model outputs. The fundamental model was used in the implementation of different control strategies, including a multivariable control strategy using model predictive control (MPC) software Predict Pro, via an inferential control strategy to treat missing process measurements. *Copyright © 2006 IFAC*

Keywords: Dynamic modeling, distillation, model-based control, multivariable control.

1. INTRODUCTION

Distillation is clearly the largest energy-consuming separation process used in chemical industries to recover products, by-products and unreacted raw materials. Improving its process efficiency is an on-going goal of the chemical processing and refining industries, given recent increase in energy prices. The use of dynamic modeling software in chemical and refining applications has been intensified with the adoption of commercial process modeling software and increased computer processing capabilities. The modeling software is used in a broad range of applications like parameter estimation, process optimization, and control. Most modern control methods require some kind of process model to predict future process outputs but industrial applications typically do not link fundamental dynamic models in commercial software with the control software. Some model-based control and optimization techniques are based on steady state physical models that account for the physical drifting of the process itself (such as fouling of a heat exchanger, temperature fluctuation of the feed, etc.) or changes in market demands and economic conditions, information that can be used to modify product specifications and plant schedules.

Azeotropic distillation is a process widely used to separate non-ideal binary mixtures. This separation technique uses another component, known as an entrainer. Depending on the mixture, the entrainer forms an azeotrope with one of the components in the binary mixture or breaks an existing azeotrope in the binary mixture. There are three azeotropic distillation configurations: homogeneous azeotropic distillation, heterogeneous azeotropic distillation and extractive distillation.

Azeotropic distillation presents multiple challenges in design and operation due to the presence of non-idealities, phase splitting, possible multiple steady states, and distillation boundaries. When designing these systems, it is important to keep in mind that distillation boundaries cannot be crossed. For these reason, in order to isolate two pure components which lie in two different distillation regions, it is necessary to have two different feed compositions (one from each of the two regions) and two distillation columns (Doherty and Caldarola 1985). Published experimental work on azeotropic distillation has often analyzed the behavior of the azeotropic distillation systems in laboratory scale sieve columns (Baur, *et al*., 1999; Müller and Marquardt, 1997; Springer and Krishna, 2001; Wang, *et al*., 1998); in addition, some experimental studies (Chien, *et al*., 2000) have also implemented

different control strategies using temperature as the controlled variable.

This work presents the results from experimental validation of dynamic models of an azeotropic distillation system of methanol, normal pentane and cyclohexane. All the experiments were performed in the homogeneous region without liquid phase splitting. The model was validated with experimental data from a pilot-scale size packed distillation unit configured at finite reflux. The approach presented in this work links the process fundamental dynamic model (HYSYS) with the control software used in the process. The model is modified online using a feedback configuration to eliminate the difference between the process and model outputs. The model is used in the implementation of different control strategies to infer process variables that cannot be determined with field instrumentation. Two different variable pairings are studied and the results from individual control loop configurations are compared with a multivariable control strategy using model predictive control (MPC).

The dynamic model was developed using HYSYS from Aspen Technologies. The model was linked to Emerson Process Management's DeltaV digital automation system. The experiments were developed in the pilot plant of The Separation Research Program (SRP) at The University of Texas at Austin.

2. EXPERIMENTAL SYSTEM

The chemical system selected for the experiments performed in this research was a ternary mixture of methanol, pentane and cyclohexane. The ternary mixture diagram is presented in Figure 1 where the two azeotropes in the system are apparent. The two azeotropes divide the diagram into two distillation regions. Figure 1 also identifies the feasible product region for a particular feed point using the intersections between the distillation and material balances lines.

The highest pentane purity achievable in the distillate product was the azeotropic composition, which was a viable objective in both regions; however, the bottom composition objective changed from pure cyclohexane in the first region to pure methanol in the second.

The plant where the experiments were carried out is located at the Separation Research Program, at the University of Texas at Austin. The column used in the experiments is a 6 in stainless steel column with 30 ft of contacting height packed with 0.7 Nutter rings metal random packing.

The system is well-instrumented with state of the art sensors and actuators from Fisher-Rosemount. The experimental plant is operated with Emerson Process Management's DeltaV digital automation system. The simulation was implemented in HYSYS from Aspen Technologies at the application station in the DeltaV system and connected to the controllers through an interface with the digital control system. Additional details on the equipment configuration can be found at the SRP website http://uts.cc.utexas.edu/~utsrp/.

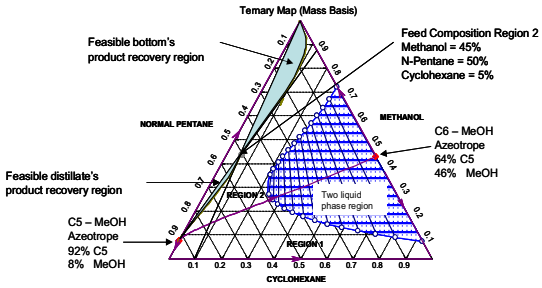


Fig. 1. Ternary map (mass basis) for cyclohexane, normal pentane and methanol. P = 6 psig. Property Package: Split from Aspen Tech.

An analytical procedure for the analysis of the samples collected from the system was developed and implemented in two HP 5890 gas chromatographs. The error in the measurement was calculated to be less than 3%.

3. AZEOTROPIC DISTILLATION DYNAMIC MODEL

The fundamental model was developed following a methodology that included five steps: (1) the system physical and thermodynamic behavior was identified and appropriate physical property relationships were determined; (2) different modeling approaches were studied and compared with process data to determine the most suitable method to model the system; (3) the model was developed and validated with process data; (4) the model parameters to be updated on-line were selected; and (5) the model updating method was implemented.

Steady state simulation and ternary and binary diagrams were used to study the system's thermodynamic behavior. After analyzing the ternary diagram and validating the simulated data with experimental data, it was concluded that the process did not display multiple steady states. The existence of multiple steady states could be determined based on the geometry of the distillation region boundaries and product paths in the ternary diagrams (Bekiaris *et al*., 1993; Bekiaris and Morari, 1996).

Some studies have concluded that rate-based or non-equilibrium models are necessary to obtain a good description of the azeotropic system (Repke *et al*., 2004; Springer and Krishna, 2001), while others have validated azeotropic distillation equilibrium models experimentally (Müller and Marquardt, 1997; Kumar *et al*., 1984). These studies suggest that the equilibrium approach can perform very well in modeling of azeotropic distillation systems. In order to determine whether or not equilibrium models could be used to accurately predict the azeotropic system behavior, non-equilibrium and equilibrium steady state models were developed, and their model predictions were compared with a wide range of experimental data. Both models used the same

equipment configuration, operating conditions, and thermodynamic properties. Conditions from the two distillation regions were simulated, and their results were validated experimentally. It was concluded that the equilibrium model did accurately predict the azeotropic behavior and therefore a dynamic equilibrium model was developed.

Table 1. Column Configuration for Steady State Simulation.

Number of Theoretical Stages	24 (Without condenser and reboiler)
Feed Stage	18
Condenser Type	Total (Stage 1)
Reboiler Type	KETTLE (Stage 26)
Valid Phases	Vapor-Liquid-Liquid
Internal Type	Packed (Nutter Ring Metal Random No. 0.7)
Stage Packing Height [in]	13.84615
Stage Vol [ft3]	0.226557121
Diameter [in]	6
Void Fraction	0.977
Specific Surface Area [sqft/cuft]	68.8848
Robbins Factor	11.8872

The column configuration is summarized in Table 1. The activity coefficient model NRTL was used as the main property method for the liquid phase while the Redlich-Kwong equation-of-state was used for the calculations in the gas phase.

Analysis of the process dynamic responses indicated that in this particular system temperature was not a good choice as a controlled variable. The temperature response to feed disturbances and changes in the steam and reflux flow rates displayed a highly nonlinear behavior. In addition, in some regions the gain was very small and the changes in temperature due to changes in the manipulated variables were within the noise level.

4. ONLINE MODEL RECONCILIATION

The dynamic model was intended to be used as a tool to determine process parameters that could not be measured directly in the field, such as composition, a variable needed to implement the control strategies. For this reason, the model predictions had to be accurate and track the process behavior throughout the entire operating region.

Traditionally, model reconciliation is performed using a steady-state model and the parameter estimates are obtained off-line using an optimization algorithm, such as the weighted least square (WLS) formulation, where the objective is to find estimates that minimize the squared error between the model predictions and the measurements, normalized by the measurement covariance. Usually before the parameters are estimated, the measurement data are first validated with some conservation equations, and then reconciled such that the model parameters and the adjusted data satisfy the process model equations (Seborg, *et al*., 2004).

In this work we used a reconciliation module to calculate the model parameters that minimize the error between plant measurement and model variables. The algorithm used in the reconciliation module is based on the gradient approach for model-reference adaptive control (Åström and Wittenmark, 1995). The objective is to modify the parameters in the model so that the error between the outputs of process and reference model is driven to zero. In the gradient approach the parameter is obtained as the output of an integrator. A quicker adaptation could also be achieved by adding a proportional adjustment to the integral action. The control law then takes the form of (1), which can be implemented in the plant using PI controller software where the constants γ_1 and γ_2 represent the proportional and integral gains respectively (Åström and Wittenmark, 1995).

$$u(t) = \gamma_1 e(t) + \gamma_2 \int_0^t e(\tau) d\tau \qquad (1)$$

Initially, the model parameters selected to be updated online were overall column heat transfer coefficient and HYSYS dynamic efficiency (Abouelhassan and Simard, 2003). Given that the packing HETP value was obtained from experimental data, the efficiency value should not change considerably from the value of one. However, it was expected to have some variations given different flooding conditions mainly due to the system's non-ideal behavior. The model efficiency value was modified to match the process distillate C5 composition. Because the error in the measurement was about 3%, the parameter was modified if the model output was off by more than 3% from the process output. After data from the experiments was analyzed, it was concluded that the efficiency value was fairly constant at a value of 0.7 in the distillation region rich in cyclohexane and pentane and 0.5 in the distillation region rich in methanol and normal pentane. Although the reconciliation module was used to determine these values, the efficiency parameter was not longer modified on-line to reconcile the model on-line.

The column's external surface heat transfer coefficient directly influences the heat loss experienced by the column. The model developed in this work used a simple heat loss equation, where the heat loss is calculated from the parameters specified by the user: overall heat transfer coefficient U and ambient temperature T_{amb}. The heat transfer area A and the fluid temperature T_f are calculated for each stage by the model. The heat loss is calculated using (2).

$$Q = UA\left(T_f - T_{amb}\right) \qquad (2)$$

During the model validation phase, U was modified until the mass balance in the model matched the experimental mass balance, that is distillate and bottoms flow rates were the same in the model and the experiment when all the other conditions in the model where set to match the conditions in the experiment. T_{amb} was introduced given the conditions of the experiment, but was not continuously upgraded. The heat transfer coefficient was updated online using the reconciliation module during the control experiments described in the next section. Its value

increased up to 5% as the liquid flow in column decreased and vice versa. The heat transfer coefficient is dependent upon the physical properties of the fluid and the physical conditions of the experiment, since both fluid composition and process conditions changed with the operation region the heat transfer coefficient also changed. In addition, the heat transfer coefficient was reflecting the variations in the ambient temperature given that this value was not measured continuously nor automatically upgraded during the experiments. Although this variation was found to be small, it shifted the model from the process outputs.

5. CONTROL STRATEGIES

Stabilizing the basic operation of the column was achieved by inventory (level), flow and pressure controls. The control loops in this level were configured with independent PID controllers (See Table 2).

Table 2. Basic Configuration

Manipulated Variable	Controlled Variable
Feed Flow Valve Position	Feed Flow Rate
Preheater Steam Flow Valve Position	Feed Temperature
Reflux Flow Valve Position	Reflux Flow Rate
Distillate Flow Valve Position	Distillate Flow Rate
Bottom Flow Valve Position	Bottom Flow Rate
Reboiler Steam Flow Valve Position	Steam /Duty Flow Rate
Nitrogen Flow Splitter Valve Position	Column Pressure

To control the product composition in the column, two different configurations were considered based on the relative gain array analysis (RGA). RGA was used to get an initial understanding on how to pair variables in the inventory and separation control. The gain matrix was calculated using the step responses from the dynamic model developed in HYSYS. Different step changes were performed in the manipulated variables, using different magnitudes and directions, and then the results were averaged. The analysis indicated that two configurations were viable (see Table 3).

Table 3. Composition Manipulated and Controlled Variable Configurations.

	Manipulated Variable	Controlled Variables	Λ (RGA)	
1	Reflux Flow Rate (R)	DC – R	0.972	0.028
	Steam Flow Rate (Q)	BC – Q	0.028	0.972
2	Distillate Flow Rate (D)	DC – Q	-0.004	1.004
	Steam Flow Rate (Q)	BC – D	1.004	-0.004

DC = Distillate Composition; BC = Bottom Composition

The results from the RGA analysis were consistent with the traditional control configuration used in ordinary distillation (pairing 1) and the results from studies in azeotropic distillation where the opposite pairing (pairing 2) gave less loop interaction than the traditional variable pairing used in distillation (Chien, *et al*., 2000; Tonelli, *et al*., 1997).

As mentioned previously, the process had two feasible distillation regions. The data presented in this paper includes experimental data only from region one (feed composition with high concentration of cyclohexane and normal pentane). The control objective was to maintain the pentane/methanol azeotrope in the distillate and maximum recovery of cyclohexane in the bottom stream. For this reason the key components selected for control were normal pentane for the distillate stream and cyclohexane for the bottom stream. The manipulated variables were selected between the same options as for inventory control: distillate, reflux, steam, and bottom flow rate.

The level in the reflux drum was paired with the distillate flow rate in the first configuration (pairing 1) and with the reflux flow rate in the second configuration (pairing 2). The column level was paired with the bottom flow in the two control configurations.

The dynamic model was connected online to the DCS and provided estimates for variables where instrumentation was not available. Since the plant did not have an online measurement of composition, this configuration provided the controlled variable estimates. During experimentation, samples of distillate and bottom products were collected after mass balance was achieved in the process and compared with the values provided by the simulation. The difference between measured and estimated values was within ± 3%. Samples of the feed were collected every half hour and the values introduced in the model.

PID controllers were configured in the experimental plant to control the composition in the distillate and bottom streams using the pairings described in Table 3. The tuning of the PID controller was performed using the advanced control module DeltaV Tune, which implements a relay oscillation test based on the Aström-Hägglund algorithm for calculating the tuning parameters of a process control loop (Seborg, *et al*., 2004). The results are given in Table 4.

Table 4. Composition Controller Tuning.

	Pairing 1		Pairing 2	
	DC – R	BC – Q	DC – Q	BC – D
Ultimate κ	10.90	10.55	6.42	4.98
Ultimate T	207.00	699.50	663.50	277.50
Process θ	28.45	85.91	99.46	42.98
Process κ	0.72	0.82	1.14	1.43
Process τ	257.42	957.85	766.61	311.92
Suggested Tuning Parameters:				
PID	P: 2.31 I: 191.1 D: 30.58	P: 1.66 I: 654.16 D: 104.67	P: 0.86 I: 369.39 D: 59.1	P: 1.15 I: 227.13 D: 36.34
θ Dominant	P: 2.72 I: 52.78	P: 2.64 I: 178.37	P: 0.79 I: 165.24	P: 1.25 I: 70.76
Implemented Tuning Parameters:				
	P: 2 I: 191 D: 30	P: 2 I: 654 D: 104	P: 0.5 I: 369 D: 59	P: 1 I: 227 D: 36

κ = Gain. θ = Dead Time. τ = Time constant. T = Period.

Although the steam loop exhibited a considerable dead time that could limit the effectiveness of the controllers, based on experimentation it was

determined that the best PID tuning parameters were close to values suggested in the literature. The response with the dead time dominant configuration was more aggressive and exhibited oscillatory behavior.

5.1 PID Controller Performance

Figure 2 illustrates the PID controller performance after a series of step changes in the distillate and bottoms composition set points.
Both controllers drove the controlled variables to the desired set point. Pairing 2 gave fast responses but presented poor rejection to interaction. Figure 3 illustrates the closed-loop responses to disturbances in the feed temperature.

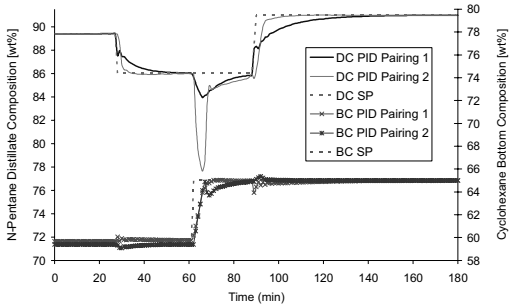


Fig. 2. Closed-loop composition control using PID controllers.

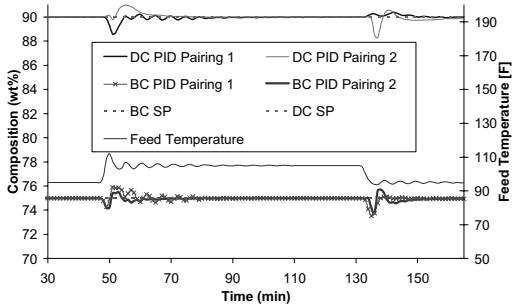


Fig. 3. PID controller closed-loop response to disturbances in the feed temperature.

5.2 Model-based control

Linear MPC was implemented using the commercial advanced control module Predict Pro from DeltaV. The process model used by the controller was identified online using the process model identification tool included in the module. Although with DeltaV PredictPro it is possible to run an automated test on the process, a manual test was performed for each input variable to generate the data for model identification. DeltaV PredictPro uses step response modeling for the generation of the MPCPro controller.
The step responses are generated using two types of models: Finite Impulse Response (FIR) and Auto-Regressive (ARX). The FIR model is used to identify the process delay used in the ARX model. The identified step responses are presented in Table 5. The MPC variables were selected based on best result from the PID study.

The gain (κ) is dimensionless because it is normalized by the transmitter range. The controller in the MPC algorithm is designed as an online-horizon optimization problem that is solved subject to the given constraints.

Table 5. MPC Step response models.

	Distillate C5 Composition	Bottom C6 Composition
	$\kappa = 3.8$	$\kappa = -1.4$
Reflux Flow rate	$\theta = 16$ s	$\theta = 8$ s
	$\tau = 689.23$ s	$\tau = 172.31$ s
	$\kappa = -3.2$	$\kappa = 3.5$
Steam Flow rate	$\theta = 48$ s	$\theta = 40$ s
	$\tau = 1828.95$ s	$\tau = 1899.86$ s
	$\kappa = -0.2$	$\kappa = 0.2$
Feed Temperature	$\theta = 16$ s	$\theta = 88$ s
	$\tau = 344.62$ s	$\tau = 190.67$ s
	$\kappa = 0.4$	$\kappa = -0.2$
Feed Flow rate	$\theta = 24$ s	$\theta = 16$ s
	$\tau = 689.23$ s	$\tau = 221.54$ s

κ = Gain. θ = Dead Time. τ = First order time constant. Time to reach steady state= 960 s.

For MPC based on linear process models, both linear and quadratic objective functions can be used (Qin and Badgwell 2003). Equation (3) represents the control law that minimizes a quadratic objective function.

$$\Delta U(k) = \left(S^T Q S + R\right)^{-1} S^T Q \hat{E}^0(k+1) \quad (3)$$

The vector $\hat{E}^0(k+1)$ corresponds to the predicted deviations from the reference trajectory when no further control action is taken; this vector is known as the predicted unforced error vector. The matrices Q and R are weighting matrices used to weight the most important components of the predicted error and control move, vectors respectively (Seborg, *et al.,* 2004). In DeltaV Predict Pro the elements of Q are known as penalty on error while the entries of R are the "penalty on move". The MPC controller is tuned by modifying the values of the matrices Q and R. R offers convenient tuning parameters because increasing the values of its elements reduces the magnitude of the input moves, providing a more conservative controller.
Figure 4 illustrates the linear MPC performance in the experiment after a series of step changes in the distillate and bottoms composition set points. Both output errors were assigned a penalty of one. The penalty on move was set to 25 for the steam flow rate and 20 for the reflux flow rate. In the experiment the optimizer was also configured to maximize the concentration of C5 in the distillate. The SP was allowed to change 0.5% for both controlled variables.
Given that the system is nonlinear, the tuning parameters in the multivariable controller were set up to provide robustness and eliminate oscillation in the response. The main difficulty occurred due to changing gains in the process. The gains related to the distillate composition were smaller when the azeotropic composition was reached in the distillate composition than in other regions with lower pentane recovery in the overhead product.

Figure 6 illustrates MPC responses inside and outside the azeotropic region with different tuning parameters.

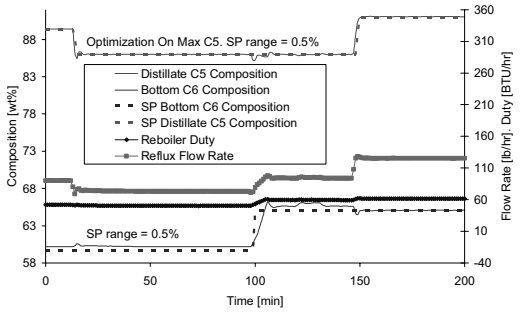


Fig. 4. Experimental composition control using linear MPC. Feed Flow Rate configured as manipulated variable.

Controller tuning 1 has a higher penalty on move (PM) for both manipulated variables than controller tuning 2. The parameters used in controller tuning 2 were the values suggested by DeltaV Predict Pro. These values are calculated based on the assumption that the system is linear. A higher penalty on move improved system stability in the region with higher gains. The penalty on error was set to 1 for both controlled variables. From Figure 5 it is observed that controller tuning 2 produces an unstable closed loop response.

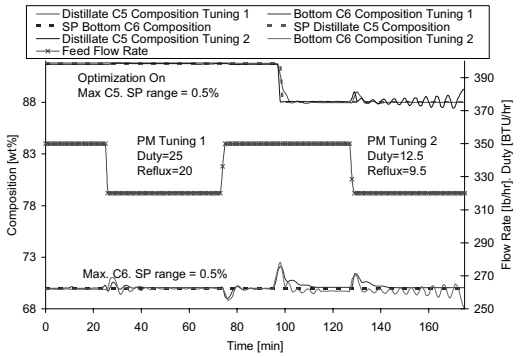


Fig. 5. Experimental MPC behavior using different tuning parameters.

6. CONCLUSIONS

Analysis of the process steady and dynamic models indicated that equilibrium models accurately predict the distillation column behavior. In multicomponent azeotropic distillation, temperature measurements do not offer accurate indications of composition, hence commercial dynamic simulation software were used to obtain an inferential control solution. Linear MPC gives excellent performance when the composition is used directly as a controlled variable and the appropriate tuning is used.

REFERENCES

Abouelhassan M., Simard A. (2003). The Scoop on Tray Efficiency. *Distillation Series, Aspentech Documentation.*

Åström K.J and Wittenmark B. *Adaptive Control* (1995). Addison-Wesley Publishing Co. Boston, MA.

Baur, R., Taylor, R., Krishna, R. and Copati, J. A. (1999). Influence of mass transfer in distillation of mixtures with a distillation boundary, Trans IchemE, Part A, *Chem Eng Res Des.* 77, 561–565.

Bekiaris, N., G.A. Meski, G.A. Raudu and M.Morari (1993). Multiple Steady States in Homogeneous Azeotropic Distillation. *Ind. Eng. Chem. Res.* 32(9), 2023–2038.

Bekiaris, N. and M. Morari (1996). Multiple Steady States in Distillation: 1=1 Predictions, Extensions, and Implications for Design, Synthesis, and Simulation. *Ind. End. Chem. Res*. 35, 4264–4280.

Chien, I-L., Wang, C. J., Wong, D. S. H., Lee, C.-H., Cheng, S.-H., Shih, R. F., Liu, W. T. and Tsai, C. S. (2000). Experimental investigation of conventional control strategies for a heterogeneous azeotropic distillation column. *Journal of Process Control*, 10(4), 333-340.

Doherty, M. F. and Caldarola G. A. (1985). Sequencing of Columns for Azeotropic and Extractive Distillation. *Ind. Eng. Chem. Fundam.* 24, 474-485.

Kumar, S., Wright, J. D. and Taylor, P. A. (1984). Modeling and dynamics of an extractive distillation column. *Canadian Journal of Chemical Engineering.* 62(6), 780-9.

Müller, D. and Marquardt, W. (1997). Experimental Verification of Multiple Steady States in Heterogeneous Azeotropic Distillation. *Ind. Eng. Chem. Res.* 36, 5410-5418

Qin, S.J. and Badgwell, T.A. (2003) A Survey of Industrial Model Predictive Control Technology. *Control Eng. Practice*. 11, 733.

Repke, J-U, Villain, O. and Wozny, G. (2004). A nonequilibrium model for three-phase distillation in a packed column: modelling and experiments. *Computers & Chemical Engineering.* 28(5), 775-780.

Seborg, D. E., T. F. Edgar, and D. A. Mellichamp (2004). *Process Dynamics and Control,* Wiley, New York.

Springer, P. A. M. and Krishna, R. (2001). Crossing of boundaries in ternary azeotropic distillation: Influence of interphase mass transfer, *Int Commun Heat Mass*. 28, 347–356.

Tonelli, S. M., Brignole, N. B. and Brignole, E. A. (1997). Modeling and control of an industrial azeotropic train. *Computers & Chemical Engineering* 21(Suppl., Joint 6th International Symposium on Process Systems Engineering and 30th European Symposium on Computer Aided Process Engineering, 1997), S559-S564.

Wang, C. J., Wong, D. S. H., Chien, I-L., Shih, R. F., Liu, W. T. and Tsai, C. S. (1998) Critical Reflux, Parametric Sensitivity, and Hysteresis in Azeotropic Distillation of Isopropyl Alcohol + Water + Cyclohexane. *Ind. Eng. Chem. Res.* 37, 2835-2843.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

RUN-TO-RUN CONTROL OF MEMBRANE FILTRATION PROCESSES [1]

Jan Busch, **Wolfgang Marquardt** [2]

Lehrstuhl für Prozesstechnik, RWTH Aachen University,
Turmstr. 46, 52064 Aachen, Germany

Abstract: Membrane ltration processes are often operated cyclically, where one cycle comprises a ltration and a backwashing phase. Due to the complex mechanisms involved, these ltration processes are mostly operated with xed values of the manipulated variables. In this paper, a model-based process control approach is introduced, which is based upon run-to-run control theory. To evaluate the controller, a suitable model of submerged membrane ltration in wastewater applications is developed, which describes the main process phenomena while being computationally inexpensive. The model-based controller is then tested in a simulation environment employing a validated reference model. Excellent results with respect to prediction quality and optimality are obtained.

Keywords: run-to-run control, self-adaptive control, online control, online optimization, ltration, membrane ltration

1. INTRODUCTION

Filtration, and more recently, membrane ltration, are well-known and established technologies for the separation of particles, macromolecules or even dissolved molecules from uids. Depending on the size of the separable substances, di erent technologies are known as ltration, micro ltration (MF), ultra ltration (UF), nano ltration (NF), and reverse osmosis (RO). This paper addresses ltration technologies where the separation principle is based on the di erence in size of macromolecules/particles and of the diameter of the pores of the ltration medium. This includes regular ltration applications as well as MF and UF. For simplicity, all lters belonging to this broad class will subsequently be termed *membranes*. For these applications, the driving force facilitating ltration is usually a pressure di erence across the membrane, that drives those particles through the membrane pores which are small enough to pass. Together with the solvent uid, they leave the system as *permeate*, while particles larger than the pores are held back as *retentate*.

In most applications, the repelled particles concentrate on the feed side of the membrane and build a lter cake, which increases the ltration resistance (organic fouling). Furthermore, pores can be blocked by intruding particles (pore blocking). Finally, microorganisms can grow on the membrane and pore surfaces, leading to bio lms, which decrease the performance and can also damage the membrane (biofouling). When repelled by the membrane, soluble substances concentrate on the feed side of the membrane, and after reaching maximum solubility, they crystallize and add to cake layer formation (scaling, anorganic fouling).

[1] The financial support by the DFG (German Research Foundation) in the project "Optimization-based process control of chemical processes" is gratefully acknowledged.
[2] Corresponding author: marquardt@lpt.rwth-aachen.de

All of these phenomena, which are known to contribute to *membrane fouling*, can be counteracted by membrane and module design as well as by appropriate process control strategies. In this paper, the focus will be on the process control aspect.

There are three main concepts for limiting membrane fouling. Firstly, high cross- ow velocities along the membrane surface (perpendicular to the pores) decrease the deposition of substances. Secondly, the ow direction through the membrane is periodically reversed, such that the membrane pores are ushed with uid (usually permeate). The third measure is to chemically or mechanically clean the membranes, which is usually performed at a much lower frequency.

1.1 Filtration process control: State-of-the-art

State-of-the-art process control for ltration processes usually employs xed values for the manipulated variables, which are only adjusted to meet the required net ux

$$J_{\text{net}} = \frac{J_f \quad t_f \quad J_b \quad t_b}{t_f + \quad t_b}. \tag{1}$$

The manipulated variables are the permeate and the backwashing uxes J_f and J_b and the ltration and backwashing durations t_f and t_b, respectively. A further manipulated variable is the cross- ow velocity u_c.

The reason for the rather simple control strategies lies in the high complexity of the ltration process. It is characterized by the periodic change between ltration and backwashing, the drift of the membrane permeability due to irreversible membrane fouling, and the typically non-steady-state operation. Furthermore, only very limited measurement information is available in industrial installations.

1.2 Membrane ltration modeling

The rigorous modeling of ltration processes is a highly complex task due to the many physical and possibly chemical and biological phenomena, which take place on very di erent time scales. There are numerous works in literature, which deal with the detailed modeling of various aspects of ltration processes. At the same time, there are several approaches to describe ltration processes only from a phenomenological point of view using simple, empirical correlations.

From a model-based process control point of view, both the mechanistic and the empirical models have advantages and drawbacks. If the uncertainty can be su ciently reduced by measurements, mechanistic models can yield a much higher prediction quality. Simpler, possibly empirical models have a lower computational demand and can often be identi ed from less data.

1.3 Outline of the paper

Our aim is to operate the ltration process at its economical optimum at every point in time while regarding safety constraints. In the framework of nonlinear model predictive control (NMPC), this objective is achieved by repeatedly solving a nonlinear, dynamic, and constrained optimization problem on a moving horizon. Its objective function resembles the operational cost, and its constraints re ect operational limitations. In the general case of measurement and process uncertainty, the optimization model needs to be regularly updated with current state information. Furthermore, the model has to be adapted to current process behavior, which is usually achieved by tting the parameters to past measurements on a suitably chosen estimation horizon. The success of the approach strongly depends on the ful llment of the following objectives:

- Satisfactory prediction and optimization,
- online applicability,
- robustness against disturbances, and
- adaptation to process drift and changes.

The key idea pursued throughout this paper is the following: A simple model is required to ful ll the online requirements of low computational cost and su cient identi ability. The lack of prediction precision is overcome by a frequent adaptation to plant measurements. This allows decent predictions at least in the vicinity of the current operating point. The ltration process is divided into ltration and backwashing phases. One ltration phase followed by one backwashing phase makes up one ltration cycle. The sequence of cycles can be exploited to update the process model after each cycle based on the available measurement data from the last cycle. In order to make the approach widely applicable in the process industry, only the transmembrane pressure (TMP) across the entire membrane module is assumed to be measured. The manipulated variables are then optimized for each upcoming cycle based on a model identi ed on the previous cycle. This concept is known as *run-to-run control*. It is introduced and adapted to ltration processes in Section 2. The resulting controller is evaluated in Section 3.

2. RUN-TO-RUN CONTROL FOR FILTRATION PROCESSES

Run-to-run process control is the strategy of applying one control action between two batches (cycles) in a process, while continuous control actions

during the cycle are taken by base controllers. The task of the run-to-run controller is to issue set-points for the base controllers. Fig. 1 illustrates the embedding of the run-to-run controller into the control system. The run-to-run controller is activated only once between two cycles. The parameter update of the model is performed employing the measurement information of the previous cycle. The updated model is used to determine optimal set-points for the next cycle.

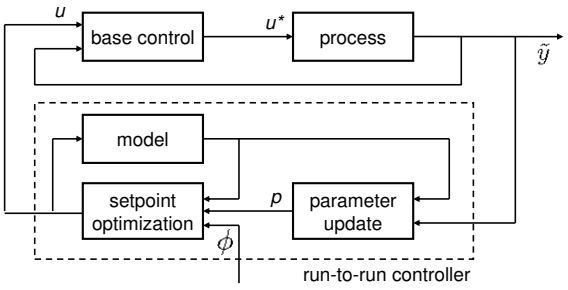


Fig. 1. Run-to-run control

An extensive review of theory and applications of run-to-run theory is provided by del Castillo and Hurwitz (1997). For the ltration systems treated in this paper, a very general problem formulation is required, which also has to account for the fact that each cycle is divided into a ltration (index f) and a backwashing phase (index b). First, the optimal control problem is formulated. To correctly represent the repeated solution of the problem on a moving horizon, the cycle index k should be introduced for every variable. The parameters $\mathbf{p}$ should be stated as $\mathbf{p}_{k|k\ 1}$, indicating that the parameters used in cycle k were estimated on the measurements of cycle $k\ \ 1$. However, to simplify the notation, this correct indexing is omitted:

$$\min_{\mathbf{u}_j, t_{j,e}} \ \phi \tag{P1}$$

$$\text{s.t. } \mathbf{f}_j\left(\dot{\mathbf{x}}_j, \mathbf{x}_j, \mathbf{y}_j, \mathbf{u}_j, \mathbf{p}_j, \mathbf{d}_j, t\right) = \mathbf{0}, \tag{2}$$

$$\mathbf{g}_j\left(\mathbf{x}_j, \mathbf{y}_j, \mathbf{u}_j, \mathbf{p}_j, \mathbf{d}_j, t\right) \leq \mathbf{0}, \tag{3}$$

$$\mathbf{h}_{j,\mathrm{eq}}\left(\mathbf{x}_j, \mathbf{y}_j, \mathbf{u}_j, \mathbf{p}_j, \mathbf{d}_j, t_{j,e}\right) = \mathbf{0}, \tag{4}$$

$$\mathbf{h}_j\left(\mathbf{x}_j, \mathbf{y}_j, \mathbf{u}_j, \mathbf{p}_j, \mathbf{d}_j, t_{j,e}\right) \leq \mathbf{0}, \tag{5}$$

$$\Gamma\left(\mathbf{x}_f\left(t_{f,e}\right), \mathbf{x}_b\left(t_{b,0}\right)\right) = \mathbf{0}, \ \mathbf{x}\left(t_0\right) = \mathbf{x}_0, \tag{6}$$

$$t_0 = t_{f,0} \leq t_{f,e} = t_{b,0} \leq t_{b,e} = t_e, \tag{7}$$

$$t \in [t_0, t_e], \ j = \begin{cases} f & \text{for } t \in [t_{f,0}, t_{f,e}], \\ b & \text{for } t \in [t_{b,0}, t_{b,e}]. \end{cases} \tag{8}$$

$\mathbf{x}$ are di erential and $\mathbf{y}$ are algebraic variables, $\mathbf{d}$ are disturbances, and t is the time. ϕ is the objective function representing the operational cost. $\mathbf{f}_j$ is the set of di erential-algebraic equations of index 1 describing the respective process, and $\mathbf{g}_j$, $\mathbf{h}_j$, and $\mathbf{h}_{j,\mathrm{eq}}$ represent equality (endpoint) and inequality (path and endpoint) constraints. Eq. (6) states initial conditions and linking conditions between the di erential states at the end of the ltration phase and the beginning of the backwashing phase. Eqs. (7)-(8) de ne the optimization horizon and the phase durations.

In a similar fashion the parameter estimation problem is formulated. With the additional assumption of white process and measurement noise, the parameter estimation problem reduces to a least squares optimization problem, whose formulation is omitted here for brevity.

The run-to-run control framework up to this point is generic for a wide range of ltration applications. In the following, its application to submerged MF/UF membrane ltration in wastewater applications is demonstrated.

2.1 Process model

The model proposed in the following is based on simple descriptions of the main phenomena of MF/UF membrane ltration processes. The transmembrane pressure p is commonly described using Darcy's law,

$$p = J\eta R, \tag{9}$$

where J is the ux, η is the uid's viscosity, and R is the membrane resistance (e.g. Broeckmann *et al.*, 2005). While J is a manipulated variable, η depends on the feed suspension properties. As the TMP is assumed to be measurable, Eq. (9) represents the system's output equation. In the model proposed in the following, the resistance is described by di erent state equations for the ltration and the backwashing phase.

Filtration phase During ltration, the membrane resistance R_f can be described by

$$\frac{dR_f}{dt} = m{J_f}^{\alpha}{u_c}^{\beta}, \ R_f\left(t_{f,0}\right) = R_f^0. \tag{10}$$

R_f^0 is the initial membrane resistance. Assuming that the ltration ux J_f and the cross- ow velocity u_c are constant, a linear increase of membrane resistance results. It describes the cake layer formation, which is the dominating e ect on this timescale and which strongly depends on the ux and on the cross- ow. m, α, and are parameters to adapt the model to a particular process.

Backwashing phase While often a linear increase of membrane resistance can be observed during ltration, its decrease during backwashing takes rather an exponential form, which converges to an irreversible resistance R_b^{∞}:

$$\frac{dR_b}{dt} = \frac{n{J_b}^{\gamma}}{\tau_b {J_b}^{\delta}} \ Re^{\frac{t \ \ t_{f,e}}{\tau_b {J_b}^{\delta}}} \tag{11}$$

$$R_b\left(t_{b,0}\right) = n{J_b}^{\gamma} \ \ R + R_b^{\infty}, \quad R = R_f\left(t_{f,e}\right) \ \ R_b^{\infty}. \tag{12}$$

Eqs. (11) and (12) are formulated such that a simple analytical expression for R_b can be obtained (Section 3). R describes the reversible resistance. The initial resistance $R_b(t_{b,0})$ is the sum of the irreversible and the reversible resistance, but just like the resistance R_b it depends on the ux J_b due to unmodeled e ects. n, τ_b, γ, and are parameters.

Cost function Finally, those operating cost are described that can be in uenced by the process control system. They consist of the cost for electrical energy to provide the TMP and the cross- ow and the cost for membrane replacement. The rst two are given by

$$\frac{dE_E}{dt} = \frac{|\ \ p_j J_j A|}{\eta_P} + e_c,\ E_E(t_0) = 0, \qquad (13)$$

where A is the membrane area, η_P is an e ciency factor of the permeate pump, and e_c is the necessary power to provide the cross- ow. The cost for membrane replacement E_R cannot be described as straightforwardly as the energy cost. In fact, there is no quantitative insight to describe the in uence of the manipulated variables on the membrane lifetime. Depending on the ltration system under consideration, di erent models for E_R have to be developed. For MF/UF membranes in wastewater applications, it has been observed in practice that a strong increase of the resistance within a ltration cycle indicates an overstraining of the membrane. Therefore, its gradient is penalized:

$$E_R = \xi \frac{dR_f}{dt} = \xi m {J_f}^{\alpha} {u_{c,f}}^{\beta}, \qquad (14)$$

where ξ is a parameter that needs to be speci ed for each application based on process experience. The overall objective function ϕ comprising the power consumption and the penalty term E_R is

$$\phi(t_e) = \frac{E_E(t_e)}{t_e \quad t_0} + E_R. \qquad (15)$$

2.2 Run-to-run controller

In this section, the run-to-run controller is designed. First the estimation problem is considered, then the optimal control problem, and nally the control algorithm itself.

Estimation In industrial practice, only the TMP across the membrane is measured. In order to make the proposed approach widely applicable, it is therefore assumed that only this TMP is available as measurement. Since the uxes J_f and J_b and the cross- ow u_c are set to constant values for each phase, α, , γ, and cannot be estimated on a horizon of one cycle due to the missing excitation. This is referred to as the *dual control problem* (Wittenmark, 1995). The concerned parameters are estimated offline using historical data from several cycles and then set constant in the run-to-run control scheme.

The estimation problems for the ltration and the backwashing phase are coupled through Eq. (12). In order to simplify the problem and decrease the computational demand, they are, however, solved sequentially. Since the model structure is simple enough, the di erential equations are solved analytically. The discretized estimation problem for cycle k using the measurement data from cycle k 1 for the ltration phase is then

$$\min_{m,R_f^0} \sum_{l=1}^{n_{f,l}} \frac{1}{2} (\ \ \tilde{p}_{f,l} \qquad p_{f,l})^2 \qquad \text{(P2)}$$

$$\text{s.t.} \quad p_{f,l} = J_f \eta R_{f,l}, \qquad (16)$$

$$R_{f,l} = R_f^0 + m J_f^{\alpha} u_c^{\beta} t_l, \qquad (17)$$

where $\tilde{p}_{f,l}$ are discrete measurements at the sampling points t_l, $l \in \{1, ..., n_{f,l}\}$, in cycle k 1, and $p_{f,l}$ are the corresponding simulated TMP samples.

The parameters of the backwashing model are estimated from

$$\min_{n,\tau_b,R_b^{\infty}} \sum_{l=1}^{n_{b,l}} \frac{1}{2} (\ \ \tilde{p}_{b,l} \qquad p_{b,l})^2 \qquad \text{(P3)}$$

$$\text{s.t.} \quad p_{b,l} = J_b \eta R_{b,l}, \qquad (18)$$

$$R_{b,l} = R_b^{\infty} + \quad R n J_b^{\gamma} e^{\frac{t_l \quad t_{n_{f,l}}}{\tau_b J_b^{\delta}}}, \qquad (19)$$

$$R = R_f \quad t_{n_{f,l}}) \quad R_b^{\infty}. \qquad (20)$$

Optimal Control The control problem, which is solved based upon the updated parameters, is

$$\min_{J_f, J_b, u_c, t_{f,e}, t_{b,e}} \phi \qquad \text{(P4)}$$

$$\text{s.t.} \quad p_j = J_j \eta R_j, \qquad (21)$$

$$R_f = R_f^0 + m J_f^{\alpha} u_c^{\beta} t, \qquad (22)$$

$$R_b = R_b^{\infty} + \quad R n J_b^{\gamma} e^{\frac{t \quad t_{f,e}}{\tau_b J_b^{\delta}}}, \qquad (23)$$

$$R = R_f(t_{f,e}) \quad R_b^{\infty}, \qquad (24)$$

$$J_{\text{net}} = \frac{J_f(t_{f,e} \quad t_0) \quad J_b \ (t_e \quad t_{b,0})}{t_e \quad t_0}, \qquad (25)$$

$$R_b(t_{b,e}) \le \nu R_b^{\infty},\ \nu \ge 1, \qquad (26)$$

$$J_f \le J_b, \qquad (27)$$

$$p_{\min} \le \quad p \le \quad p_{\max}, \qquad (28)$$

$$\mathbf{u}_{\min} \le \mathbf{u} \le \mathbf{u}_{\max}, \qquad (29)$$

$$t_0 = t_{f,0} \le t_{f,e} = t_{b,0} \le t_{b,e} = t_e, \qquad (30)$$

$$t \in [t_0, t_e],\ j = \begin{cases} f & \text{for } t \in [t_{f,0}, t_{f,e}], \\ b & \text{for } t \in [t_{b,0}, t_{b,e}]. \end{cases} \qquad (31)$$

The net ux J_{net} is considered a set-point speci ed by the operator or an upper level controller. Eq. (26) forces the nal resistance $R_b(t_{b,e})$ to be close to the irreversible resistance R_b^{∞} at the end of

the cycle. The backwashing ux J_b is forced to be at least equal to the ltration ux J_f (Eq. (27)), which is a safety measure to limit pore blocking. Eqs. (28) and (29) give bounds on the TMP and on the manipulated variables J_f, J_b, $t_{f,e}$, $t_{b,e}$, and u_c. ϕ is de ned as in Section 2.1.

Algorithm Ideally, the model identi cation and optimization takes place between two cycles k 1 and k, and the optimized values for the manipulated variables are applied at the beginning of the new cycle k. This would require zero calculation time. Hence, a delay in the implementation of the new set-points is inevitable. The reader is referred to Findeisen and Allgöwer (2003) for a rigorous discussion of possible stability problems due to computational delay in NMPC applications.

3. CASE STUDY - SUBMERGED MF/UF IN WASTEWATER TREATMENT

In order to evaluate the proposed model and control algorithm, it is tested against simulated data from a rigorous membrane ltration model, which describes MF/UF with submerged membranes in a wastewater treatment plant. The uid feed consists of water, in which a variety of organic and inorganic particles and dissolved substances are present. Organic fouling, biofouling, and pore blocking are therefore the dominating fouling e ects. Usually hollow bre membranes or plate modules with nominal pore sizes around 1 m are employed. The cross- ow is realized with air bubbles that are periodically injected at the bottom of the modules.

In order to study the highly complex process, a rigorous model has been developed, which is discussed in detail by Broeckmann *et al.* (2005) and Cruse (2006). It has been shown to adequately represent real plant behavior, and is used as a reference model in the following.

The model proposed in Section 2.1 is adapted to re ect the speci c characteristics of the given process. The cross- ow velocity u_c is usually not explicitely available as manipulated variable, since air is injected with a constant, yet intermitted volume ow Q. The periodically changing intervals with and without aeration have the lengths t_{on} and t_{off}. u_c is then heuristically described as

$$u_c = Q\frac{t_{\text{on}}}{t_{\text{on}} + t_{\text{off}}}, \tag{32}$$

and t_{off} is chosen as manipulated variable. The power for the aeration e_c is expressed as

$$e_c = \frac{QTR_g\gamma_{\text{a}}\left[(1+p_{\text{a}})^{\frac{\gamma_{\text{a}}}{\gamma_{\text{a}}\ 1}} \quad 1\right]t_{\text{on}}}{v_{\text{a}}\left(\gamma_{\text{a}} \quad 1\right)\left(t_{\text{on}} + t_{\text{off}}\right)\eta_A}, \tag{33}$$

assuming that the compression is a polytropic process. T is the ambient air temperature, v_{a} is the molar volume of air, R_g is the gas constant, $\gamma_{\text{a}} = 1.4$ is the polytropic coe cient, p_{a} is the pressure di erence across the compressor (in bar), and η_A is an e ciency factor.

3.1 Results

Three aspects are analyzed in the following: the quality of the TMP prediction, the adaptation to process changes, and the predicted optimal solutions. The simulation results based on the reference model will be referred to as *measurements*.

TMP prediction Fig. 2 depicts a snap-shot of the simulated controlled process. It shows the measured and the predicted TMP for cycles k and $k+1$, between which the ux is increased. Each cycle comprises a ltration (positive TMP) and a backwashing phase (negative TMP). The parameters α, , γ, and , which are not estimated online, have been tted a priori. During ltration the predicted and the measured TMP are almost identical, and only small errors are observed during backwashing. The relative deviation is below 1%. The same results are achieved with respect to changing backwashing uxes, ltration and backwashing durations, and cross- ow intensities. This shows the excellent prediction capability of the model.

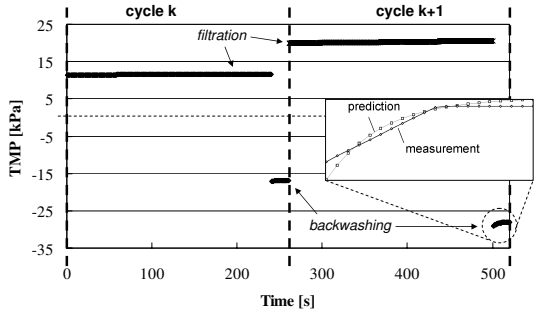


Fig. 2. TMP measurement and prediction

Model adaptation Next, the performance in the presence of unforeseen process changes is evaluated. In Fig. 3, the ltration ux after cycle k is reduced by 20%. This could be caused by an unexpected problem with a pump. The TMP prediction is false in cycle $k+1$, but the controller adapts by solving the estimation problems (P2) and (P3) with data from cycle $k+1$. Reliable predictions are provided from cycle $k+2$ on.

Control In the following an updated model is assumed to be available, and the optimization for the next cycle is carried out for di erent

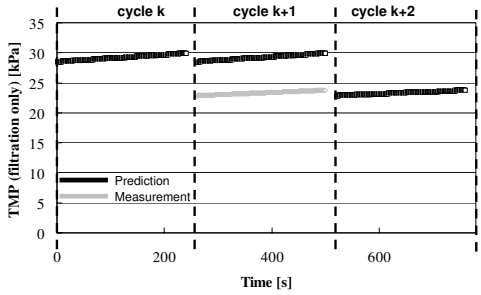


Fig. 3. Adaptation to a sudden process change

required net uxes. Fig. 4 presents the results for the ltration ux and the aeration pause. The ltration ux increases almost linearly with higher net uxes. The ltration time is at its upper bound of 600s. The backwashing ux always equals the ltration ux, and together with the minimum backwashing time of 15s, the constraint on the minimum resistance removal (Eq. (26)) is always met. The aeration pause becomes smaller with increasing ux.

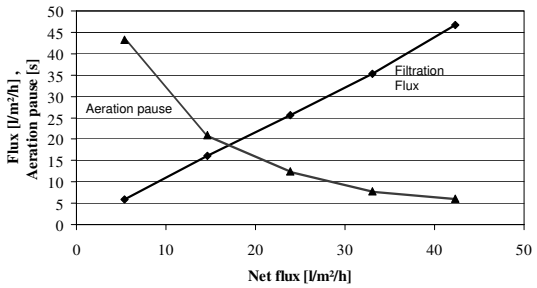


Fig. 4. Variation of the net ux

Finally, the performance of the proposed controller is compared against manual operation with xed set-points. A typical choice in an industrial installation is e.g. $J_f = 40\frac{\mathrm{l}}{\mathrm{m^2h}}$, $J_b = 50\frac{\mathrm{l}}{\mathrm{m^2h}}$, $t_f = 240$s, $t_b = 20$s, and $t_{\mathrm{off}} = 6$s, which gives a net ux of $J_{\mathrm{net}} = 33.1\frac{\mathrm{l}}{\mathrm{m^2h}}$. For the same net ux, the optimized solution depicted in Fig. 4 requires 20% less energy despite employing a 14% higher aeration.

3.2 Discussion

Assuming a decent choice and adaptation of the parameters α, , γ, and in the ltration models, an excellent prediction of the TMP is achieved. Furthermore, the controller quickly adapts to unexpected changes in the process.

The interpretation of the optimization results is straightforward. Low uxes with long ltration times are preferred over small ltration periods with high uxes. This is in line with current observations in MBR installations. Instead of placing an upper bound on the ltration time, a more sophisticated approach could be designed, which e.g. establishes a link between the upper bound and the ux, if according process knowledge is available. The shortening of the aeration pauses with increasing uxes is clearly due to the penalty term, which prevents long-term damage to the membranes. In the case study, the backwashing intensity is at its lower bounds, yet this depends on the cleaning e ciency of the process, which is detected in the parameter estimation step. The remaining tuning parameter ξ (Eq. (14)) re ects the balance between short-term (energy) and long-term (replacement) cost.

Finally, employment of the approach does not only promise substantial economical bene t, but also implies continuous adaptation of the membrane's operation to process drifts and changes. This enables not only optimal, but also safe process operation.

4. CONCLUSIONS

A methodology for the model-based control of membrane ltration processes is proposed. It is based on run-to-run control concepts and employs a newly developed process model. The proposed controller is tested in a simulation scenario describing submerged membrane ltration in wastewater applications. It is shown to achieve excellent results concerning the prediction quality, the adaptation to process changes, and the process optimization with respect to power consumption and membrane replacement cost. Its performance is currently experimentally veri ed in an industrial pilot plant.

REFERENCES

Broeckmann, A., J. Busch, T. Wintgens and W. Marquardt (2005). Modeling of pore blocking and cake layer formation in membrane ltration for wastewater treatment. *Accepted for Desalination.*

Cruse, A. (2006). Modellierung und optimierungsbasierte Prozessführung von kommunalen Abwasseraufbereitungsanlagen mit getauchten Membranmodulen. PhD thesis. RWTH Aachen University.

del Castillo, E. and A. M. Hurwitz (1997). Run-to-run process control: Literature review and extensions. *Journal of Quality Technology* **29**(2), 184–196.

Findeisen, R. and F. Allgöwer (2003). Computational delay in nonlinear model predictive control. In: *Proc. Int. Symp. Adv. Control of Chemical Processes, ADCHEM'03.*

Wittenmark, B. (1995). Adaptive dual control methods: An overview. In: *IFAC Adaptive Systems in Control and Signal Processing.*

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

MODEL PREDICTIVE CONTROL OF A CATALYTIC FLOW REVERSAL REACTOR WITH HEAT EXTRACTION

Fuxman, A.M. * **Forbes, J.F.** *,1 **Hayes, R.E.** *

* *Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2G6*

Abstract:
This paper presents the formulation of a controller for a Catalytic Flow Reversal Reactor (CFRR) with heat extraction. The controller is based on the Model Predictive Control (MPC) concept. The MPC scheme uses a model that assumes plug flow and neglects radial gradients in the reactor but accounts for the two phases within the reactor. The prediction of the future output behavior from the model is obtained by using the Method of Characteristics as proposed by Shang *et al.* (2004) for convection dominated distributed parameter systems. The formulated controller is applied to a CFRR unit for the catalytic oxidation of fugitive lean methane mixtures. The objective of the control algorithm is to maintain stable reactor operation, while extracting the maximum amount of useful energy by hot gas removal from the mid-section of the reactor. Simulations are used to show the performance of the designed controller.

Keywords: Reverse Flow Reactor, Model Predictive Control, Method of Characteristics.

1. INTRODUCTION

Catalytic Flow Reversal Reactor (CFRR) technology has received much attention in recent years (Matros and Bumimovich, 1996) and has been proposed for many applications including: methane combustion, oxidation of volatile organic compounds (VOC), oxidation of sulphur dioxide (SO2) and the synthesis of methanol.

CFRR has recently been suggested for the combustion of lean methane streams (Hayes, 2004). Fugitive lean methane streams are common in the oil and gas industry and are a great source of pollutant emission. Sources of methane emissions include leaks in natural gas transmission facilities such as pipelines and compression stations and upstream oil and gas production facilities, especially from solution gas. These methane emissions are typically available at ambient temperatures, where catalytic reaction rate is very slow, but the use of reverse flow technology has been shown to be feasible technology to achieve sufficiently high reactor temperatures (Hayes, 2004).

The primary advantage of the technology is that the thermal capacity of the solid material within the reactor acts as a regenerative heat exchanger, allowing authothermal operation without the use of heat exchangers. For exothermic reactions, switching the flow direction periodically creates a heat trap effect. This effect can be used to achieve and maintain an enhanced reactor temperature

[1] Email: fraser.forbes@ualberta.ca, phone: (780) 4492-0873, fax: (780) 492-2881

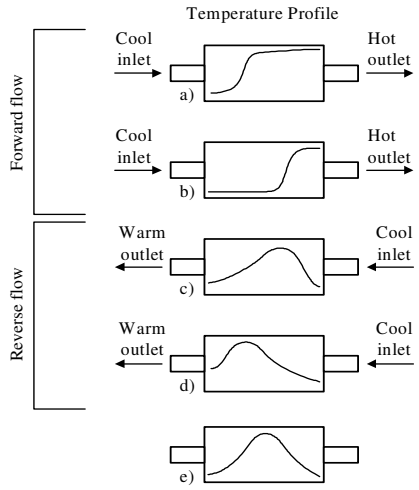


Fig. 1. Illustration of the heat trap effect for reverse flow operation.

compared to a single flow direction mode of operation.

The principle of the heat trap effect is illustrated in Figure 1. Figure 1(a) illustrates a reactor temperature profile that might be observed in a standard uni-directional flow operation for a combustion. If a temperature pattern, shown in Figure 1(a) and (b) is established, the reverse flow operation can then be used to take advantage of the high temperatures near the reactor exit to pre-heat the reactor feed. A quasi-steady state operation may be achieved in which the reactor temperature profile has a maximum value near the centre of the reactor, which slowly oscillates as the feed is switched between the two ends of the reactor, as shown in Figure 1(c-e).

The control of the CFRR is a particularly challenging problem. In addition to the complexities of any distributed parameter tubular reactor (i.e. nonlinear distributed dynamics and limited on-line measurement information), the CFRR presents periodic change of feed flow direction.

When controlling the CFRR, the main objective is to maintain the reactor operating in a region where the temperature in the active sections (catalyst sections) is below a critical value to avoid overheating or deactivation of catalyst and is above the extinction temperature of the reaction. The CFRR system is open-loop is stable. However, disturbances (i.e. inlet concentration), if sufficiently large, can extinguish the reaction or burn the catalyst. Different designs and control measures have been proposed to control the CFRR (Nieken *et al.*, 1994) including: bypassing the flow in the mid-section of the reactor, withdrawing of gas to a external cooler and cooling of the entire mid-section by using a heat exchanger.

The first work on feedback control for the CFRR was done by Budman *et al.* (1996). Two control strategies for a CO oxidation unit were discussed in their work: a PID feedback loop used to control the outlet concentration by manipulation of the cooling rate in the mid-section of the reactor and a feed-forward scheme that measures inlet concentration and select optimal cycle period and cooling rate from a parametric map. Barresi and Vanni (2002) discuss the use of a feedback logic controller, with cycle period as manipulated variable, to avoid extinction of the reaction in a volatile organic compounds (VOC) CFRR unit. More recently, a Model Predictive Control (MPC) was proposed to control a VOC combustor with flow reversal operation (Dufour and Couenne, 2003; Dufour and Toure, 2004) by using a power supply at the core of the reactor and inlet dilution. In the MPC formulation, a linear model obtained from the linearization of a nonlinear distributed parameter system model about a fixed operating point was used.

The aim of this paper is to present a control scheme that can be used to maintain the CFRR at a stable operational conditions, while extracting the maximum amount of hot gas from the reactor.

We investigate the use of heat removal by mass extraction from the middle section of the reactor as a manipulated variable. Extraction of a hot stream has an additional benefit of providing energy that can be used for many purposes such as heating and power generation (Kushwaha *et al.*, 2005). Hot gas withdrawal has been proposed in the literature to avoid overheating of the CFRR unit, but none of the control strategies published in the literature for the CFRR use the gas withdrawal from the mid-section as a control variable.

With this work, we contribute with the application of a Model predictive Control to a distributed parameter flow system with periodic oscillation of the flow direction. The controller is designed using the MPC scheme proposed by Shang *et al.* (2004) for convection-dominated distributed parameter systems where the Method of Characteristics (Arnold, 1988) is used to predict future output behaviour of the controlled plant. By applying this scheme to a CFRR unit we are extending its application to a distributed parameter system with output constraints. The method of characteristics for convection dominated systems (hyperbolic partial differential equation models) is simple and systematic and provides a geometric way of viewing the solution structure and can help in providing insight into the future evolution of the process output.

The work presented here is focused on the development of a candidate MPC scheme that will produce a high level of performance for the CFRR and to investigate the computational challenges inherent in this problem.

2. CONTROL SYSTEM DESIGN

In this work, a model of the CFRR unit to be controlled is used to predict the future behavior of the plant. The performance of the plant is optimized over a future finite horizon according to the current state of the plant. A sequence of manipulated variable adjustments is determined by optimizing an open-loop performance objective on a time interval extending from the current time through a specified prediction horizon. The computed settings for the manipulated variables are implemented and kept constant until the next control interval. Feedback is incorporated by using the measurements to correct modeling errors and update the disturbance estimate in the optimization problem for the next time step.

2.1 Modelling

The reactor consists of two parallel sections with an internal diameter of 0.2 metres mounted side by side and connected by a U-bend at the bottom (total length = 2.73 m). The reactor internals consist of a combination of open spaces, inert (monolith) sections and active catalyst (packed-bed) sections, as shown in Figure 2. A heterogeneous one-dimensional model is used to predict the future output behavior of the CFRR. The model is a simplified version of the two-dimensional heterogeneous model developed by Salomons *et al.* (2004). The basic equations for the mass and energy balance in the CFRR reactor, assuming plug flow, are:

$$\frac{\partial(Y_f)}{\partial t} + \alpha v_s \frac{\partial(Y_f)}{\partial x} = k_m a_v (Y_s - Y_f) \quad (1)$$

$$\frac{\partial(T_f)}{\partial t} + \alpha v_s \frac{\partial(T_f)}{\partial x} = \frac{h a_v}{\rho_f C p_f}(T_s - T_f) \quad (2)$$

$$k_m a_v C_f (Y_f - Y_s) = (1-\epsilon)\eta k_R C_s \quad (3)$$

$$\frac{\partial(T_s)}{\partial t} = \frac{\eta k_R C_s Y_s (-\Delta H_R)}{\rho_s C p_s} + h(T_f - T_s) \quad (4)$$

with boundary conditions $Y_f(t,0) = Y_{f_0}$ and $T_f(t,0) = T_{f_0}$, where Y_f and T_f are the mole fraction of methane and temperature of the fluid phase, Y_s and T_f are their counterpart in the solid phase, $(1-\alpha)$ is the fraction of mass extracted and v_s is the superficial velocity of the gas stream. Values for the various parameters in the model are given in Salomons *et al.* (2004).

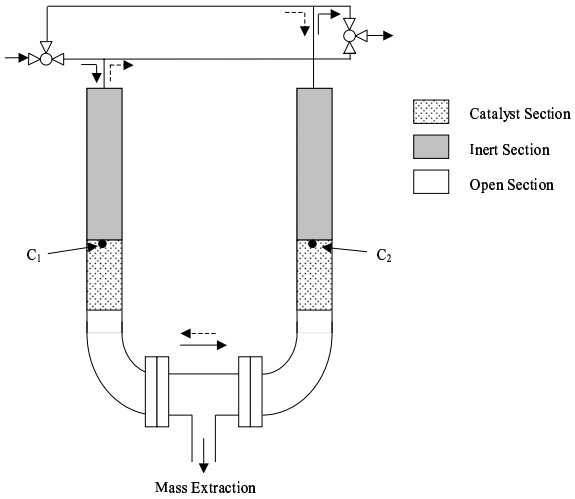


Fig. 2. Schematic picture of the CFRR reactor.

The dynamic behaviour of the CFRR is dominated by the energy balance in the solid phase, equation 4, while the dynamics of the mole and energy balance in the fluid phase are fast owing to the short residence time of the fluid in the reactor and the low thermal mass of the fluid. The resulting system of equations (1)-(4) can be solved using the method of characteristics (Acrivos, 1956). Equations (1) and (2) can be described by a system of ODEs along the characteristic curve ξ_1:

$$\xi_1 = \frac{dx}{dt} = \alpha v_s \quad (5)$$

Along ξ_1, the state variables Y_f and T_f are described by:

$$\frac{dY_f}{dt} = k_m a_v (Y_s - Y_f) \quad (6)$$

$$\frac{dT_f}{dt} = \frac{h a_v}{\rho_f C p_f}(T_s - T_f) \quad (7)$$

On the other hand, the energy balance equation in the solid phase varies along the time axis only, and its solution is described along a constant second characteristic line, ξ_2, by:

$$\frac{dT_s}{dt} = \frac{\eta k_R C_s Y_s (-\Delta H_R)}{\rho_s C p_s} + h(T_f - T_s) \quad (8)$$

The future output is predicted by numerically integrating the system of equations:

$$\frac{dx}{dt} = \alpha v_s \quad (9)$$

$$\frac{dt}{dt} = 1 \quad (10)$$

$$\frac{dY_f}{dt} = k_m a_v (Y_s - Y_f) \quad (11)$$

$$\frac{dT_f}{dt} = \frac{h a_v}{\rho_f C p_f}(T_s - T_f) \quad (12)$$

$$\frac{dT_s}{dt} = \frac{\eta k_R C_s Y_s (-\Delta H_R)}{\rho_s C p_s} + h(T_f - T_s) \quad (13)$$

$$k_m a_v C_f (Y_f - Y_s) = (1-\epsilon)\eta k_R C_s \quad (14)$$

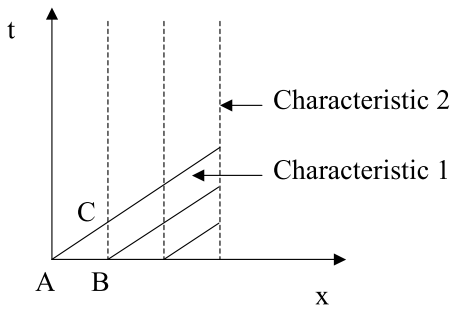


Fig. 3. Schematic picture of the CFRR reactor.

Predictions of future output values are obtained by discretizing the initial state at a finite number of spatial points, projecting the characteristic curves from each of these points and then computing the values of the state variables at the intersection points. Figure 3 illustrates the calculation of the state variables at point C from the values at points A and B. The segment AB is the domain of dependence of point C, given that the values of state variables at point C are completely defined by the state-variable values on the segment AB. By varying point C and repeating the procedure, the values of the state variables at different grid points and different future times can be calculated. The value of the state variables at the intersection points is obtained by integrating the differential equations using, for example, the Euler or mplicit Euler method and solving the resulting system of nonlinear equations. Using the output prediction method described above, the value of the output for a prediction horizon time is obtained for specific control actions. To use the predictions in the MPC scheme, the predicted output is expressed in a locally linear form (Shang *et al.*, 2004):

$$\hat{\mathbf{y}} = \hat{\mathbf{y}}_\mathbf{0} + \mathbf{S}\boldsymbol{\Delta}\mathbf{u} \tag{15}$$

$$\hat{\mathbf{y}}_\mathbf{0} = \mathbf{y}_0 + \mathbf{S}[u_{-1}, u_{-1}, \cdots, u_{-1}]^T \tag{16}$$

$$\boldsymbol{\Delta}\mathbf{u} = [u_0 - u_{-1}, \cdots, u_{H_c-1} - u_{-1}]^T \tag{17}$$

where $\hat{\mathbf{y}}_\mathbf{0}$ is the vector of predicted outputs due to past control actions in the prediction horizon time, $\mathbf{y_0}$ is the vector of initial values of the state variable, $\boldsymbol{\Delta}\mathbf{u}$ is the vector of future control increments ($u \triangleq \alpha$; $\Delta u_i = \alpha_i - \alpha_{-1}$ *for* $i = 0 \cdots H_C - 1$), $\hat{\mathbf{y}}$ is the vector of the predicted outputs and $\mathbf{S}$ is the rate of output variation about past control actions ($\mathbf{u}_{-1}$). The elements of $\mathbf{S}$ are updated at each sampling time and are computed via perturbation:

$$\mathbf{S} = \left(\frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{u}}\right)_0 = \frac{\hat{\mathbf{y}}|_{\mathbf{u}_{-1+\delta}} - \hat{\mathbf{y}}|_{\mathbf{u}_{-1}}}{\delta} \tag{18}$$

where δ is a numerical perturbation on past input $\mathbf{u}_{-1}$, $\hat{\mathbf{y}}|_{\mathbf{u}_{-1+\delta}}$ and $\hat{\mathbf{y}}|_{\mathbf{u}_{-1}}$ are the predicted future output under the control actions $\mathbf{u}_{-1+\delta}$ and $\mathbf{u}_{-1}$. The future output is predicted up to an appropriate prediction horizon (H_P).

2.2 Optimization: Control Objective and Constraints

The control objective is to maintain the reactor temperatures within an appropriate range so that overheating and/or reaction extinction are not possible. In this work we focus the attention on the extinction phenomena. This control objective is met by manipulating the flow of hot gas from the mid-section of the reactor. An additional control objective involves the extraction of the maximum amount of energy from the reactor. To achieve the control objective, the following finite horizon problem is solved at each sampling time (k):

$$\min_{u(k+i|k)} \; J(u(k+i|k)) = \sum_{i}^{H_C-1} \left[\frac{\Delta u(k+i|k) - (u_{-1} + u_{min})}{(u_{max} - u_{min})}\right]^2 \tag{19}$$

such that

$$u_{min} \leq u \leq u_{max}$$
$$\Delta u_{min} \leq \Delta u \leq \Delta u_{max}$$
$$T_{min} \leq T_s \leq T_{max}$$

The constraints are arranged in a linear vector inequality form and are softened by penalizing the ∞-norm of the constraint violations. Feedback is incorporated by comparing the actual measurement of the plant and the predicted output from the model. The resulting constrained quadratic optimization is solved using the active set method.

3. SIMULATIONS

To evaluate the performance of the designed controller, simulations of the CFRR unit and controller were implemented in the Matlab® environment. For all the simulation cases, an initial temperature profile, Figure 4, a set of boundary conditions $Y_f(t, x = 0) = 0.5\ mol\ \%$ $T_f(t, x = 0) = 298K$ and inlet flow velocity ($v_s(t, x = 0) = 0.2 \cdot m/sec$) are chosen. A fixed time for the flow reversal is chosen and setted to 300 sec. For the controller, the following parameters are used: $T_{sampling} = 50(sec)$, $H_P = 18T_{sampling} = 900(sec)$, $H_C = 1$, $0.6 \leq \alpha \leq 1$ and $|\Delta\alpha| = 0.05$.

The controller is then used to control the minimum temperature in the active catalyst sections (to avoid extinction of the reaction) while extracting the maximum amount of heat from the mid-section of the reactor. The minimum temperature can be chosen as the minimum temperature required to avoid extinction of the reaction or as the temperature that will give a high conversion

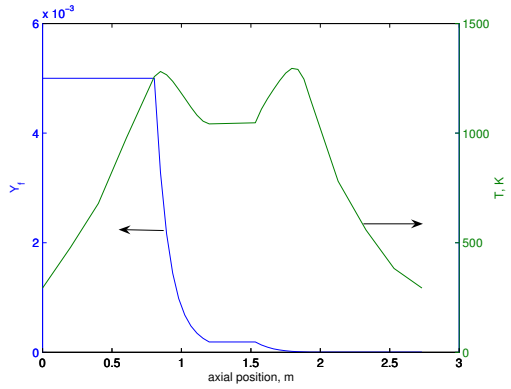


Fig. 4. Initial temperature ($T_f(0,x) = T_s(0,x)$) and concentration distribution along the axial distance of the reactor.

of reactants. If the latter is not known, then a constraint on the maximum allowable mol fraction of methane (Y_f) on the exit of each catalyst section can be added. In this work, a minimum temperature of 950 K is arbitrarily chosen for simulation purposes. The infinite dimensional process variables at the current time are discretized into $m = 60$ points along the axial direction of the reactor. It is assumed that all process variables can be measured at these discrete points.

The control performance is first evaluated with a simulated plant that matches exactly the the model used to predict the future output behavior of the CFRR unit (i.e. equations (1) - (4)). Figure 5(a)-(b) shows the output behaviour of the temperature at two points (marked as C_1 and C_2 in Figure 2). The temperature at the inlet of the catalytic sections are of great importance since most of the reaction takes place near the entrance to these sections. It can be seen from Figure 5(a)-(b) that the controller is able to drive the output to a new stationary state where the minimum temperature is above the desired temperature. Figure 6 shows the optimal fraction of total mass flow of hot gas that is extracted to achieve the desired control performance.

The control performance is also evaluated with a plant simulated with a highly detailed model . The new plant consist of a 2-dimensional heterogeneous model developed by Salomons *et al.* (2004) that is solved using the finite element method in Femlab®. The main structural difference of simplified 1-D model and the full 2-D model is the effect of the thermal insulation (thickness of insulation = 0.28 metres), which has been shown to be important for small diameter reactors and low air velocity conditions (Aube and Sapoundjiev, 2000; Salomons *et al.*, 2004). Figure 7(a)-(b) shows the output behavior of the temperature at the two points marked in Figure 2. It can be seen from Figure 7(a)-(b) that the controller is able to drive the output to a new stationary

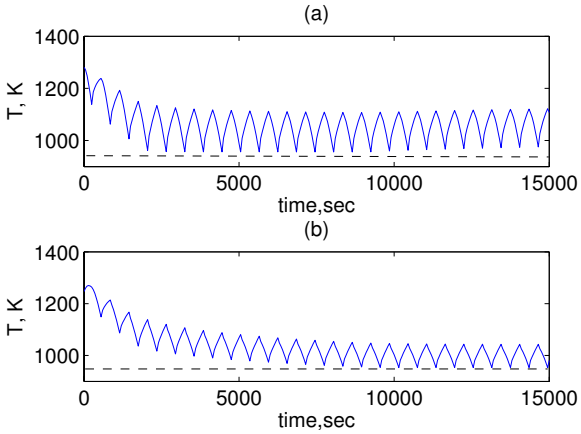


Fig. 5. (a) Temperature at point C_1 (see Figure 2) with control. (b) Temperature at point C_2 (see Figure 2) with control. The dashed line indicates the temperature lower bound.

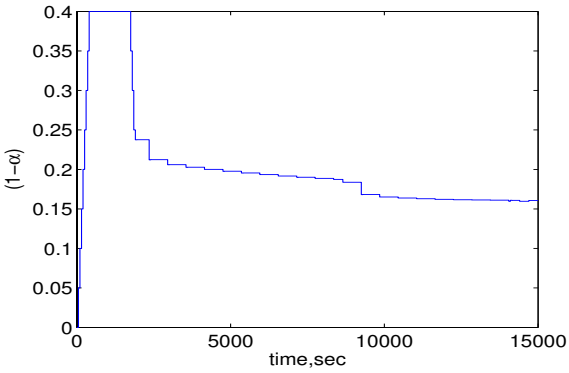


Fig. 6. trajectory of the manipulated variable: fraction of total mass flow of hot gas that is extracted

state where the minimum temperature at the at the inlet of the catalytic sections is above the threshold value. Figure 7(c) shows the optimal fraction of total mass flow of hot gas that is extracted to achieve the desired control performance. The optimal value is now higher than the value obtained in Figure 6 and this is expected since the plant includes the effect of the external heat transfer resistant that is given by the insulation. By simulations (not shown) we observed that the oscillations in the adjustments of the manipulated variable can be decreased by increasing the number of discrete points in the controller (m). However, a finer discretization comes at a higher computational load.

4. CONCLUSION

In this paper we presented the formulation of a controller to a distributed parameter flow system with periodic oscillation of the flow direction. The controller uses the Model Predictive Control concept and is based on the Model Predictive control

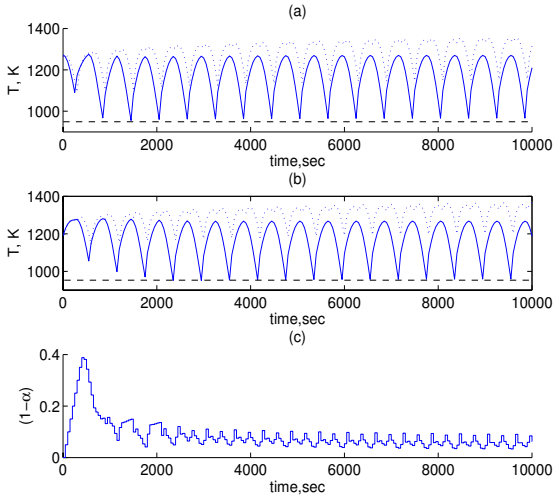


Fig. 7. (a) Temperature at point C_1 (see Figure 2) with control (solid line) and without control (dotted line). (b) Temperature at point C_2 (see Figure 2) with control (solid line) and without control (dotted line). The dashed line indicates the temperature lower bound.(c) Trajectory of the manipulated variable: fraction of total mass flow of hot gas that is extracted

scheme for convection-dominated distributed parameter systems proposed by Shang *et al.* (2004). We applied the control scheme to a CFRR unit for the combustion of methane to maximize the amount of energy that can be extracted from the reactor without extinguishing the reaction.

Simulations are used to show the ability of the controller to find the optimal extraction of hot gas extraction while keeping the minimum temperature at the inlet of the active catalyst sections above a minimum threshold value that guarantee stable operation (no deactivation of catalyst sections).

NOMENCLATURE

v_s superficial velocity (m/s)
α fraction of total inlet mass flow in the reactor (-)
k_m mass transfer coefficient (m/s)
a_v surface area per unit volume (m^2/m^3)
h heat transfer coefficient
ρ density (kg/m^3)
Cp heat capacity $(J/kg \cdot K)$
k_r first-order rate constant (s^{-1})
η effectiveness factor
ΔH enthalpy of reaction of methane (J/mol)
H_P prediction horizon
H_C control horizon
$T_{sampling}$ sampling time
m number of discrete points in the controller

REFERENCES

Acrivos, A.A. (1956). Method of characteristics technique: Application to Heat and Mass Transfer Problems. *Industrial and Engineering Chemistry* **48**, 703–710.

Arnold, V.I. (1988). *Geometric methods in the theory of ordinary differential equations.* Springer-Verlag. New York.

Aube, F. and H. Sapoundjiev (2000). Mathematical model and numerical simulations of catalytic flow reversal reactors for industrial applications. *Computers and Chemical Engineering* **24**, 2623–2632.

Barresi, A.A. and M. Vanni (2002). Control of catalytic combustors with periodical flow reversal. *AICHE Journal* **48**(3), 648–652.

Budman, H., M. Kzyonsek and P. Silveston (1996). Control of Nonadiabatic packed bed reactor under periodic flow reversal. *Canadian Journal of Chemical Engineering* **74**, 751–759.

Dufour, P. and Y. Couenne, F.and Toure (2003). Model predictive control of a catalytic reverse flow reactor. *IEEE Transactions on Control Systems Technology* **11**(5), 705–714.

Dufour, P. and Y. Toure (2004). Multivariable model predictive control of a catalytic reverse flow reactor. *Computers and Chemical Engineering* pp. 2259 – 2270.

Hayes, R.E. (2004). Catalytic solutions for fugitive methane emissions in the oil and gas sector. *Chemical Engineering Science* **59**(19), 4073–4080.

Kushwaha, A., M. Poirier, R.E. Hayes and H. Sapoundjiev (2005). Heat extraction from a flow reversal reactor in lean methane combustion. *Chemical Engineering Research and Design* **83**, 205–213.

Matros, Y.S. and G.A. Bumimovich (1996). Reverse-flow operation in fixed bed catalytic converters. *Catalytic Reviews: Science and Engineering* **38**, 1–36.

Nieken, U., G. Kolios and G. Eigenberger (1994). Control of the ignited steady state in autothermal fixed-bed reactors for catalytic combustion. *Chemical Engineering Science* **49**, 5507–5518.

Salomons, S., R.E. Hayes, M. Poirier and H. Sapoundjiev (2004). Modelling a reverse flow reactor for the catalytic combustion of fugitive emissions. *Computers and Chemical Engineering* **28**, 1599–1610.

Shang, H., J.F. Forbes and M. Guay (2004). Model predictive control for quasilinear hyperbolic distributed parameter systems. *Industrial and Engineering Chemistry Research* **43**, 2140–2149.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

NMPC WITH STATE-SPACE MODELS OBTAINED THROUGH LINEARIZATION ON EQUILIBRIUM MANIFOLD

Stephan Koch, Ricardo G. Duraiski, Pedro Bolognese Fernandes, Jorge O. Trierweiler

Group of Integration, Modelling, Simulation, Control and Optimization of Processes (GIMSCOP)
Department of Chemical Engineering, Universidade Federal do Rio Grande do Sul (UFRGS).
Rua Luiz Englert, s/n, ZIP CODE: 90040-040, Porto Alegre - RS - Brazil.
{stephan, ricardo, pedro, jorge}*@enq.ufrgs.br*

Abstract: This paper presents a Nonlinear Model Predictive Control approach for nonlinear state-space models obtained with the modelling and identification technique recently proposed in literature as Linearization on the Equilibrium Manifold (LEM). The predictive controller that will be applied to the LEM uses the Local Linearization on the Trajectory algorithm (LLT) which simulates the nonlinear plant and calculates optimal control actions based on local linearizations around the simulated trajectory by online minimization of an objective function. The proposed combination of the LEM and LLT techniques is tested with a nonlinear SISO system.

Keywords: nonlinear systems, identification, linearization, predictive control, trajectories

1. INTRODUCTION

The importance of nonlinear approaches to control systems in industrial chemical processes has been rising significantly during the last few years and will continue to do so in the future. The high demands of today's economy in terms of process yield and the obedience of environmental standards require an increased efficiency that cannot always be achieved with linear control concepts. At the same time, the availability of nonlinear dynamic models has been recognized in the literature as one of the main obstacles, if not the most important, for the application of nonlinear control strategies. High cost and complexity of nonlinear approaches often impose restrictions on practical usability.

This situation calls for methods that take into account the well-developed linear control theory, extending it for usage with nonlinear processes. One possibility, which can be termed "grey-box" modelling, is the use of "local models", understood as approximations of the original system in a limited sub-region of the operating domain in order to construct a nonlinear model. The underlying principle is that the system behavior is "simpler" locally than globally and as a result local models can be identified more easily. Examples of this methodology are the local linear models tree (Nelles, 1997) and the identification through the decomposition into operating regimes (Johansen and Murray-Smith, 1997).

The Linearization on the Equilibrium Manifold (LEM) approach (Bolognese Fernandes and Engell, 2005) proposes a way of constructing a nonlinear model by interpolating the equilibrium manifold and the linear behavior of the system between different operating points. It has been shown that various problems of local modelling techniques can be avoided by using this method, making it an appealing way of obtaining a nonlinear model with less than the effort necessary for a first-principles modelling.

A suitable control strategy for this kind of system would also use linear models for determining control actions, as these are already available and in use for the construction of the global model. The Local Linearization on the Trajectory algorithm LLT (Duraiski, 2001) is a predictive control strategy for nonlinear models which uses local linearizations at the current point of the system state. In the following, a combination of the LLT with LEM models will be proposed and compared to the already well-tested combination of the LLT with a first-principles model.

This paper is structured as follows: section 2 presents the basics of the LEM method in the general (MIMO) form, section 3 explains the LLT control strategy. Section 4 proposes a combination of the two methods, which is evaluated with numerical experiments in section 5 using a nonlinear SISO example system. Concluding remarks and proposals for further investigations can be found in Section 6.

2. LINEARIZATION ON THE EQUILIBRIUM MANIFOLD (LEM) MODELS

Consider a continuous MIMO nonlinear dynamic system of the form

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{r}(\mathbf{x},\mathbf{u}) \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}) \end{aligned} \tag{1}$$

where $\mathbf{r}: X \times U \to \Re^n$ is at least once continuously differentiable on $X \subseteq \Re^n$, $U \subseteq \Re^m$, and $\mathbf{h}: X \to \Re^p$ is at least once continuously differentiable. The output equation will be frequently omitted in the sequel for shortness. The equilibrium manifold of (1) is defined as the family of constant equilibrium points

$$\Xi = \left\{ (\mathbf{x}_s, \mathbf{u}_s, \mathbf{y}_s) \in \Re^n \times \Re^n \times \Re^n : \; \mathbf{r}(\mathbf{x}_s, \mathbf{u}_s) = \mathbf{0}, \;\; \mathbf{y}_s = \mathbf{h}(\mathbf{x}_s, \mathbf{u}_s) \right\}. \tag{2}$$

Similarly, the family of linearizations of (1) at the set of equilibrium points determined by (2) is given in the usual way as

$$\dot{\mathbf{x}} = \left[\frac{\partial \mathbf{r}(\mathbf{x},\mathbf{u})}{\partial \mathbf{x}} \right]_{\mathbf{x}_s,\mathbf{u}_s} (\mathbf{x} - \mathbf{x}_s) + \left[\frac{\partial \mathbf{r}(\mathbf{x},\mathbf{u})}{\partial \mathbf{u}} \right]_{\mathbf{x}_s,\mathbf{u}_s} (\mathbf{u} - \mathbf{u}_s) \tag{3}$$

and similarly for the output equation. Under the condition that the rank of $[\partial \mathbf{r}(\mathbf{x}_s,\mathbf{u}_s)/\partial \mathbf{x}]$ is n for the set Ξ (Wang and Rugh, 1987, Bolognese Fernandes 2005), the equilibrium manifold and consequently the family of linearizations of (1) will be specified by m among the $n+m$ variables $(\mathbf{x},\mathbf{u})$. Therefore, if this matrix is full rank, the set of inputs fully parameterize both families of equilibrium points and linearizations.

Calling the steady-state map $\mathbf{\Omega}: \Re^m \to \Re^n$, such that $\mathbf{r}(\mathbf{\Omega}(\mathbf{u}),\mathbf{u}) = \mathbf{0}$ (that is, the function $\mathbf{\Omega}$ gives the steady-state $\mathbf{x}_s$ corresponding to a constant input $\mathbf{u}_s$), the input-parameterized linearization the equilibrium manifold (LEM) of (1) is defined as the system (Bolognese Fernandes and Engell, 2005)

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{u})(\mathbf{x} - \mathbf{\Omega}(\mathbf{u})) \tag{4}$$

$\mathbf{A}(\mathbf{u})$ represents the evaluation of the Jacobian matrix $[\partial \mathbf{r}(\mathbf{x},\mathbf{u})/\partial \mathbf{x}]$ on $(\mathbf{\Omega}(\mathbf{u}),\mathbf{u})$. The focus on input parameterization is due to the fact that identification experiments to obtain $\mathbf{A}(\mathbf{u})$ and $\mathbf{\Omega}(\mathbf{u})$ from process data are carried out by exciting the plant with a designed input signal. The output equation can be linearized in an analogous way, considering the stationary output mapping: $\mathbf{\Psi}: \Re^m \to \Re^p$.

The model (4) has to be interpreted as *a* (state-affine) nonlinear system that possesses the same family of equilibrium points (2) and the same linearization family (4) as the nonlinear system (1). Following the discussion in Bolognese Fernandes (2005), the LEM system can be a good approximation of (1) in transient regimes away from the equilibrium manifold depending on the degree of nonlinearity, in that way substituting a first-principles model. Obviously, other representations that are equivalent on the equilibrium manifold can be constructed on the basis of any m distinct parameters. Moreover, these representations can be easily interchanged, provided that the inverses of the corresponding elements in $\mathbf{\Omega}(\mathbf{u})$ and $\mathbf{\Psi}(\mathbf{u})$ exist. For further information about how to obtain the equilibrium manifold and the dynamic matrix $\mathbf{A}$, please refer to Bolognese Fernandes (2005).

3. LOCAL LINERIZATIONS ON THE TRAJECTORY (LLT)

In the following section a control strategy for the model introduced above will be presented. It was developed by Duraiski (2001) and consists of a model predictive control algorithm which works in the following way: the control actions applied to the manipulated variables are obtained by optimizing an objective function of control costs using a nonlinear internal model to predict the future system outputs. The control actions, however, are determined in each iteration through the use of a set of linear models in the step response form, obtained through local linearizations around the trajectory of the system, previously obtained in the last iteration. This ensures that the optimization problem is quadratic as it is in the case of Linear Model Predictive Control, and thus easy to solve.

3.1 Algorithm description

The LLT algorithm (Duraiski, 2001) consists of the following iterative calculation steps:

1) The first solution is based on a linearized model at the current operating conditions. Using this trajectory it is possible to simulate the nonlinear model which is used to calculate a sequence of linear models for the next iteration.
2) With the sequence of linearized models on the trajectory a new control action trajectory is calculated.

3) This sequence of control moves is applied to the simulation of the open-loop response of the internal model.
4) Based on the new trajectory, it is possible to determine a new set of linearized models in the same way as it is done in the first step. Then, this set of models is used in the next iteration step.
5) The steps 2, 3 and 4 are sequentially carried out until the algorithm converges, i.e., when the i-th control action trajectory (calculated in the current iteration) does not differ too much from the (i-1)-th, satisfying the maximum norm convergence criterion $\|u_i - u_{i-1}\|_\infty \leq TolU \in \Re$. In case the algorithm does not converge after a given time and number of iterations, e. g. when the setpoint is unattainable, the best of all calculated control actions in this time step will be applied.

.

3.2 Linearized Step Response Model

In this part the linear step response model used for predicting the future system output will be developed. Primarily, the following discrete time state space equation is considered.

$$(\mathbf{x}_k - \mathbf{x}^B_{k-1}) = \mathbf{A}_{k-2} \cdot (\mathbf{x}_{k-1} - \mathbf{x}^B_{k-2}) + \mathbf{B}_{k-2} \cdot (\mathbf{u}_{k-1} - \mathbf{u}^B_{k-2}) \quad (5)$$

$$(\mathbf{y}_k - \mathbf{y}^B_{k-1}) = \mathbf{C}_{k-1} \cdot (\mathbf{x}_k - \mathbf{x}^B_{k-1}) + \mathbf{D}_{k-1} \cdot (\mathbf{u}_k - \mathbf{u}^B_{k-1}) \quad (6)$$

The matrices $\mathbf{A}_{k-2}$, $\mathbf{B}_{k-2}$, $\mathbf{C}_{k-1}$ and $\mathbf{D}_{k-1}$ are obtained by *discretization* of the continuous linear state space system resulting from the Taylor linearization of equation (1). They are not to be confused with the traditional notation for the continuous state space matrices (i.e., **A**, **B**, **C** and **D)**. The variables $\mathbf{x}^B$, $\mathbf{u}^B$, $\mathbf{y}^B$ represent the variables **x**, **u**, **y** at the point of linearization. Equations (5) and (6) can now be applied iteratively for the time steps from 0 up to the simulation horizon P, yielding an output equation for each discrete time step. With this, an equation for the output **Y** from the time instant 0 to P can be constructed. Written in a compact matrix form, the following equation is obtained:

$$\mathbf{Y}_{\left[\begin{smallmatrix}P\\0\end{smallmatrix}\right]} = \mathbf{Su} \cdot \boldsymbol{\delta}\mathbf{U}_{\left[\begin{smallmatrix}P\\0\end{smallmatrix}\right]} + \mathbf{Sx} \cdot \boldsymbol{\delta}\mathbf{x}_0 + \mathbf{Y}^B_{\left[\begin{smallmatrix}P-1\\-1\end{smallmatrix}\right]} \quad (7)$$

Equation (7) will be used with some alterations within the calculation and optimization of the objective function. For details please refer to Duraiski (2001).

3.3 Objective function

The optimization problem consists of the minimization of a quadratic objective function with penalty terms for setpoint deviations and control actions, in the most general form being

$$J = \min_{\Delta U_{\left[\begin{smallmatrix}M\\0\end{smallmatrix}\right]}} \left(\sum_{i=0}^{P} (\gamma_i \cdot (y_i - r_i))^2 + \sum_{i=0}^{M} (\lambda_i \cdot \Delta u_i)^2 \right) \quad (8)$$

In the case of the LLT method, the input difference variable $\Delta u_k = u_k - u_{k-1}$ is replaced by the deviation variable $\delta u_k = u_k - u_{k-1}^B$. Furthermore, a penalty term for soft constraints $\varphi|s|$ ($s \geq 0$ being a scalar slack variable that is only nonzero while the constraints are violated) and for the deviation of the manipulated variable from a given target z_i are introduced. With these alterations, the objective function is

$$J = \min_{\delta U_{\left[\begin{smallmatrix}M\\0\end{smallmatrix}\right]}} \left(\begin{array}{l} \sum_{i=0}^{P} (\gamma_i \cdot (y_i - r_i))^2 + \\ \sum_{i=0}^{M} (\lambda_i \cdot ((\delta u_i + u^B_{i-1}) - (\delta u_{i-1} + u^B_{i-2})))^2 + \\ \sum_{i=0}^{M} (\psi_i \cdot ((\delta u_i + u^B_{i-1}) - z_i))^2 + (\varphi |s|)^2 \end{array} \right) \quad (9)$$

The parameters γ_i, λ_i, ψ_i and φ_i are to be determined by common MPC parameter tuning methods. For the actual implementation, equation (9) can be rewritten in a matrix form. Further details will not be discussed here and can be found in Duraiski (2001).

4. COMBINING A LEM MODEL WITH AN LLT CONTROLLER

Now a combination of models obtained through the LEM technique with a nonlinear model predictive controller using the LLT method will be proposed. In general, two possibilities exist to achieve this goal: first, using the LEM as a nonlinear model for the LLT algorithm as it is, deriving the needed Jacobians **A**, **B**, **C**, **D** through analytic or numerical differentiation of the LEM itself. A second possible approach is altering the LLT algorithm in a way that it can deal directly with the dynamic matrix **A** and the stationary manifold vector **Ω**(**u**) of the LEM. In this work only the first possibility will be investigated, as it demonstrates the feasibility of the approach with fairly low effort in terms of implementation. The control performance and computational effort of the LEM+LLT combination will be compared to an LLT controller with a nonlinear model.

It will be assumed in the sequel that a nonlinear LEM model in the autonomous state space form as stated in equation (4) has been constructed by identifying the matrix **A** and the equilibrium manifold **Ω**(**u**) using appropriate techniques. Further details about model construction can be found in Bolognese Fernandes (2005). Equation (4) will now be incorporated into the LLT algorithm, using it as a description of the nonlinear process. As can be seen in equations (5) and (6), it is necessary to derive a general linearization of the process for later discretization and the calculation of control actions by the LLT. This is achieved by a straight-forward Taylor linearization of the LEM model around an arbitrary point ($\mathbf{x}^B$, $\mathbf{u}^B$) in state space. Note that this has to be a dynamic linearization as we cannot assume the system to be at an equilibrium state at all times. However, the resulting *bias* caused by the term $\mathbf{r}(\mathbf{x}^B, \mathbf{u}^B) = \mathbf{A}(\mathbf{u}^B)(\mathbf{x}^B - \boldsymbol{\Omega}(\mathbf{u}^B))$ will cancel out in the differential equation, as shown in Duraiski (2001), Appendix B.

Linearizing equation (4) around an arbitrary point ($\mathbf{x}^B$,$\mathbf{u}^B$) in state space yields:

$$\begin{aligned}\Delta\dot{\mathbf{x}} &= \frac{\partial(\mathbf{r}(\mathbf{x},\mathbf{u})}{\partial\mathbf{x}}\Big|_{\mathbf{x}^B\mathbf{u}^B}\cdot\Delta\mathbf{x} + \frac{\partial(\mathbf{r}(\mathbf{x},\mathbf{u}))}{\partial\mathbf{u}}\Big|_{\mathbf{x}^B\mathbf{u}^B}\cdot\Delta\mathbf{u} \\ &= \mathbf{A}(\mathbf{u}^B)\Delta\mathbf{x} + \frac{\partial\mathbf{A}(\mathbf{u})}{\partial\mathbf{u}}\Big|_{\mathbf{u}^B}\cdot\mathbf{x}^B\cdot\Delta\mathbf{u} \\ &\quad - \frac{\partial\mathbf{A}(\mathbf{u})}{\partial\mathbf{u}}\Big|_{\mathbf{u}^B}\cdot\boldsymbol{\Omega}(\mathbf{u}^B)\cdot\Delta\mathbf{u} - \mathbf{A}(\mathbf{u}^B)\cdot\frac{\partial\boldsymbol{\Omega}(\mathbf{u})}{\partial\mathbf{u}}\Big|_{\mathbf{u}^B}\Delta\mathbf{u}\end{aligned} \tag{10}$$

With this the following matrixes are obtained:

$$\begin{aligned}\mathbf{A}^B &= \mathbf{A}(\mathbf{u}^B) \\ \mathbf{B}^B &= \frac{\partial\mathbf{A}(\mathbf{u})}{\partial\mathbf{u}}\Big|_{\mathbf{u}^B}\cdot\left(\mathbf{x}^B - \boldsymbol{\Omega}(\mathbf{u}^B)\right) - \mathbf{A}(\mathbf{u}^B)\cdot\frac{\partial\boldsymbol{\Omega}(\mathbf{u})}{\partial\mathbf{u}}\Big|_{\mathbf{u}^B}\end{aligned} \tag{11}$$

The output equation from (1) can also be linearized in a straight-forward way, yielding matrices $\mathbf{C}^B$ and $\mathbf{D}^B$ for the linear state space model.

$$\Delta\mathbf{y} = \frac{\partial(\mathbf{h}(\mathbf{x}))}{\partial\mathbf{x}}\Big|_{\mathbf{x}^B}\cdot\Delta\mathbf{x} + \frac{\partial(\mathbf{h}(\mathbf{x}))}{\partial\mathbf{u}}\Big|_{\mathbf{x}^B\mathbf{u}^B}\cdot\Delta\mathbf{u} \tag{12}$$

$$\mathbf{C}^B = \frac{\partial(\mathbf{h}(\mathbf{x}))}{\partial\mathbf{x}}\Big|_{\mathbf{x}^B}, \quad \mathbf{D}^B = \mathbf{0}. \tag{13}$$

Equations (11) and (13) will be discretized for each instant of time, yielding matrixes $\mathbf{A}_k$, $\mathbf{B}_k$ and $\mathbf{C}_k$. As we assume the output equation to be only dependent on $\mathbf{x}$, $\mathbf{D}_k$ will always be 0. With this, equations (5) and (6), which serve the purpose of determining the control actions of the predictive controller, can be easily constructed. The trajectory simulation is performed with the original nonlinear model from equation (4).

5. CASE STUDY

To prove the applicability of the proposed combination of the LEM and LLT methods, a nonlinear SISO system will be considered as an example.

5.1 Methodology and control objectives

The system under consideration will be tested and compared in four different forms:

a) Nonlinear model. A nonlinear differential equation derived from first-principle modelling techniques. This model also represents the real plant that is to be controlled. This holds for all the five cases a)-d).

b) Analytic LEM model. A nonlinear LEM model is constructed by analytic calculation of the dynamic matrix **A** and the equilibrium manifold $\boldsymbol{\Omega}(u)$.

c) Interpolated LEM model. A nonlinear LEM model is constructed by spline interpolation of the dynamic matrix **A** and the equilibrium manifold $\boldsymbol{\Omega}(u)$ between different operating points. These operating points and their linear dynamics are determined analytically.

d) Linearized model. One of the linear models from b) at one point of operation only, without LEM. This yields actually a purely linear MPC problem.

For each of the mentioned models, numerical experiments with various LLT controllers in different operation domains are conducted. The controller parameters are determined using the RPN methodology developed by Trierweiler and Farina (2003), the control objective being a decrease of the closed-loop rise time to 1/6 of the open-loop rise time. For testing the closed-loop system, a series of random set point changes is applied to the controlled variable of the system. Furthermore, the total cost J_{total} accumulated during the simulation time is compared, as well as the total necessary iterations.

5.2 The example system: isothermal CTSR reactor with Van de Vusse reaction scheme

The Van de Vusse reaction scheme is a well-known benchmark problem for nonlinear control algorithms and has been studied extensively by various researchers. A detailed model for this system was presented in Engell and Klatt (1993). For shortness, only the differential equations will be shown here.

$$\begin{aligned}\dot{x}_1 &= -k_1x_1 - k_3x_1^2 + (x_{1,in} - x_1)u \\ \dot{x}_2 &= k_1x_1 - k_2x_2 - x_2u \\ y &= x_2\end{aligned} \tag{14}$$

In these equations x_1 is the concentration of component *A*, x_2 is the concentration of component *B* and $x_{1,in}$ is the feed concentration of *A*, assumed to remain constant. The parameter values are k_1=15.0345 h^{-1}, k_2 = 15.0345 h^{-1}, k_3 = 2.324 $l\cdot mol^{-1}\cdot h^{-1}$ $x_{1,in}$ = 5.1 $mol\cdot l^{-1}$ (Engell and Klatt, 1993). In this example only the operating range of $3 < u_s < 35$ h^{-1} is investigated.

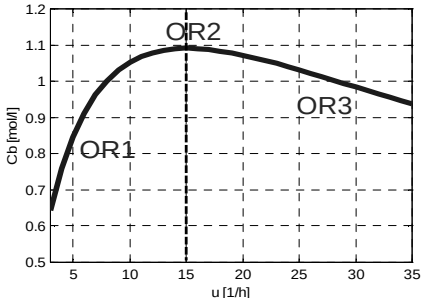


Fig. 1: Steady state C_B (x_2) concentration vs. dilution rate u for $x_{1,in}$ = 5.1 $mol\cdot l^{-1}$

A particularity of the Van de Vusse reaction is the division of the operating domain in two parts with different dynamic behaviours. Fig. 1 shows the steady state C_B solutions as a function of the dilution rate u with the three typical operating regions (i.e., OR1, OR2, and OR3). For OR1 (i.e., u_s < 15 h^{-1}) a non-minimum phase behaviour can be observed, while for OR3 (i.e., u_s > 15 h^{-1}) the behaviour is

minimum-phase. Close to the peak the zero gets close to the origin, intensifying the non-minimum phase behaviour of the left side and making the controller design more difficult. At the peak, the zero is null as well as the static gain, which is positive for OR1 and negative for OR3.

5.3 Numerical results

Details on obtaining the different LEM systems are not discussed here and can be found in Bolognese Fernandes (2005). First, OR1 and OR2 are investigated. In this region the proposed control goal could only be achieved by accepting excessive control action and inverse responses. Thus, the closed-loop rise time will only be reduced to about 3/4 of the open-loop value. The following LLT parameters are used:

Table 1: non-minimum-phase LLT parameters

Prediction / Control horizon (P / M)	136 / 34
Output / Input variable weight (Γ / Λ)	0.7 / 0.06
Sampling time (Ts)	0.15 min

The system is subjected to a series of setpoint changes in the region $0.7M < c_B < 1.11M$.

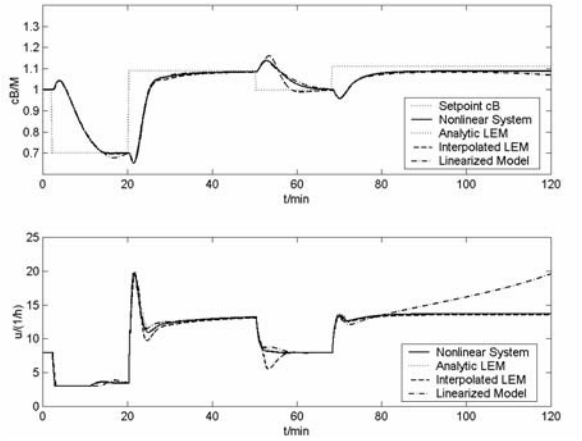


Fig. 2: time response in OR1 and OR2

In the first 20 minutes of the simulation the closed-loop velocity is limited by the lower constraint of $u = 3h^{-1}$. Up to this point, the behaviour of all the systems is very similar. The next setpoint is c_B=1.09M which is the theoretical maximum of the concentration c_B (OR2, see Fig. 1) and causes a rapid almost step-like control action. Note that the response of any input-parameterized LEM system to a step in the manipulated variable is comparable to the behaviour of the linear model at the equilibrium point corresponding *to the new input* (Bolognese Fernandes 2005). All four systems approach the equilibrium state with a similar velocity. From the 75th minute, the system is subjected to a setpoint of $c_B = 1.11h^{-1}$ that is not attainable (see Fig. 1). With the two LEM model versions the controller stabilizes the system at the maximum value of $c_B = 1.09h^{-1}$. This is a clear advantage in comparison to the linear MPC shown in Fig. 2, or any other linear (e.g. PI) controller, which is not able to keep the system at the point of maximum yield and trespasses gradually into OR3.

Now the region OR3 for $1M < c_B < 1.09M$ is investigated. Minimum-phase behaviour causes a better overall performance. First, the controller shown above is tested with a series of setpoint changes of random magnitude to prove the usability of one set of control parameters for the whole operation domain. Fig. 3 shows these results for OR3, the linearized model being designed for OR1, while Fig. 4 shows the same setpoint changes applied to OR1. Table 2 below summarizes the quantitative performance of the different models.

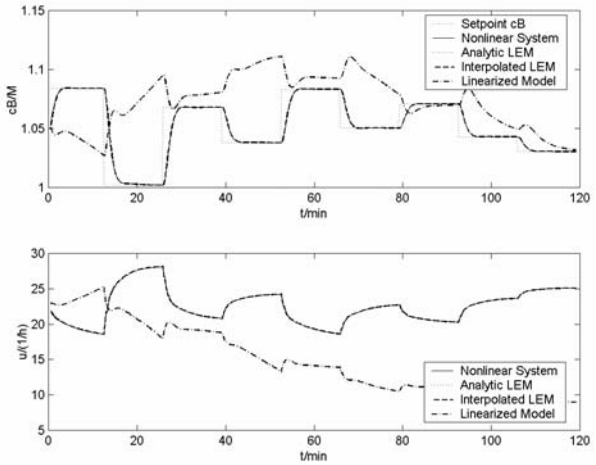


Fig. 3: minimum-phase responses, $1M < c_B < 1.09M$.

Fig. 3 shows clearly that the linear controller for OR1 cannot operate in OR3. All the other models follow the perturbations in the same way.

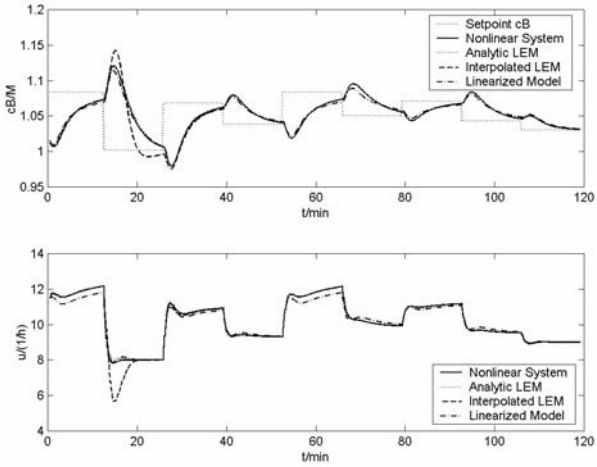


Fig. 4: non-minimum-phase responses, $1M < c_B < 1.09M$

Table 2: minimum-phase and non-minimum-phase performance, $1M < c_B < 1.09M$

	NL	ANA	INT	LIN
Iterations, min.-phase	1055	1055	1053	1030
J_{total}, m.p.	131.82	131.73	131.99	$3.33 \cdot 10^4$
Iterations, n.min.phase	831	831	858	1047
J_{total}, n.m.p.	183.37	183.69	177.48	$7.01 \cdot 10^6$

Finally, a controller designed especially for the minimum-phase region will be tested in order to see how the performance can be improved when the right tuning parameters are applied. Table 3 summarizes the new controller parameters.

Table 3: minimum-phase LLT parameters

Prediction / Control horizon (P / M)	3 / 1
Output / Input variable weight (Γ / Λ)	0.7 / 0.02
Sampling time (Ts)	0.15 min

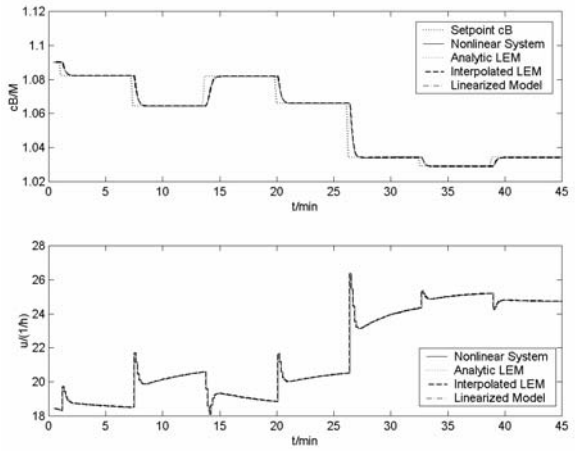


Fig. 5: minimum-phase responses, faster controller

It is evident that this controller performs much better than the one designed for non-minimum-phase behaviour, which makes it the preferable option for operation near the maximum yield of c_B=1.09M. Even the linear MPC controller shows almost exactly the same behaviour as the controller with the true nonlinear model. However, it has to be assured that the system does not trespass into the non-minimum-phase region, since in this case it would cause excessive inverse responses. This can be achieved by defining a 'target' for the manipulated variable in the minimum-phase domain within the LLT algorithm.

Table 4: minimum-phase performance, $1M < c_B < 1.09M$

	NL	ANA	INT	LIN
Iterations	810	810	810	814
J_{total}	4.2630	4.2630	4.2324	6.3172

As expected, the LEM models perform well in all the shown cases with only slight deviations from the original nonlinear model.

6. CONCLUSIONS AND OUTLOOK

In this paper the proposal for a combined use of the nonlinear modelling and identification technique LEM and the nonlinear predictive control algorithm LLT was made and tested with a SISO example. Analytical considerations and numerical results suggest a good applicability of this combination to lower order SISO systems, performing fairly well where purely linear methods fail completely. Losses in control performance are mainly due to the inevitable deviation of the identified and interpolated equilibrium manifold and the associated dynamics from the true nonlinear system's dynamics. The necessary iterations for convergence of the algorithm, as well as the accumulated values of the objective function are in all cases in the same scale, they can vary slightly because of the described deviations. It is clear that large efforts have to be made to obtain good approximations for the equilibrium manifold and the associated dynamics. Taking this into account, it can be concluded that further investigation of the demonstrated combination appears promising. Future work will be done on the testing of the technique with MIMO models and models of higher order, the goal being the proof of applicability to real industrial processes. The second possibility of implementation mentioned in section 4, the alteration of the LLT algorithm to be directly suitable for the LEM structure, will also be explored.

ACKNOWLEDGMENTS

The authors thank PETROBRAS and FINEP for their financial support. The first author thanks the HEINRICH-BOELL-FOUNDATION, the third author thanks the GERMAN ACADEMIC EXCHANGE SERVICE (DAAD).

REFERENCES

Duraiski, R.. G.; (2001a) Controle Preditivo Não Linear Utilizando Linearizações ao Longo da Trajetória; M. Sc. Thesis, UFRGS, Brazil.

Engell, S. and K. U. Klatt (1993). Nonlinear Control of a Non-Minimum-Phase CSTR. *Proceedings of the American Control Conference*, pp. 2041-2045, San Francisco, California, June 1993.

Bolognese Fernandes, P. (2005). *The input-parameterized linearization around the equilibrium manifold approach to modeling and identification*. Phd Thesis, University of Dortmund (to be published).

Bolognese Fernandes, P., S. Engell (2005). Continuous Nonlinear SISO System Identification using Parameterized Linearization Families. *Proc. of the XVI IFAC World Congress, Prague, Tchech Republic*.

Johansen, T.A. and R. Murray-Smith (1997). The Operating Regime Approach to Nonlinear Modelling and Control. *Multiple Model Approaches to Nonlinear Modelling and Control* (R. Murray-Smith and T.A. Johansen. (Eds)), pp. 3-72. Taylor & Francis, London.

Nelles, O. (1997). LOLIMOT- Lokale, lineare Modelle zur Identifikation nicht-linearer, dynamischer Systeme. *Automatisierungstechnik*, **4**, 163-174.

Trierweiler, J. O., Farina, L. A. (2003), RPN tuning strategy for model predictive control. *Journal of Process Control*, v. 13, p. 591-598, 2003.

Wang, J. and J. W. Rugh (1987). Parameterized Linear Systems and Linearization Families for Nonlinear Systems. *IEEE Transactions on Circuits and Systems*, **34**, 650-657.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

MULTI MODEL APPROACH TO MULTIVARIABLE LOW ORDER STRUCTURED-CONTROLLER DESIGN

M. Escobar, J.O. Trierweiler

Laboratory of Process Control and Integration (LACIP),
Group of Integration, Modelling, Simulation, Control and Optimization of Processes (GIMSCOP)
Department of Chemical Engineering, Federal University of Rio Grande do Sul.
E-mail: escobar@enq.ufrgs.br/ jorge@enq.ufrgs.br

Abstract: The method presented here offers an effective and time saving tool for robust low order multivariable controller design. The relation between controller complexity and closed loop performance can easily be evaluated. The method consists of five steps: 1. A desired behavior of the closed loop system is specified. Considering the nonminimum phase part of the process model the closed loop attainable performance is determined. 2. The process model and the attainable performance are scaled by the RPN-scaling procedure. 3. This defines an "ideal" scaled controller, which is usually too complex to be realized. 4. The frequency response of the ideal scaled compensator is approximated by a simpler one with structure and order chosen by the user. 5. Since the approximation in frequency response is performed with the scaled system, it is necessary to return to the original system's units. This procedure can be implemented using a multi-model approach, what increase the robustness of synthesized controller. *Copyright © 2006 IFAC*

Keywords: multivariable project design, frequency domain, multi models, low order controllers

1. INTRODUCTION

The increase of the complexity of the modern plants promoted an increase of the interaction among the variables of the process increasing the number of necessary control loops to maintain the conditions of desired operations and the quality of the obtained products.

Restrictions on the feedback compensator structure are often encountered in chemical plants, when several control stations are provided only with local measurements. Such decentralized information structures result in block-diagonal compensator matrices. Decentralized controllers are also attractive because the information about the feedback is concentrated in the diagonal blocks. This means they are easier to understand and to put into operation and more easily made failure tolerant than general multivariable control systems.

Even for plants with strong interaction, a decentralized controller can be attractive from a performance viewpoint, since depending on the disturbance direction and the model uncertainty can exhibit a better performance to disturbance rejection than a centralized one. Usually to improve the performance to set-point change is interesting to include some degree of decoupling between the main interacting loops. All these situations imply and require a structured controller.

The controller order is another point to be considered, since it is strongly related to implementation easiness. Low order controllers (e.g. PID) are much simple and easy to implement and maintain in industrial control systems (DCS) than a high order state space centralized controller

Due the uncertainties associated to the model and the need of working at different operating points (OPs) with different dynamic behaviours, it is required that the controller must exhibit certain robustness degree. Usually, it is common to design a controller for each OP separately, or to tune for the worst case and to test it to the other OPs, which in general does not produce the best achievable result.

The design of robust decentralized controllers remains a demanding problem; standard methods for robust design cannot be used for structured compensators. The standard techniques for robust full controller design (e.g., Hinf,μ) cannot be directly applied to design a robust structured low order controller. In this paper it is proposed a new methodology to solve this problem, which conciliates design simplicity with DCS implementation easiness.

The proposed approach is based on the multi-model system representation and on the frequency domain approximation. The basic idea of this approach is to approximate the high order full controller that achieves the desired attainable closed loop response by a low order structured controller.

2. METODOLOGY

Consider that the block diagram in figure 1 requires the closed-loop behavior to be a predetermined transfer function chosen $T_0(s)$. Given the model G, mathematically the requirement to make the process closed-loop exactly equal to $T_0(s)$ is satisfied if, and only if

$$C(s) = G^{-1}(s)\left(T_0^{-1}(s) - I\right)^{-1} \quad (1)$$

$C(s)$ is the "ideal" controller that can be a high order controller, since no restriction is used in (1). Although the ideal controller is usually not realizable, it provides the designer with the necessary information about the desired controller frequency response. The basic idea is to approximate in frequency domain the ideal controller (C_0) by a low order structured controller. Since we want that the approximated controller performs so close as possible to the ideal one, it is better to approximated the closed-loop frequency response, i.e. $\Delta T = T - T_0$, instead of approximating $\Delta C = C - C_0$ directly.

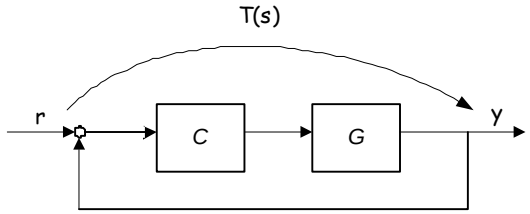


Fig. 1: Standard Feedback Configuration.

In this paper, the proposed methodology will use a two degree-of-freedom control loop configuration shown in Figure 2, where the controller C is separated into four blocks: C_{PI}, C_{PV}, C_{SP} and C_F.

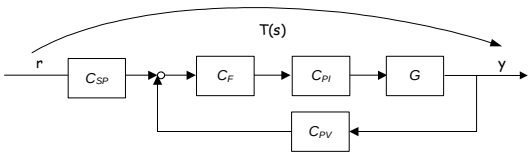


Fig.2: Two degree-of-freedom feedback control

The C_{PI} block is a PI controller whose structure is always fixed and always given by (2), whilst C_{SP} and C_{PV} are dependent on the PID controller parameterization (e.g., series, parallel, ISA-form).

$$C_{PI} = K_C\left(1 + \frac{1}{T_I s}\right) \quad (2)$$

As discussed by Faccin and Trierweiler (2004), the advantage to use the 2DOF control configuration is threefold: (a) It divides a typical nonconvex optimization problem (when the standard configuration is used) into two convex problems. (b) It consists in a common base, in which all possible industrial PID parameterization can be converted. In Faccin (2004) it is shown this conversion for several industrial PID parameterizations. (c) The controller order can be easily increased and implemented in modern DCS. For example, process filters for noise averting can be synthezed and incorporated into C_{PV}. The conversion of different PID or other control forms are very simple, since the algorithm relates the control action (? u) with the variable process (? y) and the variable of reference (? r), i.e., $\Delta u = M(s)\Delta y + N(s)\Delta r$

$$C_{PV}(s) = C_{PI}^{-1}(s)M(s) \;, C_{SP}(s) = -C_{PI}^{-1}(s)N(s) \quad (3)$$

When more than a PID is desired to control the system, it can be done using the block C_F. This block is also diagonal with elements given by the orthogonal serie:

$$C_F(s) = \sum_{k=1}^{n=ordem} (T_F)_k \, j_k(s)$$
$$j_k(s) = j_1(s)\prod_{i=1}^{k}\frac{s-l}{s+l} \qquad j_1(s) = \frac{\sqrt{2l}}{s+l} \quad (4)$$

Where l is the frequency point where the fit of the curve $\Delta T / s$ is more precarious and the coefficients T_F are the decision variables of the optimization problem.

2.1 Optimization Problem

After algebraic manipulation

$$\Delta T(s) = T - T_0 = S(s)[G(s)C_{PI}(s)\left(C_{SP}(s) - C_{PV}(s)T_0(s)\right) - T_0(s)] \quad (5)$$

If $S \cong S_0$ ($S_0 \cong I - T_0$), and C_{SP}, C_{PV}, C_F are diagonal blocks ,the problem can be seen as that the j-th column of ΔT is only influenced by the j-th column of ΔC, so the problem is independent in the column, and can be solved separately. The objective function (6) consists of the Euclidian norm of the step response of the transfer function ΔT on frequency domain for N frequency points.

$$FO = \min_{PID} \sum_{s=jw_1}^{s=jw_N} \left\| T(s) - T_0(s) \right\|^2 \quad (6)$$

The problem is solved in an iterative and sequential way. In the initiation, C_{SP}, C_{PV}, $C_F = I$, and the parameters from the PI block is determined. In agreement with the selected algorithm, the C_{SP} and C_{PV} blocks are determined fixing the PI. A new iteration starts always fixing the knowns parameters from then previous iteration. This procedure is executed until that the stop approach is satisfied.

When the PID is determined in this sequential and iterative method, it is fixed and the C_F block is solved, to determining T_F.

These problems can be formulated as a least squares problem for each model. If it is desired a control design using the multi model representation, each model generates the same kind of problem. So, the whole problem can be solved as a weighted least squares problem, and these weights are selected by the project designer.

All this procedure is performed in a very fast way, but it is just an approximation because if difference ΔC is not sufficiently small, S deviates from S_0, and the computation error of the column-by-column optimization may be large. The controller can be improved by a non-linear optimization, which considers the closed-loop resulting directly.

The cost function in the non-linear optimization is

$$FO_{Global} = \sum_{k=1}^{ni}\sum_{i=1}^{no}\sum_{l=1}^{N}\left\|\frac{\Delta T_{ik}(j.w_l)}{j.w_l}\right\|^2 \qquad (7)$$

Where *no* and *ni* are the number of outputs and inputs of the system respectively and ***N*** is the number of frequencies in the frequency vector. The controller from the column-by-column optimization is used as a starting point for the non-linear optimization. The following equation can be formulated

$$\begin{aligned} &\min g \\ &g, x \in \Re \\ &subject \quad to: \quad FO_i(x) - w_i g \le 0 \end{aligned} \qquad (8)$$

Where g is an auxiliary variable and w_i is the weight for each FO_i calculated for the model *i*.

2.2 General procedure

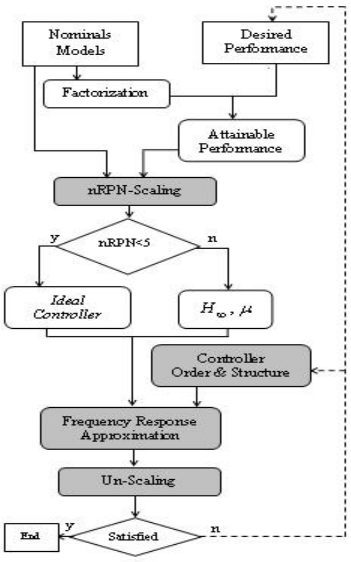


Fig.3: General Procedure.

Fig. 3 shows the general procedure. The desired performance is established to each output through specifications in the time domain (rise time and maximum % of overshoot) that are mapped into a second order transfer function. The models must be factorized to insert some restrictions in the performance like RHP - zeros and -poles and time delay to maintain the internal stability of the feedback system (Trierweiler *et al.*, 2000).

The RPN (robust performance number) (Trierweiler and Engell, 1997) and nRPN (Farenzena and Trierweiler, 2004) when it is working with multi-model are calculated. Small values indicate a good performance using this method. Diagonal matrices that minimize the condition number of the system at the frequency that RPN assumes its maximal value, are used to scale the models. With the controller structure and order, the frequency response approximation is used to calculate the blocks (C_{PI}, C_{PV}, C_{SP} and C_F). The controller is returned to the original units and if the simulation shows a poor performance, the desired performance or its structure and order can be modified.

3. CASE STUDY

The case study consists of a six spherical tank plant. The unit is composed by six level tanks interacting to each other, two control valves, one recycle tank and one pump. The objective is to control the levels h_3 and h_6, manipulating the two valves defining the flow rates F_1 and F_2.

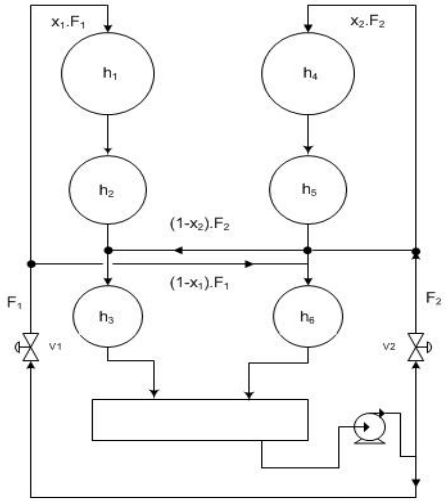


Fig. 4: The six spherical tanks process.

For this process the simplified model expressed by (9) was developed. Where g is the constant of gravity, a_i is the section area of the discharge pipe from the tank *i*, and D_i is the diameter of the tank *i*.

After linearizing the model and transforming into Laplace domain at the operating point (h_{1s},h_{2s},h_{3s},h_{4s}) the corresponding transfer matrix is given by (11).

$$
\begin{aligned}
A_1\frac{dh_1}{dt} &= f_1 = x_1\cdot F_1 - R_1\sqrt{h_1}\\
A_2\frac{dh_2}{dt} &= f_2 = R_1\sqrt{h_1} - R_2\sqrt{h_2}\\
A_3\frac{dh_3}{dt} &= f_3 = (1-x_2)\cdot F_2 + R_2\sqrt{h_2} - R_3\sqrt{h_3}\\
A_4\frac{dh_4}{dt} &= f_4 = x_2\cdot F_2 - R_4\sqrt{h_4}\\
A_5\frac{dh_5}{dt} &= f_5 = R_4\sqrt{h_4} - R_5\sqrt{h_5}\\
A_6\frac{dh_6}{dt} &= f_6 = (1-x_1)\cdot F_1 + R_5\sqrt{h_5} - R_6\sqrt{h_6}
\end{aligned}
\qquad (9)
$$

where

$$A_i = p\left(D_i h_i - h_i^2\right) \quad \text{and} \quad R_i = a_i\sqrt{2g} \qquad (10)$$

$$
\begin{bmatrix} h_3(s) \\ h_6(s) \end{bmatrix} =
\begin{bmatrix}
\dfrac{x_1 c_1 e^{-0.9s}}{\prod_{i=1}^{3}(t_i s+1)} & \dfrac{(1-x_2)c_1 e^{-0.3s}}{(t_3 s+1)} \\
\dfrac{(1-x_1)c_2 e^{-0.3s}}{(t_6 s+1)} & \dfrac{x_2 c_2 e^{-0.9s}}{\prod_{i=4}^{6}(t_i s+1)}
\end{bmatrix}
\begin{bmatrix} F_1(s) \\ F_2(s) \end{bmatrix} \qquad (11)
$$

with

$$c_1 = \frac{2\sqrt{h}_{3s}}{R_3},\ c_2 = \frac{2\sqrt{h}_{6s}}{R_6} \qquad (12)$$

$$t_i = \frac{2A_i\sqrt{h_{is}}}{R_i} \qquad (13)$$

When the sum of x_1 and x_2 is greater than one, the system has a RHP-zero. If $x_1+x_2=1$, the system has a zero located at the origin and as greater goes this sum, the zero is moved away of the origin along the positive axis.

Table 1: Process Parameters.

Parameters	Value
D_1 D_4 [cm]	35
D_2 D_5 [cm]	30
D_3 D_6 [cm]	25
R_1 R_4 [$cm^{2.5}min^{-1}$]	1690
R_2 R_5 [$cm^{2.5}min^{-1}$]	1830
R_3 R_6 [$cm^{2.5}min^{-1}$]	2000

Table 2: Operating Points.

Variables	OP 1	OP 2	OP 3	OP 4
h_3 [cm]	4.8400	17.0156	8.4100	11.7306
h_6 [cm]	3.2400	11.3906	8.1225	5.4056
F_1[L/min]	4	7.5	4	7.5
F_2[L/min]	4	7.5	7.5	4
x_1,x_2	0.7,0.6	0.7,0.6	0.7,0.6	0.7,0.6
RHP-zero	1.0246	0.1915	0.3818	0.3158
y_z (output zero direction)	$\begin{bmatrix}0.68\\-0.79\end{bmatrix}$	$\begin{bmatrix}0.58\\-0.81\end{bmatrix}$	$\begin{bmatrix}0.52\\-0.85\end{bmatrix}$	$\begin{bmatrix}0.69\\-0.72\end{bmatrix}$
u_z (input zero direction)	$\begin{bmatrix}-0.58\\0.59\end{bmatrix}$	$\begin{bmatrix}-0.77\\0.63\end{bmatrix}$	$\begin{bmatrix}-0.65\\0.75\end{bmatrix}$	$\begin{bmatrix}-0.85\\0.51\end{bmatrix}$

Table 1 shows the parameters used in the model, while Table 2 summarizes the steady-state and operating conditions of the studied OPs.

The four OPs have different dynamics and RHP-zero. The model 2 (M_2) is considered as the nominal model and it has the slowest dynamic. The model 1 (M_1) is the critical point OP, since the dynamic differs on most.

This process is difficult to control due to the time delay and the RHP-zero (which limit the achievable closed loop performance making the response slower). Figure 5 shows the step response to the models. The RGA (*Relative Gain Array*) in the channel (1,1) from all models is equal to 1.4 indicating some interaction and the correct choice to decentralized project designs.

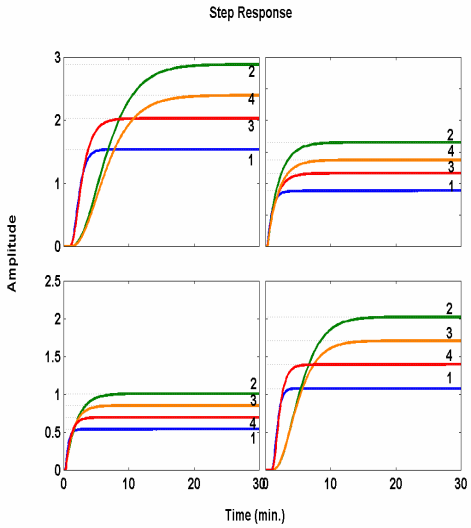


Fig. 5: Step Response of the Models.

4. RESULTS

Table 3 shows the desired performance applied to design the controllers. In the frequency domain, the RHP-zeros (and pure time delays) constraint the bandwidth up to which effective disturbance attenuation is possible. The largest bandwidth is determined by the RHP-zero closest to the origin (0.1915 in this case). The performance was established considering the limitation imposed by this zero, making the performance as fast as it is possible.

Table 3: Desired Performance (T_d).

Characteristic	Td
Rise time [y1,y2] (min.)	10,7
Overshoot %	10,10

The model must be factorized since it has an RHP-zero and time delay. The zero with the same output direction and the factorable time delay must be present in the closed loop transfer function to keep the internal stability of the feedback system. The time delay that cannot be factored out is approximated by Padé. For a second order Padé

approximation, the zero is moved to 0.1731 indicating that the nonfactorable time delay have a unfavorable, but not significant, influence on the system controllability.

To analyse the controller's performance it was used a set point change (servo problem) in opposite directions, which is the worst situation than the controller can face according with the output zero direction as shown in Table 2. Similarly it was used as regulatory problem the unitary change to a at u_1 and -a at u_2 according the input zero direction.

It was designed three full controllers to the nominal model with the desired performance from Table 3. The simulation is presented in Figure 6. PID +F2 is used to indicate a PID with a second order filter.

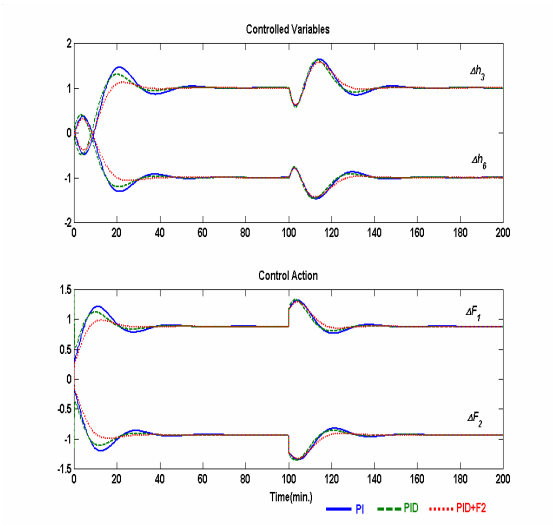


Fig. 6 Servo (r=[1 -1]) and regulatory (d=0.4.[1 -1]) response to the centralized controllers PI, PID,PID+F2 for the nominal model using the desired performance Td.

The figure shows how the RHP-zero can limit the speed of the control loop. Even the PI controller can present less overshoot making the controller slower. On the other hand, the increase of the order has a stabilizing effect on the performance. It allows doing the controller faster without harming its performance

In figure 7, three decentralized controllers were designed to the nominal model using the same performance. In this case the order increase has less effect because the PI controller shows a slow, but satisfactory, performance. Comparing these results with figure 6, it can be concluded that due the interaction (as indicated by a RGA analysis) the decentralized controllers with the same order are slower and presents a larger interaction, even though present good results.

The best controllers (decentralized/full) designed to the nominal model was simulated using the model linearized at the OP_1 (the smallest gain) and it indicated a poor performance (very slow).

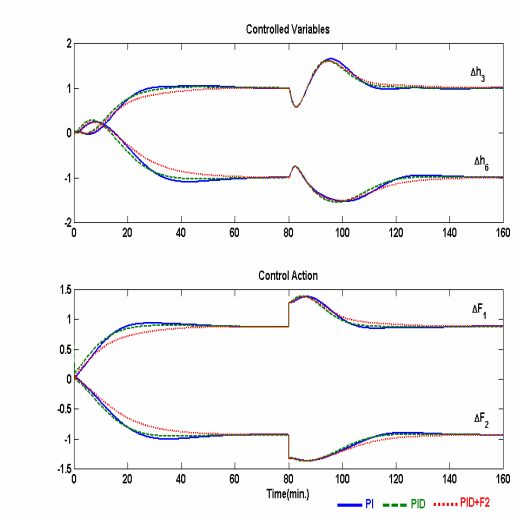


Fig. 7 Servo (r=[1 -1]) and regulatory (d=0.4.[1 -1]) response to the decentralized controllers PI, PID, PID+F2 to the nominal model using the desired performance Td.

Figure 8 shows the performance to the nominal model using all four models with the same weight to design three full controllers.

This is the problem choosing the critical case (high gain) to design only one controller to the whole process. The performance in the region of low gain can be made very slow, but not use the critical point can affect the stability making the process unstable in the high gain region.

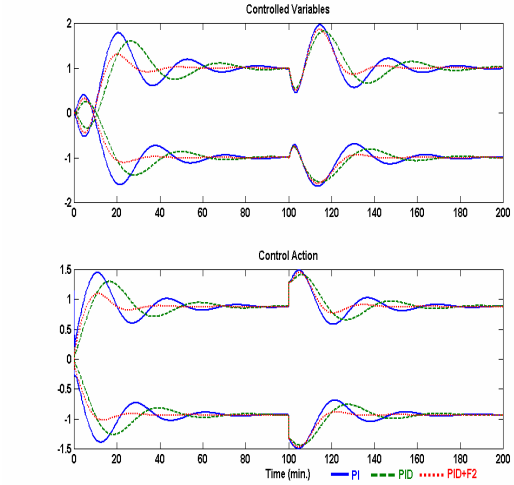


Fig. 8 Servo and regulatory response of the nominal model with the controller designed with all four models.

The results show that the increase of the order has more stabilizing effect when the controller is designed for multi-model case than the nominal model one. Also, it is important to realize that increasing the order of the controller using the polytope make the performance slow, but it is more significant in the region of high gain (M_2).

It was selected two full controllers that shown the best performance using the nominal model (C_n) and the polytope (C_p) respectively. This controller in both

cases were a PID with a second order filter controllers.

Figure 9 compares the results to a servo response of C_n and C_p. The polytope controller is faster in all OPs, and it shows a performance as good as the nominal controller even in the nominal OP.

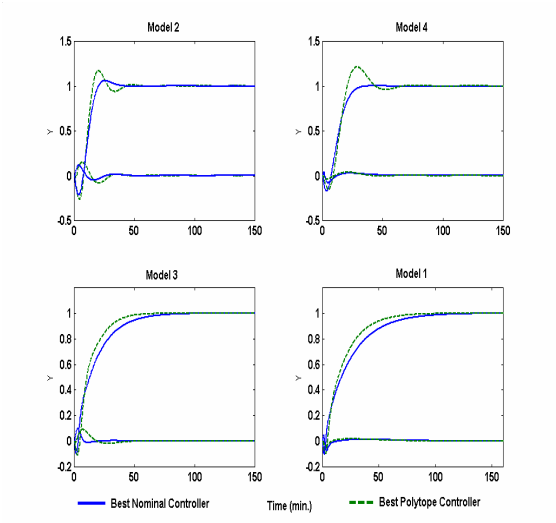


Fig. 9: Step response of C_n and Cp to all OPs.

Both controllers were simulated with the nonlinear model. The simulation starts in the OP2 and the process is changed to the OP4, OP1, OP3 successively until the time of 400 minutes where a set point change in opposites directions (the worst situation that the controller can face it) and at the time 500 the values of x1 and x2 are inverted (x1=0.6 and x2=0.7) and so the process return to the OP2. The simulation results are shown in figure 10 demonstrated that the performance of the polytope controller is better than the nominal controller to the most changes because the first one consider all the OPs into the design. Also, choosing the weights allows improving the performance in a given region.

In table 4 are presented the parameters of C_n and C_p. The equation used to derivative action is given by

$$C_{PV}(s) = C_{SP}(s) = \frac{T_D s + 1}{a T_D s + 1} \; (series) \qquad (13)$$

Table 4: Controller's parameters.

	Controller	C_n/C_p	
Parameter	i \ j	1	2
K_P	1	0.178/0.022	-0.226/-0.194
	2	-0.159/-0.073	0.112/0.038
T_I	1	3.130/0.339	6.126/4.362
	2	6.164/2.502	1.5926/0.463
T_D	1,2	0.993/2.554	0.994/0.990
T_F	1,2	$\begin{bmatrix}0.93\\0.41\end{bmatrix}/\begin{bmatrix}0.93\\0.49\end{bmatrix}$	$\begin{bmatrix}0.82\\0.29\end{bmatrix}/\begin{bmatrix}0.97\\0.49\end{bmatrix}$
l	1,2	1.054/0.873	

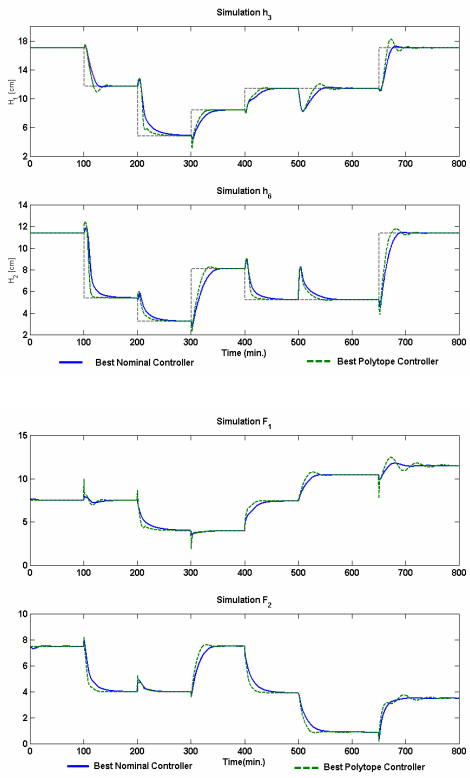


Fig. 10: Nonlinear simulation with C_n e C_p.

5. CONCLUSIONS

It was presented a fast and efficient method to design and to evaluate alternative multivariable control structures. The methodology is very flexible, allowing its use even in complex process with RHP-zero and time delay. Moreover, the controller designed with all OPs can provide a trade-off between the performance and robustness.

REFERENCES

Faccin, F. (2004). Uma abordagem inovadora para o projeto de controladores PID (In Portuguese). Master Thesis. Federal University of Rio Grande do Sul, Chemical Eng. Department, Brazil.

Faccin &Trierweiler (2004), A Novel Tool for multi-model PID Controller Design. . In: Dycops- (7th IFAC Symposium on Dynamics and Control of Process Systems), 2004, Boston. CD: Dycops 2004, 2004. v. 1. p. 176-186.

Farenzena, M.; Trierweiler, J.O. (2004), System Nonlinearity Measurement Based on the Rpn Concept. In: Dycops2004 (7th IFAC Symposium on Dynamics and Control of Process Systems), 2004, Boston. CD DYCOPS 2004, 2004. v. 1. p. 181-191

Trierweiler, J.O., R. Müller and S. Engell (2000), Multivariable Low Order Structured Controller design by frequency response approximation. In: *Brazilian Journal of Chemical Engineering* Vol.17, No 04-07, pp 793-807, Brazil.

Keynote 11

On Data Processing and Reconciliation: Trends and the Impact of Technology

J.A. Romagnoli, P.A. Rolandi, Y.Y. Joe, and K.V. Ling
Louisiana State University

Keynote 12

Iterative Learning Control Applied to Batch Processes

J. H. Lee and K. S. Lee
Georgia Institute of Technology

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ON DATA PROCESSING AND RECONCILIATION: TRENDS AND THE IMPACT OF TECHNOLOGY

J.A. Romagnoli[a,1], P.A. Rolandi [b], Y.Y. Joe [c], Z.Q. Ding [c], K.V. Ling [d]

[a] *Department of Chemical Engineering, Louisiana State University, Baton Rouge, LA 70803, USA*
[b] *Process Systems Enterprise Limited, UK*
[c] *Singapore Institute of Manufacturing Technology, Singapore*
[d] *Nanyang Technological University, Singapore*

Abstract: The developments in technologies are expanding the boundaries and broadening the domain of what is technically and economically feasible to achieve in the application of data reconciliation activities in manufacturing plants. They also, naturally, incorporate additional issues and open the opportunities for new research activities. For example, recent developments on model-centric technologies to support plant operations based on advanced process modelling technologies opened the opportunities for performing large-scale parameter estimation – data reconciliation applications in complex dynamic industrial environments. On the other hand, new sensor technologies are becoming available based on recent advances in microprocessor-based instrumentation and digital communications. They provide opportunities for the realization of novel sensor network architectures towards a truly distributed environment for data processing and reconciliation. In this presentation we will discuss current research activities combining efforts in these areas towards the future operation of manufacturing plants.

Keywords: model-centric technologies, support of process operations, data processing, data reconciliation, sensor network, intelligent sensor.

1. INTRODUCTION

Rational use of the large volume of data generated by manufacturing plants requires the application of suitable techniques to improve their accuracy and to extract useful information about the operational status of the process. A number of technologies (data reconciliation, trend analysis, fault diagnosis, etc) have been the subject of active research during last 20 years with important advances. Among these data processing strategies, Data Reconciliation (DR) is one of the most typical approaches to obtain a consistent set of data.

Recent technological developments are expected to have a strong impact, broadening the domain of applications of data processing and reconciliation activities in manufacturing plants. In this presentation we will restrict the scope and will focus on the impact of two key technologies: the recent progress made towards model-centric approaches for support of manufacturing activities and the developments on sensor/sensor networks technologies expanding the capabilities of existing sensors.

During the last decade, general-purpose modelling tools have reached a level of maturity that allows the definition and solution of model-based problems of unprecedented complexity. Nowadays, state-of-the-art modelling, simulation and optimisation environments (MSOEs) have expanded their languages to account not only for the definition and solution of dynamic simulation activities, but also the declaration of dynamic optimisation and parameter estimation/data reconciliation activities with comparable generality and flexibility. However, while commercial and academic modelling technologies have largely engaged in developing frameworks and methodologies for tackling the model development process, complementary

[1] Corresponding author, e-mail: jose@lsu.edu

frameworks and mechanisms to help conceptualising and implementing "models" of process-engineering problems to support plant operation remain virtually unexplored. Progress in this direction, unquestionably, will expand the scope of what is technically feasible to achieve in the application of data processing and reconciliation activities in manufacturing plants. This will provide opportunities for performing large-scale applications within complex dynamic industrial environments and, additionally, integrating these capabilities with other activities for support of process operations into a single and consistent model-centric framework. In this work we will present a series of initiatives towards this vision.

On the other hand, recently sensors have received greater attention than in the past. This is due to: greater demands placed on all aspects of plant operation and improvements in technology. In terms of plant operation, competition has resulted in higher product quality and plant efficiency. Safety standards are constantly rising and measurements are the primary means of identifying potentially hazardous circumstances. In terms of improvements in technology, new sensor technologies are becoming available (extending the properties that can be measured, the environment in which they can be sampled). Microprocessors-based instrumentation and digital communications are having profound effect on the capability and/or functionalities of the sensor. These developments provide the opportunity for the realization of federated sensor network architectures towards a truly distributed environment for plant operation. In this presentation we will discuss a new conceptual model for the next generation of sensor devices, which incorporates activities such as DR at the sensor level thus improving diagnosis/classification and reducing the computational load at the controller levels. This type of architecture encompasses the extra capabilities required for the next generation of sensors and sensor networks and accommodates the additional demands required for modern manufacturing.

2. MODEL-CENTRIC TECHNOLOGIES AND DATA RECONCILIATION/ PARAMETER ESTIMATION

2.1 Background

Throughout the last decades, the computer-aided process engineering (CAPE) community made considerable progress in two strategic areas: the technical development and commercialisation of general-purpose modelling, simulation and optimisation environments; and the standardisation of open interface specifications for component-based process simulation. High-level equation-oriented declarative modelling languages have gained increased acceptance as the most appropriate framework to tackle the modelling process when full control over the scope and detail of the process model is required (Foss et al., 1998) because they provide the modeller with a series of sophisticated tools and mechanisms that contribute enormously to increase the efficiency of the modelling process.

An important advantage of equation-oriented modelling languages is the intrinsic independence between mathematical models and solution methods. By segregating the mathematical definition of any given model from structural, symbolic or numerical solution algorithms, a single model description can be used to accommodate for a potentially large number of complementary activities. Another major advance was the creation of high-level declarative languages to describe a wide range of advanced model-based problems such as dynamic optimisation and parameter estimation with a degree of generality and flexibility comparable to existing dynamic simulation languages. These days, commercial modelling languages have evolved into multi-purpose process-engineering modelling tools which we shall denote as "modelling, simulation and optimisation environments" (MSOEs).

As the CAPE community continues developing and validating individual process models, the incentive behind developing and implementing model-based technologies grows. In the mid 1990s, developers and end-users were confronted with the reality that the accessibility and usability of model descriptions embedded within modelling environments was very limited. To address this problem, the CAPE community initiated the CAPE-OPEN (CO) and Global CAPE-OPEN (GCO) projects. CO focussed on providing standard mechanisms to support a two-fold long term vision according to which: process modelling components (PMCs) built or wrapped upon the standard could be incorporated into process modelling environments (PMEs) straightforwardly; and model descriptions declared within PMEs supporting the standard would be accessible to external modelling tools.. This way, developers would be able to assemble software components from heterogeneous sources to solve complex model-based problems. The GCO consortium continued revising and updating existing standards and creating new ones for technologies beyond modelling and simulation. Within the scope of this work, the CO standards will be used as an enabling paradigm to support the creation of the advanced framework proposed later in this paper and as a point of reference to inspire some of its most innovative features.

2.2 *Model-Centric Framework for Support of Manufacturing Activities*

Following the previous discussion, it is clear that the creation of a model-centric framework that supports the definition of rigorous model-based activities and promotes the transfer of knowledge between complementary model-based software applications will extend the viability of model-centric technologies. In a series of papers, Rolandi and Romagnoli (2006a) presented a framework of such

characteristics that enables the definition and implementation of model-based process-engineering problems typical of industrial environments.

The conceptual core of the framework was conceived according to the following vision. The framework was tailored to use mathematical models of process systems derived on the basis of mechanistic descriptions of natural phenomena. Although in principle the framework is applicable to a widespread of process systems, plant-wide models of industrial manufacturing plants were the main motivation of this work. Of course, as a result of rigorous mechanistic modelling of plant-wide industrial systems, the framework was tailored to deal with complex large-scale process models. In this work, the idea of modelling for multiple purposes was pursued, so that several model-based components were able to use a single fundamental model of the process to solve a widespread range of problems, implementing the notion of a model-centric framework. This integration crystallised the vision of a consistent solution of process-engineering problems, seeding synergistic interactions across model-based activities due to a consistent model formulation among the software components. Last but not least, the framework addressed a series of problems of relevance to industrial manufacturing operations, such as: model-based process simulation and optimisation, parameter estimation, data reconciliation and advanced process control. It is worth to emphasising, though, that the estimation/reconciliation component of interest to this work is just one of the modules of the entire framework discussed in Rolandi and Romagnoli (2006a).

Figure 1 provides a conceptual representation of how the different model-based components of the proposed model-centric framework are expected to support the operation of an industrial process system. As expected, the data pre-processing environment precedes all modules that make use of raw plant data, since it is imperative to obtain a consistent set of data by reconstruction of the process trajectories for the robust execution of any subsequent tasks. The estimation environment incorporates dynamic parameter estimation and dynamic data reconciliation activities, which make use of consistent data sets for the estimation of process operating parameters and evaluation of process measurement biases. The information gained from these activities is presented to the decision-makers, who then have a chance to make informed decisions on issues such as process instrumentation and equipment maintenance and inventory analysis. Consistent data sets are also provided to the simulation environment, which extracts meaningful information from past operating conditions. These process analysis activities are complemented by process improvement tasks such as process optimisation, transition planning studies and constrained real-time model-based optimisation and control, in a sequence of execution that that reshapes raw plant data into useful process knowledge and, hence, levers the chances for informed operative and supervisory decisions (see Rolandi and Romagnoli, 2006a).

In this work, we suggest extending the software architecture proposed by the CO standards (CAPE-OPEN Consortium, 2000) by introducing a new software object: the Problem Definition Environment (PDE). As sketched in Figure 2, the PDE manages the definition of advanced model-based problems by interacting with both the Process Modelling Executive (PMEs) and the user, while the PME performs the corresponding model-based activity by coordinating the calls to several Process Modelling Components (PMCs). These PMCs contain the mathematical description of the process model, and they also provide other services such as physical property calculations and numerical solution algorithms (Braunschweig et al., 2000). While the standardisation of open interfaces of the PME and PMCs has been the focus of the CO/GCO projects, the communication between the PDE and other elements of the architecture is regulated by a series of mechanisms intrinsic to the framework described in this work. These mechanisms entail the manipulation of the so-called "Data Model Templates" (DMTs) and "Data Model Definitions" (DMDs) (Rolandi and Romagnoli (2006a).

In the software architecture shown schematically in Figure 2, the MSOE (a PME and several PMCs) can be seen as a software tool for managing the development of mathematical models and, ultimately, coordinating the execution of the model-based activity, i.e. the MSOE is essentially a model builder and activity executive. On the other hand, the PDE is conceived as a software tool for supporting the definition of model-based problems, i.e. how to use plant data and the process model in the context of realistic process-engineering problems, which requires additional skills and expertise; in other words, the PDE is basically a problem builder.

2.3 *A Framework for Joint Parameter Estimation and Data Reconciliation*

As discussed above, a novel paradigm for the definition of rigorous model-based problems is now possible through the introduction of the PDE. The PDE manipulates the so-called Data Model Templates (DMTs) and Data Model Definitions (DMDs). In this section, we will briefly discuss the structure and purpose of the two data models relevant for the definition of hybrid data-driven/model-based parameter estimation data reconciliation problems. These data structures are the so-called Process Data Object data model (PDO) and the Dynamic Estimation Problem data model (DEP).

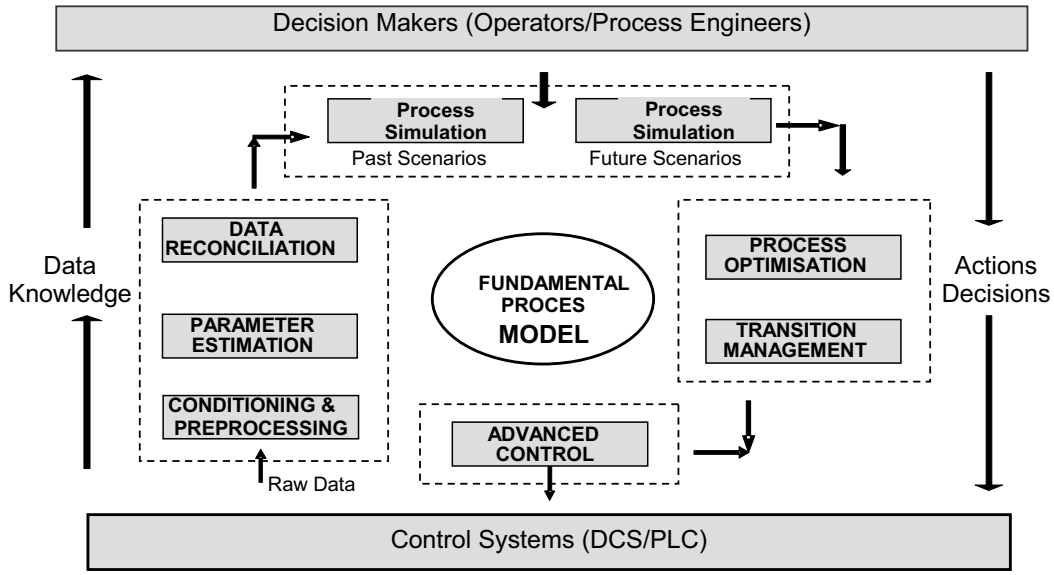


Fig. 1. The conceptual definition of the integrated framework for model-centric support of process operations.

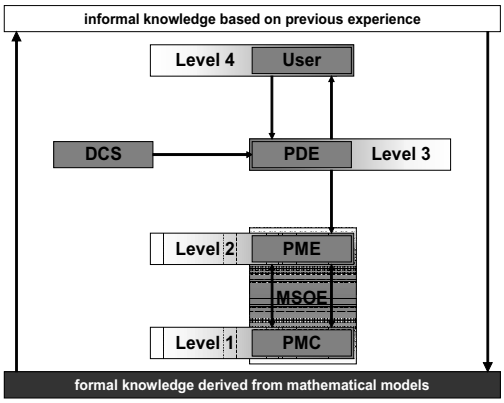


Fig. 2. The architecture of the framework.

- *DEP model:* Contains data determining the structure of a general dynamic estimation problem; for example, a valid instance of this data model links explicitly model variables to control/measured (input/output) process variables and corresponding process instrumentation, and determines the form of the objective function (estimator).

- *PDO model:* Contains data representing raw experimental process data in a form fit for hybrid empirical/mechanistic modelling; for instance, a valid instance of this data model associates the reconstructed dynamic trajectories of the input/output process variables (retrieved from process instrumentation) to model variables.

The distinction between these two structures is substantiated by the fact that it is not convenient to associate an experimental data set with a given estimation case-study, and vice versa. Overall, the DMT/DMD mechanism creates an innovative means to embed process knowledge and expertise on the definition of model-based problems, as well as increased opportunities for documentation and re-use of case-studies. The manipulation of these DMTs/DMDs is the scope of the PDE software object. In addition, the PDO and DEP data models encapsulate corporate expertise on the use of high-level declarative modelling languages, the process model and the process system, and they make possible the definition and execution of rigorous estimation/reconciliation problems of interest to operations personnel.

As described by Rolandi and Romagnoli (2005), an environment for the definition of generic dynamic estimation problems of the characteristics described above shares common characteristics with the simulation and optimisation environments of the integrated framework. Effectively, similarities with the simulation environment are due to the fact that estimation- reconciliation problems are based on experimental plant data and, therefore, there is a need for data pre-processing, conditioning and reconstruction of process trajectories. Similarities with the optimisation environment are given by the fact that dynamic estimation experiments are a particular type of dynamic optimisation problems and, hence, additional structural information (i.e. information apart from that contained in the mathematical model of the process) is needed to fully determine the nature of the estimation/reconciliation experiment/case-study. This is a point-of-synergy which can be exploited during the design, implementation and use of an integrated model-centric system for support of process operations.

2.4 *A Case Study: Application to the Pulping Section of a Pulp and Paper Mill*

The challenge associated to the joint parameter estimation and data reconciliation case-study proposed in this section lies on the complexity of both the industrial process system and the actual

model-based problem. Unfortunately, the points-of-synergy between the solution of this process-engineering problem and other analysis and improvement tasks supported by the integrated model-centric framework is out of the scope of this manuscript.The goal of this case-study is to reconcile the process model and the plant data focussing on the closure of the general mass balance of the continuous pulping system. The estimation horizon is 1440min (24hr) and the window for reconstruction of process trajectories (Rolandi and Romagnoli, 2006b) is 30min for both input and output variables. A subset of 26 input process variables is used to imitate the input behaviour of the continuous process system. Among them 21 are controlled variables (set-points of PID control loops) and 5 are uncontrolled measured variables (disturbances).

The wood chip impregnation factor is a measure of the flowrate of steam condensate bounded to the interstitial void space between wood chips after the atmospheric pre-steaming step at the chip bin and before entering to the chip meter. Conventionally, the magnitude of this parameter would be obtained from the P&ID; however, changes in wood handling operations and operating conditions of the chip bin will change its nominal value. Since the magnitude of this parametric variable affects the closure of the mass balances, it will be chosen as a decision variable of the joint parameter estimation and data reconciliation problem. Additionally, we will estimate the magnitude of the pre-multiplier of the fundamental kinetic model of the Kraft pulping reactions occurring within the continuous cooking digester. Finally, we will also estimate the magnitude of the bias of three flow measurement devices: overall white liquor addition; wash filtrate addition to the digester's bottom; and black liquor extraction from the upper screens of the digester (see Table 1). The measurements of eight sensors are used for the purpose of estimation (Rolandi and Romagnoli, 2006b). The potential for model-based joint parameter estimation and data reconciliation of a large-scale complex industrial process system is demonstrated in this case-study: the problem results in the estimation of five parametric process variables (three of them are measurement biases) from an experimental data pool of eight measured variables and twenty-six control variables.

Table 1. Parametric variables of the continuous pulping area

DCS Tag	Variable Description
EE212.KinPreMult	Kinetic pre-multiplier
EE103.ChipImpFctr	Wood chip impregnation factor
FT212A.MB	Overall white liquor addition flow
FT212H.MB	Wash filtrate addition flow to digester bottoms
FT212C.MB	Upper extraction screen extraction flow

Table 2 shows the optimal estimates, confidence intervals and lower and upper bounds for the parametric variables. From this information we can calculate that the coefficient of variation for a 95% confidence on the individual estimates of the parametric variables EE212.KinPreMult and EE103.ChipImpFctr are 1.4% and 6.0% respectively; for all practical purposes, this is an indication of a satisfactory accuracy of estimation. The coefficient of variations of FT212A.MB and FT212H.MB based on a 95% confidence (Table 2) are reasonably small (3.4% and 4.0%, respectively), which is an indication of satisfactory accuracy of the estimates. On the other hand, the coefficient of variation corresponding to FT212C.MB is fairly large (46.7%), indicating a large uncertainty in the determination of this measuring device bias. In spite of this, the process variable can still be successfully estimated given the data pool used in this case-study. Figure 3 shows the fulfilment of the general mass balance of the continuous pulping system before and after reconciliation.

From a practical viewpoint, it was our aim to estimate those biases which have a strong impact on inventory analysis, or whose quantification is vital for other operational purposes (e.g. inferential soft-sensing). In the case of an industrial continuous pulping system, the most significant sources of revenue and expenses are likely to be the production of pulp, the cost of chip consumption and the cost of evaporation of weak black liquor (Rolandi and Romagnoli, 2006b). Fortunately, the cost of evaporation of weak black liquor can be partially reconciled from the estimate of the bias of the upper-screen extraction flow measurement. Interestingly, the 6.4% error of this process measurement (see Table 2) is associated to a material stream which accounts for nearly 32% of the overall weak black-liquor extraction flow from the continuous cooking digester at this nominal production level (~3.1m3/min). Additionally, the treatment of the black liquor in the evaporation area comprises approximately 56% of the variable costs of operation of the continuous pulping area (~ 88$/min). Hence, a 6.4% measurement error on such a critical process stream is equivalent to a production accounting miscalculation of approximately 0.50 million US$ per year, or an inventory analysis error of roughly 32 thousands cubic meters per year. This analysis demonstrates the economic incentive for advanced dynamic data reconciliation.

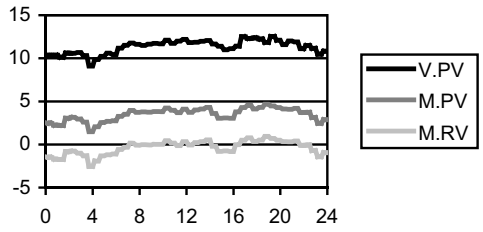


Fig. 3. Fulfilment of general mass balance (relative error [%] vs time [hr]); volumetric flow (measured process variables) and mass flow (calculated and reconciled process variables).

Table 2. Optimal estimates and confidence intervals

	EST	CI(95%)	LB	UB	CV [%]	ERR [%]
EE212.KinPreMult [adim]	3.34E-01	4.73E-03	3.29E-01	3.39E-01	1.42	n/a
EE212.ChipImpFctr [m3/kg]	1.83E-01	1.09E-02	1.72E-01	1.94E-01	5.96	n/a
FT212A.MB [m3/min]	7.32E-02	2.52E-03	7.06E-02	7.57E-02	3.44	7.1
FT212H.MB [m3/min]	1.69E-01	6.76E-03	1.62E-01	1.75E-01	4.01	8.2
FT212C.MB [m3/min]	6.51E-02	3.04E-02	3.47E-02	9.55E-02	46.7	6.4

Recently, Romagnoli and co-workers (Wang and Romagnoli, 2003) have presented a complete framework for Robust Data Reconciliation based on the generalized-t (GT) distribution, which can accommodate a family of distribution and thus providing extra flexibility as well as efficiency and optimality. The drawback of this approach is that actual statistical characterization of each sensor needs to be estimated, thus increasing the computation load if done at a centralized level.

In our proposed sensor network (Joe et al, 2005), some of these tasks can be delegated to the individual sensor (self-learning) in a decentralized manner, reducing the overall load at the higher level (control system) and at the same time providing extra functionalities for the sensor allowing the implementation of robust self-checking strategies at the local level. Two key things to be formulated in realising the sensor network are the architecture and the intelligence (or task allocation) of the sensor network. The following two subsections will discuss these, with a focus on GT-based data reconciliation. The last subsection presents application of the proposed sensor network to an integrated pilot-scale plant.

2.5 *Architecture of the Sensor Network*

The architecture of the sensor network defines the interconnection structure and functional relationships among the intelligent nodes in the network. The federated processing architecture is adopted in our proposed sensor network. Federated means that certain responsibilities are allocated to nodes at higher tier, but many functions are performed autonomously by nodes at lower tier (Sastry and Iyengar, 2005). Figure 4 shows the proposed network architecture. The network hierarchy consists of two tiers, each corresponding to an information processing level. Clusters are formed, i.e. each upper level node manages a few lower level ones. Each cluster corresponds to a process unit, i.e. the lower level nodes are none other than sensors measuring the variables of the process unit. We termed the lower level nodes as cluster members and the upper level ones as cluster heads. Nodes at the upper level can be considered as virtual sensors, in that they do not sense any physical phenomena, but mainly function as information processing units, performing tasks that are multivariate in nature such as data reconciliation. As such, cluster heads must collect information from their members in order to carry out their tasks. The cluster head and its members may require different computational capabilities.

The communication scheme is depicted by the lines connecting the nodes, i.e. cluster members communicate with their respective cluster heads, while cluster heads also communicate with one another. This results in different requirements of communication and networking capability and interface of the cluster heads as compared to the cluster members.

2.6 *Intelligence of the Sensor Network: Distributed Data Rectification*

The federated processing in a sensor network provides a potentially more efficient alternative implementation of the GT-based DR strategy than the centralized scheme. The GT-based DR strategy comprises two procedures that can be distributed in the federated sensor network: statistical characterization of sensor data (using GT distribution) and reconciliation of data using the obtained statistical characteristics. Since each sensor node (cluster member) is intelligent, the statistical characterization of data can be performed at the sensor level, resulting in self-*learning* of each sensor. Besides relieving the higher level from the computational burden and compressing the data to be communicated, self-learning also provides a signal model which is useful in reducing uncertainty in the measurement data and is the basic information that can be used for further processing that is not limited to DR only. The steps involved in sensor self-learning include the collection of a set of data points from which the sensor characteristics are extracted, and the estimation of the parameters of the statistical distribution of the data itself. This estimation can be mathematically expressed as:

$$\{\mu, p, q, \sigma\} = \arg\max \sum_{i=1}^{n} \log f_{GT}(u_i, p, q, \sigma) \qquad (1a)$$

where:

$$u_i = y_i - \mu \qquad (1b)$$

$$f_{GT}(u, p, q, \sigma) = \frac{p}{2\sigma q^{1/p} B(1/p, q)\left[1 + |u|^p / \sigma q^{1/p}\right]^{q+1/p}} \qquad (1c)$$

y is the i-th data point, n is the number of data points in the current data set, μ is the estimate of the process

variable and {p, q, σ} are the parameters of the GT distribution function f_{GT}, i.e. the statistical characteristics of the sensor.

The second step of the distributed GT-based DR strategy, i.e. the reconciliation step, is performed at the level of cluster head (process unit) due to its multivariate nature. The cluster head is therefore responsible for consolidating data from each member sensor in the cluster, and subsequently performing the computation to reconcile the data. Mathematically, this computation can be expressed as:

$$\underset{x}{\text{Max}} \; -\log f_{\text{GT}}(\text{u} \mid \text{p}, \text{q}, \sigma) \text{ s.t. } g(x) = 0 \qquad (2)$$

where u = y-x, y is the measurements, T is the estimates of the p process variables and g(x) denotes the set of conservation equations. Note that the values of {p,q, σ} used in this estimation are the self characteristics communicated by each individual sensor.

2.7 *A Case Study: Application to an Integrated Pilot-Scale Plant*

Experiment Environment. An experimental platform comprising plant simulator, sensor (cluster member) simulator, and cluster head simulator is constructed.

- *Plant Simulator:* The virtual version of a process unit within an integrated pilot-scale plant is developed. This unit, a continuous stirred tank reactor (CSTR) with a cooling coil, is simulated using Matlab/Simulink. Measured variables include: Fin (feed flow rate), Tin (feed temperature), F (effluent flow rate), T (effluent temperature), Fc (cooling water flow rate), Tcin (inlet cooling water temperature), Tc (outlet cooling water temperature), Trx (reaction vessel temperature).

- *Cluster Member (Sensor) Simulator:* The cluster member consists of two parts: physical sensing and data processing. Accordingly, the simulator consists of noise generator and saturation function to mimic sensing, and a self-learning module to realize the data processing segment. For each of the eight measured variables of the CSTR, a cluster member is assigned. A cluster head manages and monitors these eight cluster members. To realize the mapping between cluster heads and cluster members, each cluster member is labelled by unique identification. The cluster members will transmit this identification and its estimated self/ signal characteristics {μ, p, q, σ} to the respective cluster heads and cluster head will use this information to perform data reconciliation.

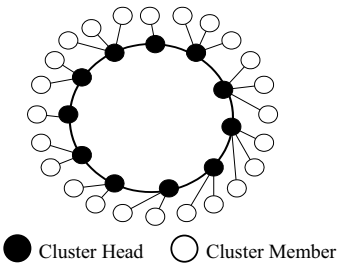


Figure 4: Federated sensor network architecture

- *Cluster Head Simulator:* The cluster head consists of communication controller, task controller and task modules. The communication controller is responsible for both external communication, i.e. with the cluster members and with the user interface, and internal communication, i.e. with the task controller. The task controller oversees all the tasks assigned to the cluster head. As such, the distributed data reconciliation as described in Section 3 is one of the underlying intelligence of the task controller. In our original proposal (Joe et al, 2005), besides data reconciliation, the intelligence also includes fault diagnosis and sensor reconstruction. The implemented framework is shown in Figure 5.

Experiment Results

- *Self-learning:* The results of self-learning for a few different noises are depicted in Figure 6, demonstrating considerably accurate characterization of the sensor data.

- *Data Reconciliation:* The formulated distributed GT-based data reconciliation is performed by cluster head using the characteristics obtained by self-learning in each cluster member. Figure 7 compares the estimation accuracy (ratio of absolute error of reconciled to that of measured data) of the proposed distributed method with the conventional centralized one according to Wang and Romagnoli (2003). Comparable performance is observed, hence demonstrating the viability of the distributed scheme.

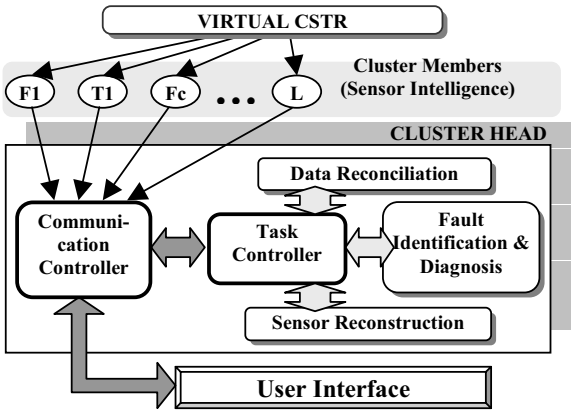
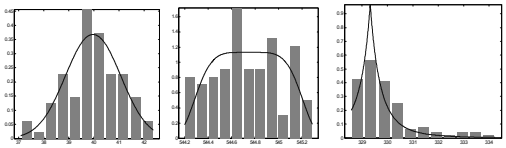


Figure 5: Overview of modules in the experimental setup



(a) F1, Gaussian (b) T1, Uniform (c) Fc, Laplacian

Figure 6: Statistical characterization of sensor data

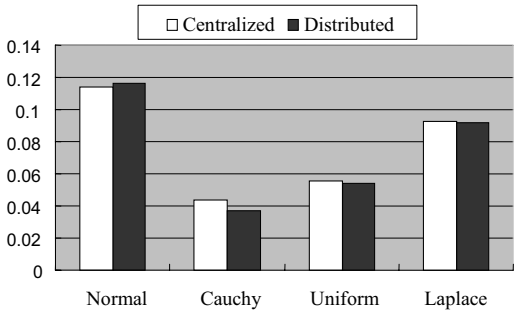


Figure 7: Estimation errors of the distributed and centralized data reconciliation approach

3. CONCLUSION & FUTURE WORK

The impact of emerging technologies on what is technically and economically feasible to achieve in the application of data reconciliation activities in manufacturing plants has been discussed and demonstrated.

Recent developments on model-centric technologies to support plant operations based on state-of-the-art process-engineering software tools were shown to provide the opportunities for performing large-scale parameter estimation – data reconciliation applications in complex dynamic industrial processing systems. The results of the industrial case-study were assessed from the perspective of production accounting and inventory analysis, and found a great incentive for the Process Industries to benefit from these advanced methodologies for plant data management. An environment for the definition of generic dynamic estimation problems of the characteristics described above shares common features with complementary simulation and optimisation environments. This is a point-of-synergy which can be exploited during the design, implementation and use of an integrated model-centric system for support of process operations.

New sensor technologies based on recent advances in microprocessor-based instrumentation and digital communications provided opportunities for the realization of novel sensor network architectures towards a truly distributed environment for data processing and reconciliation. The proposed federated architecture of intelligent sensor networks provides a flexible underlying framework for distributed data processing and rectification. We have presented a reformulation of the conventionally centralized robust partially adaptive data reconciliation scheme into a distributed scheme based on the federated sensor network. Experiment results showed highly comparable performance in terms of estimation efficiency, hence confirming the feasibility of the distributed scheme. Furthermore, in decentralizing the data reconciliation task, intelligence in the form of self-learning is incorporated into sensors. The presented distributed scheme serves as the first step towards a holistic distributed treatment of sensor data using the federated sensor network. This includes, but is not limited to, multi-resolution sensor data modeling, robust filtering and missing sensor data reconstruction.

REFERENCES

Braunschweig, B.L., Pantelides, C.C., Britt, H.I. and Sama, S. (2000). Process modeling: The promise of open software architectures. *Chemical Engineering Progress*, **Vol.96**, pp.65-76.

Global CAPE-OPEN Consortium (2000). Cape-open conceptual design document (cdd2). Available from: http://www.global-cape-open.org/02_CO_Conceptual_Design_Document_CDD2.pdf. Accessed on: 2004 Dec. 1.

Foss, B.A., Lohmann, B. and Marquardt, W. (1998). A field study of the industrial modeling process. *Journal of Process Control*, **Vol.8**, pp.325-338.

Joe, Y.Y, Ding, Z.Q, Ling, K.V. and Romagnoli, J.A. (2005). An intelligent sensor network for distributed data rectification and process monitoring. In *Proceedings of the 3rd International Conference on Industrial Informatics (INDIN 2005)*, IEEE, Perth, Australia.

Rolandi, P.A. and Romagnoli, J.A. (2005), An integrated environment for support of process operations, *AIChE Annual Meeting*, Cincinnati.

Rolandi, P.A. and Romagnoli, J.A. (2006a). Integrated model-centric framework for support of manufacturing operations. Part i: The framework. *Computers and Chemical Engineering*, Submitted for publication.

Rolandi, P.A. and Romagnoli, J.A. (2006b). Integrated model-centric framework for support of manufacturing operations. Part iii: Joint parameter estimation and data reconciliation. *Computers and Chemical Engineering*, Submitted for publication.

Sastry, S. and Iyengar, S.S. (2005). A taxonomy of distributed sensor networks. In *Distributed Sensor Networks* (Iyengar, S.S., Tandon, A., Brooks, R (Ed.)), pp.29-43. CRC Press, Boca Raton, Florida.

Wang, D. and Romagnoli J.A. (2003). A framework for robust data reconciliation based on a generalized distribution function. *Ind. Eng.Chem. Res.,* **Vol.42**, pp.3075-3084.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ITERATIVE LEARNING CONTROL APPLIED TO BATCH PROCESSES: AN OVERVIEW

Jay H. Lee * **Kwang S. Lee** **

* *School of Chem. and Biomolec. Eng., Georgia Institute of Technology*
311 Ferst Dr. NW, Atlanta, GA 30332-0100, USA
** *Dept. of Chem. and Biomolec. Eng., Sogang University*
1 Shinsoodong, Mapogu, Seoul 121-742, Korea

Abstract: With the recent emphasis on batch processing by the emerging industries like the microelectronics and biotechnology, the interest in batch process control has been renewed. In this paper, we present an overview of the Iterative Learning Control (ILC) technique, which can be used to improve tracking control performance in batch processes. We present the fundamental concepts and review the various ILC algorithms, with a particular focus on a model-based algorithm called Q-ILC and an application involving a Rapid Thermal Processing (RTP) system. The study indicates that one can solve a seemingly very difficult multivariable nonlinear tracking problem with relative ease by combining the ILC technique with basic process insights and standard system identification techniques. We also bring forth some related techniques in the literature with the hope of unifying them and also suggest some remaining challenges.

Keywords: tracking control, iterative learning control, model based control, batch process control

1. INTRODUCTION

Batch processes have historically lagged continuous processes in terms of development and deployment of advanced optimization and control tools. Whereas significant developments have occurred during the past few decades in the industrial practice of continuous process control (Qin and Badgwell, 1996; Morari and Lee, 1999), the same has not been the case for batch processes, which have continued to rely on old techniques like ladder-logics and PID control. Part of this can be attributed to the comparatively lower production volume through batch processing. Another reason for this may be that batch processes present a set of challenges uncommon in continuous processes, including nonstationary operating recipes, the consequent exposure to process nonlinearity, and significant variations in the initial charge condition (Berber, 1996). These challenges are not easily met by the standard linear optimal control theories and tools, which are widely adopted for continuous industrial process control today.

However, the role of batch processing is ever-increasing in today's diversified manufacturing environment. Besides the fine or specialty chemicals, new industries that have emerged from the VLSI technology, bio-technology, and material science are mostly batch-processing-oriented. In accordance with its increased importance, its operation support tools need to be upgraded. Such a shift in the trend has already started taking place, as evidenced by the extensive use of run-to-run control and multivariate monitoring in some of the new industries. We believe, however, that much more can be done, even with the existing technologies today. For example, Iterative Learning Control

(ILC), the topic of this paper, has not enjoyed a serious look by the practitioners thus far despite its vast potentials for improving tracking control performance in batch processes.

This paper presents an overview of ILC in the context of trajectory tracking problems in batch processes. Although basic theories of ILC have been firmly laid out in the literature, it is not always straightforward to apply them to achieve success in practice. We present ILC in the context of a multiple point temperature tracking problem in a Rapid Thermal Processing (RTP) system. By doing so, our objective is to bring forth the unique capabilities of ILC for batch process control and at the same time some of the subtle challenges one may face in applying the technique. Fortunately, such challenges are not insurmountable and the standard linear ILC technique can provide an excellent performance for what appears to be a very difficult nonlinear trajectory tracking problem. We also point out some related techniques like Run-to-Run Control and Repetitive Control, highlighting the similarities and differences. Finally, we point to some of the open issues left for future research.

2. EXEMPLARY PROBLEM: TEMPERATURE TRACKING CONTROL FOR A RTP SYSTEM

In the operation of Rapid Thermal Processing (RTP) systems, one of the most important challenges is to achieve uniform temperature distribution across the wafer surface while tracking a reference trajectory with a temperature range of several hundred degrees. From the system theoretic viewpoint, it is a nonlinear, multivariable control problem involving a batch system with fast dynamics and noisy measurements. Added to this are the facts that a single RTP system can be used for different wafer fabrications demanding different temperature trajectories to be followed and the characteristics of a RTP system may vary significantly by reasons like contamination. All these factors combine to make reliable modeling of the system very difficult (Cho and Gyugyi, 1997).

In the *experimental* RTP equipment, the silicon wafer is heated by an array of 38 bar-type tungsten-halogen lamps, the maximum power of which is 1 Kw each. The lamps are assembled together by two, four or six to comprise a total of ten independent groups, as shown in the figure. The chamber wall is cooled with circulating cooling water. The electric power inputs to the ten groups, denoted by $u_1, \cdots, u_{10}$, are therefore the manipulated inputs. Wafer temperature was measured at eight points with K-type thermocouples (TC's) glued on the backside of the wafer surface. As a consequence, the experimental RTP equipment is configured as an 8×10 MIMO system. In the commercial RTP operation, however, such a wafer with embedded thermocouples would be available only for testing purposes. In the actual production runs, in-situ temperature measurements would have to be provided by pyrometers, and for economic reasons, the number of such sensors per equipment may be limited. Hence, in addition to the full 8×10 system, we investigate the possibility of limiting the temperature measurements to just three locations. In this case, selection of the measurement points becomes an important issue.

Radiative heat transfer equations can be used to construct a fundamental or semi-empirical model representing heat balances. Due to the space limitation, we refer the readers to the open literature for the details of such models (Lee *et al.*, 2001*b*; Lee *et al.*, 2003). Given the idiosyncratic designs of individual equipments, however, it is more realistic to try to develop a control model from system identification, which is the approach we adopt here.

3. ITERATIVE LEARNING CONTROL

ILC is a general technique for improving transient tracking performance of a system that executes a same operation repeatedly. In its basic formulation, a target system has the following characteristics: i) Each run lasts for a fixed length of time; ii) the reference trajectories (to be followed by the outputs) remain the same from run to run; iii) the process state is reset to a same value at the start of each operation. ILC techniques developed under such assumptions can be used effectively on a process with some disturbances and initialization errors as well as occasional changes in the reference trajectories, however. Temperature tracking problems in many chemical batch processes can be tailored to fit into this category and hence the relevance to batch process operations.

3.1 *Historical Account*

It seems that the first technical contribution on ILC was the patent work by Garden (1971) three decades ago. Although a few independent contributions followed after that (Miller and G. T. Mallick, 1978; Uchiyama, 1978), it seems to have gone unnoticed by the larger control community until Arimoto *et al.* (1984) proposed the so-called D-type learning algorithm as a teaching mechanism for robot manipulators. This seminal work launched ILC into the mainstream control community and established it as a new branch of control technology. Significant body of work followed after that, which we summarize below.

However, we do not claim the list of references to be complete or unbiased.

3.2 *Basic Formulation*

Let $\mathbf{e}_k = \mathbf{r} - \mathbf{y}_k$ and $\mathbf{u}_k$ represent the output error trajectory and the manipulated input trajectory for the k^{th} run. For sampled **data** systems, these trajectories can be represented by finite dimensional vectors as follows:

$$\begin{aligned} \mathbf{e}_k &\triangleq \left[\, e_k(1)^T \;\; e_k(2)^T \;\; \cdots \;\; e_k(N)^T \,\right]^T \\ \mathbf{u}_k &\triangleq \left[\, u_k(0)^T \;\; u_k(1)^T \;\; \cdots \;\; u_k(N-1)^T \,\right]^T \end{aligned} \tag{1}$$

where N represents the number of sample points in each run. Note that the sampling interval do not need to be same for the inputs and the outputs or even uniform throughout the batch interval. This assumption is made here just for the convenience of exposition.

The objective in the ILC design can be simply stated as:

$$\|\mathbf{e}_k\| \to \mathbf{0} \;\; \text{as} \;\; \mathrm{k} \to \infty \tag{2}$$

A basic rule for updating the input trajectory on a run-to-run basis is the so-called first-order learning algorithm, which is represented by

$$\mathbf{u}_k = \mathbf{u}_{k-1} + \mathbf{H}\mathbf{e}_{k-1} \tag{3}$$

Here, $\mathbf{H}$ is called the learning filter matrix, which in general can be any map that transforms the finite-length error trajectory to a trajectory whose length and dimension are equal to those of the input trajectory. The learning filter can be designed as a dynamic filter, $H(s)$ or $H(z)$, operating on the time signal $e(t)$, depending on the underlying time domain for the system representation. Since $\mathbf{H}$ operates on the error trajectory of the previous run, it is *not* limited to *causal* maps, however. This is what gives ILC the distinct ability to overcome hindrances from dynamic elements like time delays to provide perfect tracking.

Note that the above learning algorithm has an integral action over the run index k. Hence, one can intuitively argue that (2) can be fulfilled with an appropriately chosen $\mathbf{H}$, just as the integral action in a PID controller can remove offset in the time domain. Since batch processes have no dynamics carried over from one run to next, pure integral control such as in (3) is sufficient in achieving the convergence. Nevertheless, a high-order algorithm like

$$\mathbf{u}_k = \mathbf{u}_{k-1} + \mathbf{H}_1\mathbf{e}_{k-1} + \cdots + \mathbf{H}_p\mathbf{e}_{k-p} \tag{4}$$

has also been studied as a generalization of (3) (Bien and Huh, 1989). When all the states of a batch process are reset to same values at the start of each run and disturbances do not vary from batch to batch, there is no benefit to be gained from the high-order generalization. However, when errors do not carry over completely from one batch run to next due to run-specific disturbances, measurement noises, and model errors, the high-order algorithm can deliver a superior performance owing to its ability to filter the error trajectories by using the results of several runs.

3.3 *Model-Based Formulation*

With the above form of the learning algorithm, the problem of ILC design is reduced to the design of the learning filter. In the initial period of development, researchers focused on model-free approaches, where a certain generic structure is presupposed on $\mathbf{H}$ and the parameters are tuned to achieve the convergence. D-type(Arimoto *et al.*, 1984) and PID-type(Bondi *et al.*, 1988) algorithms are such examples.

Alternatively, model-based algorithms were introduced in order to address more complex problems (e.g., MIMO systems) and to bring more insights into the technique. Early approaches were based on direct model inversion, *i.e.*, $\mathbf{H} = \mathbf{G}^{-1}$, and its variants(Togai and Yamano, 1985; Oh *et al.*, 1988; Lucibello, 1992; Moore, 1993; Lee *et al.*, 1994), where $\mathbf{G}$ represents the input-output map of the concerned process (*i.e.*, $\mathbf{y}_k = \mathbf{G}\mathbf{u}_k$). Note that $\mathbf{G}$ contains information about the batch process *dynamics* and is similar to the concept of *dynamic matrix* used in model predictive control (MPC). Though we are using a linear map here, note that the underlying dynamics are not limited to be linear time-invariant; time-varying dynamics (approximating nonlinear dynamics for instance) may be easily incorporated into the $\mathbf{G}$ matrix. In the case that $\mathbf{G}$ is exactly known, this particular choice of $\mathbf{H}$ eliminates the error completely after one iteration, which can be easily verified by multiplying $\mathbf{G}$ on both sides of (3). Nonminimum-phase dynamics do not cause any problem in the inversion here as $\mathbf{H}$ is not restricted to be causal. In practice, however, the inverse-model-based learning filter can give many problems. For a typical over-damped system, of which the inverse has increasingly higher gains with the frequency, the filter can be very sensitive to high frequency components of $e_k(t)$ producing extremely spiky input profiles. Also, since high frequency dynamics typically carry large model errors, the high filter gains in high frequency region can cause divergence.

Furthermore, the objective in (2) cannot always be satisfied for general MIMO systems. When the number of output variables is larger than that of input variables or when constraints become active, the error may not be made zero in general. An alternative objective may be to try to converge to an input trajectory that minimizes the output error:

$$\|\mathbf{e}_k\| \rightarrow \min_{\mathbf{u}} \|\mathbf{e}\| \quad \text{as} \quad \mathrm{k} \rightarrow \infty \tag{5}$$

where $\|\cdot\|$ is some vector norm. The 2-norm is the typical choice.

Moore (1993) proposed to solve

$$\min_{\mathbf{u}_k} \|\mathbf{e}_k\|^2 \tag{6}$$

before the start of the k^{th} run, based on the error from the $k-1^{\text{th}}$ run. For a linear system, $\mathbf{y} = \mathbf{G}\mathbf{u}$, the error model can be written as

$$\mathbf{e}_k = \mathbf{e}_{k-1} - \mathbf{G}(\mathbf{u}_k - \mathbf{u}_{k-1}) \tag{7}$$

The least squares solution is

$$\mathbf{u}_k = \mathbf{u}_{k-1} + \mathbf{G}^{+}\mathbf{e}_{k-1} \tag{8}$$

where the superscript $^{+}$ represents the pseudo-inverse. Alternatively, Amann *et al.* (1996) and Lee *et al.* (1996) independently proposed to solve

$$\min_{\Delta\mathbf{u}_k} \left\{\|\mathbf{e}_k\|^2_{\mathbf{Q}} + \|\Delta\mathbf{u}_k\|^2_{\mathbf{R}}\right\} \tag{9}$$

In the above, $\Delta\mathbf{u}_k = \mathbf{u}_k - \mathbf{u}_{k-1}$ and the notation of $\|\mathbf{x}\|^2_{\mathbf{P}}$ denotes $\mathbf{x}^T\mathbf{P}\mathbf{x}$. The resulting input for the linear system, $\mathbf{y} = \mathbf{G}\mathbf{u}$, is given as

$$\mathbf{u}_k = \mathbf{u}_{k-1} + (\mathbf{G}^T\mathbf{Q}\mathbf{G} + \mathbf{R})^{-1}\mathbf{G}^T\mathbf{Q}\mathbf{e}_{k-1} \tag{10}$$

As $k \rightarrow \infty$, input profiles that result from (8) and (10) converge to the same limit $\mathbf{u}^*$ that satisfies $\|\mathbf{e}(\mathbf{u}^*)\| = \min_{\mathbf{u}} \|\mathbf{e}\|$. What happens with (10) is that the convergence is retarded by the input change penalty term. In particular, high-frequency type changes in the input profile, which tend to be large due to low system gains, are penalized heavily. In the course of learning, therefore, much smoother input profiles that drive the error to a near-minimum (*i.e.*, minimum except for the high frequency components for which the learning gains are much too low compared to the system gains due to the input penalty term) are obtained by (10). In practice, the learning can be stopped before the input profiles become very spiky or divergence behavior starts setting in. For simplicity, we call the algorithm based on (9) Q-ILC (Quadratic criterion-based ILC) hereafter.

Later, Lee *et al.* (2000) considered a batch process subject to stochastic disturbances and proposed an observer-based Q-ILC algorithm. This extension provided a more convenient and intuitive tuning knob to control the rate of convergence and the ability to take a more systematic account of disturbances and noises. A robustness study of Q-ILC indicated that convergence can be achieved in the presence of a fairly large model error (Lee *et al.*, 2000; Kim *et al.*, 2000). They also pointed out that, when the constraints are given as linear inequalities, the optimal input profile respecting the constraints can be obtained through a quadratic programming technique. Convergence of the constrained algorithm with an observer was proved in Lee and Lee (2000).

4. PROTOTYPICAL MODEL-BASED ILC TECHNIQUE

There are several similar versions of Q-ILC (Amann *et al.*, 1996; Lee *et al.*, 2000; Lee *et al.*, 1999; Chin *et al.*, 2004). The basic idea for all these versions is the same, but they differ in the way the filtering of error trajectories (or detuning of the learning gain) is accomplished, and whether and how real-time feedback control (RFC) signal is added to the feedforward ILC signal. Here we present a particular version of Q-ILC, which is based on the work by Chin *et al.* (2004) and was found to be particularly suited to the RTP application.

A general form of ILC combined with RFC is

$$\begin{aligned} u_k(t) = \; & u_{k-1}(t) + H_1(t)e_{k-1}(1:N) \\ & + H_2(t)e_k(1:t) \end{aligned} \tag{11}$$

where $(i:j)$ means data from $t = i$ to j and H_1 and H_2 represent the gains for the ILC and RFC, respectively. The role of RFC is to further modify $u_k(t)$ based on the real-time error feedback of $e_k(1:t)$, which contains information on a new disturbance occurring during the on-going k^{th} run. However, under this scenario, the updated input for the next run would also be affected by the new disturbance, which may not show up in the next run. The following technique attempts to separate $u(t)$ into the ILC- and RFC-related terms so that disturbances specific to a current run would have minimal effect on the next run.

4.1 *Model Formulation*

Suppose that the system dynamics are represented by

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + Kv(t) \\ y(t) &= Cx(t) + v(t) \end{aligned} \tag{12}$$

In (12), $v(t)$ is a zero-mean independent, identically distributed (i.i.d.) sequence in time but $v_k(t)$

for different k's may show correlation. In fact, $v_k(t)$ may even exhibit drifting behavior along k and not just random fluctuations around a stationary mean. Such behavior can be reasonably described by the equation

$$\begin{aligned} \bar{v}_k &= \bar{v}_{k-1}(t) + n_k(t) \\ v_k(t) &= \bar{v}_k(t) + \hat{v}_k(t) \end{aligned} \tag{13}$$

where $\hat{v}_k(t)$ and $n_k(t)$ are zero-mean i.i.d. sequences with respect to both k and t. Such a model can be obtained through system identification techniques like N4SID (Overschee and Moor, 1994), as will be demonstrated later.

Now, using the superposition principle, we decompose $u_k(t)$ into $\bar{u}_k(t)$ (the ILC input) and $\hat{u}_k(t)$ (the RFC input), and also separate (12) into two parts, one that is driven by $\bar{u}_k(t)$ and $\bar{v}_k(t)$, and the other by $\hat{u}_k(t)$ and $\hat{v}_k(t)$ as follows:

$$\begin{aligned} \bar{x}_k(t+1) &= A\bar{x}_k(t) + B\Delta\bar{u}_k(t) + Kn_k(t) \\ \bar{y}_k(t) &= C\bar{x}_k(t) + \bar{y}_{k-1}(t) + n_k(t) \end{aligned} \tag{14}$$

$$\begin{aligned} \hat{x}_k(t+1) &= A\hat{x}_k(t) + B\hat{u}_k(t) + K\hat{v}_k(t) \\ \hat{y}_k(t) &= C\hat{x}_k(t) + \hat{v}_k(t) \end{aligned} \tag{15}$$

where $\Delta\bar{u}_k \triangleq \bar{u}_k - \bar{u}_{k-1}$. Naturally, we have

$$y_k(t) = \bar{y}_k(t) + \hat{y}_k(t) \tag{16}$$

4.2 *Algorithm*

The Q-ILC algorithm follows the following steps:

(1) *Information Gathering*: After the $k-1^{\text{th}}$ run, $\{e_{k-1}(t),\ \bar{u}_{k-1}(t), \hat{u}_{k-1}(t)\}$ are available.
(2) *ILC Signal Computation*: Before starting the k^{th} run, compute $\bar{u}_k(t),\ t \in [0, \cdots, N-1]$ off-line according to a chosen cost function involving $\bar{e}_{k|k-1}(t)$.
(3) *RFC Signal Computation*: At each t during the k^{th} run, calculate $\hat{u}_k(t)$ such that the predicted error for the current run is minimized. Then, apply $u_k(t) = \bar{u}_k(t) + \hat{u}_k(t)$ to the process.

4.2.1. *Details of the ILC Signal Computation*

Define

$$\begin{aligned} \bar{\mathbf{e}} &\triangleq \left[\, \bar{e}(1)^T\ \bar{e}(2)^T\ \cdots\ \bar{e}(N)^T \,\right]^T \\ \Delta\bar{\mathbf{u}} &\triangleq \left[\, \Delta\bar{u}(0)^T\ \Delta\bar{u}(1)^T\ \cdots\ \Delta\bar{u}(N-1)^T \,\right]^T \end{aligned} \tag{17}$$

and similarly for other variables where $\bar{e} \triangleq r - \bar{y}$. Expanding equations (14) and (15) gives

$$\begin{aligned} \bar{\mathbf{e}}_k &= \bar{\mathbf{e}}_{k-1} - \mathbf{G}\Delta\bar{\mathbf{u}}_k + \mathbf{w}_k \\ \mathbf{e}_k &= \bar{\mathbf{e}}_k - \mathbf{G}\hat{\mathbf{u}}_k + \mathbf{m}_k \end{aligned} \tag{18}$$

where

$$\mathbf{G} \triangleq \begin{bmatrix} CB & 0 & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-1}B & CA^{N-2}B & \cdots & CB \end{bmatrix} \tag{19}$$

$\mathbf{w}_{k+1}$ and $\mathbf{m}_k$ are appropriately defined signals that are independent in k.

$\Delta\bar{\mathbf{u}}_k$ is determined according to

$$\min_{\Delta\bar{\mathbf{u}}_k}\ \frac{1}{2}\left\{\|\bar{\mathbf{e}}_{k|k-1}\|^2_{\mathbf{Q}} + \|\Delta\bar{\mathbf{u}}_k\|^2_{\mathbf{R}}\right\} \tag{20}$$

where $\bar{\mathbf{e}}_{k|k-1}$ is the optimal prediction of $\bar{\mathbf{e}}_k$ based on the information available at the start of the k^{th} run, which is given by the Kalman filter applied to (18). The unconstrained solution to (20) is

$$\begin{aligned} \Delta\bar{\mathbf{u}}_k &= \mathbf{H}\bar{\mathbf{e}}_{k-1|k-1} \\ &= (\mathbf{G}^T\mathbf{Q}\mathbf{G} + \mathbf{R})^{-1}\mathbf{G}^T\mathbf{Q}\bar{\mathbf{e}}_{k-1|k-1} \end{aligned} \tag{21}$$

The Kalman filter estimates $\bar{\mathbf{e}}_{k-1|k-1}$ is computed as

$$\begin{aligned} \bar{\mathbf{e}}_{k-1|k-1} &= \bar{\mathbf{e}}_{k-1|k-2} + \mathbf{K}(\mathbf{e}_{k-1} + \mathbf{G}\hat{\mathbf{u}}_{k-1} - \bar{\mathbf{e}}_{k-1|k-2}) \\ \bar{\mathbf{e}}_{k|k-1} &= \bar{\mathbf{e}}_{k-1|k-1} - \mathbf{G}\Delta\bar{\mathbf{u}}_k \end{aligned} \tag{22}$$

where $\mathbf{K}$ is the optimal gain matrix computed using the model of (18).

4.2.2. *Details of RFC Signal Calculation*

For the calculation of RFC signal, it is tempting to think that one should use (15) to control just the $\hat{y}$ component of y. However, it turns out to be beneficial to try to reduce the entire error in y as it helps to achieve faster convergence. It is convenient to consider a model with the input $\Delta\bar{u}_k(t)$ term taken out from the model. For this, we consider the following deterministic model that represents the output by $\Delta\bar{u}_k(t)$

$$\begin{aligned} a_k(t+1) &= Aa_k(t) + B\Delta\bar{u}_k(t) \\ y_{a,k}(t) &= Ca_k(t) \end{aligned} \tag{23}$$

Subtracting (23) from (14) eliminates $\Delta\bar{u}_k(t)$ from the equation:

$$\begin{aligned} \bar{x}_{a,k}(t+1) &= A\bar{x}_{a,k}(t) + Kn_k(t) \\ \bar{y}_k(t) &= C\bar{x}_{a,k}(t) - Ca_k(t) + \bar{y}_{k-1}(t) + n_k(t) \end{aligned} \tag{24}$$

We replace $\bar{y}_{k-1}(t)$ with $\bar{y}_{k-1|k-1}(t)$ which is given by (22). Combining (15) and (24) yields

$$\begin{aligned} &\begin{bmatrix} \hat{x}_k(t+1) \\ \bar{x}_{a,k}(t+1) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}\begin{bmatrix} \hat{x}_k(t) \\ \bar{x}_{a,k}(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix}\hat{u}_k(t) \\ &\qquad + \begin{bmatrix} K\hat{v}_k(t) \\ Kn_k(t) \end{bmatrix} \\ &y_k(t) + Ca_k(t) - \bar{y}_{k-1|k-1}(t) \\ &\qquad = \begin{bmatrix} C & C \end{bmatrix}\begin{bmatrix} \hat{x}_k(t) \\ \bar{x}_{a,k}(t) \end{bmatrix} + \hat{v}_k(t) + n_k(t) \end{aligned} \tag{25}$$

Denote the above as

$$z_k(t+1) = \Phi z_k(t) + \Gamma \hat{u}_k(t) + \zeta_k(t) \quad (26)$$
$$y_k(t) + Ca_k(t) - \bar{y}_{k-1|k-1}(t) = \Sigma z_k(t) + \eta_k(t)$$

Then, $\hat{u}_k(t)$ is determined based on the quadratic criterion

$$\min_{\hat{u}_k(\cdot)} E\Big\{ \|e_k(N)\|_M^2 + \sum_{t=0}^{N-1} \|e_k(t)\|_Q^2 + \|\hat{u}_k(t)\|_R^2 \Big\}$$
$$\text{subject to (26)} \quad (27)$$

Enforcing $e_k(t) \to 0$ is equivalent to steering $y_k(t) + Ca_k(t) - \bar{y}_{k-1|k-1}(t)$ to $r(t) + Ca_k(t) - \bar{y}_{k-1|k-1}(t)$. (27) is a standard LQ servo problem for the output of (26) to follow $r(t) + Ca_k(t) - \bar{y}_{k-1|k-1}(t)$. The solution is standard and given in the following form:

$$\begin{aligned} & \hat{u}_k(t) = -L_{\text{fb}}(t) z_k(t|t) + L_{\text{ff}}(t) b_k(t) \\ \to \quad & u_k(t) = \bar{u}_k(t) + \hat{u}_k(t) \end{aligned} \quad (28)$$

Detailed forms of $L_{\text{fb}}(t)$, $L_{\text{ff}}(t)$ and $b_k(t)$ can be found in textbooks like (Lewis and Syrmos, 1995), or in (Lee *et al.*, 2001*b*).

5. RESULTS OF APPLICATION TO AN EXPERIMENTAL RTP SYSTEM

The previously described Q-ILC algorithm was applied to the experimental RTP system introduced earlier. The length for each experimental run was 40 seconds and the sampling time was chosen as 0.5 second. The reference temperature trajectory was comprised of a holding zone at 400°C for 7 seconds and a ramping zone with 30°C/sec followed by another holding zone at 700°C for 25 seconds. Control of the wafer temperature for the first run was carried out using the regular MPC and then Q-ILC was applied from the 2nd run on.

5.1 *Model Identification*

The identification experiments were conducted while maintaining the wafer temperatures at around 650°C. Independent PRBSs were simultaneously added for 2,000 seconds to the respective steady state input values of the 10 lamp groups and the resulting temperature responses at the eight locations were taken. Suitable choice for the minimum clock period of the PRBS signals was found to be 15 seconds.

Examining the radiative heat transfer terms leads us to consider that an RTP system would be better represented by a linear dynamic model that relates $u(t)$ to $T_w^4(t)$, rather to $T_w(t)$ as previously done in (Lee *et al.*, 2001*b*) and (Lee *et al.*, 2003). Here T_w represents a vector containing the wafer temperatures at the eight locations. A linear dynamic model between $u(t)$ and $y(t) \triangleq T_w^4(t)$ can be obtained using a standard identification method, *e.g.*, N4SID (Overschee and Moor, 1994). To remove the bias effect ('trend'), we pretreated the input and output data using a difference filter of $F(q^{-1}) = \frac{1-q^{-1}}{1-fq^{-1}}$, $0 \le f < 1$. We chose $f = 0.974$. Let the filtered variables be

$$u_f(t) \triangleq F(q^{-1})u(t), \quad y_f(t) \triangleq F(q^{-1})y(t) \quad (29)$$

Processing $\{u_f(t)\}$ and $\{y_f(t)\}$ using N4SID yields a linear stochastic state space model in the following innovation form:

$$\begin{aligned} x_f^r(t+1) &= \bar{A}x_f^r(t) + \bar{B}u_f(t) + \bar{K}v(t) \quad (30) \\ y_f(t) &= \bar{C}x_f^r(t) + v(t) \end{aligned}$$

where $\{v(t)\}$ is a zero-mean i.i.d. sequence referred to as the *innovations*. (30) can be rewritten in terms of the original input and output as follows:

$$\begin{bmatrix} x^r(t+1) \\ x^d(t+1) \end{bmatrix} = \begin{bmatrix} \bar{A} & \bar{K} \\ 0 & I \end{bmatrix} \begin{bmatrix} x^r(t) \\ x^d(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} \bar{K} \\ (1-f)I \end{bmatrix} v(t)$$
$$y(t) = \begin{bmatrix} \bar{C} & I \end{bmatrix} \begin{bmatrix} x^r(t) \\ x^d(t) \end{bmatrix} + v(t) \quad (31)$$

Note that this model is in the form of (12), which was the basis for the Q-ILC algorithm development.

5.2 *Results and Discussion*

The result of applying the Q-ILC algorithm to the 8×10 system with T_w^4 as the output is shown in Table 2. The tracking error is continuously decreased with the run number and converges approximately after 7 runs. The temperature gap given in the table is defined as the maximum difference among the eight temperatures. This was reduced down to 5°C in the constant temperature region.

One of the main feature of the observer based Q-ILC algorithm is that batch-specific disturbances are filtered out to have minimal effect on the learning for subsequent runs. To demonstrate this, we lowered the initial wafer temperature to 370°C in the 8^{th} run and then returned it to the nominal value in the following run. From the results in Table 1, we can see that the performance of the 7th run is almost completely restored in the 9th run. Though not shown, it was observed that the ILC input for the 9^{th} run was hardly affected by the disturbance that occurred during the 8^{th} run.

In Table 2, the temperature from T_w^4 model-based control is compared with that of the conventional

linear model-based control where $y(t) \triangleq T_w(t)$. We can see that the temperature uniformity could be remarkably improved by the use of T_w^4 as the output of the control model. Note that the temperature gap was reduced significantly on the average. This shows that such simple process knowledge can be a huge factor if used appropriately.

We also attempted to limit the measurements to three locations, leading to a 3×10 system. The three measurement locations were decided to minimize the estimation error covariance. For comparison, we show in Table 3. the results when only the three temperatures were controlled. The three controlled points could achieve excellent tracking and regulation, but the temperature gap among the eight points, which were just monitored, was observed to be quite large. As an alternative, we decided to estimate the temperatures of the remaining five points based on the three measured temperatures. Linear estimation was employed for simplicity. The estimated temperatures along with the measured ones were used as if they were all measured, as in the 8×10 system case. Performance similar to that for the full measurement case could be achieved, as can be seen from Table 4.

6. RELATED TECHNIQUES

There are several techniques in the literature that share strong similarities with ILC. For example, repetitive control is a technique applied to a continuous system with periodic characteristics, which are also seen in repeated batch runs and thus form a basis for ILC. Given such strong similarities, cross-breeding of the techniques from these fields should be possible. Here we briefly review the related techniques and discuss whether they provide any new insight into ILC and vice versa.

6.1 *Run-to-Run Control*

Run-to-run control is a popular method to control product qualities in processes where direct in-situ measurements of the quality variables are impractical and off-line product analysis results must be utilized instead. Such processes are common in semi-conductor and polymer manufacturing. The basic form of linear model used for run-to-run control is

$$q_k = Mp_k + \frac{1}{1-q^{-1}} v_k \tag{32}$$

where q_k is the vector containing the end quality variables, p_k is the vector containing the recipe parameters, and v_k is an i.i.d. sequence. Here, the disturbance is modeled as an integrated white noise to account for jumps and slow drifts. Other disturbance models such as $\frac{1-\alpha q^{-1}}{1-q^{-1}}$ or double integrators can be used if deemed more appropriate.

The optimal prediction model for (32) is expressed as

$$q_{k|k-1} = q_{k-1} + M\Delta p_k + v_k \tag{33}$$

and Δp_k representing the recipe adjustment can be computed by minimizing $\|r - q_k\|_P^2$ or $\|r - q_k\|_P^2 + \|\Delta p_k\|_Q^2$ as before where r represents the desired quality values. Of course, nonlinear recipe models can be used, which may or may not be combined with a nonlinear state observer.

Examining the above, we see that there is little difference between ILC and run-to-run control except that the former addresses trajectory tracking problems whereas the latter addresses end quality control problems. This similarity was noticed by Chin *et al.* (2000), who combined ILC and run-to-run control concepts into an integrated technique called QBMPC intended for simultaneous trajectory tracking and end quality control. An added feature of QBMPC was that the end quality variables for an on-going batch run could be inferred from the on-line process measurements, thereby giving the controller the capability to make more immediate adjustment with respect to disturbances.

6.2 *Batch-to-Batch Optimization*

Batch-to-batch optimization (BBO) refers to the use of nonlinear programming technique on a real batch process. The key is to evaluate the objective function with actual process measurements rather than a model. The idea was once called EVOP (EVolutionary OPtimization) (Wilde and Beightler, 1967) and popular in the 60's in optimizing operating conditions in the process industries. The technique was recently brought back to the attention of the process control community by a number of researchers, mostly in the context of optimizing semiconductor manufacturing processes(Zafiriou and Zhu, 1990; Zafiriou *et al.*, 1995).

Consider an optimization problem

$$\min_{\mathbf{u}} \quad \phi(\mathbf{y}, \mathbf{u}) \tag{34}$$

Suppose the batch process's dynamics are represented by the nonlinear input/output map, $\mathbf{y} = \mathcal{N}(\mathbf{u}, \mathbf{d})$, where $\mathbf{d}$ is a disturbance trajectory. Based on the model, the optimization may be recast as

$$\min_{\mathbf{u}} \; J(\mathbf{u}) \; (= \phi(\mathbf{y}, \mathbf{u})) \tag{35}$$

In the above, J can be evaluated through ϕ where $\mathbf{y}$ is provided by the prediction model

$$\mathbf{y}_{k+1|k} = \mathbf{y}_k + \left(\frac{\partial \mathcal{N}}{\partial \mathbf{u}}\right)_k \Delta \mathbf{u}_{k+1} \tag{36}$$

Note that the actual process measurements of $\mathbf{y}_k$ is used, which provides some robustness when the dynamic map $\mathcal{N}$ and the disturbance $\mathbf{d}$ are not perfectly known.

A general search algorithm for $\mathbf{u}$ can be written as

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \alpha_k \mathbf{g}_k \tag{37}$$

where $\mathbf{g}$ represents a search direction. From

$$\begin{aligned} J(\mathbf{u}_{k+1}) &\approx J(\mathbf{u}_k) + \nabla J(\mathbf{u}_k)^T (\mathbf{u}_{k+1} - \mathbf{u}_k) \\ &= J(\mathbf{u}_k) + \alpha_k \nabla J(\mathbf{u}_k)^T \mathbf{g}_k \end{aligned} \tag{38}$$

it can be seen that $J(\mathbf{u}_{k+1}) < J(\mathbf{u}_k)$ for a sufficiently small α_k as long as $\mathbf{g}_k$ is given such that $\nabla J(\mathbf{u}_k)^T \mathbf{g}_k < 0$. In the steepest descent algorithm, $\mathbf{g}_k$ is given by $-\nabla \tilde{J}(\mathbf{u}_k)$, where $\tilde{J}$ represents ϕ evaluated with the prediction model (36). Alternatively, one can also solve for $\Delta \mathbf{u}$ minimizing the objective function ϕ. Note that a significant model error can be allowed in J (or equivalently in $\mathcal{N}$) before the convergence is violated.

From the use of (36), it is easy to see that the above described BBO is very closely related to ILC. In fact, the standard ILC can be interpreted as a special case of BBO where the objective function $\phi(\mathbf{y}, \mathbf{u})$ is quadratic in the tracking error and the input change and the underlying dynamic input-output map is linear. Hence, the BBO approach provides a natural extension of the standard ILC to the case of nonlinear model.

In Bonvin *et al.* (2001), BBO methods are classified into four categories according to how the model information is used. In their paper, the term BBO is used in a wider sense to include a model-based method with recursive identification, and reference-based and data-based model-free methods. The BBO method described in the above belongs to the (uncertain) fixed-model-based method according to the classification. In addition, they proposed a special BBO method called invariant-based optimization where some invariant structure of the input profile is determined first off-line using a coarse process model, the input is parameterized based on the invariant structure, and the true invariants are found using the process measurements.

6.3 *Repetitive Control*

Repetitive control (RC) is a control technique for canceling a periodic disturbance or tracking a periodic reference signal in a continuous process. Repetitive controllers are originally constructed using a delay loop based on the internal model principle(Hara *et al.*, 1988). However, the parallel with the ILC is clear that the error trajectory tends to repeat from cycle to cycle. The main difference is that, unlike in batch systems, the system state is not reset at the beginning. In the ILC literature, such a case is referred by the name of *no-reset ILC*, which is a new branch of study in ILC(Sison and Chong, 1996). In no-reset ILC, transient effect of previous cycles carried over is not taken into account and is simply left to die out with cycles.

Lee *et al.* (2001*a*) noted the similarity between the two problems and proposed an MPC technique called RMPC (Repetitive MPC), which is based on a periodically varying system description and has a strong parallel to Q-ILC. The RMPC technique was successfully applied to the control of a simulated bed chromatography system(Natarajan and Lee, 2000; Erdem *et al.*, 2004). More recently, Lee and Gupta (2005) proposed an extension of this to add robustness to mismatch in the period length. The robustness is achieved by penalizing the input change in a higher order difference form. Though the period mismatch is an important issue for batch processes as well, it is not clear how the idea would extend to the batch system case.

7. CONCLUSIONS

7.1 *Summary*

Iterative Learning Control (ILC) has great potentials for improving tracking control in batch processes. Though initially developed as a heuristic method for improving trajectory tracking performance of robot manipulators, two decades of research has laid solid theoretical foundations and generated insights needed for successful use in general tracking problems in batch processes. In particular, model-based algorithms like Q-ILC can address complex multivariable constrained systems and can be designed for significant robustness to model errors. We demonstrated the potentials and subtle challenges by presenting a case study involving an experimental RTP system. As turned out, one can effectively solve what appeared to be a very difficult multivariable, nonlinear tracking problem by combining the model-based ILC technique with some sound engineering judgment and creativity.

7.2 *Future Research Directions*

An important assumption behind all current ILC algorithms is that the run length is fixed and the

reference trajectory remains same. In many industrial batch processes, however, this assumption is oftentimes violated. Even though each batch run may slightly different, the basic pattern of the trajectory, such as hold-ramp-hold may not change. The main question is how one can translate the error trajectory from previous batch runs into an error trajectory for a new run, which may have a different length and reference trajectory.

Another important issue is more systematic accounting for model errors in the ILC design. In particular, when the model error can be described quantitatively such as polytopic bounds for the dynamic gain matrix, we would like to be able to use such information directly in the design. Because the batch system can be viewed as a simple integrating system along the batch index, derivation of robust ILC algorithms using the usual formulation like min-max optimization may prove to be more tractable. Some initial ideas along this direction can be found in Lee *et al.* (2000).

Finally, the use of a nonlinear model within the existing ILC algorithms has been studied extensively but it is beyond the scope of this paper.

8. REFERENCES

Amann, N., D. H. Owens and E. Rogers (1996). Iterative learning control for discrete-time systems with exponential rate of convergence. *IEE Proc.-Control Theory and Applications* **143**, 217–224.

Arimoto, S., S. Kawamura and F. Miyazaki (1984). Bettering operation of robots by learning. *J. Robotic Syst.* **1**, 123–140.

Berber, R. (1996). Control of batch reactors: a review. *Trans IChemE.* **74**, 3–20.

Bien, Z. and K. M. Huh (1989). Higher-order iterative learning control algorithm. *IEE Proc.* **136**, 105–112.

Bondi, P., G. Casalino and L. Gambradella (1988). On the iterative learning control theory for robotic manipulators. *IEEE J. Robot Autom.* **4(1)**, 14–22.

Bonvin, D., B. Srinivasan and D. Ruppen (2001). Dynamic optimization in the batch chemical industry. *Proc. of CPC-VI, Tucson, AZ.*

Chin, I., S. J. Qin, K. S. Lee and M. Cho (2004). A two-stage ilc technique combined with real-time feedback for independent disturbance rejection. *Automatica* **40**, 1913–1922.

Chin, I. S., K. S. Lee and J. H. Lee (2000). A technique for integrated quality control, profile control, and constraint handling for batch processes. *Ind. Eng. Chem. Res.* **39**, 693–705.

Cho, Y. M. and P. J. Gyugyi (1997). Control of rapid thermal processing: A system theoretic approach. *IEEE Trans. Contr. Sys. Tech.* **5**, 644.

Erdem, G., S. Abel, M. Morari, M. Mazzotti, M. Morbidelli and J. H. Lee (2004). Automatic control of simulated moving bed. *Ind. Eng. Chem. Res.* **43**, 405–421.

Garden, M. (1971). Learning control of actuators in control systems. *US Patent 3555252.*

Hara, S., Y. Yamamoto, T. Omata and N. Nakano (1988). Repetitive control system: a new type servo system for periodic exogeneous signals. *IEEE Trans. on A.C.* **33**, 659–668.

Kim, W. C., I. S. Chin, K. S. Lee and J. Choi (2000). Analysis and reduced-order design of quadratic criterion-based iterative learning control using singular value decomposition. *Comp. and Chem. Eng.* **24**, 1781–2039.

Lee, J. H. and M. Gupta (2005). Period robust repetitive model predictive control. *Journal of Process Control.*

Lee, J. H., K. S. Lee and W. C. Kim (2000). Model-based iterative learning control with a quadratic criterion for time-varying linear systems. *Automatica* **36**, 641.

Lee, J. H., S. Natarajan and K. S. Lee (2001*a*). A model-based predictive control approach to repetitive control of continuous processes with periodic operations. *J. Process Control* **11**, 195–207.

Lee, K. S. and J. H. Lee (2000). Convergence of constrained model predictive control for batch processes. *IEEE Trans. on A.C.* **45**, 1928–1932.

Lee, K. S., H. J. Ahn, D. R. Yang and J. H. Lee (2003). Experimental application of a quadratic optimal iterative learning control method for control of wafer temperature uniformity in rapid thermal processing. *IEEE Trans. Semicond. Manuf.* **16**, 36.

Lee, K. S., J. H. Lee, I. S. Chin and H. J. Lee (1999). Model predictive control technique combined with iterative learning control for batch processes. *AIChE J.* **45**, 2175–2187.

Lee, K. S., J. Lee, I. S. Chin, J. Choi and J. H. Lee (2001*b*). Control of wafer temperature uniformity in rapid thermal processing using an optimal iterative learning control technique. *Ind. Eng. Chem. Res.* **40**, 1661.

Lee, K. S., S. H. Bang and K. S. Chang (1994). Feedback-assisted iterative learning control based on an inverse process model. *J. Process Control* **4**, 77–89.

Lee, K. S., W. C. Kim and J. H. Lee (1996). Model-based iterative learning control with quadratic criterion for linear batch processes. *J. Cont. Autom. Sys. Engng.* **2**, 148–157.

Lewis, F. L. and V. L. Syrmos (1995). *Optimal Control.* John Wiley and Sons. New York.

Lucibello, P. (1992). Learning control of linear systems. *Proc. of Amer. Control Conf., Chicago, IL* pp. 1888–1892.

Miller, R. C. and Jr. G. T. Mallick (1978). Method for controlling an automatic machine tool. *US Patent 4088899*.

Moore, K. L. (1993). *Iterative Learning Control for Deterministic System*. Springer-Verlag. New York.

Morari, M. and J. H. Lee (1999). Model predictive control: past, present, and future. *Comp. Chem. Eng.* **23**, 667–682.

Natarajan, S. and J. H. Lee (2000). Repetitive model predictive control applied to a simulated bed chromatography system. *Comp. Chem. Eng.* **24**, 1127–113.

Oh, S. R., Z. Bien and I. H. Suh (1988). An iterative learning control method with application for the robot manipulator. *IEEE J. Robot Autom.* **4(5)**, 508–514.

Overschee, P. Van and B. D. Moor (1994). 4sid: Subspace algorithms for the identification of combined deterministic-stochastic system. *Automatica* **30**, 75.

Qin, S. J. and T. A. Badgwell (1996). An overview of industrial model predictive technology. *Proc. CPC-V, Lake Tahoe, CA* pp. 232–256.

Sison, L. G. and E. K. P. Chong (1996). No-reset iterative learning control. *Proc. of IEEE Conf. on Decision and Control, Kobe, Japan* pp. 3062–3063.

Togai, M. and O. Yamano (1985). Analysis and design of an optimal learning control scheme for industrial robots: a discrete system approach. *Proc. 24th IEEE Conf. on Decision and Control, Ft. Lauderdale, FL* pp. 1399–1404.

Uchiyama, M. (1978). Formation of high speed motion pattern of mechanical arm by trial. *Trans. the Society of Instrum. and Control Engineers* **19**, 706–712.

Wilde, D. J. and C. S. Beightler (1967). *Foundations of Optimization*. Prentice-Hall. Englewood-Cliffs.

Zafiriou, E. and J. M. Zhu (1990). Optimal control of semi-batch processes in the presence of modeling error. *Proc. of Amer. Control Conf., San Diego, CA* pp. 1644–1649.

Zafiriou, E., R. A. Adomaitis and G. Gattu (1995). Approach to run-to-run control for rapid thermal processing. *Proc. of Amer. Control Conf., Seattle, WA* pp. 1286–1288.

Table 1: Performance of full-order Q-ILC. After 7 runs, the profiles converged. The initial wafer temperature, T_0 was dropped to 370^oC in the 8^{th} Run and then returned to 400^oC in the next run. (Temperature gap is defined to be the largest difference among the 8 temperatures. Mean square error here means $\frac{1}{8}\sum_{i=1}^{8}(y_i(t)-r)^2$. For both, 'Max' and 'Min' entries represent the maximum and minimum values over the course of a run. 'Mean' entry represents the average over time.)

	Temperature Gap			Mean Square Error		
Run	Max.	Min.	Mean	Max.	Min.	Mean
1^{st}	41.35	4.357	22.35	8250	2.170	3136
3^{rd}	17.49	4.707	10.97	563.8	3.739	77.29
7^{th}	8.116	1.893	4.968	91.80	1.667	14.97
8^{th}	9.880	4.429	7.030	880.5	2.433	97.62
9^{th}	7.613	4.189	6.265	95.20	1.821	7.374

Table 2: Comparison of the gap temperature between T_ω-model-based and T_ω^4-model-based Q-ILC.

	Temperature Gap		
Model	Max.	Min.	Mean
T_ω	12.02	5.371	7.740
T_ω^4	8.116	1.893	4.968

Table 3: Performance of Q-ILC with 3 point measurements and 3 point control. Results for (a) three controlled points and (b) 8 monitored points.

	Temperature Gap			Mean Square Error		
7^{th} *Run*	Max.	Min.	Mean	Max.	Min.	Mean
(a)	7.840	0.110	1.625	197.9	0.032	7.890
(b)	27.53	14.90	20.29	437.8	39.99	102.3

Table 4:Performance of Q-ILC with 3 point measurements and 8 point control under explicit inference of the unmeasured temperatures with different initial temperature,T_0.

8^{th} Run	Temp Gap			Mean Square Error		
$T_0/^oC$	Max.	Min.	Mean	Max.	Min.	Mean
350	10.96	5.553	7.776	2463	4.354	256.5
400	12.87	6.282	8.903	106.2	5.450	22.97
450	15.86	6.702	9.509	2516	4.601	422.1

Session 8.1

Optimization and Control of Petrochemical Systems

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

APPLICATION OF PLANTWIDE CONTROL TO LARGE SCALE SYSTEMS. PART I - SELF-OPTIMIZING CONTROL OF THE HDA PROCESS

Antonio Araújo * **Marius Govatsmark** *
Sigurd Skogestad *,[1]

* *Department of Chemical Engineering*
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

Abstract: This paper describes the application of self-optimizing control to a large scale process, the HDA plant. The idea is to select controlled variables which when kept constant lead to minimum economic loss. In order to avoid the combinatorial problem common to the selection of outputs/measurements for such large plants, applications of singular value decomposition (SVD) based methods are used which, although not guaranteeing optimality, give a consistent and practical way for selection. A controllability analysis is carried out to compare the dynamic performance of the selected sets of controlled variables and the conclusion is that the expected performances of the proposed control structures are essentially the same.

Keywords:
HDA process, self-optimizing control, selection of controlled variable, sequential design, SVD, RGA, controllability analysis.

1. INTRODUCTION

This paper deals with the selection of controlled variables for the HDA (hydrodealkylation of toluene) process. One of the main objective is to discuss different approaches to tackle the combinatorial control structure problem that can be found in such large-scale problems.

The HDA process, due to McKetta (1977), was first presented in a contest the American Institute of Chemical Engineers arranged for the industry to find enhanced solutions to typical design problems. It has been exhaustively studied by several authors, e.g Stephanopoulos (1984), Brognaux (1992), Cao and Rossiter (1997), Wolff (1994), Herrmann *et al.* (2003), Ng and Stephanopoulos (1996), Ponton and Laing (1993), Brekke (1999), Luyben *et al.* (1998), Luyben (2002), and Konda *et al.* (2005) (see Table 1) with different objectives, such as steady state design, controllability and operability of the dynamic model and control structure selection and controller design.

[1] Corresponding author: sigurd.skogestad@chemeng.ntnu.no

2. SELF-OPTIMIZING CONTROL

Definition: *Self-optimizing control is when one can achieve an acceptable loss with constant setpoint values for the controlled variables without the need to re-optimize when disturbances occur (real time optimization).*

To quantify this more precisely, we define the (economic) loss L as the difference between the actual value of the cost function and the truly optimal value, i.e.

$$L(u,d) = J(u,d) - J_{opt}(d) \quad (1)$$

Table 1. Selection of controlled variables.

Stephanopoulos (1984)								
Brognaux (1992), Cao *et al.*, Wolff (1994), and Herrmann *et al.* (2003)								
Ng and Stephanopoulos (1996)								
Ponton and Laing (1993)								
Brekke (1999)								
Luyben et al. (1998) and Luyben (2002)								
Konda *et al.* (2005)								
This paper								
Number of steady-state economic controlled variables	**7**	**6**	**5**	**7**	**12**	**9**	**8**	**14**
Fresh toluene feed rate	x			x			x	x
Recycle gas flow rate	x							
Recycle gas hydrogen mole fraction	x				x			x
Recycle gas methane mole fraction			x	x		x	x	
Compressor power					x			x
Compressor outlet pressure						x		
Total toluene flow rate to the reaction section						x		
FEHE by-pass flow rate					x			x
Reactor inlet temperature	x	x			x	x		
Separator temperature		x		x	x		x	x
Separator pressure		x	x	x	x		x	
Hydrogen to aromatics ratio at the reactor inlet		x					x	x
Hydrogen mole fraction in the reactor outlet				x				
Overall toluene conversion in the reactor							x	
Quencher flow rate	x							
Quencher outlet temperature						x	x	x
Purge flow rate	x							
Hydrogen mole fraction in the distillate of the stabilizer			x					
Benzene mole fraction in the distillate of the stabilizer				x	x			
Boil-up flow rate in the stabilizer					x			
Ratio benzene in feed to benzene in the distillate of the stabilizer								x
Product purity in the distillate of the benzene column	x	x	x		x	x	x	x
Production rate (benzene column distillate flow rate)		x			x			
Temperature in an intermediate stage of the benzene column						x		
Ratio toluene in feed to toluene in the bottom of the benzene column								x
Toluene mole fraction in the bottom of the benzene column								x
Ratio benzene in feed to benzene in the bottom of the benzene column								x
Ratio toluene in feed to toluene in the distillate of the toluene column					x			x
Toluene column reflux drum level						x		
Temperature in an intermediate stage of the toluene column						x		
Distillate flow rate from the toluene column					x			
Toluene mole fraction in the bottom of the toluene column				x				
Toluene mole fraction in the distillate of the toluene column			x					x

NB! In addition, separator level, pressure and reflux drum and bottom sump levels of all columns are controlled.

Truly optimal operation corresponds to $L = 0$, but in general $L > 0$. A small value of the loss function L is desired as it implies that the plant is operating close to its optimum. The main issue here is not to find the optimal set points, but rather to find the right variables to keep constant. The precise value of an acceptable loss must be selected on the basis of engineering and economic considerations.

Skogestad (2000) recommends that a controlled variable c suitable for constant set point control (self-optimizing control) should have the following requirements:

- **R1.** The optimal value of c should be insensitive to disturbances, i.e., $c_{opt}(d)$ depends only weakly on d.
- **R2.** The value of c should be **sensitive** to changes in the manipulated variable u, i.e., the gain $y = Gu$ should be large (equivalently, because $\partial J^2/\partial^2 c = G^{-T}\ \partial J^2/\partial^2 u\ G^{-1}$, the optimum should be flat with respect to the variable c, i.e., $\partial J^2/\partial^2 c$ should be small).
- **R3.** For cases with two or more controlled variables, the selected variables in c should not be closely correlated.
- **R4.** The variable c should be easy to measure and control.

In the present paper, the loss is to be evaluated based on the maximization of the minimum singular value (Skogestad, 2000), which is a combination of the requirements above. This rule states that: *assuming each candidate controlled variable c has been scaled such that the expected variation in $c - c_{opt}$ is of magnitude 1 (including the effect of both disturbances and control error), then select the variables c that minimize the norm of G^{-1} (where G is the scaled steady state gain matrix formed by considering the unconstrained degree of freedom only), which in terms of the two-norm is the same as maximizing the minimum singular value of G, $\underline{\sigma}(G)$.* This condition is computationally attractive, but because it only provides local information (based on one equilibrium point), it can be very misleading in some cases, e.g. where the minimum occurs very close to infeasibility.

3. HDA PROCESS

The HDA process (see Figure 1) is used to manufacture benzene by thermal dealkylation of toluene. This is a high-temperature, noncatalytic process in which toluene and hydrogen react to form benzene and methane, with minor amounts of by-product. Excess hydrogen must be used to suppress side reactions and coke formation. The reaction products must be separated, by-products rejected, unreacted toluene recovered and recycled, and the benzene product clay treated and distilled to the proper level of purity.

The model used in this paper and implemented in $MATLAB^{TM}$ is a slight modified version of the model developed by Brognaux (1992) and later used by Wolff (1994). The difference lies in the introduction of a quencher to cool down the reactor effluent and in the adjustment of the equations used to describe the reactions, as pointed by Cao *et al.* (1998). A simplified model for the separation section is used for optimization purposes since it is assumed that the dynamics of the distillation train is much slower than the remaining of the plant (Brognaux, 1992). It is also assumed the distillation columns have large number of stages leading to high-purity products and product purity has little effect on the cost. In addition, three loops are closed for stability and flexibility reasons.

Details of the process model used in this paper are available on-line at Sigurd Skogestad's home page under "Publication list".

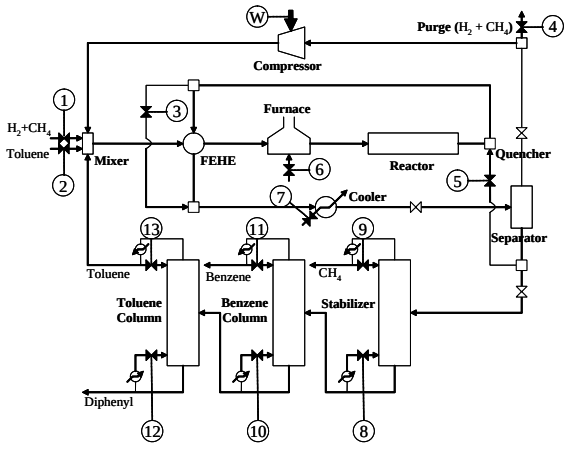


Fig. 1. HDA process flowsheet.

4. RESULTS

4.1 Step 1. Degree of freedom analysis

Table 2 summarizes the degree of freedom analysis for the HDA process considered in this paper. There it is shown the number of steady state operational degrees of freedom (N_{ss}) for the process units.

In addition, there are seven liquid levels (reflux drums and bottom sumps of the distillation columns and separator level) with no steady state effect (they must be stabilized) and the pressures in the distillation columns, all of them consume ten **dynamic** degrees of freedom. One steady state degree of freedom must be selected to keep the quencher outlet temperature at its equality constraint and we are left with thirteen degrees of freedom that can be used for steady state optimization.

Table 2. Steady state degree of freedom analysis

Process units	**Manipulations**	**DOF**
External feed streams: feed rate	Valves 1 and 2	2
Splitters: n-1 (n is the number of exit streams)	Valves 3 to 5	3
Compressor: duty	Source W	1
Adiabatic flash(*)	None	0
Gas phase reactor(*)	None	0
Heat exchangers: duties	Valves 6 and 7	2
Distillation columns: LV (or DB) configuration	Valves 8 to 13	6
Equality constraint		
Quencher outlet temperature		-1
Degrees of freedom at steady state		**13**

(*) No extra valve is assumed: pressure is "given" by the surroundings.

138 candidate controlled variables were identified for this process which gives $\binom{138}{13} = \frac{138!}{13!125!} = 5.9 \cdot 10^{17}$ (!) possible sets of controlled variables. Clearly, this number is intractable for any further computation/consideration. To reduce it, one has first to determine the active constraints which are to be controlled.

4.2 Step 2. Definition of optimal operation

The following profit function ($M\$/year$) based on Douglas (1988)'s economic potential (EP) is to be maximized:

$$-J = (p_{ben}D_{ben} + \sum_{i=1}^{nc} cF_{f,i}) - (p_{tol}F_{tol} + p_{gas}F_{gas} + p_{fuel}Q_{fuel} + p_{cw}Q_{cw} + p_{pow}W_{pow} + p_{stm}Q_{stm}) \quad (2)$$

subject to

1. Production rate:

$$D_{benzene} \geq 265\, lbmol/h \quad (3)$$

2. Hydrogen to aromatic ratio in reactor inlet:

$$\frac{F_{H_2}}{(F_{benzene} + F_{toluene} + F_{diphenyl})} \geq 5 \quad (4)$$

3. Bound on toluene feed rate:

$$F_{toluene} \leq 300 lbmol/h \quad (5)$$

4. Reactor pressure:

$$P_{reactor} \leq 500\, psia \quad (6)$$

5. Reactor outlet temperature:

$$T_{reactor} \leq 1300^o F \quad (7)$$

6. Quencher outlet temperature:

$$T_{quencher} = 1150^o F \quad (8)$$

7. Product purity in the benzene column distillate:

$$x_{D_{benzene}} \geq 0.9997 \quad (9)$$

8. Separator inlet temperature:

$$95^o F \leq T_{separator} \leq 105^o F \quad (10)$$

9. Mole fraction non-negative constraints:

All mole fractions corresponding to the products in the distillation columns are constrained to be non-negative, $x_{purity} \geq 0$.

In addition, all manipulated variables are bounded.

Note that it is assumed that all emissions (purge, stabilizer distillate, and toluene column bottom) are sold as fuel.

4.3 Step 3. Identification of important disturbances

The disturbances considered in this paper are listed in Table 3.

Table 3. Disturbances

Disturbance	**Unit**	**Nominal**	**Lower**	**Upper**
Bound on toluene feed flow rate	lbmol/h	300	285	315
Fresh toluene feed temperature	°F	100	80	120
Gas feed composition	mol% of H_2	95	90	100
Benzene price	$/lbmol	9.04	8.34	9.74
Gas feed temperature	°F	100	80	112
Inlet cooling water temperature to cooler	°F	59	50	70
Downstream pressure after the purge	psia	350	300	400
Energetic value of fuel to the furnace	MBTU/lbmol	0.1247	0.12	0.13
Relative volatility of hydrogen in stabilizer		36	32,4	39,6
Relative volatility of benzene in benzene column		2.7	2.43	2.97
Relative volatility of toluene in toluene column		10	9	11
Toluene recycle temperature	°F	212	202	230

4.4 Step 4. Optimization

Eight constraints are active at the optimal point, namely:

1. Product purity (lower bound)

2. Benzene mole fraction non-negative constraint in the distillate of stabilizer (lower bound)
3. Benzene mole fraction non-negative constraint in the bottom of benzene column (lower bound)
4. Toluene mole fraction non-negative constraint in the bottom of benzene column (upper bound)
5. Toluene mole fraction non-negative constraint in the distillate of toluene column (upper bound)
6. Fresh toluene feed rate (upper bound)
7. Separator inlet temperature (lower bound)
8. Hydrogen to aromatic ratio in reactor inlet (lower bound)

All of them must be controlled to achieve optimal operation, at least nominally (active constraint control).

Consequently, the number of unconstrained degrees of freedom is found to be **five**. This reduces the number of possible sets of controlled variables to $\binom{129}{5} = \frac{129!}{5!124!} = 275,234,400$. However, this is still too large a number to be further considered in the analysis.

4.5 Step 5. Identification of candidate controlled variables - local analysis

As seen before, the number of possible sets of controlled variables is very large and impossible to be handled. Skogestad and Postlethwaite (2005) note that the number of combinations has a combinatorial growth, so even a simple input-output controllability analysis becomes unmanageable if there are too many alternatives. One way of avoiding this combinatorial problem is to base the selection directly on the "big" linear model G_{all} of the plant (in the present case, G_{all} is the 129×5 matrix, where the active constraints and the sole equality constraint are not considered). One may consider the singular value decomposition and relative gain array of G_{all} as discussed later in this section. This rather crude analysis may be used, together with physical insight, rules of thumb and simple controllability measures, to perform a pre-screening in order to reduce the possibilities to a manageable number. These candidate combinations can then be analyzed more carefully.

The matrix G_{all} is scaled such that each candidate controlled variable has the expected variation from its optimal ($c - c_{opt}$) of magnitude 1 for all disturbances and the inputs all have the same effect on the cost function J (Skogestad and Postlethwaite, 2005). According to the maximum singular value rule, one should select controlled variables with large gains from the inputs to the outputs (maximization of the minimum singular value).

It is possible to find the optimal combination of outputs as outlined in Section 4.5.4 below. However, this is rather time consuming, so it will be first considered three related methods that do not require much use of computation. The three methods are all based on an SVD of $G_{all} = U_r \Sigma_r V_r^T$ (economy size SVD), where r represents the rank of G_{all}, and make use of the output singular vector U_r (in general, one wants to select outputs with large elements in U_r):

4.5.1. Sequential SVD selection: The idea is to select sequentially the output that corresponds to the largest element of the first column of U_r (corresponding to the largest singular value), remove this variable by closing the loop between this output and one input (the choice of the input does not matter for this analysis), and obtain the new matrix G_{all} with one input (and output) less until only one candidate controlled variable remains.

4.5.2. "One-shot" RGA selection: Another simple yet effective screening tool for selecting inputs and outputs, which avoids the combinatorial problem, is the relative gain array (RGA) of the "big" scaled transfer matrix G_{all} with all candidate inputs and outputs included, $\Lambda = G_{all} \otimes G_{all}^{\dagger_T}$ (where $\dagger$ is the pseudo-inverse operator). Essentially, the method is an SVD-type since the sum of the elements of row i in the RGA matrix is equal to the 2-norm of row i in U_r, i.e. $\sum_{j=1}^{n} \lambda_{i,j} = \left\| e_i^T U_r \right\|_2^2$ (Cao and Rossiter, 1997). So, it is preferred to select outputs corresponding to rows in the RGA where the sum of the elements is larger.

4.5.3. Sequential RGA selection: At each step in this method, the output with the largest RGA row sum is selected. This is the same as the previous method, except that it is done sequentially as for the sequential SVD method.

The results are shown in Table 4 and 5. The overall matrix with all outputs (the 129 outputs $\times 5$ inputs matrix) has $\underline{\sigma}(G_{all}) = 37$. The sequential SVD and sequential RGA plants both have the same set of unconstrained controlled variables.

4.5.4. Optimal selection: A branch-and-bound algorithm based on the maximization of the minimum singular value was used to calculate the optimal set(s) of controlled variables. Five sets were identified and their minimum singular values differ only slightly from the sequential SVD and RGA: all five sets have $\underline{\sigma}_{opt}(G_{5\times5}) = 14.89$. On the other hand, the optimal set s of variables differ quite a lot from the local methods (SVD and RGA). The optimal sets are also shown in Table 5.

Now, the selected sets of controlled variables are to be used in the evaluation of the loss.

4.6 Step 6. Evaluation of loss

The evaluation of the loss for alternative combinations of controlled variables is done by computing the loss imposed by keeping constant set-points when there are

Table 4. Selected controlled variables.

	Selected controlled variables
1	Hydrogen mole fraction in mixer outlet
2	Diphenyl mole fraction in mixer outlet
3	FEHE hot side outlet temperature
4	Flow rate through bypass in FEHE
5	Recycle gas hydrogen mole fraction
6	Compressor power
7	Separator pressure
8	Hydrogen mole fraction in benzene column bottom
9	Methane mole fraction in benzene column bottom
10	Toluene mole fraction in benzene column bottom
11	Diphenyl mole fraction in benzene column bottom
12	Boil-up flow rate in toluene column
13	Reflux flow rate in toluene column
14	Toluene mole fraction in toluene column distillate

Table 5. Results for the selection of outputs.

Method	Variables	Minimum singular value of the G_{5x5} matrix
Sequential SVD	4, 5, 6, 10, 14	13.91
"One-shot" RGA	3, 4, 6, 10, 11	$5.74 \cdot 10^{-8}$
Sequential RGA	4, 5, 6, 10, 14	13.91
Optimal selection - Set 1	1, 2, 3, 8, 12	14.89
Optimal selection - Set 2	1, 2, 7, 9, 12	14.89
Optimal selection - Set 3	1, 2, 7, 8, 12	14.89
Optimal selection - Set 4	1, 2, 9, 12, 13	14.89
Optimal selection - Set 5	1, 2, 8, 12, 13	14.89

disturbances or implementation errors. The average losses show basically no difference between the sets although the loss for the "one-shot" RGA is the largest one. The other authors' selection shown in Table 1 will give larger losses as their selections were not based on optimal assumptions.

4.7 Step 7. Final evaluation and selection

The analysis up to now has been based purely on steady state economics and nothing has been said about implementation of the proposed controlled variables. Obviously, this is also an important issue, as one choice of controlled variables might result in a system that is easy to control whereas another might result in serious control problems, for example, caused by unstable (RHP) zeros (the multivariable extension of inverse response behavior). The truly optimal approach would be to solve the entire problem as one big optimization problem, taking into account both economics and control. However, this is intractable for most real problems, and the approach taken in this paper is therefore preferred. Here, candidate sets of controlled variables with acceptable steady-state economics are firstly identified. The (input-output) controllability of the best alternative is then check. If it is acceptable, then a viable solution has been found. If it is not, the remaining candidate sets are checked. If none of these turns out to be controllable, then the requirements on the steady state economics must be relaxed and more candidate sets must be considered.

4.7.1. Controllability analysis of the eight 14×14 sets of controlled variables: A controllability analysis based on Skogestad and Postlethwaite (2005) on page 253 was carried out on each set of candidate outputs and essentially no performance difference was found.

5. DISCUSSION

As expected, benzene purity at the outlet of the process is kept at its bound for economic reasons. Moreover, fresh feed toluene is maintained at its maximum flow rate to maximize the profit ($D_{benzene} > 265$). The separator inlet temperature is kept at its lower bound in order to maximize the recycle of hydrogen and to avoid the accumulation of methane in the process. Luyben's rule of keeping all recycle loops under flow control seems to lose its meaning in this process since it is economically optimum to leave the recycle flows fluctuate.

The selection of disturbances used in this paper was based on the work by Brognaux (1992) and some heuristics. Alstad (2005) approaches this subject in a more systematic way aiming to optimize the solution of the problem. Not all disturbances are of importance ($\|g_d(j\omega)\|_2 \leq 1, \ \forall\omega$) in a steady state point of view. The change in the price of benzene is the most important disturbance considered, but in practice nothing can be done to mitigate it.

The number of measurements is really very large, 138, but in practice not all of them can be regarded for a possible use due to operational limitations or impediments, f. e. composition measurements are rather difficult and very costly. The engineer's judgment must come at this stage in order to specify the number of degrees of freedom that can really be considered for the analysis. This pre-screening can substantially reduce the dimension of the problem and thus the number of controlled variable combinations. But there might be situations where the remaining number of possibilities is still very large, in which case one can try to perform a local analysis (based on an equilibrium point) that can lead to a good selection which can be found optimal by using optimization technique like branch-and-bound algorithms or some sub-optimal approach like the SVD-based calculations used in this paper. They are not guaranteed to give the best solution but due to their practicality and ease of use, they become very attractive in practice.

In summary, all the selected sets generate stable (no RHP-poles) plants and inverse responses are not expected (no RHP-zeros). Moreover, input saturation is expected for set point changes but not for disturbance

rejection and it can be concluded that all alternatives, including the optimal selection of controlled variables, are equally easy to control.

From an implementation standpoint, "the best" set of variables to be controlled would be the one found by either the sequential SVD or RGA methods based on local analysis (corresponding to $\underline{\sigma}(G_{5\times5}) = 13.91$). One possible control structure is shown in Figure 2. It is assumed that all lower layer loops are closed (regulatory control layer), e.g. level of the reflux drums and bottoms of all distillation columns as well as of the separator.

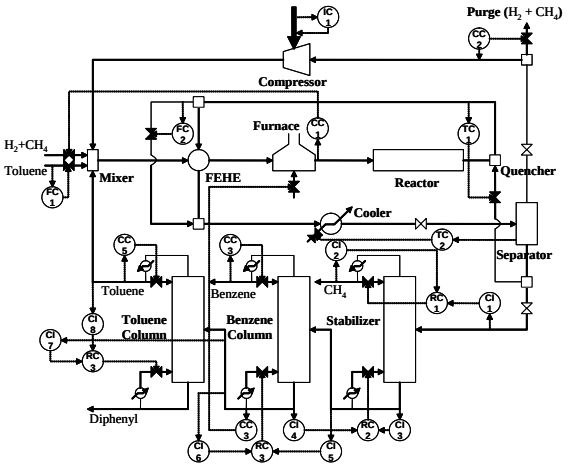


Fig. 2. Proposed control structure with controlled variables determined by the sequential SVD method.

6. CONCLUSIONS

This paper discussed the selection of controlled variables for the HDA process using the self-optimizing control procedure. The large number of variable combinations made it a very challenging problem in the sense that new approaches had to be used to decide for suitable outputs. Eight candidate sets were found by local analysis based on the SVD of the "big" scaled linear plant G_{all}. These easy-to-use tools for the selection of outputs produced a sub-optimal result which is not far away from the optimum found by applying integer optimization methods, namely branch-and-bound techniques. A controllability analysis showed that the dynamic performances of the proposed sets of controlled variables were essentially the same.

REFERENCES

Alstad, V. (2005). Studies on selection of controlled variables. PhD thesis. Norwegian University of Science and Technology, Trondheim, Norway.

Brekke, K. (1999). Optimization and control of the HDA process. Master thesis. Norwegian University of Science and Technology, Trondheim, Norway.

Brognaux, C. (1992). A case study in operability analysis: The HDA plant. Master thesis. University of London, London, England.

Cao, Y. and D. Rossiter (1997). An input prescreening technique for control structure selection. *Computers and Chemical Engineering* **21**(6), 563–569.

Cao, Y., D. Rossiter, D.W. Edwards, J. Knechtel and D. Owens (1998). Modelling issues for control structure selection in a chemical process. *Computers and Chemical Engineering* **22**(Suppl.), S411–S418.

Douglas, J. M. (1988). *Conceptual Design of Chemical Processes*. McGraw-Hill. USA.

Herrmann, G., S. K. Spurgeon and C. Edwards (2003). A model-based sliding mode control methodology applied to the hda-plant. *Journal of Process Control* **13**, 129–138.

Konda, N. V. S. N. M., G. P. Rangaiah and P. R. Krishnaswamy (2005). Simulation based heuristics methodology for plant-wide control of industrial processes. In: *Proceedings of 16th IFAC World Congress*. Praha, Czech Republic.

Luyben, W. L. (2002). *Plantwide dynamic simulators in chemical processing and control*. Marcel Dekker, Inc.. New York, USA.

Luyben, W. L., B. D. Tyréus and M. L. Luyben (1998). *Plantwide process control*. McGraw-Hill. USA.

McKetta, J. J. (1977). *Benzene design problem*. Encyclopedia of Chemical Processing and Design. Dekker, New York, USA.

Ng, C. and G. Stephanopoulos (1996). Synthesis of control systems for chemical plants. *Computers and Chemical Engineering* **20**, S999–S1004.

Ponton, J.W. and D.M. Laing (1993). A hierarchical approach to the design of process control systems. *Trans IChemE* **71**(Part A), 181–188.

Skogestad, S. (2000). Plantwide control: The search for the self-optimizing control structure. *Journal of Process Control* **10**, 487–507.

Skogestad, S. and I. Postlethwaite (2005). *Multivariable Feedback Control: Analysis and Design*. John Wiley & Sons. Chichester, UK.

Stephanopoulos, G. (1984). *Chemical process control*. Prentice-Hall International Editions. New Jersey, USA.

Wolff, E. A. (1994). Studies on control of integrated plants. PhD thesis. Norwegian University of Science and Technology, Trondheim, Norway.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DYNAMIC REAL-TIME OPTIMIZATION OF A FCC CONVERTER UNIT

Euclides Almeida Neto[1], Argimiro R. Secchi[2]

Grupo de Modelagem, Simulação, Controle e Otimização de Processos (GIMSCOP)
Departamento de Engenharia Química – Universidade Federal do Rio Grande do Sul
Rua Sarmento Leite, 288/24 – CEP: 90050-170 – Porto Alegre –RS – Brazil
Phone: +55-51-3316-3528 – Fax: +55-51-3316-3277
E-mail: {[1]eanet,[2]arge}@enq.ufrgs.br

Abstract: Fluidized-bed Catalytic Cracking (FCC) is a process subject to frequent variations in the operating conditions and changes in the feed quality and feed rate, due to the attempts to maximize LPG and gasoline. This fact makes the FCC converter unit an excellent opportunity for real-time optimization. The present work aims to apply a dynamic real-time optimization (D-RTO) into a simulation of an industrial FCC converter unit, using a mechanistic dynamic model. The algorithms that solve D-RTO problems need to deal with large-scale problems due to the full or partial system discretization along the optimal trajectory. In this work a simultaneous approach, present in the IPOPT solver, was used to discretize the system and solve the resulting large-scale NLP problem.

Keywords: Dynamic Optimization, FCC Orthogonal Collocation, D-RTO.

1. INTRODUCTION

The Fluidized-bed Catalytic Cracking Unit (FCC) is one of the most profitable process units of a petroleum refinery. The FCC converter is part of the reaction section of the unit, where it transforms the low-value raw-materials into commercial products of high-aggregated value.

The FCC converter is a flexible equipment, where the operating conditions can be adjusted to obtain higher yields of LPG (liquefied petroleum gas). When the price of the gasoline is favorable, it can be adjusted to maximize the yield of cracked naphtha, whereas LCO (light oil of recycle) is maximized when the "spread" is favorable to the diesel production. Due to the high profitability of this unit, it should be used its maximum capacity, operating at its maximum feed rate, pushing the big machines, as the gas compressor and air blower, to their upper limit.

Frequent operating-points transitions occur in the converter due to variations in the feed quality, such as variations in the raw-material quality or blends of different streams (coke of gasoil, naphtha, or atmospheric residue) to compose the feed. Frequent changes also happen in the production planning, moving LCO for gasoline or gasoline for LPG, in order to maximize the profitability of the supply chain of the refinery. This process unit is also subject to disturbances in the environmental conditions and limitations of equipments capacity in other process areas.

These facts suggest the use of dynamic real-time optimization (D-RTO) of this system trying to find interesting solutions to optimize the unit, subject to frequent changes in the process operating conditions and production planning. This study seeks to analyze the benefits and limitations of applying dynamic optimization to address this kind of problem. Besides, the critical factors of success of the use of D-RTO should be evidenced to obtain the whole financial potential of this process unit.

Control and optimization of FCC converters has been subject of many studies. The optimization of these processes has been made through MPC's (Odloak *et al.* 1995) and steady-state RTO's (Chitnis and Corropio 1998; Zanin *et al*., 2000a). NMPC has also been applied (Ali and Elnashaie, 1997) and other strategies of RTO, as optimization in the same layer of advanced control (Odloak *et al*., 2002; Gouvêa and Odloak, 1998). Zanin *et al.* (2000b) made a comparative study of the use of different optimization strategies in FCC converters. Recently Kadam *et al*. (2005) have been studying dynamic optimization using as example an FCC unit.

2. PROCESS MODEL

The FCC conversion area is composed by a furnace for feed pre-heating, a system of reactor-regenerator, air blower, main fractionating tower, and gas compressors. The cracking process constitutes of breaking heavy molecules in a high-temperature tubular reactor, producing fuel gas, LPG, cracked naphtha (gasoline), LCO, decanted oil and coke. During the reaction, deposit of coke occurs in the catalyst surface causing its deactivation and, therefore, its regeneration is mandatory, making part of the process. During the regeneration process there is a heat recovery used to heat the feed up to the cracking reaction temperature.

The FCC converter model used in this work, and developed by Secchi *et al*. (2001), is constituted of the following parts: riser model, separator model, gas compressor model, regenerator model, and valves and controllers models. These models describes a FCC UOP stacked converter, Figure 1, used by PETROBRAS in the Alberto Pasqualini refinery (REFAP S/A). The model was adjusted to the operating conditions of this process unit, describing reasonable well its dynamic behavior.

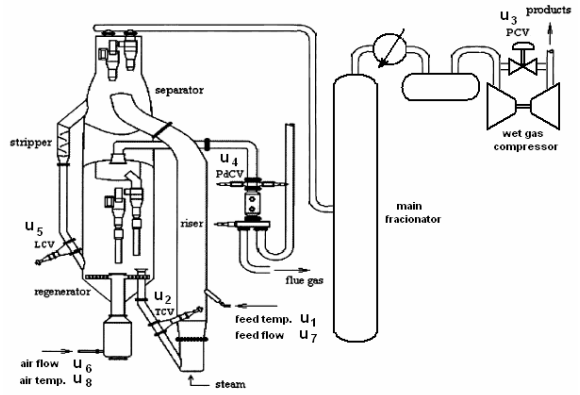


Fig. 1 FCC UOP Stacked Converter (Secchi *et al.*, 2001).

2.1 Riser Model

The Riser is modeled as an adiabatic plug flow reactor, with the kinetics described by the ten lumps model of Jacob *et al*. (1976), and using catalyst deactivation and coke formation tendency functions.

The dynamic model of the riser is represented by the mass balance of each lump and coke, using the reaction kinetics of formation of each species, and the energy balance. The resulting partial differential equation was discretized using backward finite-difference technique, with a log-scale non-uniform mesh. A mesh of 20 points was shown satisfactory.

2.2 Separator Model

The separator is assumed to be a continuous stirred tank, where catalyst and vapor products (hydrocarbons) are separated. The model of this equipment, based on mass and energy balances, focuses on the prediction of the catalyst level in the separator, the coke content in the spent catalyst, and the catalyst temperature in the separator. The pressure dynamics in the separator is established by a momentum balance.

2.3 Gas Compressor Model

The gas compressor is modeled as a single stage centrifugal compressor, driven by a constant speed. The polytropic flow model predicts the suction pressure of the compressor that establishes the pressure in the main fractionating tower and in the separator. There is a recycle stream around the compressor to control the suction pressure, and the mass balance is given by assumed dynamics.

2.4 Regenerator Model

The catalyst regeneration is carried out by burning the coke in the catalyst in a fluidized-bed reactor. The fluidized bed is modeled as emulsion and bubble phases that exchange mass and heat. The bubble phase is assumed to be at the pseudo steady-state condition. The disengagement section is modeled as two serial continuous well-mixed tank reactors, corresponding to the diluted and flue gas phases, according to the Figure 2.

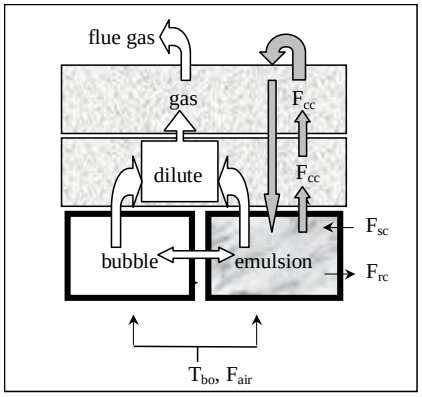


Fig. 2 Regenerator phases (Secchi *et al.*, 2001).

In the regeneration kinetics is used the assumption that the combustion reactions of coke occur in the emulsion, diluted, and gas phases. Component mass balances for O_2, CO, CO_2, H_2O, and coke describe the dynamic behavior of these reactions, resulting in five state equations for each phase of the regenerator. The catalyst inventory in the regenerator is modeled by the overall mass balance for catalyst. The pressure change behavior in the regenerator is obtained through the global mass balance in the gas phase. To predict the dynamic behavior of the temperatures in the regenerator, energy balance was applied in each phase. Considering that the catalyst loss in the regenerator is negligible, the whole catalyst that enters in the regenerator is accumulated or sent to the riser. The coke content in this catalyst is burned mainly in the emulsion phase, but it also suffers reaction in the diluted and gas phases.

2.5 Valves and Controllers Models

This type of FCC converter has four degrees of freedoms to provide stability to the system. These degrees of freedom are eliminated by placing regulatory PI controllers in their respective positions:

- Compressor suction pressure controller, using a control valve (PCV) in the compressor recycle stream;
- Reaction temperature controller, using a control valve (TCV) in the stand-pipe to the riser;
- Pressure drop controller between the reactor and the regenerator, using a control valve (PdCV) in the hole chamber of the combustion gases of the regenerator;
- Catalyst level controller in the separator, using a control valve (LCV) in the stand-pipe to the regenerator;

The dynamics of the valves openings are determined by their respective time constants. Additionally, each PI controller has one state equation to describe the integral action. The reaction temperature control will be done by the dynamic optimizer, through a supervisory action directly on the slide-valve. Therefore, only three PI controllers were used.

2.6 Empirical Correlations for Product Yields

The FCC converter model does not supply directly all the outputs of interest to analyze and optimize the process. Usually the predictions of products yield and conversion are important to carry out these studies. Empiric correlations were used to supply such desired information. In this case, the volumetric conversion and the yields of fuel gas, LPG, gasoline (GLN), light oil of recycle (LCO), decanted oil (OCLA) and coke (CK).

3. FORMULATION OF DYNAMIC OPTIMIZATION PROBLEM

The dynamic optimization problem of a process has the following general form:

$$\min_{u(t)} \varphi\left(z(t), y(t), u(t)\right) \tag{1}$$

subject to:

Dynamic Model (EDO):

$$F\left(\frac{dz(t)}{dt}, z(t), y(t), u(t)\right) = 0 \tag{2}$$

Algebraic Equations (EA):

$$G\left(z(t), y(t), u(t)\right) = 0 \tag{3}$$

Initial Conditions:

$$z(0) = z^0 \tag{4}$$

Bounds:

$$z^L \le z(t) \le z^U$$
$$y^L \le y(t) \le y^U$$
$$u^L \le u(t) \le u^U \tag{5}$$

3.1 Solution of Dynamic Optimization Problem

The infinite dimension dynamic optimization problem can be solved through variational methods, using the Pontryagin's maximum principle and solving the resultant two-point boundary value problem (TPBVP), or approximating to a finite formulation, with predefined functional forms for the control variables. In this last case, the resultant NLP problem can be solved by sequential or simultaneous approaches. In the sequential approach only the control variables are discretized or parameterized, while in the simultaneous approach the whole system is discretized in the time domain, usually using orthogonal collocation techniques. See the work of Biegler *et al*. (2002) for a deeper revision on these methods.

In this work the simultaneous strategy has been used, where the continuous problem is converted in a nonlinear programming problem (NLP) when approximating the state and control profiles by a family of orthogonal polynomials on finite elements (Cervantes, 1998). For first order differential equations the approximation results in:

$$z(t) = z_{i-1} + h\sum_{q=1}^{ncol} \Omega_q\left(\frac{t - t_{i-1}}{h_i}\right)\frac{dz}{dt_{i,q}} \tag{6}$$

The control profiles and algebraic equations are approximated in a similar way and the equation takes the following form:

$$y(t) = \sum_{q=1}^{ncol} \Psi_q\left(\frac{t - t_{i-1}}{h_i}\right) y_{i,q} \tag{7}$$

$$u(t) = \sum_{q=1}^{ncol} \Psi_q\left(\frac{t - t_{i-1}}{h_i}\right) u_{i,q} \tag{8}$$

Usually, in dynamic optimization problems the number of control variables is small and the number of state variables is very large. In this case the rSQP algorithm (reduced SQP) is efficient (Waanders *et al*., 2002). The solution of these problems is also efficient using the interior point algorithms, however they require improvements, and many of them have been proposed. The following ones can be highlighted: the use of the preconditioned conjugated gradient method (PCG) to update the control variables (Cervantes and Biegler, 2001), and the introduction of a filter in the strategy of the line search, where the objective function compete with the infeasibility of the problem (Wächter, 2002). In the interior point algorithm the original NLP problem can be written as:

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & c(x) = 0 \\ & x \ge 0 \end{aligned} \tag{9}$$

The barrier function is added to reduce the dimension of the problem, and then the problem takes the form:

$$\min \quad \varphi_\mu(x) = f(x) - \mu \sum_{i=1}^{n} \ln\left(x^i\right) \qquad (10)$$
$$s.t. \quad c(x) = 0$$

All of these features were implemented in the IPOPT algorithm developed by Carnegie Mellon University (CAPD Report, 2003; Lang and Biegler, 2005).

3.2 Configuration of the Objective Function

In the optimization of FCC converters there are some concurrent production objectives. The maximization of the operational profit is a common objective; however there are moments where some specific product needs to be maximized. This is due to the optimization of the refinery supply chain. There are situations where the local optimum of an isolated unit of process is not the global optimum of the supply chain. In order to attend these situations, multiple objectives can be adopted. In this dynamic optimization problem, the integral of different factors along a day (t_f = 24 h) were maximized. The most common situations are the following ones:

Maximization of the operational profit:

This is the most common objective function; however it is more difficult for the operators to analyze the results from the optimizer.

$$Profit = Revenue - Costs$$

$$Revenue = m_{FG}\,\mathrm{Pr}_{FG} + m_{CK}\,\mathrm{Pr}_{CK} + V_{LPG}\,\mathrm{Pr}_{LPG} + V_{GLN}\,\mathrm{Pr}_{GLN} + V_{LCO}\,\mathrm{Pr}_{LCO} + V_{OCLA}\,\mathrm{Pr}_{OCLA} \qquad (11)$$

$$Costs = V_{Feed}\,\mathrm{Pr}_{Feed} + m_{Cat}\,\mathrm{Pr}_{Cat} + Q_{PreH}\,\mathrm{Pr}_{Q} + Q_{Proc}\,\mathrm{Pr}_{Fuel} + Pot_{Blwr}C_{Blwr} + Pot_{Compr}C_{Compr} \qquad (12)$$

$$FObj_1 = -\int_0^{t_f} Profit.dt \qquad (13)$$

Maximization of the total conversion:

The maximization of the total conversion leads to the use of the maximum capacity of the converter, breaking the molecules into more important products. The disadvantage of this approach is that the conversion does not focus in highest price products. The average conversion is calculated in the following form:

$$FObj_2 = -\frac{\int_0^{t_f} Conv_v V_{Feed}.dt}{\int_0^{t_f} V_{Feed}.dt} \qquad (14)$$

Maximization of LPG production:

This objective is adopted when there is a clear advantage in the maximum conversion in LPG product.

$$FObj_3 = -\int_0^{t_f} \frac{\eta_{LPG}}{100} V_{Feed}.dt \qquad (15)$$

Maximization of gasoline production (GLN):

This objective is adopted when there is a clear advantage in the maximum conversion in gasoline.

$$FObj_4 = -\int_0^{t_f} \frac{\eta_{GLN}}{100} V_{Feed}.dt \qquad (16)$$

Maximization of LCO production:

The maximization of production of LCO is adopted when there is a clear advantage in the maximum conversion in LCO. In this case the LCO is an intermediary product and it is incorporated to the diesel pool. When LCO is a diluent, there is only interest in maximize it when displace some part of kerosene from the diluent pool to the jet fuel pool.

$$FObj_5 = -\int_0^{t_f} \frac{\eta_{LCO}}{100} V_{Feed}.dt \qquad (17)$$

The specific productions objectives are mutually exclusive. When you need to maximize the production of a specific stream, the weights of the other objectives should be zero.

Objective function formulation:

In general case each production objective can be represented in the following way:

$$FObj_i = -\int_0^{t_f} OBJ_i.dt \qquad (18)$$

The multi-objectives problem can be written as a weighted sum of each specific objective:

$$\varphi = \sum_{i=1}^{n} k_i FObj_i \qquad (19)$$

The constraint based multi-objectives strategies (ε-constrained and goal attainment) will be studied in future works, and was partially adopted here.

The integral in each specific objective is manipulated by differentiating the original objective function and creating a new state φ added to the set of differential equations. Therefore, the objective function assumes the following form:

$$\frac{d\varphi}{dt} = -\sum_{i=1}^{n} k_i OBJ_i \qquad (20)$$

3.3 Additional Constraints

Besides the constraints usually imposed to the states and the control variables, supplementary constraints were added that represent bounds in the production objectives to guarantee the feasibility of the solution in the optimization problem. The additional constraints are the following ones:

Minimum daily profit:

$$Profit \geq Profit_{\min} \tag{21}$$

Minimum conversion in the riser:

$$Conv_v \geq Conv_{\min} \tag{22}$$

Minimum and maximum LPG production:

$$V_{LPG}^{\min} \leq \frac{\eta_{LPG}}{100} V_{Feed} \leq V_{LPG}^{\max} \tag{23}$$

Minimum and maximum GLN production:

$$V_{GLN}^{\min} \leq \frac{\eta_{GLN}}{100} V_{Feed} \leq V_{GLN}^{\max} \tag{24}$$

Minimum and maximum LCO production:

$$V_{LCO}^{\min} \leq \frac{\eta_{LCO}}{100} V_{Feed} \leq V_{LCO}^{\max} \tag{25}$$

4. CASE STUDIES AND RESULTS

The dynamic optimization problem has been solved applying the IPOPT algorithm through the software of dynamic optimization DynoPC developed by Carnegie Mellon Univ. (Lang and Biegler, 2005). The several alternatives of production objectives studied in this work are presented in Table 1.

Table 1. Case studies.

Case	***Production Objective***
1	Maximum Profit
2	Maximum Feed Rate
3	Maximum Conversion
4	Maximum LPG Production
5	Maximum Gasoline (GLN) Production
6	Maximum LCO Production
7	Maximum Profit with Max. Conversion
8	Maximum Profit with Max. Feed Rate
9	Maximum Profit with Max. LPG Prod.
10	Maximum Profit with Max. GLN Prod.
11	Maximum Profit with Max. LCO Prod.

The maximization of the profit is the more common production objective, and it is usual to associate it to a specific objective, as maximum conversion or some product that the scheduling people defines as priority. This prioritization can also be made putting bounds in secondary objectives, for example, the minimum conversion (ε-constrained approach).

4.1 *Dimension of the Optimization Problem*

As the definition of the problem described above, the number of variables involved in the problem are given in Table 2.

Table 2. Number of variables in the formulation.

Number of differential variables (nz)	274
Number of algebraic variables (ny)	21
Number of control variables (nu)	8
Number of finites elements (ne)	40
Number of collocation points/element (ncol)	3
Total number of discretized variables	47642
Total number of constraints	47594
Total number of lower bounds	14440
Total number of upper bounds	14440

4.2 *Case Studies*

Due to space limitation, the obtained results are analyzed for the maximization feed rate case. The optimization problem was solved in an Intel Pentium IV, 2.8 MHz computer and spent 30 - 45 min of CPU time. This time consumption is compatible with the interval per control action (around 4 to 6 per day).

To obtain accurate and numerically stable results, it was necessary to tune the discretization parameters as the number of collocation points and finite elements. Also, in order to reduce the number of control actions some finite elements were grouped (Lang and Biegler, 2005). This procedure provided a more robust solution of the optimization problem and with a better performance.

Case 2 – Maximum Feed Rate

The maximization of the feed rate is prioritized when it is necessary to use the total capacity of the process unit. In this case, it is reached the limit of catalyst circulation or the limit of capacity of a main machine (air blower or gas compressor). Notice that the optimizer increased the feed flow rate, opened the catalyst valve to the maximum, dropped the suction pressure of the gas compressor, and reduced the pressure drop between the reactor and regenerator (Figs. 3 to 6). In order to supply the additional energy demanded by the system, the regenerator and riser temperatures were increased (Figs. 7 and 8).

As result of the dynamic optimization, the profit operation was increased by the order of 5.5 thousand dollars a day ($0.20/bbl). It is the normal potential of benefit with the advanced control and RTO applications (Fig. 9). It also can be observed that there was an increase of volumetric conversion (Fig. 10) and yields of gasoline (GLN) and LCO and a reduction in the decanted oil yield (OCLA), which is a less valuable product (Fig. 11).

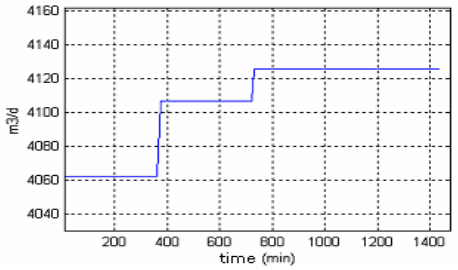


Fig. 3 Feed flow rate (u_1).

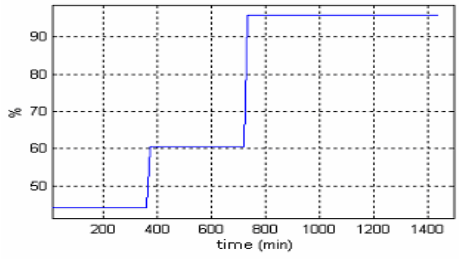


Fig. 4 TCV control signal (u_2).

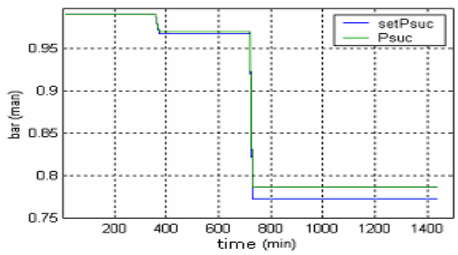


Fig. 5 Suction pressure of gas compressor (u_3).

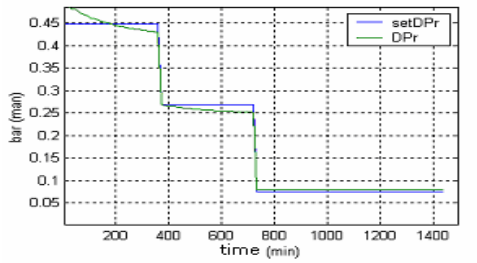


Fig. 6 Differential pressure reactor/ regenerator (u_4).

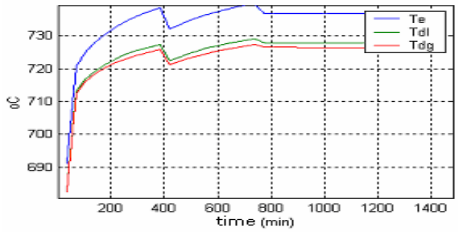


Fig. 7 Regenerator's temperatures.

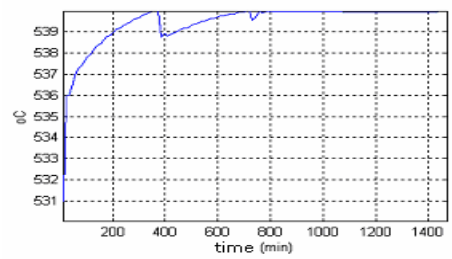


Fig. 8 Riser temperature (reaction).

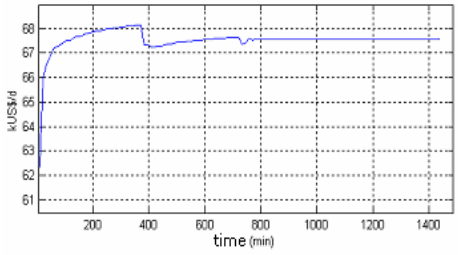


Fig. 9 Operation profit.

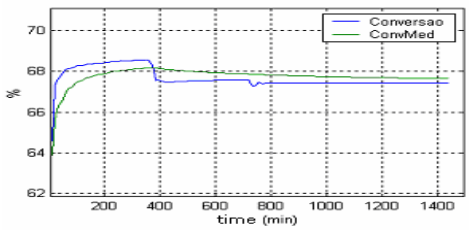


Fig. 10 Volumetric conversion.

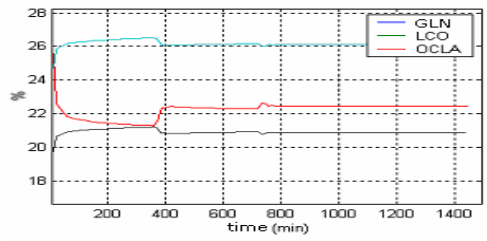


Fig. 11 Volumetric yields.

5. CONCLUSIONS

The dynamic optimization of the FCC converter has obtained coherent results with the expected in an industrial unit. The results demonstrate that the application of D-RTO in this kind of unit can bring significant benefits. The simultaneous approach has shown to be effective for the solution of the problem, but it demanded a lot of time to tune the discretization parameters of the control variables. The strategy of grouping intervals for the control variables was the one that presented better performance. Due to space limitation, the other analyzed cases will be presented in the symposium.

REFERENCES

Ali, E.E.. and S.S.E.H Elnashaie (1997). Non-linear Model Predictive Control of Industrial type IV Fluid Catalytic Cracking (FCC) units for maximum gasoline yield. *Ind. Eng. Chem. Res.*, **36**, 389-1007.

Biegler, L.T., A.M. Cervantes and A. Wächter (2002). Advances in Simultaneous Strategies for Dynamic Process Optimization. *Chem. Engng Sc.*, **57,** 575–593.

CAPD *Report* (2003). Center for Advanced Process Decision-Making. *CAPD Report*. March.

Cervantes, A. and L.T.Biegler (1998). Large-Scale EAD Optimization Using a Simultaneous NLP Formulation. *AIChE J*., **44** (5), 1038-1050.

Cervantes, A. and L.T. Biegler (2001), Optimization Strategies for Dynamic Systems. *Encyclopedia of Optimization*. Kluwer, **4**, 216-227.

Chitnis, U.K. and A.B. Corropio (1998). On-line optimization of a Model IV catalytic cracking unit. ISA Trans., **37**, 215-226.

Gouvêa, M.T. and D. Odloak (1998). One-layer real time optimization in the FCC unit: procedure, advantages and disadvantages. *Comp. Chem. Engng.*, **22**, S191-S198.

Jacob S.M., B. Gross, S.E. Voltz and V.M. Weekman (1976). A Lumping and Reaction Scheme for Catalytic Cracking. *AIChE J*., **22** (4), 701-713.

Kadam, J., M. Schlegel, B. Srinivasan, D. Bonvin and W. Marquardt (2005). Dynamic Real-Time Optimization: from off-line Numerical Solution to Measurement-based Implementation. *IFAC World Congress*, Prague.

Lang, Y.D. and L. T. Biegler (2005). A Software Environment for Simultaneous Dynamic Optimization, submitted to *Comp. Chem. Engng*.

Odloak, D., A.C. Zanin, and M.T.D. Gouvêa (2002). Integrating real-time optimization into the model predictive controller of FCC. *Control Engng. Practice*, Londres, **10** (8) 819-831.

Odloak, D., L.F.J.R. Moro and D. Spandri (1995). Constrained Multivariable Control Of Fluid Catalytic Cracking Converters, A Practical Application. In: *AIChE Spring Meeting*, Houston. II., 84-90.

Secchi, A.R., M.G. Santos, G.A. Neumann and J.O. Trierweiler (2001). A Dynamic Model for a FCC UOP Stacked Converter Unit. *Comp. Chem. Engng*., **25**, 851-858.

Waanders, B.B., R. Bartlett, K. Long, P. Boggs and A. Salinger (2002). Large-Scale Non-Linear Programming for PDE Constrained Optimization. *SAND2002-3198,Sandia National Laboratories*.

Wächter, A. (2002). An Interior Point Algorithm for Large-Scale Nonlinear Optimization with Applications in Process Engineering, *Ph.D. thesis*.

Zanin, A.C., M.T. Gouvêa and D. Odloak (2000a). Industrial implementation of a real-time optimization strategy for maximizing production of LPG in a FCC unit. *Comp. Chem. Engng.*, **24**, 525-531.

Zanin, A.C., M.T. Gouvêa and D. Odloak (2000b). Comparing different real-time optimization strategies for the FCC catalytic converter. In: *ADCHEM 2000*, Pisa.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

INFERENTIAL CONTROL BASED ON A MODIFIED QPLS FOR AN INDUSTRIAL FCCU FRACTIONATOR

TIAN Xuemin*, TU Ling, and DENG Xiaogang

College of Information and Control Engineering, China University of Petroleum, Dongying, Shandong 257061, China

Abstract: A modified Quadratic Partial Least Squares (MQPLS) algorithm based on nonlinear constrained programming is proposed. Sequential Unconstrained Minimization Technique (SUMT) is employed to calculate the outer input weights and the parameters of inner relationship. It was found that MQPLS can not only explain more of the underlying variability of the data, but also has improved modelling and predictive ability. An inferential control system is implemented on the Distribute Control System (DCS) of a fluid catalytic cracking unit (FCCU) main fractionator. A soft sensor MQPLS–based was developed to estimate solidifying point of diesel oil. The controller was established via constrained Dynamic Matrix Control (DMC) algorithm. Real time application results demonstrated the performance of the inferential control system based on MQPLS was much better than the original tray temperature control system. This resulted in a 1.0% increase in production rate, and a significant increase in profit. *Copyright © 2006 IFAC*

Keywords: Partial Least Squares, Soft Sensor, Dynamic Matrix Control, Inferential Control.

1. INTRODUCTION

Many variables, which characterize the 'quality' of the final product in chemical processes, are often difficult to measure in real-time, and hence cannot be directly used in a feedback control. Most online quality analyzers, like gas chromatographs and NIR (Near-Infrared) analyzers, suffer from large measure delays and high investment and maintenance costs. Under these circumstances, a common alternative is to set up soft sensors to infer the product properties (primary variables) by employing some auxiliary measurements (secondary variables), and then build an inferential control scheme.

Statistic regression techniques have been extensively used in establishing soft sensing models from historical data. Among other related regression techniques, PLS has been proved to be a powerful tool for problems where data is noisy and highly correlated and where there are only a limited number of observations (Berglund and Wold, 1997; MacGregor et al., 1991). The power of PLS lies in the fact that it projects the input-output data down into a latent space, extracting a number of principle components with an orthogonal structure, while capturing most of the variance in the original data. Therefore, PLS can overcome the limitation that when dealing with highly correlated multivariate data, the traditional Least Squares (LS) regression will result in singular solution or imprecise parameter estimations.

However, in many practical situations, industrial processes exhibit significant nonlinear behaviors. As a linear regression method, PLS is inappropriate for modeling nonlinear systems.

Hence various kinds of nonlinear PLS (NLPLS) methods have been proposed in the literature which extend the PLS model structure to capture non-linearities of systems. A successful step towards

*Corresponding author
Tel：+86-546-8391245
Email address: tianxm@hdpu.edu.cn

nonlinear PLS modeling was the quadratic PLS (QPLS) proposed by Wold et al. (1989). In QPLS, second order polynomial (quadratic) regression is used to fit the function between each pair of input and output score vector, namely, the inner relation. The other 'generic' nonlinear PLS (NLPLS) such as spline PLS (SPLS) (Wold, 1992), neural networks PLS (NNPLS) (Qin and McAvoy, 1992) and Fuzzy PLS (FPLS) (Yoon et al., 2003) were developed. As their names suggest, SPLS uses spline function (quadratic or cubic) as inner model and NNPLS uses neural networks inner model. FPLS uses TSK (Takagi-Sugeno-Kang) fuzzy model as the inner model. All the algorithms above are developed from the nonlinear iterative partial least squares (NIPALS) algorithm (Geladi and Kowalski, 1986), which is called the 'engine' of the PLS methodology.

The problem of the input weight updating in NLPLS was firstly considered by Wold et al. (1989) and the benefit achieved by applying an updating procedure to the parameters of the NLPLS model was also proved. It has attracted the interests of many researchers. Especially, by modifying the input weight updating procedure of Wold et al., an error-based input weight updating approach was presented by G.baffi et al. (1999a, 1999b and 2000). In this paper, the input weight updating procedure is summarized to a constrained nonlinear optimal problem. Sequential Unconstrained Minimization Technique (SUMT) is utilized to calculate the outer input weights and the parameters of inner relation. It can make remedies of the shortcomings of the pseudo-inverse and large calculation burden that exist in the error-based input weight updating approach. Although this new kind of weight updating method is applicable to any nonlinear PLS algorithm, the new updating method is only combined with the original QPLS in this paper, leading to a modified quadratic Partial Least Squares (MQPLS) algorithm.

The paper is organised as follows. In Section 2, the basic principle of the NLPLS is introduced, and the error based input weight updating procedure by G. Baffi et al. (1999a) is briefly reviewed. Section 3 proposed a new input weight updating method and highlighted the details of the corresponding modified QPLS algorithm. Section 4 introduced the main structure of an inferential control system, in which the soft-sensor was built based on the modified QPLS to estimate diesel oil solidifying point, and the controller was established via a simplified Dynamic Matrix Control (DMC) algorithm. Section 5 gives the conclusions.

2. QUADRATIC PARTIAL LEAST SQUARES

PLS algorithm decomposes $\mathbf{X}$ and $\mathbf{Y}$ by projecting them to the directions (input weight w and output weight c) to extract several pair of input score vector t_h and output score vector u_h .The decomposition, known as the PLS outer relation, is formulated as follows:

$$\mathbf{X} = \sum_{h=1}^{k} t_h p_h^T + \mathbf{E}_h \tag{1}$$

$$\mathbf{Y} = \sum_{h=1}^{k} \hat{u}_h q_h^T + \mathbf{F}_h \tag{2}$$

Where p_h and q_h are loading vectors, $\mathbf{E}_h$ and $\mathbf{F}_h$ are residuals, and $\hat{u}_h$ is the estimator of u_h and calculated by the inner relation.

$$u_h = f_h(t_h) + e_h \tag{3}$$

$$\hat{u}_h = f_h(t_h) \tag{4}$$

The traditional linear PLS performs an ordinary LS regression between pair of corresponding score vectors, that is,

$$u_h = b_h t_h + e_h \tag{5}$$

$$b_h = t_h^T u_h / t_h^T t_h \tag{6}$$

while QPLS employs second order polynomial (quadratic) regression for inner mapping:

$$u_h = b_{0,h} + b_{1,h} t_h + b_{2,h} t_h^{\,2} + e_h \tag{7}$$

The appropriate number of components required to describe the data structure, k , is generally identified by means of cross-validation and chosen to be one which minimizes the Predictive Error Sum of Squares (PRESS). It is because most of the variance of the input and output matrixes can usually be accounted for by the first few score vectors, whilst the residuals are typically associated with the random noise in the data sets.

The problem of input weight updating procedure in NLPLS cannot be omitted. (Wold et al., 1989; Baffi et al., 1999a; Yoon et al., 2003). The input updating procedure proposed by Baffi et al. (1999a) is an error-based approach and listed as follows.

The mismatch e_h between the value of u_h , given by $u_h = \mathbf{Y} q_h$, and the value of $\hat{u}_h$, given by the nonlinear mapping, $\hat{u}_h = f_h(t_h, b_h)$, can be denoted by

$$e_h = u_h - \hat{u}_h \tag{8}$$

Based on the first-order series expansion, equation (8) can be written as

$$e_h = u_h - \hat{u}_h = u_h - f_{00} = \frac{\partial f}{\partial w_h}_{00} \Delta w_h \tag{9}$$

By combining the partial derivatives $\partial f / \partial w_h$ into a matrix $\mathbf{Z}_h$, e_h can be written as $e_h = \mathbf{Z}_h \Delta w_h$ and the correction Δw_h can be regressed directly as follows

$$\Delta w_h = (\mathbf{Z}_h^T \mathbf{Z}_h)^{-} \mathbf{Z}_h^T e_h \tag{10}$$

$(\mathbf{Z}_h^T \mathbf{Z}_h)^{-}$ in equation (10) is the pseudo-inverse of the matrix $(\mathbf{Z}_h^T \mathbf{Z}_h)$. Then the input weight is updated

$$w_h = w_h + \Delta w_h \tag{11}$$

And check convergence on t_h . The updating procedure is completed if a new input score vector t_h ($t_h = \mathbf{X} w_h$) is stable; otherwise repeat the steps mentioned above.

3. MODIFIED QPLS

In this section, a new input weight updating procedure based on nonlinear programming is presented. The new weight updating procedure combined with QPLS leads to QPLS based on nonlinear programming, whose NIAPLS algorithm is also given detailed.

3.1 A new input weight updating procedure

There are three points of the error based input weight updating procedure worthy to be investigated.

Firstly, $\mathbf{Z}_h$ in equation (10) is rank deficient under two conditions. One is input dimension is lager than number of samples, the other is the partial derivatives of the inner relation being linearly correlated with themselves or alternatively with the inner relation $f_h(\cdot)$ itself. In this case, the correction Δw_h cannot be obtained directly by equation (10). So the pseudo-inverse is necessary and numerical techniques are needed to evaluate the pseudo-inverse $(\mathbf{Z}_h{}^T\mathbf{Z}_h)^-$.

Secondly, w_h is updated iteratively until the input score vector t_h is converged, which result in large computation burden.

Thirdly, by applying the error based input weight updating procedure, the NLPLS model can catch larger output cumulative variance, but smaller input cumulative variance. It was also pointed out by Yoon et al. (2003).

In this paper, a new input weigh updating procedure was proposed on the basis of the method proposed by G.baffi et al. The core of the method is as follows.

The objective of the error based weight updating procedure by G..baffi et al. is to find proper input weights and parameters of nonlinear inner relation which can minimize the regression SSE of the each nonlinear inner relationship. It can be classified as a constrained nonlinear programming problem. In QPLS, the optimal weights and polynomial coefficients of inner relationship can be derived from nonlinear programming methods. The optimization problem, including the objective function and the constraints, can be described as follows:

$$\min\left\{(u_h-\hat{u}_h)^T(u_h-\hat{u}_h)\right\} \qquad (12)$$
$$\text{s. t.} \quad \|w_h\|=1$$

where $\hat{u}_h=[1 \quad t_h \quad t_h{}^2]^T\boldsymbol{b_h}$, $t_h=X\cdot w_h$.

In this problem, w_h and b_h are the decision variables, which should be found to minimize the objective function and satisfy the constraints. Herein Sequential unconstrained minimization technique (SUMT) is used to transform problem (12) into a series of unconstrained nonlinear programming problems. Then Hook-Jeevs method is employed to solve the unconstrained nonlinear programming problems. The initial values of w_h and b_h are obtained by NIPLAS algorithm.

By applying the proposed input weight updating procedure, the optimal w_h do not need to be calculated iteratively and the steps in NIPALS algorithm are simplified accordingly. Since the weight updating method improves the fitness of inner relation by changing the spread of score vectors, the proposed one is more precise than the error based one and can catch more cumulative variance. It will be illustrated in the application in Section 4.1.

3.2 Modified NIPALS algorithm:

The new weight updating procedure combined with QPLS leads to QPLS based on nonlinear programming, which is called the modified QPLS (MQPLS). Details of the steps of modified NIPALS algorithm are shown in Table 1.

Table 1 Summary of the modified NIPALS algorithm

It is assumed that **X** and **Y** blocks have been preprocessed, i.e., scaling around zero mean and unit variance. Proper scaling prevents the score vectors from being biased towards variables with larger magnitude. For each component h :

1 Take $u_h=y_j$ (if the column of **Y** equals to 1, set u equal to **Y**)

2 Calculate the input weight $w_h{}^T=u_h{}^T\mathbf{X}/u_h{}^Tu_h$

3 Normalize $w_h=w_h/\|w_h\|$

4 Calculate the input score vector $t_h=\mathbf{X}w_h$

5 Fit the quadratic inner relationship $\boldsymbol{b}_h \leftarrow fit[u_h=[1 \quad t_h \quad t_h{}^2]^T\boldsymbol{b_h}+e_h]$

6 Calculate the nonlinear prediction of u_h $\hat{u}_h=[1 \quad t_h \quad t_h{}^2]^T\boldsymbol{b_h}$

7 Calculate the optimal input weight and parameters of inner relationship according to the new weight updating procedure described in Section 3.1

8 Calculate the new input score vector $t_h=\mathbf{X}w_h$

9 Calculate the input loading vector $p_h=t_h{}^T\mathbf{X}/t_h{}^Tt_h$

10 Normalize p_h to unit length $p_h=p_h/\|p_h\|$

11 Calculate the new nonlinear prediction of u_h $\hat{u}_h=[1 \quad t_h \quad t_h{}^2]^T\boldsymbol{b_h}$

12 Calculate the output loading vector $q_h{}^T=t_h^T\mathbf{Y}/t_h^Tt_h$

13 Normalize q_h to unit length $q_h=q_h/\|q_h\|$

14 Calculate the output score vector u_h $u_h=\mathbf{Y}q_h$

15 Calculate the input residual $\mathbf{E}_h=\mathbf{E}_{h-1}-t_hp_h{}^T$

16 Calculate the output residual $\mathbf{F}_h=\mathbf{F}_{h-1}-\hat{u}_hq_h{}^T$

17 If $h<k$ (k is the optimal number of components), step 1-17 are repeated (**X** and **Y** should be replaced by $\mathbf{E}_h$ and $\mathbf{F}_h$).

4. INFERENTIAL CONTROL OF A FCCU FRACTIONATOR

4.1 Soft sensor of diesel oil solidifying point

An industrial FCCU main fractionator is one of the key processes in modern petroleum refining. The function of the unit is to separate heavy distillates from FCC reactor like gas oils or residuals into gasoline, diesel oil and middle distillates. The MOPLS algorithms described above are applied to establish the soft sensors on the unit to predict diesel oil solidifying point.

Through mechanism analysis, fifteen process variables are chosen as secondary variables and measured online at one minute intervals. Secondary variables include top pressure，top temperature, the flow rate, temperature of the second reflux, etc. A data set including 720 samples are gathered from the DCS database of the FCCU main fractionator. The actual analysis value of product quality is only available from the lab with a frequency of 2 hours. The outliers have been removed beforehand. The data is split into a training data and a test data. Every fifth observation is placed in the test data set, totally 144 samples, and the remaining 576 observations form the training data. The optimal number of components is calculated by cross validation.

Slight nonlinearity is found in first pair of component of data gathered, which is suitable to be fit by quadratic polynomial. The cumulative variance of the $\mathbf{X}$ block and $\mathbf{Y}$ block captured by each model and their Mean Square Predictive Error (MSPE) is given in Table 2 for linear PLS, QPLS, error based QPLS and MQPLS，respectively. Figures 1-4 illustrate the final predication for the test data for the four algorithms.

The MSPE of the original QPLS is 1.3197, whilst the error based QPLS is 1.1651 and MQPLS is 1.0847. It is clearly evident that the three kinds of QPLS algorithms catch the main nonlinear characteristic in the data set. Although the predictive abilities of the error based QPLS and MQPLS are comparable, MQPLS shows a few better than the error based QPLS. The predictive results of MQPLS are used as a reference of the operators.

4.2 Predictive inferential control scheme

The product quality control of the fractionator has been a classical and difficult problem. Traditionally, the product quality is represented by tray temperature control, which has a wide application in the chemical plants. An inferential controller for quality control can be established once the solidifying point of diesel oil is available through the modified QPLS based soft sensor. Many papers (Kano et al., 2000; Kano et al., 2003) have proposed cascade inferential control system in which the set point of tray temperature controller is given by the output of quality inferential controller. However, in such control scheme, the inner temperature controller has a greater influence on the performance of the whole system, and its complex structure brings some difficulties to operators.

In this paper, a new inferential control system is proposed in which tray temperature controller and quality inferential controller can be switched without producing any disturbance. The configuration of the proposed inferential control system is showed in Figure 5. Temperature controller (denoted as TC in Figure 5) still uses the original tray temperature controller. Inferential controller (denoted as AC in Figure 5) adopts constrained Dynamic Matrix Control (DMC) algorithm.

Table2. Model comparison: Cumulative variance (%)

LV	Linear PLS		Original QPLS		Error based QPLS		MQPLS	
	X	Y	X	Y	X	Y	X	Y
1	69.75	56.37	74.40	72.47	29.42	78.61	34.56	82.70
5	70.85	62.14	93.61	75.51	37.17	80.24	53.56	84.96
10	92.45	65.06	99.40	78.24	52.08	91.73	87.26	92.57
15	100.00	68.85	100.00	78.62	62.58	92.55	88.83	93.79
MSPE	1.4687		1.3197		1.1651		1.0847	

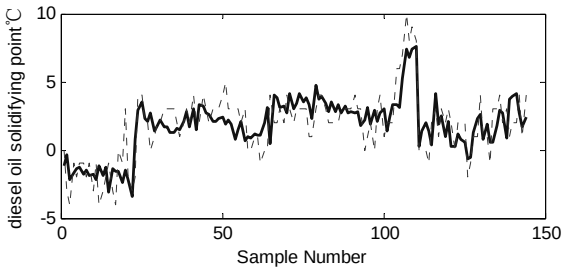


Fig.1 Actual versus Predicted values for the Linear PLS (……actual; ——predicted)

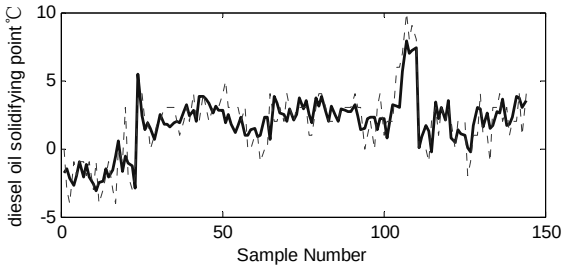


Fig.2 Actual versus Predicted values for the QPLS (……actual; ——predicted)

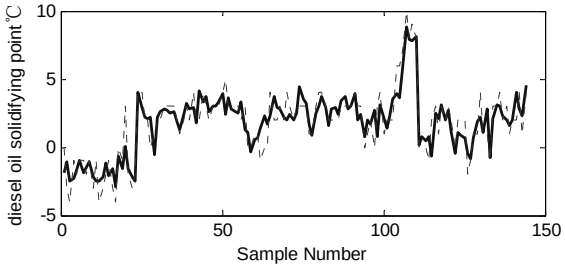


Fig.3 Actual versus Predicted values for the error based QPLS (……actual; ——predicted)

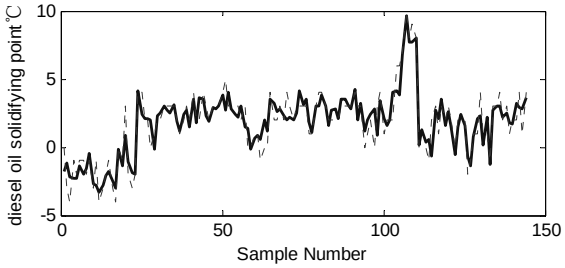


Fig.4 Actual versus Predicted values for MQPLS (……actual; ——predicted)

DMC uses the step response model for predicting the process output P step into the future in the absence of further control action. The model is also used to calculate the present and future M control actions which minimizes the following objective function:

$$J = \{[y_r(k+p) - y_c(k+p)]^2 q + \sum_{i=0}^{m-1} [\Delta u(k+i)]^2 r_i\} \quad (13)$$

s. t. $|\Delta u(k)| \le \Delta u_{\max}, u_{\min} \le u(k) \le u_{\max}$

Where $y_r(k+p)$ is the set objective value.

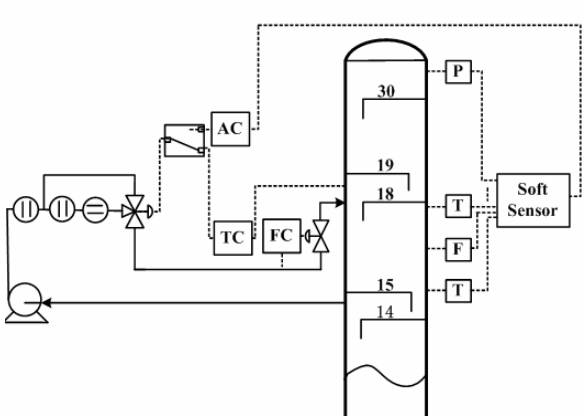


Fig. 5 Schematic diagram of a FCCU main fractionator

This kind of design scheme can make use of the original system module, and is easy to be implemented on the DCS system, and gives facilities for operating. However, there are some questions to pay attention to in practice. When step response of process is made for DMC, it must be sure that step response starts from some steady state. Also the inferential control system should consider some abnormalities from DCS and process to guarantee the safety of the process. Because the running performance of chemical plants is often in change, predictive model updating is another key point.

4.3 Real-time implementation Results

The designed soft sensor and the predictive inferential controller were implemented on the Distribute Control System (DCS) of an industrial FCCU main fractionator using CL (Control Language) programming. The sequential predicting results are shown in Figure 6, in which the dotted line is gathered from the laboratory and the solid line is computed by MQPLS soft sensor. The MSPE is 1.1055 and the predicted result is satisfactory to be used as the set point of the inferential controller.

Figure 7 compares the diesel oil quality control performance for both before and after implementing predictive inferential control system. It can be seen that the control variance decreases clearly when inferential control system is employed. Figure 8 show closed loop response of predictive inferential control system. When the set point step change of solidifying point is from -7.5℃ to -6℃, the control system can quickly trace the desired value.

Application results indicate that inferential control system has a better performance than tray temperature control system.

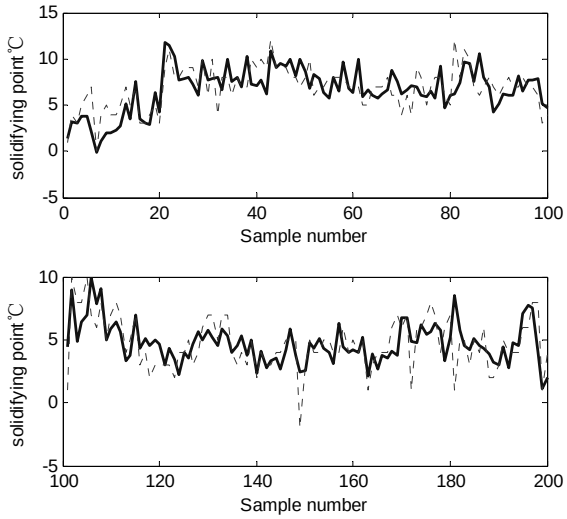


Fig. 6 Validation data set. Comparison between the actual value of solidifying point and its estimates provided by MQPLS (——predicted; ……actual)

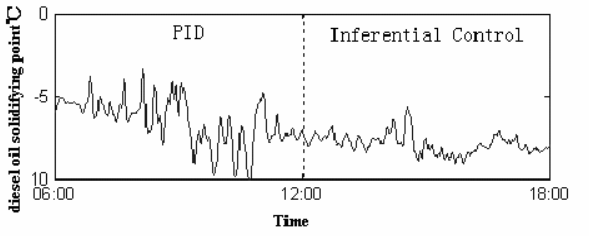


Fig 7 Comparison of the tray temperature control and inferential control system

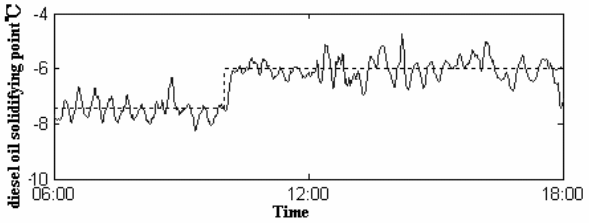


Fig 8 Closed-loop response of predictive inferential control system by step set point change

5 CONCLUSIONS

In this paper, the error based weight updating procedure of G. Baffi et al. is studied. A new weight updating procedure based on nonlinear programming is formulated. MQPLS algorithms are proposed. In comparison with existing QPLS algorithms, MQPLS can catch much higher percentage of input and output cumulative variance, avoid the problem of the pseudo-inverse of matrix and reduce the calculation burden. To realize online measurement, a soft sensor is built based on the MQPLS to estimate the solidifying point of diesel oil for an industrial FCCU main fractionator. An inferential control scheme is proposed. This control scheme can switch between usual tray temperature controller and inferential controller based on constrained DMC algorithm. The practical results obtained from an industrial plant show that the proposed system has a better performance than the traditional tray temperature control system.

6. ACKNOWLEDGEMENTS

This work is supported under “863” project of China National foundation.

REFERENCE

Berglund A.and S. Wold (1997). Implicit non-linear latent variable regression. In: *J. Chemometrics*, **11**, pp. 141-156.

MacGregor, J. F. B. Skagerbeg and C. Kiparissides (1991). *Multivariate statistical process control and property inference applied to low density polyethylene reactors.* IFAC Symposium ADCHEM’91, Toulouse, France, October 1991, pp. 131-135. Pergamon Press, Oxford.

Wold S., N. Kettanch-Wold and B. Skagerberg (1989). Non-linear PLS modeling, In: *Chemomet. Intell. Lab. Sys.* , **7**, pp. 53-65.

Wold S. (1992). Non-linear partial least square modeling, II. spline inner function, In: *Chemomet. Intell. Lab. Syst.*, **14**, pp. 71-84.

Qin S.J. and T.J. McAvoy (1992). Non-linear PLS modeling using neural networks. In: *Comput. Chem. Eng.*, **16**, pp. 379-391.

Yoon H. B., K.Y. Chang and I. B. Lee (2003). Nonlinear PLS modeling with fuzzy inference system, In: Chemomet. Intell. Lab. Syst, **64**, pp.137-155.

Geladi P. and B.R. Kowalski (1986). Partial least-square regression: a tutorial, In: *Anal. Chim. Acta* , **185**, pp. 1-17.

Baffi G., E.B. Martin, A. J. Morris, Non-linear Projection to Latent Structure Revisited: The Quadratic PLS Algorithm , In: Comput. Chem. Eng., 1999a, **3,** pp. 395-411.

Baffi G., E.B. Martin, A. J. Morris (1999b). Non-linear Projection to Latent Structure Revisited: The neural net work PLS algorithm, In: Comput. Chem. Eng., 2**3**, pp. 1293-1307.

Baffi G., E.B. Martin, A.J. Morris (2000). Non-linear Dynamic Projection to Latent Structure modelling, In: *Chemomet. Intell. Lab. Syst*, **52**, pp. 5-22.

Kano M., N. Showchaiya, S. Hasebe, et al. (2003). Inferential control of distillation compositions: selection of model and control configuration. In: *Cont. Eng. Pract.*, **10**, pp. 927 – 933.

Kano M., K. Miyazaki, S. Hasebe, et al. (2000). Inferential control system of distillation compositions using dynamic partial least squares regression, In: *J. Proc. Cont.*, **11**, pp. 157-166.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

CONTROL SOLUTIONS FOR SUBSEA PROCESSING AND MULTIPHASE TRANSPORT

Heidi Sivertsen * **John-Morten Godhavn** ** **Audun Faanes** **
Sigurd Skogestad *,[1]

* *Department of Chemical Engineering, Norwegian University of Science and Technology, Trondheim, Norway*
** *R&D, Statoil ASA, Trondheim, Norway*

Abstract: To increase the oil production for the Tordis subsea oilfield located at the Norwegian Continental Shelf, a subsea separation and boosting station will be installed. Most of the water will be injected into a subsea reservoir instead of being transported up to the platform. Several challenges conserning process control need to be addressed before the implementation process, and dynamic simulations have therefore been performed in order to develop and test different control strategies to deal with these challenges. The results from some of these simulations will be presented in this paper.

Keywords: Process control, control system design, PI controllers, cascade control, pipelines

1. INTRODUCTION

The Tordis field operated by Statoil has proved to be even more productive than anticipated when production began in 1994 (Godhavn *et al.*, 2005). To increase production and total recovery for the field in the last years of production, processing equipment is planned installed at the sea bed. This in order to separate produced water from the production stream, inject this water into a reservoir, and increase the production rate.

Subsea processing enables production from low-pressure reservoirs over long distances, and may increase the daily oil and gas production or even the total recovery from the reservoir. By injecting produced water into a reservoir, the water emission from topside to sea can be reduced, and the subsea transportation pipelines are better exploited. Compression and pumping enable a lower wellhead pressure, and hence an increased production.

However, the installation of new subsea equipment leads to several new challenges, also related to process control. There can be several ways to solve these problems, so the first question that needed answering was; which solutions are feasible and which one will solve the problems the best.

Having control of the subsea separator pressure and liquid levels are important as it determines the flow rates and compositions for the entire system. In Section 3, some solutions to achieve control of the separator will be presented. These control solutions are then expanded to achieve other benefits, such as faster well tests and control of the water rate that is transported with the oil and gas to the platform.

Under certain conditions a flow regime called riser slugging can develop in the pipelines, which is undesirable because it can introduce large pressure oscillations in the system. In the end of Section 3 it will be shown that this problem can be solved using feedback control.

[1] Author to whom correspondence should be adressed: skoge@chemeng.ntnu.no

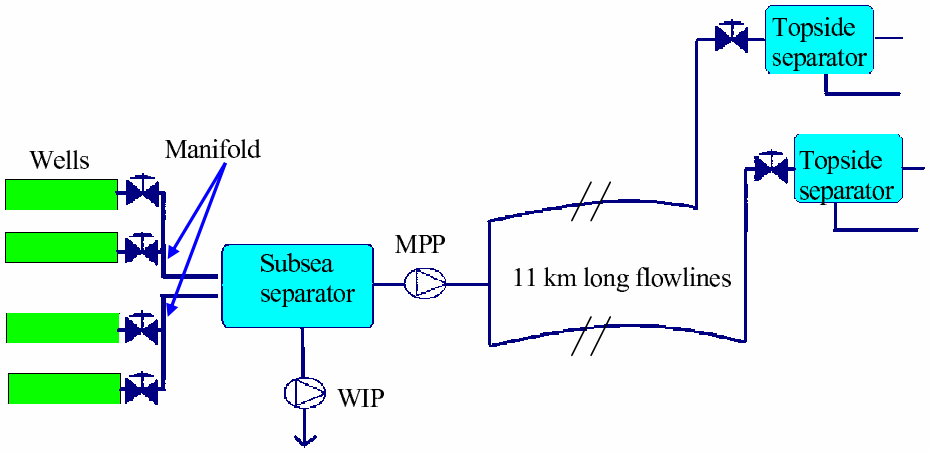


Fig. 1. Subsea processing equipment

The control solutions presented in this paper are illustrated with dynamic simulations including all equipment from the wells to the two topside receiving separators at the Gullfaks C platform (Figure 1). It is important to notice that these simulations were performed at a very early stage in the process of determining how to run the process, where the aim was to find feasible control solutions and not to find optimal control parameters. The controllers have therefore not been fine-tuned and simplified models for the equipment and pipelines have been used. This is also the reason why the absolute values for the different variables have been left out in this paper.

To simulate flow in the pipelines, OLGA 2000 dynamic multiphase simulator (`www.olga2000.com`), provided by Scandpower Petroleum Technologies (`www.scandpowerpt.com`) has been used. Most of the process equipment is simulated using Simulink. The OLGA - MATLAB toolbox enables the Simulink application to simulate multiphase flow in pipelines in OLGA together with additional process equipment and controllers modeled in Simulink.

2. SUBSEA PROCESSING EQUIPMENT

Oil, gas and water are transported from the manifold to the subsea separator through two pipelines. From the separator some of the water is to be injected into a disposal reservoir. The remaining water will be transported along with the oil and gas through two pipelines into each topside separator at the Gullfaks C platform. A multiphase boosting pump will be installed downstream the separator.

2.1 Wells

There will be a total of eight wells producing oil, water and gas to the Gullfaks C platform. The flows from the wells are merged at the manifold. Two short pipelines, each receiving the production from four wells, transport the fluid to the subsea separator.

2.2 Pipelines

To simulate the pipelines between the wells, the subsea separator and the topside separators, OLGA 2000 have been used. OLGA 2000 is a commercial available dynamic multiphase flow simulator. In our study OLGA has been run from Simulink. From OLGA, it is possible to get all the information about the flow and the equipment that is modeled in OLGA, into Simulink.

2.3 Subsea Separator

The subsea separator is illustrated in Figure 2. In the separator the water, oil and gas will separate due to gravity. The water, which is heaviest, will sink to the bottom. Most of the water is to be injected into a disposal reservoir through an outlet in the bottom of the separator. It is important that no oil enters this reservoir. The rest of the water is transported to the platform along with the gas and oil.

The thickness of the water layer and the oil layer is determined by the inlet and outlet flow rates. The multiphase pump and the water pump speed will therefore influence the thickness of these layers. The rest of the separator is filled with gas.

The separator is simulated using a simple Simulink model. It computes the separator pressure, density and composition for the flow to topside and the water and oil levels in the separator. It is assumed that the pressure is independent of gravity, that is: the pressure at the bottom is the same as in the gas layer at the top of the separator. The composition of the flow going to the platform is determined by the thickness of the water and oil layer. If the level of the water is below the outlet leading topside, no water will

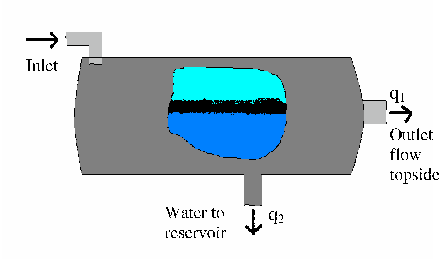


Fig. 2. Subsea separator

be transported topside. The same goes for the oil level, which depends both on the oil and water layer thickness. As already mentioned, the flow rate will be determined by the multiphase pump speed and the pressure in the separator and the pipelines.

2.4 Pumps

Multiphase pump To be able to operate the subsea separator at a low pressure despite the friction loss caused by the 11 km long pipelines to the Gullfaks C platform, pumps or compressors can be installed.

The plan is to install a multiphase boosting pump downstream the subsea separator. In this way it is possible to control the separator pressure by adjusting the pump speed and thereby the flow rate to topside, q_1.

Water pump There is also a need for a water pump to pump the water into the disposal reservoir, holding a higher pressure than the subsea separator.

The water rate through the water pump, q_2, depends on the pressure difference between the reservoir and the subsea separator, and also the pump speed. Pump speed and pressure drop over the *multiphase* pump will in the same way determine the topside production rate, but composition and density of the flow will also influence these flow rates.

2.5 Chokes

There are chokes for each of the eight wells, which make it possible to adjust the flow from each well independently. These chokes can be used for well tests, where one well after another is shut down.

At the top of each riser there are topside production chokes. They make it possible to control the flow into each of the topside separators, and can be adjusted manually or by a controller.

2.6 Measurements

Several measurements will be available, monitoring pressure, density, flow rates and other values which are necessary for controlling the different parts of the system. Measurements used directly for control are the manifold pressure, the subsea separator pressure and water level, pressure drop and density over topside production chokes, water rate out of topside separators and the pressure downstream the multiphase subsea pump. The pressure drop and density across the topside chokes are used to calculate the flow rate through the topside chokes as there are no flow measurements available.

3. CONTROL STRATEGIES

Several dynamic simulations were performed to test different control strategies for controlling the system, and some of these will be presented here. The results will be used in the design of the control system and this way serve as a basis for further studies. The solutions presented here might therefore not be the ones implemented in the end.

3.1 Control of subsea separator pressure and levels

3.1.1. Decentralized PI control of subsea separator pressure and water level To keep the oil contents in the injected water below a given limit, it is important to control the separator water level. By increasing the flow rate of the water injected into the reservoir, the water level will decrease. The flow rate through the water injection pump depends on the pressure difference across the pump and the pump speed. The speed of the pump can be set by a controller.

It is also important to control the separator pressure as this pressure will affect the wells and their production. The separator pressure can be controlled by changing the total flow rate to topside, which again is influenced by the speed of the multiphase pump. During the simulations this flow rate was set by the controller directly. The reason for this is that there was no model of the multiphase pump available at the time of the simulations.

Even though there are quite strong interactions between the level and pressure control, as will be shown, simple PI controllers were used to see how well the separator could be controlled. This is illustrated in Figure 3.

Figure 4 shows the results for a simulation where the input rates of water, gas and oil are reduced by 50% after 30 min. The pressure drops as the flow rates are reduced, but after about 15 min the pressure is back to normal due to the controller action.

What might seem surprising is that the water and liquid level start to increase at the time the inlet rates are reduced, before they decrease and end up at lower levels than they initially had. The reason for this is that the separator pressure and water level affect each other. When the separator pressure decreases due to

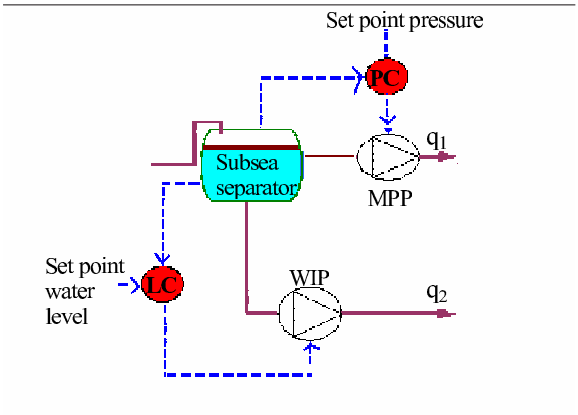


Fig. 3. PI control of subsea separator pressure and water level

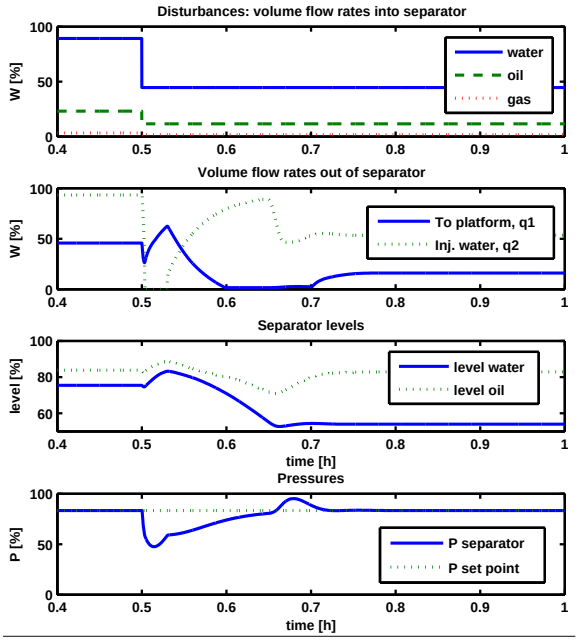


Fig. 4. Results using PI controllers to control subsea separator pressure and water level

the reduced inlet flow rates, it makes it harder for the water pump to inject water into the reservoir. Because of this, the water rate injected to the reservoir, q_2, temporarily goes down to zero, explaining the increase in levels.

In practice, a zero flow rate will cause problems for the water pump, but better tuning of the controller or other control configurations will remove this problem. Another way of avoiding this problem could be to use some other control configuration, e.g. a cascade controller where the inner loop controls the flow rate through the water pump and the outer loop controls the water level in the separator.

3.1.2. Cascade control : Control of water rate to topside At the Gullfaks C platform, the water that is transported to topside along with the gas and oil needs to be taken care of. There are limits to the amount of water the downstream process equipment can handle, and having control of this water rate can be an advantage.

By changing the water level in the subsea separator it is possible to control the water rate that is transported to the Gullfaks C platform. Figure 5 shows one way of doing this. It is an extension of the control structure presented in 3.1.1. An increased water level will lead to increased water rate topside (see Figure 2). A cascade configuration using the water rate out of the topside separator, q_3, in a slow outer loop and the water level in the inner loop, was developed to handle this.

Figure 6 shows the results from a simulation where the inlet flow rates are reduced by 50% after 1h. The set-point for the water level controller is increased when too little water is transported topside due to reduced inlet rates.

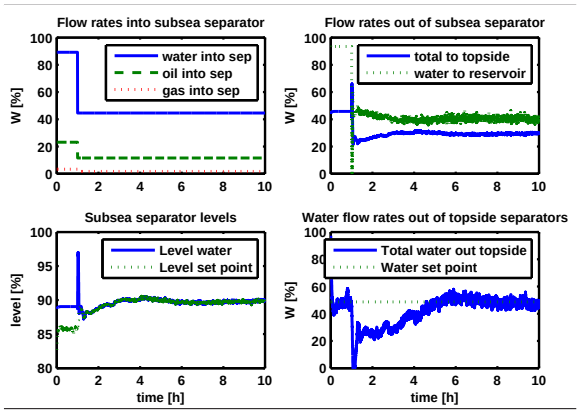


Fig. 6. Results using a cascade controller to control subsea separator water level and water rate topside

We see that after about 5 hours the water flow rate is back at its set-point, even though the flow rates into the subsea separator have been reduced substantially.

3.2 Well head pressure control

During a well test, one well after the other is shut down in order to determine the production rate from each individual well (deduction principle for tie-ins). Performing well tests is costly, as the production is reduced for the time the well test lasts. Being able to reduce the duration of a test, has therefore a large economic potential. Using active control might reduce the time needed to perform a well test.

However, when a well is shut down, the pressure drop in the pipeline will decrease due to the reduced flow rate in the pipe. This way the other wells will produce more, leading to a wrong estimate of the production from the well that is closed. Therefore, during well testing, the pressure at the *manifold* is kept constant rather than the subsea separator pressure which is normally controlled (Figure 3). There actually is a need for the subsea separator pressure to increase

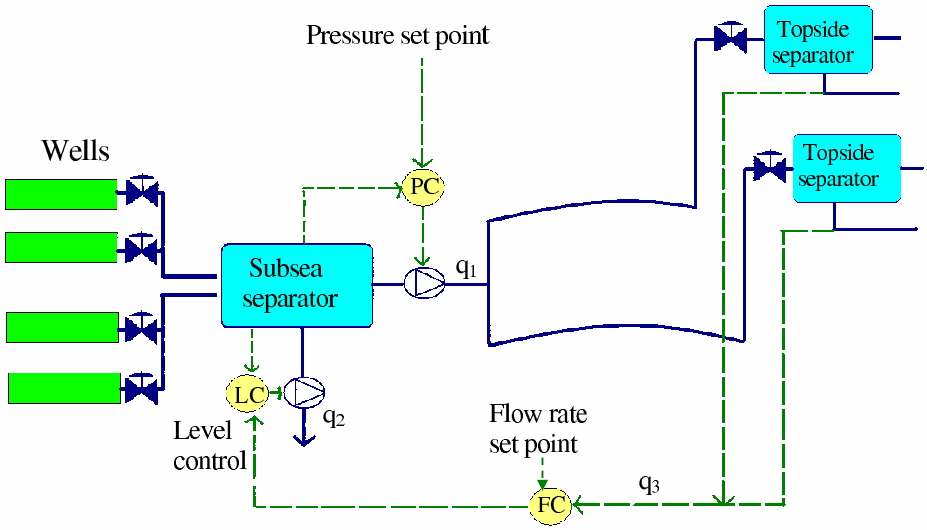


Fig. 5. Cascade controller for subsea separator water level and water rate topside

during a well test. The alternative would be to reduce the well choke openings accordingly.

There are several ways to do this. Using a cascade control configuration is one possibility. The outer loop controls the manifold pressure where the set-point is the initial pressure before the well test. The inner loop controls the subsea separator pressure. This way the set-point for the subsea separator pressure will automatically increase for every well that is shut down. The cascade control configuration is illustrated in Figure 7.

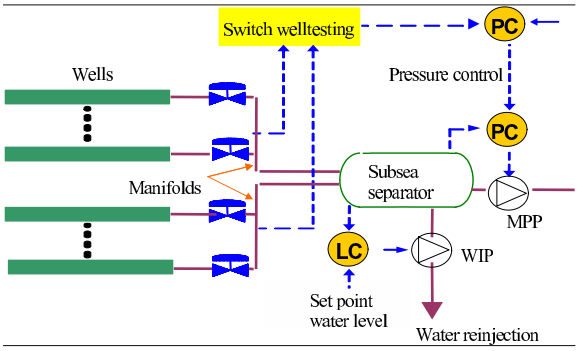


Fig. 7. Welltest using cascade configuration

Using the cascade controller for the well test, it was possible to bring the manifold pressure back to its original value. Figure 8 shows the results when three of the wells are shut down one after another. The plot at the bottom shows how the subsea separator pressure increases to counteract the effect of the reduced pressure loss in the pipelines upstream the separator.

Another way of controlling the manifold pressure is to estimate how much the manifold pressure will drop when a well is shut down, and then increase the set-point for the subsea separator pressure accordingly. This way the simple pressure PI controller described in Section 3.1.1 can be used, as long as steps in the set-point are introduced. It is important to find good estimates of how much the separator pressure need

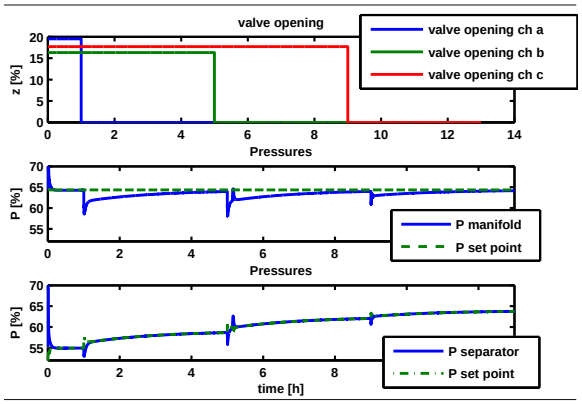


Fig. 8. Welltest results

to increase in order to use this method. Results from simulations show that it is possible to reduce the time before the manifold pressure reaches its initial value to less than 15 min. This is illustrated in Figure 9.

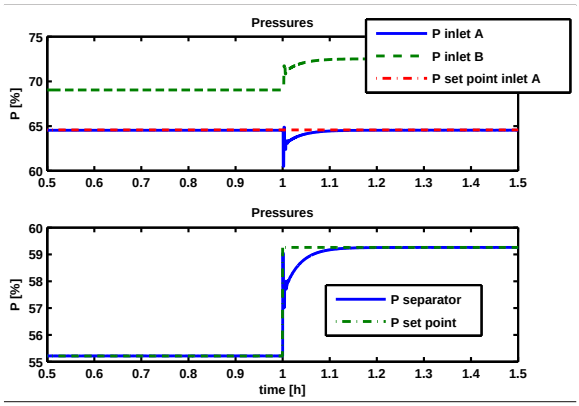


Fig. 9. Welltest results using a PI controller with setpoint changes

It is important to find a good estimate of how much the pressure drops at the manifold when a well is shut-in, in order to use this solution.

The results from the simulations show how long it takes for the manifold pressure to retain its initial

value after a well is shut down. This information can be used to predict the duration of a well test.

3.3 Slugging

Riser slugging is a well known problem offshore, where alternating bulks of liquid and gas enter the receiving facilities and cause problems due to pressure and separator level oscillations. The results are poor separation and wear on the equipment.

There are several ways to deal with the problem, but using active control has in the last years been the preferred way to avoid riser slugging, (Courbot, 1996), (Havre *et al.*, 2000), (Hedne and Linga, 1990), (Skofteland and Godhavn, 2003). Today a combination of active slug control and model predictive control (MPC) is used at Gullfaks C (Godhavn *et al.*, 2005).

A simple PI controller using the pressure upstream the flow-line ending in the riser and a control valve at the top of the riser has proved to be effective. This pressure oscillates heavily during slugging, due to the changing composition in the riser. Keeping this pressure stable forces the flow into another flow regime. In (Storkaas, 2005) control theory proves that using this measurement one is able to stabilize the flow and also to achieve good performance. This control configuration is illustrated in Figure 10.

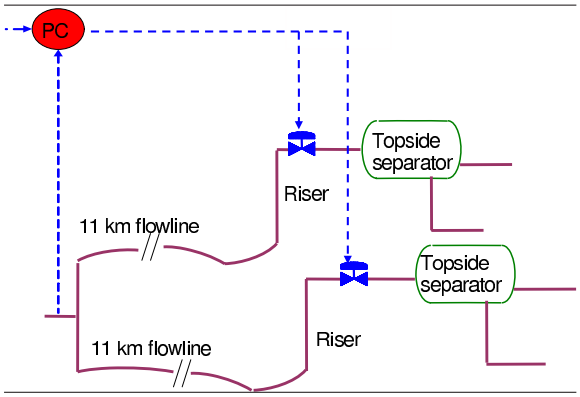


Fig. 10. Slug control applied to Tordis

Results from a simulation with the slug controller are shown in Figure 11. During the first 4 hours the controller is inactive, resulting in slugging in the pipeline and the pressure variations shown in the upper plot. When the controller starts working, the pressure stabilizes at the desired set-point.

4. CONCLUSIONS

The implementation of new subsea processing equipment to improve the productivity for a subsea oilfield is expected to introduce several new challenges regarding operation and process control that need to be addressed before the start-up. This paper presents some results from dynamic simulations performed in

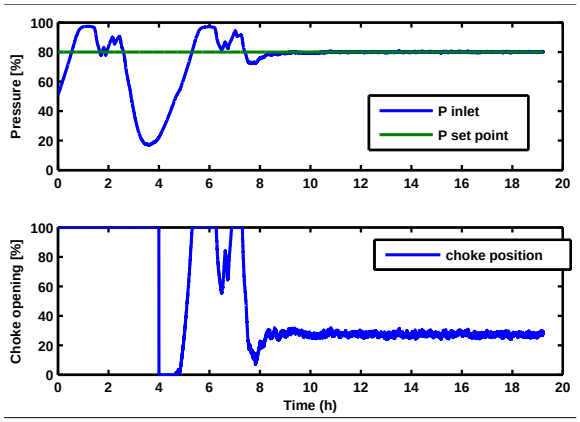


Fig. 11. Slug control results

order to investigate how the use of automatic control might deal with these challenges. For the different scenarios presented here, automatic control shows good results.

The simulations have been performed at a very early stage, before the final decisions about equipment and operation have been made. Because of this, simplified models of the pipelines and equipment were used. Also, the controllers have not been fine-tuned to get the best results at this stage. The results from this study will therefore differ from the final results. The simulations can, however, be used as a basis for later studies.

Examples of what better suited controllers can accomplish are; decreasing the time of well tests (Figure 8) and removing the effect that leads to the topside choke saturating in the first 4 hours of slug control (Figure 11).

REFERENCES

Courbot, A. (1996). Prevention of severe slugging in the dunbar 16" multiphase pipeline. *Offshore Technology Conference, May 6-9, Houston, Texas*.

Godhavn, J.-M., S. Strand and G. Skofteland (2005). Increased oil production by advanced control of receiving facilities. *IFAC'16, Prague, Czech Republic*.

Havre, K., K. O. Stornes and H. Stray (2000). Taming slug flow in pipelines. *ABB review* **4**(4), 55–63.

Hedne, P. and H. Linga (1990). Supression of terrain slugging with automatic and manual riser choking. *Advances in Gas-Liquid Flows* pp. 453–469.

Skofteland, G. and J.M. Godhavn (2003). Suppression of slugs in multiphase flow lines by active use of topside choke - field experience and experimental results. In: *Proc. of MultiPhase '03, San Remo, Italy, 11-13 June 2003*.

Storkaas, E. (2005). Stabilizing control and controllability: Control solutions to avoid slug flow in pipeline-riser systems. PhD thesis. Norwegian University of Science and Technology.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ACTIVE CONTROL STRATEGY FOR DENSITY-WAVE IN GAS-LIFTED WELLS

Laure Sinègre * **Nicolas Petit** *
Thierry Saint-Pierre ** **Pierre Lemétayer** **

* *CAS, École des Mines de Paris, France*
** *CSTJF, TOTAL Exploration-Production, Pau, France*

Abstract: We focus on the control of gas-lifted wells in the context of instable flows. Two cases are considered: casing-heading and density-wave. While it is known that active control can stabilize the casing-heading phenomenon, (passive) hardware upgrading solutions are sometimes preferred. In this paper, we advocate active control solutions in contrast to these strategies. Our aim is to stress that density-wave, which is a complicated issue not addressed by hardware solutions yet, can also be stabilized by the same simple control strategies that proved successful against casing-headings.

Keywords: Process Control, Gas-Lifted Well, Density-wave, Stabilization.

1. INTRODUCTION

Producing oil from deep reservoirs and lifting it through wells to surface facilities often requires activation to maintain oil output at a commercial level. In the gas-lift activation technique (Brown, 1973), gas is injected at the bottom of the well through the injection valve (point C in Figure 1) to lighten up the fluid column and to lower the gravity pressure losses. High pressure gas is injected at well head through the gas valve (point A), then goes down into the annular space between the drilling pipe (casing, point B) and the production pipe (tubing, point D) where it enters. Oil produced from the reservoir (point F) and injected gas mix in the tubing. They flow through the production valve (point E) located at the surface.

As wells and reservoirs get older, liquid rates begin to decrease letting wells be more sensitive to flow instabilities commonly called headings. These induce important oil production losses (see (Hu and Golan, 2003)) along with possible facilities damages. The best identified instability is the "casing-heading". It consists of a succession of pressure build-up phases in the casing without production and high flow rate phases due to intermittent gas injection rate from the casing to the tubing (see (Jansen *et al.*, 1999) or (Torre *et al.*, 1987) for a complete description). Yet, keeping the gas injection constant in the tubing does not always prevent the instability. It has been pointed out in (Hu and Golan, 2003) that headings still occur on wells equipped with NOVA valves, i.e. valves maintaining the flow critical. In such a case one refers to the density-wave instability. In details, even though the gas injection rate through valve C is kept constant, self-sustained oscillations, confined in the tubing D can occur. Out-of-phase effects between the well influx and the total pressure drop along the tubing are usually reported at the birth of this phenomenon. More details about modelling under the form of a distributed delay system can be found in (Sinègre *et al.*, 2005).

Interestingly, almost all casing-heading control strategies aim at maintaining the gas flow rate injected in the tubing at a given set-point. In practice, under the assumption of a constant well

head gas (in-)flow rate, stabilizing the casing head pressure achieves this goal. One can find details in (der Kinderen *et al.*, 1998) and also in (Eikrem and Golan, 2002) where the more advanced case of two interconnected wells is addressed.

Hardware upgrades to the NOVA valves are sometimes preferred to such active feedback control strategies. Technically, the valves track a critical flow point. This implies that flow does not depend on downstream pressure. Decoupling is thus achieved, and casing-heading stabilization is guaranteed.

Yet, further feedback control strategies have emerged. Another idea is to stabilize the pressure at the bottom of the well. As measurements at such depths are often not reliable and sometimes even not available at all, the need for estimators is critical. In (Eikrem *et al.*, 2004) example of stabilization relying on downhole pressure estimation is given. The controller relies on downhole pressure measurement and can handle sensor failures. Up to the authors' knowledge, when the well head pressures are the only measured variables, controlling the casing head pressure is the only proposed strategy.

We believe that even though very effective for casing-heading phenomenon, hardware upgrading solutions do not address all the instabilities of gas-lifted wells yet. To illustrate our point, we focus on the density-wave phenomenon. While it is known since (Hu and Golan, 2003) that density-wave on NOVA valve equipped wells can occur, we demonstrate that the original simple feedback control strategy of casing head pressure setpoint tracking does stabilize the well.

Controlling the density-wave phenomenon is studied in (Hu and Golan, 2003) and implicitly in (Dalsmo *et al.*, 2002). In both cases manipulating the production choke is used to stabilize the downhole pressure. The promising results at Brage field are reported in (Dalsmo *et al.*, 2002). Although the density-wave is not explicitly mentioned, they state that the slugging is not caused by casing-heading. They also stress that the strategy is efficient as long as the downhole pressure sensor works properly. Unfortunately, technical issues and high cost premiums usually prevent the use of the sensors required for real-time control purposes. In this paper we aim at showing that it is possible to control the density-wave using only well head measurements. We show that the control strategy described for casing-heading, i.e. stabilization of the casing head pressure through production choke actuation, is also efficient in the density-wave case. This is the contribution of our paper.

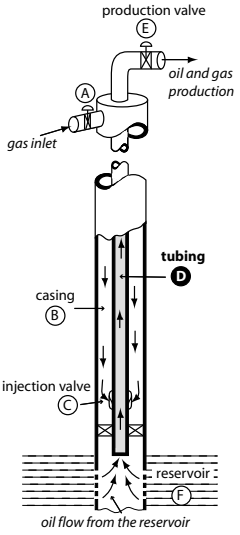


Fig. 1. Gas-lift activated well. Casing-heading involves both tubing D and casing B while density-wave takes place in the tubing D.

The article is organized as follows. In Section 2, we detail the model of controlled gas-lifted wells. This model implies an ordinary differential equation coupled to a distributed parameters system with boundary control. In Section 3, we propose a control strategy and prove local convergence. In Section 4, we give OLGA®2000 simulation results that illustrate the relevance of the approach. Conclusions and future directions are given in Section 5.

2. MODELLING

In this section, we present a gas-lifted well model. First, we detail the casing and tubing subsystems and their interconnection by feedback loops. Then, we explain through OLGA®2000 simulations why we choose the well head pressure as control variable.

2.1 Gas-lifted well modelling

Casing model The well is divided in two parts. Nomenclature is given in Table 1. The annular part, called casing, can be considered as a tank filled with gas. The dynamics is simply represented by a mass balance equation

$$\dot{x} = w_{gc} - w_{iv} \tag{1}$$

where w_{gc} is the gas inlet and w_{iv} the gas outlet. The expression of w_{iv} with respect to upstream and downstream pressures, respectively P_{ab} and P_{tb}, is given by

$$w_{iv} \triangleq C_{iv}\sqrt{\max(0, \rho_{ab}(P_{ab} - P_{tb}))}$$

Assuming that the gas is ideal and that the column is at equilibrium state, we get

$$\rho_{ab} \triangleq \alpha x \text{ and } P_{ab} \triangleq \beta x$$

where α and β are defined by

$$\beta = \alpha RT \triangleq \frac{g}{S_a} \frac{1}{1 - \exp\left(-\frac{gL_a}{RT}\right)}$$

The casing is considered as a one-dimensional system of length L, which state is the gas mass, x, with two inputs P_{tb} and w_{gc}

$$\dot{x} = w_{gc} - C_{iv}\sqrt{\max(0, \alpha x(\beta x - P_{tb}))} \qquad (2)$$

Tubing model Following (Imsland, 2002), we could use the gas and the oil masses as states and then model the tubing dynamics by two balance equations. The system resulting from the coupling of this model and the casing model accurately reports the casing-heading instability. Yet, it can not represent the density-wave phenomenon, which originates in the propagation of the gas mass fraction. For that purpose, we use the model presented in (Sinègre *et al.*, 2005).

Mass conservation laws along with proper choice of slip velocity law (see (Cholet, 2000) and (Duret, 2005)) yield the existence of a Riemann invariant (as defined in (Chorin and Marsden, 1990)) being the gas mass fraction. We assume that the gas is ideal and that no phase change occurs. Following (Asheim, 1988), we neglect transient inflow from the reservoir as well as acceleration and friction terms in Bernoulli's law. In other words we assume the flow to be dominated by gravitational effects. Furthermore, for sake of simplicity, we approximate the gas mass fraction by the gas volume fraction.

Under these assumptions, the tubing model writes under the integral form

$$P_{tb} = P_0 + \rho_l g L + \int_0^\tau k(\zeta)\left(1 - \frac{P_r - P_{tb}(t-\zeta)}{\lambda w_{iv}(x(t-\zeta), P_{tb}(t-\zeta))}\right) d\zeta \qquad (3)$$

where $\tau \triangleq L/V_g$ is the propagation delay. The right hand side is the sum of $P_0 + \rho_l g L$ which corresponds to the weight of the column full of oil, and an integral which corresponds to the lightening effect of the gas. This (convolution) integral consists of the product of the propagating gas mass fraction by a negative function k with finite support, which is proportional to the difference of density between gas and oil. The expression of k over $[0, \tau]$ is given by

$$k(t) \triangleq V_g g\left(\frac{tP_0 + (\tau - t)P_r}{\tau RT} - \rho_l\right) < 0$$

Notice that k is a strictly decreasing affine function. For sake of simplicity, we shall write from now on

$$k(t) = (k_1 t + k_2)\mathbf{1}_{[0,\tau]} \qquad (4)$$

where $\mathbf{1}_{[0,\tau]}$ is zero over the entire real line except for the interval $[0, \tau]$ where it is equal to 1.

Gas-lifted well model Coupling equations (2) and (3) gives

$$\begin{cases} \dot{x} = w_{gc} - w_{iv}(x, P_{tb}) \\ P_{tb} = P_{tb}^\star + k * \left(1 - \dfrac{P_r - P_{tb}}{\lambda w_{iv}(x, P_{tb})}\right) \end{cases} \qquad (5)$$

The state is (x, P_{tb}), where P_{tb} is a function mapping $[0, \tau]$ onto $\mathbb{R}$. The considered output is x. In practice, x is proportional to the well head casing pressure, which is actually measured. So far, the input corresponding to the production choke does not appear in model (5). Since manipulating this choke has a direct impact on the well head tubing pressure, one can assume that the input is $P_{tb}^\star \triangleq P_0 + \rho_l g L$. We stress the relevance of this approach in section 2.2.

A gas-lifted well consists of two coupled subsystems. On one hand is the casing with inputs w_{gc} and P_{tb} and output w_{iv}. On the other hand is the tubing with inputs w_{iv} and $P_{tb}^\star$ and output P_{tb}. This structure is reported in Figure 2. The two possibly positive feedback loops are at the birth of instabilities. The first loop appears in the tubing, it corresponds to the self-correlation of P_{tb} detailed in (3). This internal loop creates the density-wave. On the other hand, the casing-heading arises from the coupling of these two subsystems via the explicit feedback loop in (5).

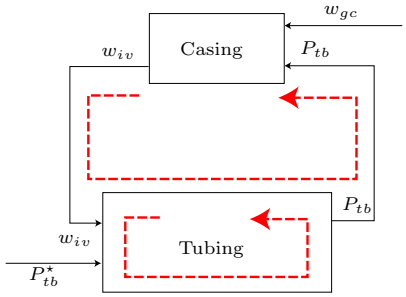


Fig. 2. Block scheme of the gas-lifted well model. The system consists of two coupled subsystem. The two arrows stand for possibly positive feedback loops, yielding instabilities.

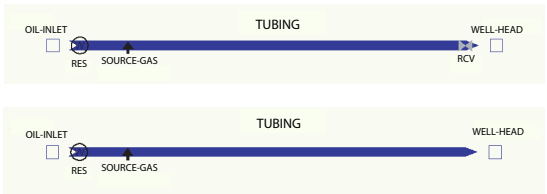


Fig. 3. Block scheme of the OLGA®2000 simulation setup. First case (top) with a production choke, second case (bottom).

2.2 Manipulated variable definition

We now investigate the role of the tubing well head pressure as input variable. With OLGA®2000 we consider two setups simulating the flow in a single vertical pipe (see Figure 3). Oil is supplied by a reservoir and gas is injected at the bottom of the pipe. In the first setup, the pipe is equipped with a production choke that we progressively open. In the second setup, there is no production choke. Instead, the tubing is modelled as a pipe with a downstream pressure boundary condition. Gradually, we decrease this boundary pressure, simulating a reduction of the well head pressure.

Figure 4 shows the steady state well head pressure values as a function of the production choke opening. Classically, our focus is on comparing the oil and gas velocities histories obtained from the two simulation setups. Figure 5 reports the static values of the oil and gas velocities as a function of the well head pressure. Over almost the whole well head pressure operating range (from 23 to 29 bar, i.e. from 0.2 to 1 choke opening), the curves coincide. It is only when the choke is almost closed that differences appear. Figure 6 shows the comparison of the step responses to an increase of the well head pressure and to a consistent decrease of the production choke opening, respectively. We notice similar undershoots of approximately 0.02 m/s. It takes between four and five noticeable oscillations for both systems to settle. This experiment suggest it is valid to consider $P_0 + \rho_l g L$ as our input variable. From now on, we denote $u \triangleq P^\star_{tb}$.

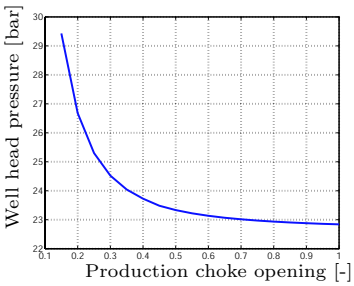


Fig. 4. Well head pressure as a function of the production choke opening.

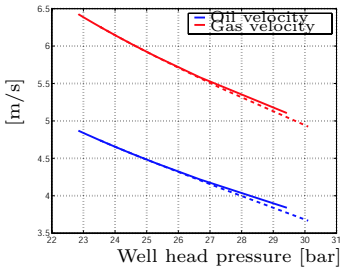


Fig. 5. Comparison of the oil and gas velocities between the fist and second cases (continuous and dashed line respectively).

Table 1. Nomenclature

Symb.		Values	Units
R	Ideal gas constant	523	S.I.
T	Temperature of the well	323	K
C_{iv}	Injection valve constant		
S_a	Casing section	0.081	m^2
α	Constant		$1/m^3$
β	Constant		$1/m/s^2$
P_r	Reservoir pressure	170e5	Pa
$P^\star_{tb}$	Pres. of the column of oil	$P_0 + \rho_l g L$	Pa
P_0	Separator pressure	22e5	Pa
g	Gravity constant	9.81	m/s^2
ρ_l	Density of oil	781	kg/m^3
V_g	Gas velocity		m/s
L	Pipe length	3000	m
λ	Constant		1/(ms)
k_p, k_i	Controller gains		
$x(t)$	Mass of gas in the casing		kg
$P_a(t)$	Casing head pressure		Pa
$P_{ab}(t)$	Casing head pressure		Pa
$\rho_{ab}(t)$	Casing gas density		kg/m^3
$w_{gc}(t)$	Gas mass flow rate		kg/s
$w_{iv}(t)$	Gas mass flow rate in the tubing		kg/s
$P_{tb}(t)$	Bottom-hole pressure		Pa
$u(t)$	Production choke opening		-

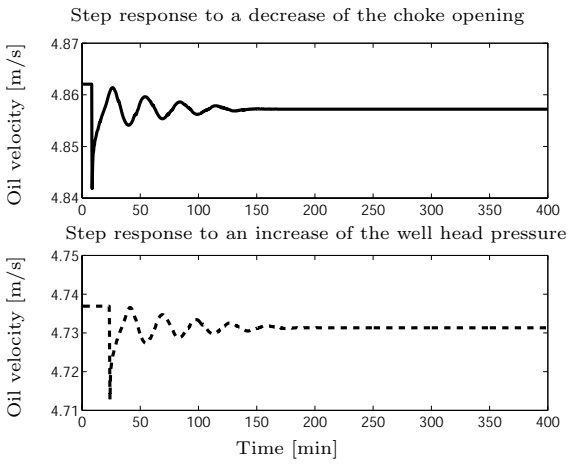


Fig. 6. Comparison of the step responses to an increase of the well head pressure and to a decrease of the production choke opening.

3. CLOSED-LOOP STABILITY ANALYSIS

We now aim at showing that it is theoretically possible to stabilize the well using a simple PI controller with the well head pressure as input and the mass of gas in the casing as output.

Linearization of equation (5) gives

$$\delta\dot{x} = -\partial_x w_{iv}\delta x - \partial_{P_{tb}} w_{iv}\delta P_{tb}$$

$$\delta P_{tb} = \delta u + \int_0^\tau k(\zeta)\left(\frac{1}{\lambda w_{iv}} + \frac{P_r - P_{tb}}{\lambda w_{iv}^2}\partial_{P_{tb}} w_{iv}\right)\delta P_{tb}(t-\zeta)d\zeta + \int_0^\tau k(\zeta)\frac{P_r - P_{tb}}{\lambda w_{iv^2}}\partial_x w_{iv}\delta x(t-\zeta)d\zeta$$

Therefore, one can rewrite

$$\begin{aligned}\delta\dot{x} &= a_1\delta x + a_2\delta P_{tb}\\ \delta P_{tb} &= \delta u + \int_0^\tau a_3 k(\zeta)\delta P_{tb}(t-\zeta)d\zeta\\ &\quad + \int_0^\tau a_4 k(\zeta)\delta x(t-\zeta)d\zeta\end{aligned}$$

which, in Laplace coordinates, leads to

$$\begin{cases}\delta\tilde{x} = \dfrac{a_2}{s-a_1}\delta\tilde{P}_{tb}\\ \delta\tilde{P}_{tb} = \delta\tilde{u} + a_3\tilde{k}(s)\delta\tilde{P}_{tb} + a_4\tilde{k}(s)\delta x\end{cases}$$

Finally, the transfer function is

$$\delta\tilde{x} = \frac{a_2}{s - a_1 - a_3 s\tilde{k}(s) + a_5\tilde{k}(s)}\delta\tilde{u}$$

with $a_5 \triangleq a_1a_3 - a_2a_4$. We now study the stability of this SISO system when closing the loop with

$$\delta u = k_p\left(1 + \frac{k_i}{s}\right)(\delta x_{sp} - \delta x),$$

where $k_i > 0$. For that purpose, one can investigate the location of the roots of

$$s - a_1 - a_3 s\tilde{k}(s) + a_5\tilde{k}(s) + a_2k_p\left(1 + \frac{k_i}{s}\right) = 0 \tag{6}$$

The following result holds

Lemma 1. There exists $k_p^* > 0$ such that for all $k_p \geq k_p^*$ the closed loop system, which characteristic equation is (6), is stable.

Proof 1. Consider $k_p > 0$, and assume that one can find a root s of the characteristic equation (6) such that $Re(s) \geq 0$. Then, $|e^{-s\tau}| < 1$. Using the mean-value inequality, $\left|\frac{1-e^{-s\tau}}{s\tau}\right| < 1$ and $\left|\frac{1-e^{-s\tau}-s\tau e^{-s\tau}}{(s\tau)^2}\right| < 1$. Therefore,

$$\begin{aligned}|\tilde{k}(s)| &= \left|k_2\frac{1-e^{-s\tau}}{s} + k_1\frac{1-e^{-s\tau}-s\tau e^{-s\tau}}{s^2}\right|\\ &< |k_1\tau^2| + |k_2\tau|\end{aligned}$$

Furthermore,

$$\begin{aligned}|s\tilde{k}(s)| &< \left|k_2(1-e^{-s\tau}) - k_1\tau e^{-s\tau} + k_1\frac{1-e^{-s\tau}}{s}\right|\\ &< 2(|k_1\tau| + |k_2|)\end{aligned}$$

Thus,

$$\begin{aligned}&|a_3 s\tilde{k}(s) - a_5\tilde{k}(s)|\\ &\quad< 2|a_3|(|k_1\tau| + |k_2|) + |a_5|(|k_1\tau^2| + |k_2\tau|)\end{aligned}$$

On the other hand, since $Re(s)$, k_i, k_p, a_2 and $-a_1$ are all positive, then

$$\begin{aligned}&\left|s - a_1 + a_2k_p\left(1 + \frac{k_i}{s}\right)\right|\\ &\quad\geq Re(s) - a_1 + a_2k_p\left(1 + k_i\frac{Re(s)}{|s|}\right)\\ &\quad\geq -a_1 + a_2k_p \geq 0\end{aligned}$$

In summary, if s is a solution of the characteristic equation (6) with positive real part then

$$\begin{aligned}&|-a_1 + a_2k_p|\\ &\quad< 2|a_3|(|k_1\tau| + |k_2|) + |a_5|(|k_1\tau^2| + |k_2\tau|)\end{aligned} \tag{7}$$

Let

$$k_p^* \triangleq \frac{2|a_3|(|k_1\tau| + |k_2|) + |a_5|(|k_1\tau^2| + |k_2\tau|)}{a_2}$$

For $k_p \geq k_p^*$, equation (7) does not hold. This proves that, for such values, one cannot find a solution of equation (6) with positive real part. Necessarily, the closed loop system is stable which concludes the proof. Finally, notice that this lower bound does not depend on k_i (which is positive by assumption). ∎

4. OLGA SIMULATIONS

4.1 Control structure

Based on the theoretical analysis of section 3 and Lemma 1 in particular, we propose the following control scheme. We use a simple P-controller on the casing head pressure, P_a, using the well head pressure P_t. Then, we derive the production choke values through the static map in Figure 4

$$\begin{aligned}P_t &= P_t^{sp} + k(P_a^{sp} - P_a)\\ u &= \frac{a}{P_t - b} + c\end{aligned}$$

where a, b and c are fit parameters.

4.2 Simulation setup

Tests of our control structure are conducted on a well simulated in OLGA®2000. We use the compositional tracking and the Matlab-OLGA link toolboxes. We consider that the gas mass flow rate injected at the casing head can be arbitrarily chosen. The reservoir has constant PI, pressure and temperature. Along the well, temperatures are kept constant as well as the separator pressure, i.e. the boundary pressure at the well head.

The following scenario is considered. In the beginning, the controller is switched on. The gas injection rate is 0.4 kg/s. This corresponds to a stable equilibrium. Then, at t = 1h the gas injection rate is decreased to 0.3 kg/s. About the corresponding steady-state, the open-loop system is unstable. When eventually the well is almost stabilized (at t = 2h50), the proportional gain is discontinuously lowered to provide a soft landing and avoid unnecessary damped oscillations. Finally at t = 13h50, the controller is switched off. As expected, the system diverges toward a self-sustained oscillatory regime. The gas injection

rate in the tubing is almost constant. The observed behavior is indeed a density-wave as shown in the fourth graph of Figure 8.

5. CONCLUSION

Our point is to demonstrate the relevance of feedback control to address the various instabilities of gas-lifted wells. Among these are the casing-heading and density-wave. The first case was already addressed in (Eikrem and Golan, 2002). The results reported here stress that, theoretically and in simulations, the density-wave phenomenon can be handled by a similar strategy. For that purpose, we use a straightforward controller. Clearly, results could be improved upon using, at least, gain-scheduling and feed-forward terms. In our approach, no extra sensors are required. It is debatable whether such performance can be achieved in actual wells, given the actuation limitations and sensor noises. This point is currently under investigation.

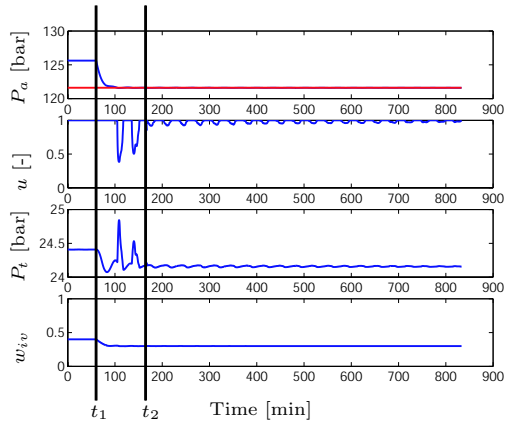


Fig. 7. Stabilization of the casing head pressure through production choke manipulations (first 840 min). At time t_1 the injection rate is switched from 0.4 to 0.3 kg/s. At time t_2 the proportional gain is reduced from 12 to 2.

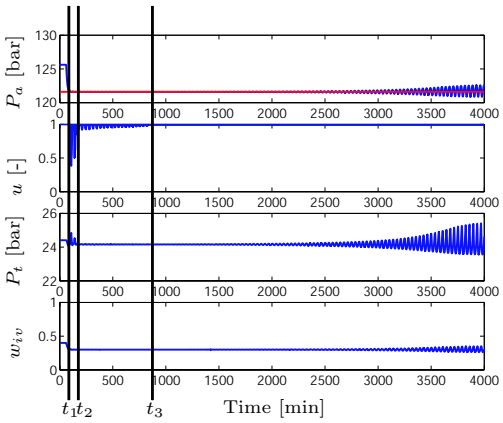


Fig. 8. Stabilization of the casing head pressure through production choke manipulations. At time t_3 the controller is switched off. The well gradually diverges towards its self-sustained oscillatory regime.

REFERENCES

Asheim, H. (1988). Criteria for gas-lift stability. *Journal of Petroleum Technology* pp. 1452–1456.

Brown, K. E. (1973). *Gas lift theory and practice.* Petroleum publishing CO., Tulsa, Oklahoma.

Cholet, H. (2000). *Well production. Practical handbook.* Editions TECHNIP.

Chorin, A. J. and J. E. Marsden (1990). *A mathematical introduction to fluid mechanics.* Springer-Verlag.

Dalsmo, M., E. Halvorsen and O. Slupphaug (2002). Active feedback of unstable wells at the Brage field. In: *SPE Annual technical Confererence and Exhibition.* number SPE 77650.

der Kinderen, W. J. G. J., C. L. Dunham and H. N. J. Poulisse (1998). Real-time aritificial lift optimization. In: *8th Abu Dhabi International Petroleum Exhibition and Conference.* number SPE 49463.

Duret, E. (2005). Dynamique et contrôle des écoulements polyphasiques. PhD thesis. École des Mines de Paris.

Eikrem, G. O., B. Foss L. Imsland B. Hu and M. Golan (2002). Stabilization of gas-lifted wells. In: *Proc. of the 15th IFAC World Congress.*

Eikrem, G. O., L. Imsland and B. Foss (2004). Stabilization of gas-lifted wells based on state estimation. In: *International Symposium on Advanced Control of Chemical Processes.*

Hu, B. and M. Golan (2003). Gas-lift instability resulted production loss and its remedy by feedback control: dynamical simulation results. In: *SPE International Improved Oil Recovery Conference in Asia Pacific.* number SPE 84917. Kuala Lumpur, Malaysia.

Imsland, L. S. (2002). Topics in Nonlinear Control - Ouput Feedback Stabilization and Control of Positive Systems. PhD thesis. Norwegian University of Science and Technology, Department of Engineering and Cybernetics.

Jansen, B., M. Dalsmo, K. Havre, L. Nøkleberg, V. Kritiansen and P. Lemétayer (1999). Automatic control of unstable gas-lifted wells. *SPE Annual technical Confererence and Exhibition.*

Sinègre, L., N. Petit and P. Ménégatti (2005). Distributed delay model for density wave dynamics in gas lifted wells. In: *Proc. of the 44th IEEE Conf. on Decision and Control, to appear.*

Torre, A. J., Z. Schmidt, R. N. Blais, D. R. Doty and J. P. Brill (1987). Casing heading in flowing wells. *SPE Production Engineering.*

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

A CONTROL STRATEGY FOR AN OIL WELL OPERATING VIA GAS LIFT

Plucenio, A. * **Mafra, G. A.** * **and Pagano, D. J.** *

* *Departamento de Automação e Sistemas, Universidade Federal de Santa Catarina, 88040-900 Florianópolis-SC, Brazil*
e-mail: {plucenio, gmafra, daniel}@das.ufsc.br

Abstract: This article proposes a control strategy for an oil well operating via gas-lift. The well model is implemented in the OLGA simulator (Scandpower) using an orifice valve (no moving parts) downhole with control in the gas lift surface valve and production choke. The dynamic identification uses the knowledge of the process static gain as the nonlinear static block of a Hammerstein model representation. An Adaptive Notch Filter was designed to damp the resonant system frequencies. Simulation results showed that the control strategy proposed was able to move the well operating point along the region of economical interest and to reject the perturbation imposed on the downstream side of the production choke.

Keywords: Nonlinear process control, oil well production, continuous gas-lift

1. INTRODUCTION

At the beginning of an oilfield development program the producing formation pressure will be sufficiently strong to push the produced fluids to surface. Each well will have an Inflow Performance Relationship curve (IPR) relating the flow rate with the pressure in front of the perforated zone. Knowledge of the well geometry, the formation fluids characteristics and the pressure at the tubing head, can be used to estimate the tubing performance which gives the pressure at the bottom of the production tubing for different flow rates . The intersection of the Tubing performance with the IPR curve will define the well operating point (Flow rate and Pressure in front of the perforated zone). The tubing head pressure can be changed with a choke to put the well in different operating points along the IPR curve. As the formation pressure declines the IPR curve changes moving the intersection point towards zero flow rate. Several artificial lift methods are employed to boost the formation fluid flow rate. The gas-lift is one of these methods. Gas is injected in the production tubing lowering the tubing performance curve permitting intersection points with higher flow rates. At the surface, the production from several wells is directed to a common separator. Gas, oil and water are separated and part or the total amount of gas leaving the separator is treated, compressed and distributed to the wells for injection. Gas-lift wells are completed with several gas lift valves distributed along the production tubing. Except for the deeper gas-lift valve, the valves are used to start the well providing gas injection in the production tubing sequentially from the shallowest to the deepest valve. After the start-up the only valve providing gas entrance to the production tubing is the deepest valve, also named operating valve. A surface valve, used to control the gas injection flow rate and a production choke are also part of a gas-lift well setup. Gas lift valves are mechanical valves normally inserted in a gas lift mandrel and can be recovered for maintenance using slick-line operations. The costs involved with the maintenance of these valves, the risks associated with slick line intervention and the need to better control the dynamic of the gas lift wells may be the motivating factors which have led to the study of new gas lift control strategies. A common characteristic of these studies is the utilization of an orifice valve as the operating valve downhole and the control made with surface actuation.

Several contributions to the solution of this problem have been published (Eikrem *et al.*, 2004), (Eikrem *et al.*, 2002),(Imsland *et al.*, 2003).

This paper is organized as follows: in Section II the gas lift control strategy is presented; then the control algorithm is discussed in Section III; The results obtained for the simulated well are shown in Section IV and finally the conclusions are drawn.

2. GAS LIFT CONTROL

A typical steady state relationship between the gas injection mass flow-rate and the wellhead formation fluid mass flow rate, considering a constant wellhead pressure is shown in figure 1. The slope of the curve is steep for low gas injection mass flow rate due to the predominance of the gravity term of the pressure drop in the production tubing. As the gas injection mass flow rate is increased, the friction term becomes important decreasing the slope until the curve reaches a maximum at point P_1. The plot of the pressure in front of the perforated zone exhibit a curve which is almost a mirror image with the minimum occurring at the same point. The control of a gas-lift well is normally realized according to an optimization strategy. Although the gas used for injection is not lost, there is a cost for the gas compression. The oil, gas and water fluid fractions produced by each well, have different economical effects. The produced water is normally treated before disposal, the gas and oil have different market values. The resources available may also constrain the operation limiting the separation, transport or compression capacity and will have an impact in the distribution of compressed gas to a group of wells. Several works have treated this problem with different optimization approaches as in (Nakashima and Camponogara, 2005) and (G.A. *et al.*, December-2002). A more general approach is to consider the reservoir recovery optimization and to treat the gas lift optimization as a sub-problem. The upper optimization layer could give, for each well the optimum pressure range in front of the perforations. The gas lift optimization would find the optimum gas allocation to comply with the upper layer while minimizing costs for a certain gas injection mass flow rate availability and installations constraints. For the control it means that the well will operate within a defined region of the curve $Q_{liq.} = f(Q_{inj.})$ as shown in figure 1. A gas lift well flow rate can become very oscillatory when changing the gas injection flow rate or letting the wellhead pressure to vary due to perturbations on the downstream equipment. This oscillatory behavior is stronger when the pressure drop in the production tubing is dominated by the gravity term. It tends to diminish as the friction term becomes comparable. This explain the reason for well operators to increase gas injection as a last resort to stabilize a gas lift well. In most cases this is not the optimum solution. On the contrary, depending on the gas availability, well

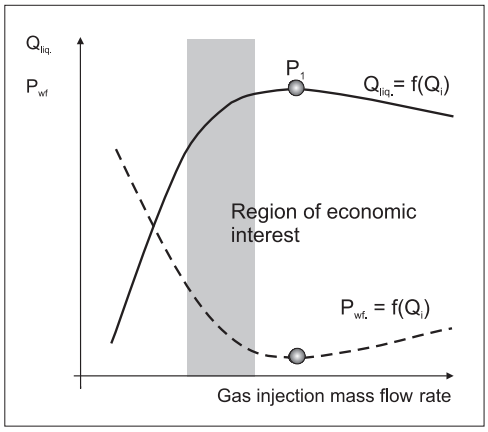


Fig. 1. Wellhead mass flow rate and Downhole pressure x Gas injection mass flow rate

production, the costs involved, the optimum operating point may be much lower than the point P_1 of figure 1. Controlling a gas lift well with an orifice valve downhole was discussed in (Plucenio, 2002) using wellhead mass flow rate as the process variable and gas injection mass-flow rate as the control input. This study was realized in a well modeled in the OLGA 2000 simulator. Identification of the dynamics relating the two mass flow-rates was done at different operating points. The transfer functions obtained although different among themselves did not present any major difficulty for control application like transport delay or non-minimum phase behavior. Unfortunately, mass flow rate of a multiphase flow is still a very expensive measurement today and the oil industry is not ready to perform it for each gas-lift well. Some oil companies have started to adopt the installation of permanent downhole pressure gages in their gas lift wells. On the other hand, well tests are regularly conducted to determine the well performance for different gas injection flow rates. Some tests are also required by the regulatory agencies (ANP in Brazil). During these tests, measurements of oil, gas and water production, downhole pressure measurements can be plotted versus gas injection mass flow-rates by directing the well production to a test separator. Downhole pressure gages can be used to estimate the pressure in front of the perforated zone, even when not installed exactly in front it. This can be done using the knowledge of the well completion geometry, the produced fluid characteristics and flow-rates. In most cases the management of an oilfield is realized by allocating desired values for this pressure along the productive life of the oilfield in order to drain the reservoir in an optimum way. The pressure in front of the perforations can be written as

$$P_{wf} = P_{wh} + P_{pt} + P_t, \tag{1}$$

where P_{wf} is the pressure in front of the perforations, P_{wh} is the pressure in the wellhead, P_{pt} is the pressure drop between the wellhead and the downhole pressure gage installation point and P_t is the pressure drop in the tail between the depth of the downhole gage instal-

lation and the perforations. The pressure measured by a downhole pressure gage is normally

$$P_{dg} = P_{wh} + P_{pt}, \tag{2}$$

The desired P_{wf} can be converted to a desired P_{dg} if one considers that in steady state the value of the pressure drop P_t can be estimated quite well with the measurements obtained during the periodic well testings. The pressure in the wellhead can be written as a sum of the separator pressure (P_{sep}), the pressure drop in the surface pipe connecting the wellhead to the separator P_{sp} and the pressure drop in the production choke (P_{pc}).

$$P_{wh} = P_{sep} + P_{sp} + P_{pc}, \tag{3}$$

Changes in the downhole pressure (P_{wf}) will change the formation fluid flow rate and consequently the pressure drop in the production tubing, production choke and surface pipe. This changes the pressure P_{wh} and the P_{wf} itself. This interaction is typical of a multivariable control problem. The strategy presented in this study considers the control of the P_{wh} acting in the production choke opening in order to keep it at a desired value P_{whd}. A cascade control is used to control the P_{pt} at a desired value P_{ptd} acting in the gas injection mass flow rate. The gas injection mass flow rate is accomplished controlling the gas injection valve opening. The desired pressure at the downhole pressure gage is obtained as

$$P_{dgd} = P_{whd} + P_{tpd}, \tag{4}$$

This strategy avoids the multivariable representation and transforms the problem into two SISO (Single Input, Single Output) problems. The response speed of the P_{wh} control is much faster than the P_{pt} loop response. Stabilizing P_{wh} and P_{pt} is equivalent to stabilizing the wellhead flow rate. The production choke nominal size should provide a minimum pressure drop when fully opened. It should operate partially closed in order to be able to compensate pressure increases in the downstream side.

The control strategy is shown in Figure 2.

Control of the wellhead pressure acting in the Production choke opening and the control of the gas injection mass flow rate acting in the surface gas injection valve will not be discussed. PI (Proportional and Integral) controllers were used for this purpose in both cases. These controllers were incorporated in the model in order to obtain an identification of the P_{pt} vs. Q_i dynamics. Figure 3 shows the steady state relation between the mass injection flow rate Qi and the pressure drop in the production tubing P_{pt}. The region of economic interest elected is shown in the figure 3. The knowledge of the process static gain was used to assembly an identification algorithm based on the Hammerstein approach where a nonlinear memoryless function is applied on the input followed by a linear dynamic model. The identification was realized between the P_{pt} variable and the transformed input variable $Qi^{'} = f(Qi)$.

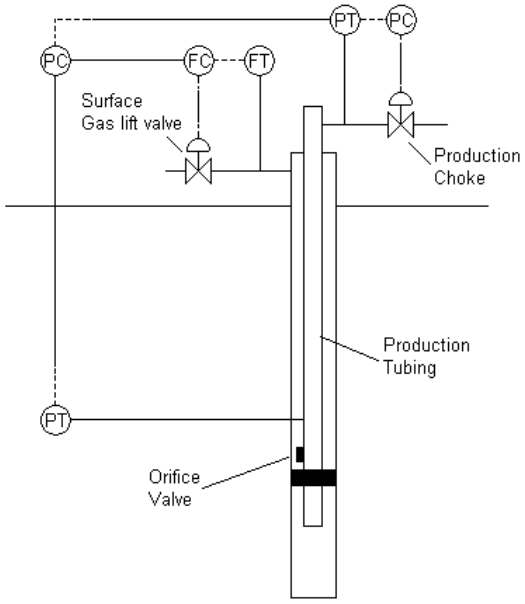


Fig. 2. Gas Lift Control Strategy

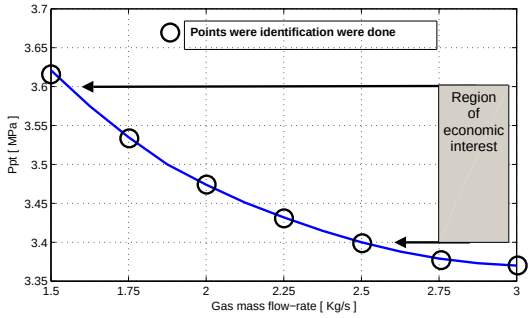


Fig. 3. Pressure drop in Production Tubing vs. Gas injection mass Flow-rate

The dynamic behavior of $P_{pt} = f(Qi^{'})$ is non-linear along all the operating region of the plant. To obtain linear models, several ARX identifications were realized between $P_{pt}\ and\ Qi^{'}$ exciting the system around the operating points indicated in figure 3. The representation of all linear models is shown as discrete transfer function in equation 5.

$$H(z) = \frac{b_1 z^2 + b_2 z}{z^3 + a_1 z^2 + a_2 z + a_3} \tag{5}$$

Figure 4 shows the identification result obtained exciting the well around $Qi = 1.5Kg/s$ using a multilevel PRBS signal.

The poles and zeros obtained with the linear models move smoothly as shown in Figure 5. The non-minimum phase characteristics is evident and can be easily explained. Gas is injected to decrease the pressure drop in the Production Tubing. For the gas to enter the tubing, the pressure on the upstream side of the orifice valve has to be increased. This has the initial effect of increasing the pressure on the downstream side of the orifice valve (Production Tubing) before

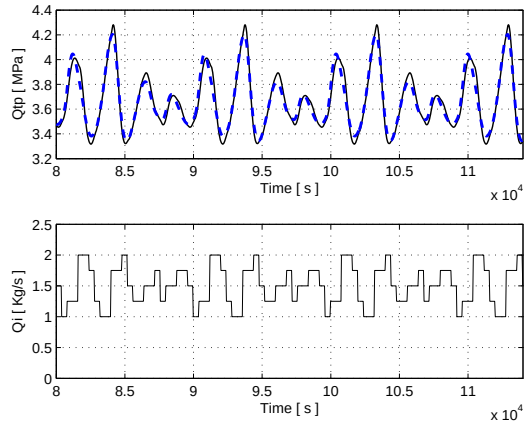


Fig. 4. Identification around Qi=1.5 Kg/s

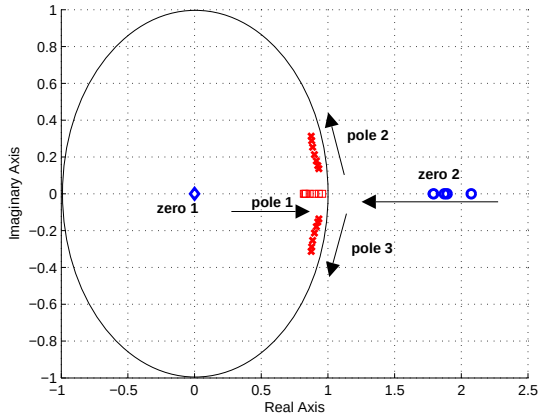


Fig. 5. Poles and Zeros of Linear Models

the benefit of the pressure drop is achieved when the gas moves up the tubing.

3. PROPOSED CONTROL ALGORITHM

The family of models obtained present one pole in the real axis plus a pair of complex conjugated poles. There is one zero at the origin and one non-minimum phase zero. The complex poles represent a resonant frequency which changes with the well operating point. It was decided to apply a control structure composed of a Reference filter, a PI (Proportional and Integral Control) with a linearizing gain look-up table plus an Adaptive Notch Filter. The control scheme is shown in Figure 6.

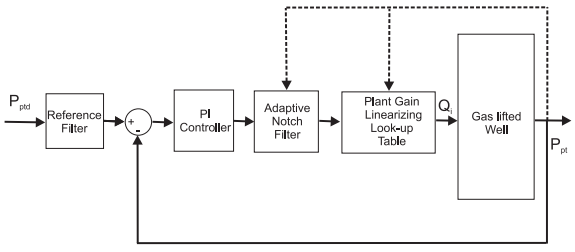


Fig. 6. Control Scheme

For every model obtained with the identification process, a Notch Filter was designed as

$$F(z) = \frac{f_1 z^2 + f_2 z + f_3}{z^2 + d_1 z + d_2}, \qquad (6)$$

where f_1, f_2 and f_3 depend on the process variable P_{pt} and are designed to provide zeros that will cancel the plant complex poles. A parameter α was found that can be used to derive any filter as a combination of the filters found at the limits of the operating region. This parameter is a function of the process variable P_{pt}. Omitting the z operator, any filter can be expressed as

$$F_{Ptp} = \alpha F_{Ptp1} + (1 - \alpha) F_{Ptp2}, \qquad (7)$$

where F_{Ptp1} and F_{Ptp2} are the filters designed for the limits of the operating range defined by the process variable P_{pt}. It is expected that the slow nature of the P_{pt} control loop will permit to adapt the filter as the process variable moves along the operating region. The Linearizing look-up table block makes the plant to appear linear to the controller as far as static gain is concerned. The look up table is built using the parameters found in the identification of $Qi^{'} = f(Qi)$ in the identification process and the expected operating range. The Reference Filter cancels the zero effect due to the PI control and defines the dynamic desired for the P_{pt} set-point changes. The zero of the PI control is chosen at the left of the leftmost real pole among all the models. Figure 7 shows the Closed Loop Root Locus when applying the PI Control and the Adaptive Notch Filter for one of the ARX models identified. Figure 8 shows the detail of the model pole

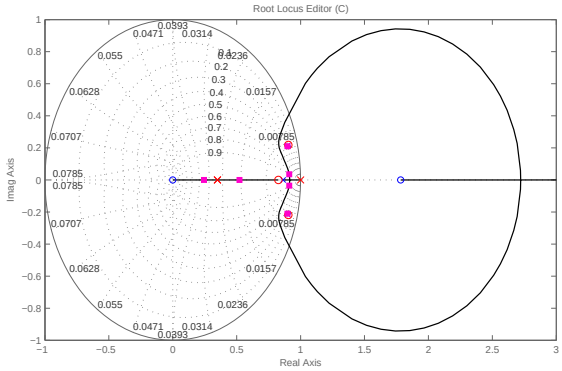


Fig. 7. Root Locus for the Closed Loop Control

cancellation due to the Notch Filter zeros and the direction to be followed for different operating points of the process.

4. RESULTS WITH OLGA SIMULATOR

This strategy was implemented in a well operating via gas lift modelled in the OLGA simulator. The model uses two constant pressure boundaries, one to represent the gas lift supply and the other to represent the separator. The gas injection is done at the mud line

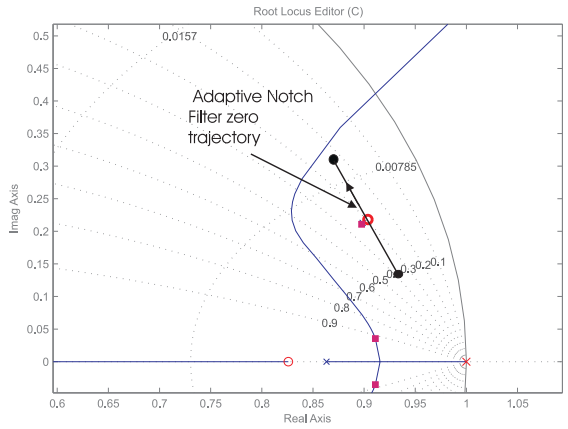


Fig. 8. Notch Filter Zero Trajectory

of an offshore well with a dry x-mass tree completion. The representation of the well is shown in figure 9 and details in table 1.

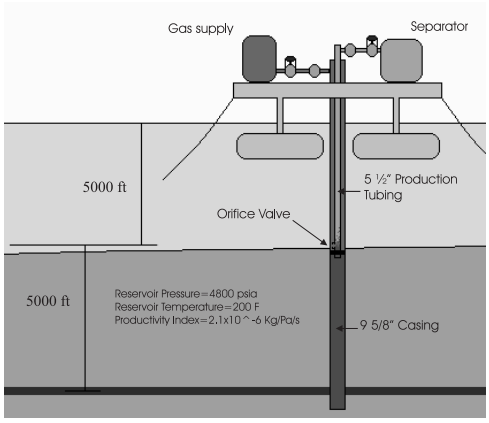


Fig. 9. gas lift well implemented in the Olga simulator

Table 1. Gas lift well implemented in the Olga simulator

$$
\begin{aligned}
Total\,depth &= 10000\,ft \\
Gas\,injection\,depth &= 5000\,ft \\
Tubing\,size &= 5\frac{1}{2}\,in. \\
Casing\,size &= 9\frac{5}{8}\,in. \\
Gas\,lift\,surface\,valve\,nom.\,size &= 1\frac{1}{2}in. \\
Production\,choke\,nom.\,size &= 2\frac{1}{2}\,in. \\
Reservoir\,pressure &= 33.0948\,Mpa\,(4800\,psi) \\
Separator\,Pressure &= 2.5855\,Mpa\,(375\,psi) \\
Wellhead\,Pressure &= 2.9992\,Mpa\,(435\,psi) \\
Reservoir\,temperature &= 93.3\circ C \\
Reservoir\,Productivity\,index &= 2.1\times 10^{-006} Kg/Pa/s
\end{aligned}
$$

The OLGA simulator ability to communicate with the Matlab environment permitted to test the control strategy performance. The well operating point was moved along different set-points within the region proposed as it can be seen in figures 10, 11 and 12. It can be noticed that the liquid flow rate moves much slower than the pressure drop in the production tubing (P_{pt}).

This behavior makes the control strategy proposed interesting since it does not require high gains for the P_{pt} control loop. In order to test the control response to perturbations, a change in the pressure at the downstream side of the production choke (separator) was imposed beginning at time 9.72hs. The pressure was increased by 10 psi in 10 minutes, kept at this value for another 10 minutes and decreased to normal value at the same rate. The production choke opening presented in the second plot of figure 12 shows the quick response of the Wellhead pressure PI controller. The Wellhead pressure changed less than 2 psi as shown in the first plot of the same figure. The effect in the P_{pt} pressure and liquid mass flow rate is nearly unnoticed.

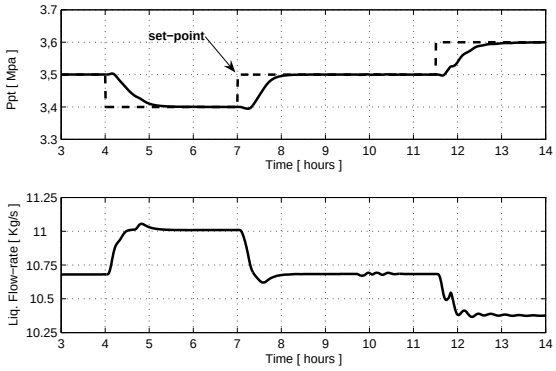


Fig. 10. Ppt and Liquid mass flow rate

Figures 13 and 14 show the response obtained without P_{pt} control but keeping the local controllers for the gas injection flow rate and Wellhead pressure. The gas injection mass flow rate were forced to the steady state values reached in the closed loop experiment. The system resonant frequencies are clearly not damped and show up in the P_{pt} and liquid flow rate. This oscillatory behavior is not acceptable on the management of an oil well. The rapid changes in the P_{pt} pressure will also be present on the pressure in front of the perforated zones. This may cause several problems, from formation damage to sand production in case of well with unconsolidated formations. The liquid flow rate oscillations will make the separation process much more difficult requiring larger separators.

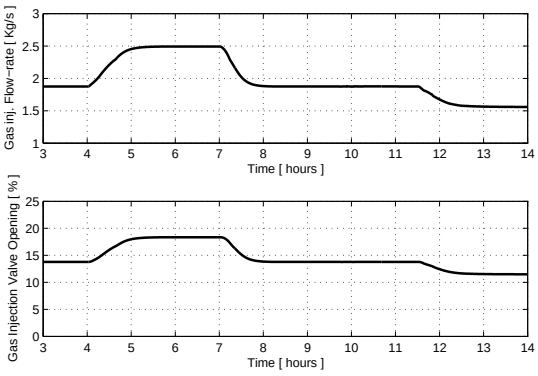


Fig. 11. Gas Injection flow rate and Gas valve opening

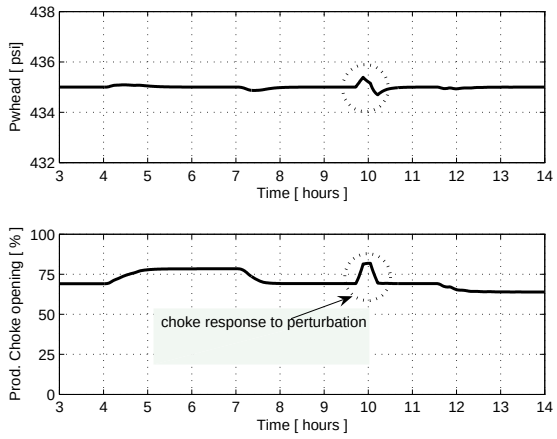


Fig. 12. Wellhead Pressure and Production choke opening

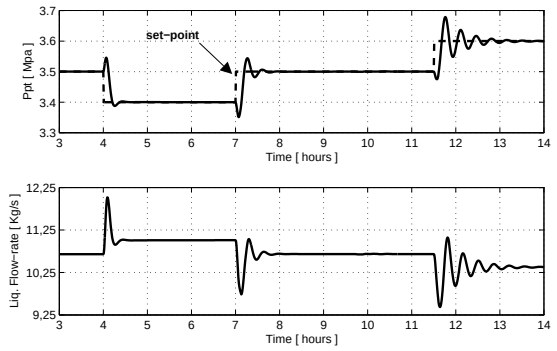


Fig. 13. Ptp, Liquid flow rate Open Loop response

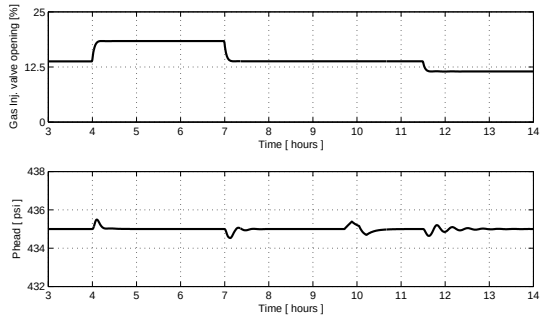


Fig. 14. Ptp, Liquid flow rate Open Loop response

5. CONCLUSIONS

In this paper, a strategy to control an oil well operating via gas-lift was presented. It uses measurements of gas injection mass flow rate, downhole pressure and wellhead pressure. The actuation is made on the surface gas lift valve and production choke openings. The strategy was tested in an oil well implemented in the OLGA simulator. It proved satisfactory to move the operating point along the region of economical interest of the well and to reject the perturbation imposed on the downstream side of the production choke. The strategy proposed can be easily implemented. It uses algorithms largely available as function blocks of industrial network control systems. The strategy was not tested to operate the well at lower gas injection mass flow-rates. The identification procedure would have to be applied to much more operating points in order to obtain the parameters needed for the Adaptive Notch Filter implementation.

6. ACKNOWLEDGMENTS

The authors were funded by Agência Nacional do Petróleo, Gás Natural e Biocombustiveis (ANP), Brazil, under project aciPG-PRH No 34 ANP/MCT. The authors would also like to acknowledge Scandpower for providing an academical OLGA software license.

7. REFERENCES

Eikrem, G. O., B. Foss and L. Imsland (2004). Stabilization of gas lifted wells based on state estimation. *IFAC ADCHEM2004, Hong Kong*.

Eikrem, G. O., B. Foss, L. Imsland, B. Hu and M. Golan (2002). Stabilization of gas lifted wells. *Proceedings of the IFAC 15th World Congress, Barcelona, Spain*.

G.A., Alarcón, Torres C. F. and Gómez E. (December-2002). Global optimization of gas allocation to a group of wells in artificial lift using nonlinear programming. *Journal of Energy Resources Technology*, $124 : 262 - 268$.

Imsland, Lars, Gisle Otto Eikrem and Bjarne Foss (2003). State feedback control of a class of positive systems: Application to gas lift control. *ECC'03, Cambridge*.

Nakashima, P. and E. Camponogara (2005). Optimization of lift-gas allocation using dynamic programming. *IEEE Transactions on Systems, Man, and Cybernetics-Part A, 2005*.

Plucenio, Agustinho (2002). Stabilization and optimization of an oil well network operating with continuous gas-lift. *SPE Annual Technical Conference and Exhibition, San Antonio, Texas*.

Session 8.2

Practical Applications of Modeling and Identification

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

MODELING FOR CONTROL OF REACTIVE EXTRUSION PROCESSES

Swapnil C. Garge[1], Mark D. Wetzel[2] and Babatunde A. Ogunnaike[1]

[1]*Department of Chemical Engineering, University of Delaware, DE 19716*
[2]*E. I. du Pont de Nemours and Co., Inc., Wilmington, DE, 19880*

Abstract: A modeling and control framework for effective control of end-use product characteristics of reactive extrusion processes is proposed. We discuss for an example process the development of two important components of the modeling scheme: an identified model that relates manipulated inputs to process outputs and a first principles process model that relates the inputs to quality variables. *Copyright © 2005 Babatunde A. Ogunnaike*

Keywords: multirate, multivariable system, nonlinearity, modeling, model-based control, identification

1. INTRODUCTION

Reactive extrusion processes have assumed significance in the polymer processing industry due to their wide-ranging applications in manufacturing neat polymers, polymer blends and, more recently, nanocomposites. The ever-tightening customer demands on product specs have necessitated comprehensive dynamics and control studies of these processes, which, until now, have mostly focused on the control of a single variable such as viscosity (Broadhead *et al.*, 1996).

The specifications are usually given for product properties such as tensile strength, melt index etc. that are rarely measured online and are obtained using measurements having variable time requirements (multirate measurements). For the purposes of effective control, it is necessary to separate these properties into product quality variables '*q*' (Melt index, viscosity, density, etc) and end-use physical characteristics '*w*' (Toughness, UV/chemical resistance, etc.). The process output variables '*y*' (Die pressure, melt temperature etc.) are measured online at a much faster rate than the product properties. In addition to the multirate nature of the system, effective control of the product properties is made difficult by the complex process mechanisms that result from the interactions between fluid mechanics, heat transfer, reaction kinetics and the extruder geometry. The ultimate objective of this work is to develop a framework for controlling product properties and assuring acceptable end use performance.

1.1 Modeling Scheme

The approach adopted for this challenging problem begins with an adequate mathematical representation of the relationships between variables across the entire processing chain. Such a representation will serve two crucial purposes: (i) provide estimates of the infrequently measured product properties at a much faster rate, and (ii) facilitate the development of a control system to meet the above mentioned objective.

Fig. 1 shows a schematic representation of the proposed modeling scheme, which consists of the following models: i) M_{uy} – a model relating the manipulated variables, u, to process output variables, y, (ii) M_{uq} – a process model relating the manipulated variables, u, to the internal product quality variables, $\hat{q}$, (iii) M_{qq} – a model relating internal quality variables, $\hat{q}$, to product quality variables, q,

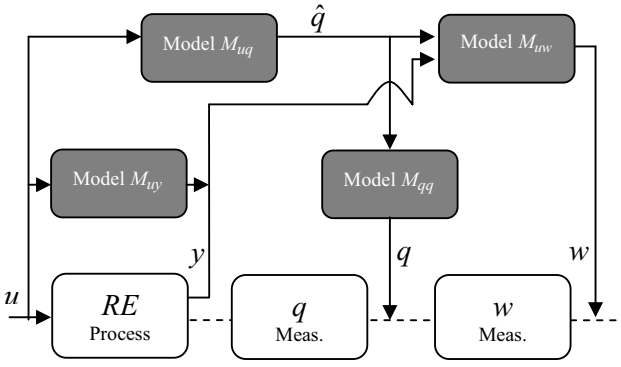


Fig. 1. Proposed modeling scheme

(iv) M_{qw} – relating internal quality variables, $\hat{q}$, to end use characteristics, and (v) M_{wz} (*not shown in Fig. 1*) – relating end-use physical characteristics, w, to product performance in end-use, z_w, a binary variable that represents acceptable performance as 1, and unacceptable performance as 0.

A modeling scheme that relates the different classes of variables sequentially, although more intuitive, is impractical in this case. This is because it is usually difficult to model the relationships between the process outputs and the measured product quality variables, since these variables based on the measurements, which are selected on practical grounds such as the availability of sensor locations on the extruder, and developing a mathematical relationship between these two classes of variables is not straightforward. To overcome this problem, an additional class of variables, called internal product quality variables '$\hat{q}$' (Composition, weight average molecular weight) that constitutes an indirect way to link 'y' with 'q', is introduced in the proposed modeling scheme.

1.2 Control Scheme

The control paradigm is predicated upon using the above network of models for two important tasks: (i) to translate the customer requirements on end-use performance to set points for process variables, and (ii) to make appropriate modifications (that is, to take control action) wherever appropriate along the manufacturing chain based on all available information. For this purpose a multivariable cascade control scheme (Fig. 2) is proposed, consisting of a fast model-based controller *C1* for the inner loop between the manipulated variables and the output variables and a slower (model based) controller *C2*, which will translate the end use performance objectives to set points for the output variables. In addition to these loops, there exists an innermost basic regulatory control loop, which ensures that the set point changes in the manipulated variables are efficiently tracked.

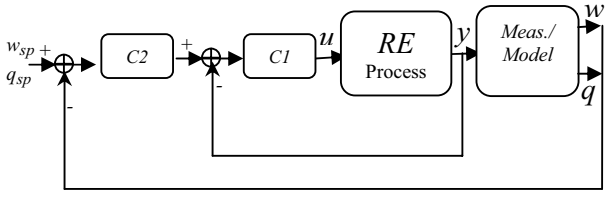


Fig. 2. Proposed control scheme. *C1* and *C2* are the inner loop and outer loop controllers respectively.

The objective of this paper is to discuss the development of two important components of the proposed modeling scheme: (i) model M_{uy} identified from input/output data, and (ii) model M_{uq} developed using first principles. The example system consists of the reaction of a functionalized ethylene co-polymer, "Elvaloy®" (Ethylene/n-Butyl Acrylate/Glycidal Methacrylate Terpolymer (E/BA/GMA)) with an acid co-polymer "Nucrel®" (Ethylene/Methacrylic Acid Copolymer (E/MAA)) in a Coperion W&P ZSK-30mm co-rotating, intermeshing twin screw extruder.

2. MODEL M_{uy}

Model M_{uy} represents the mathematical relationship between the manipulated inputs and the outputs. Due to the complexity of the interacting process mechanisms, a first-principles model is impractical for control at this level. A model identified from carefully obtained input/output data is a more practical alternative. A systematic procedure was employed to carry out the three major steps of system identification of this class of processes: (i) experimental test design and collection of input/output data, (ii) model structure and order selection and (iii) model validation.

2.1 Inputs and Outputs

The manipulated inputs (u) are the screw speed, E/MAA feed-rate, E/BA/GMA feed-rate and the barrel temperatures for the seven extruder zones. The changes in all the inputs are implemented manually. However, the barrel temperature regulatory controller loop dynamics are much slower than the process dynamics excluding the inner regulatory loops. Therefore, with the exception of a step change, any other dynamic change in the barrel temperature is impractical. The inner loops for other manipulated inputs are fast compared to the process; therefore, comparatively rapid changes can be implemented in these variables. The process outputs are E/MAA weight fraction in the melting zone, die pressure, exit melt temperature, and motor power. With the exception of the E/MAA weight fraction, which can be easily obtained from the feed-rates of the two polymers, all other outputs are measured.

2.2 Identification Tests

There are two components to the experimental test design: (i) preliminary tests (ii) final identification test. The preliminary tests were aimed at obtaining *à priori* knowledge needed for the design of the final identification test. These tests consisted of a series of step changes as well as simultaneous staircase changes in the manipulated inputs at two operating points: (A) a low melting-zone composition of E/BA/GMA (~ 1%) inducing a low extent of reaction and a relatively small change in the product viscosity as compared to the viscosity of the pure E/MAA feed, and (B) a high melting-zone composition of E/BA/GMA (~ 4 %) inducing a high extent of reaction and a significant change in the product viscosity. This selection of the operating points enabled study of the effect of the reaction on the process dynamics behavior. The key results of the preliminary test were: (i) the reaction has a strong nonlinear effect on the process dynamics, (ii) the system is ill-conditioned at the operating point A,

and (iii) the process exhibits an approximately linear behavior in the vicinity of the two operating points.

Based on these results, the final identification test, aimed at developing a linear model at each operating point, was designed using generalized binary noise (GBN) signals (Tullenken, 1990). As proposed by Zhu (2001), GBN signals with a mean switching time equal to 1/3rd of the process settling time (~360 s) were used in the *open loop* identification experiments. In recognition of unmeasured disturbances and the high measurement noise levels (low signal to noise ratio), the duration of the test was approximately 15 times the process settling time. Three uncorrelated GBN signals were administered simultaneously in the screw speed (u_1), E/MAA feed-rate (u_2), and E/BA/GMA feed-rate (u_3) at each operating point (*Fig. 3*). These uncorrelated signals are suitable for the identification of a well-conditioned process as well as identifying the high gain direction of an ill-conditioned process.

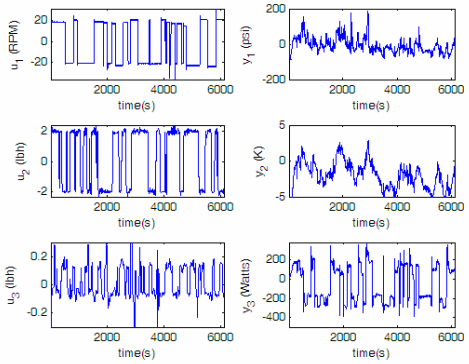


Fig. 3. GBN test data at operating point B: Inputs, left panels, outputs right panel

An additional test, based on the open loop design suggested by Zhu (2001), was used to identify the low gain direction of the process at the operating point A. The GBN signals, having correlated high amplitude periods combined with uncorrelated low amplitude periods, were administered to the inputs u_1 and u_2, while the input u_3 was unchanged.

2.3. Model Order and Structure

The purpose of model structure and order selection was to obtain a suitable linear model around each operating point. The candidate models structures included: (i) Multi Input Single Output (MISO) structure having a common model for each output, and (ii) Multi Input Multi Output (MIMO) structure having a common model for all outputs. The suitability of these model structures for representing the input/output data was tested for different parametric models (for e.g., Auto regressive Moving Average with eXogenous inputs [ARMAX], Box Jenkins [BJ] etc.). The parameter estimation was performed using the System Identification Toolbox of Matlab®. Out of these models, MISO Box Jenkins (Eq .1) model was found suitable for describing the dynamics in the input/output data.

The criteria used for selecting the model orders were: (i) percentage of output variation that is captured by the model, (ii) Akaike's final prediction error, and (iii) pole-zero diagrams to check for over parameterization. Table 1 presents the selected model orders for the MISO BJ model corresponding to each output.

Table 1 Identified BJ models at the two operating points

	MISO Model Orders				*Delay* (n_{d1}, n_{d2})		
O/I	u_1 (n_a,n_b)	u_2 (n_a,n_b)	u_3 (n_a,n_b)	*Noise* (n_c,n_d)	u_1	u_2	u_3
Operating Point A							
y_1	5,5	5,5	5,5	2,2	0,0	0,0	0,0
y_2	5,5	5,5	5,5	1,1	0,0	0,0	0,0
y_3	3,3	3,3	3,3	0,0	0,0	0,0	0,0
y_4	2,2	2,2	2,2	1,1	0,0	0,0	0,0
Operating Point B							
y_1	5,5	5,5	5,5	3,3	1,1	1,1	1,1
y_2	4,4	4,4	4,4	1,1	0,0	0,0	0,0
y_3	3,3	3,3	3,3	1,1	0,0	0,0	0,0
y_4	2,2	2,2	2,2	1,1	0,0	0,0	0,0

$$y(t) = \frac{B(q)}{A(q)}u(t-n_{d1}) + \frac{C(q)}{D(q)}e(t-n_{d2}) \qquad (1)$$

where,

$$A(q) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n_a}$$
$$B(q) = \quad b_1 q^{-1} + \ldots + b_n q^{-n_b}$$
$$C(q) = 1 + c_1 q^{-1} + \ldots + c_n q^{-n_c}$$
$$D(q) = 1 + d_1 q^{-1} + \ldots + d_n q^{-n_d}$$
$$and, \quad q^{-k}v(k) = v(t-k)$$

The overall model structure for M_{uy} is thus a collection of local linear models that are valid in each region of the input/output variable space defined by the E/BA/GMA weight fraction.

The validation data consists of a fraction of the final test(s) data that was not used for estimation, in addition to the preliminary test data. (Fig. 4).

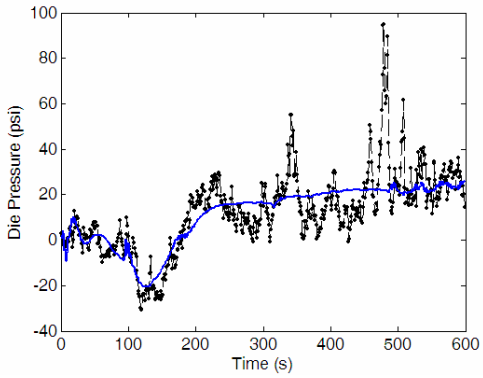


Fig. 4. Measured die pressure (dots) and the identified model predictions (solid line) for a step change in E/MAA feed-rate.

3. MODEL M_{uq}

Model M_{uq} represents the mathematical relationship between the manipulated inputs, u, and the internal quality variables, $\hat{q}$, a class of *unmeasured* variables motivated by the modeling scheme (Fig. 1). Therefore, an empirical model is not appropriate for this purpose; first-principles modeling, an alternative to the empirical models, is, therefore, an obvious choice. For validation, however, alternative quality variable measurements are essential. In this work, online viscosity measurements are used for validation.

3.1. Experimental Investigation to Obtain Fundamental Process Information

The development of such a first principles model, M_{uq}, is facilitated if the information regarding the key process mechanisms that influence the process dynamic behavior is available. The key process mechanisms may differ for dissimilar reaction systems and commonly made assumptions in the modeling of reactive extrusion processes, such as Newtonian fluid flow and isothermal conditions, may be inappropriate. Specifically, it is important to know the degree of interaction between the reaction kinetics, melt flow in the complex extruder geometry, heat transfer, and the melting of the polymers that must be captured in the first principles model. Since this information is not always available in the literature, an experimental process investigation becomes essential.

Three techniques were employed for the experimental investigation of the example process: (i) Pulse Technique: to probe the melting dynamics, (ii) Residence Time Distribution (RTD) experiments to probe the coupling between the reaction kinetics and fluid flow, and (iii) Step change experiments to study the transient process behavior. A summary of the main results of this investigation is presented below (for more details, see Garge *et. al*, 2005):

- Very little reaction occurs during the melting of the polymers.
- The melting zone location is weakly influenced by the operating conditions.
- The reaction rate is strongly influenced by the E/BA/GMA concentrations.
- The viscosity change due to the reaction does not significantly influence the residence time distribution, and in turn the flow of the polymer melt.
- The interaction between the heat transfer and the reaction kinetics is weak at the low E/BA/GMA concentration, but strong at the high E/BA/GMA concentration. Similarly, the dependence of the reaction on the flow is weak at the low E/BA/GMA concentration, while it appears to be somewhat stronger at high E/BA/GMA concentrations.

3.2. Model Components

As noted earlier, the model M_{uq} is needed for mathematically describing the relationships between the manipulated inputs, u, and the internal product quality variables, $\hat{q}$, such as the product composition, which are directly dependent on the reaction. Obviously, the process mechanisms that are related to or that affect the reaction need to be modeled appropriately, whereas the modeling of other mechanisms is not critical. The key aspects of the strategy for developing a transient process model are:

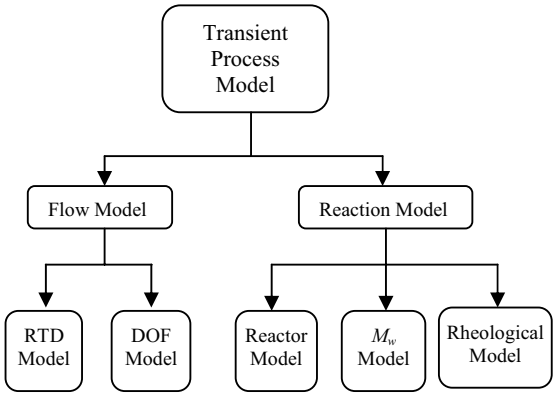


Fig. 5. Components of the transient process model

1] The experimental results suggesting that the melting does not significantly influence the reaction indicate that a simple melting model will suffice to predict the axial location where the melting process is almost complete. The location of the melting zone, however, is almost invariant for different operating conditions. Instead of developing a detailed melting model, it is proposed to demarcate the melting zone extent empirically for each operating condition and use this melting zone extent for all the other operating conditions.

2] The experimental results suggest that the flow of the polymer melt is weakly dependent on the reaction in the extruder and vice versa. Taking advantage of this fact a divided model structure consisting of distinct flow and reaction models is proposed, as illustrated in Fig. 5. Such a structure allows for the development of a simple flow model and a relatively complex reaction model.

3] The reaction kinetics are influenced by the heat transfer, at least at the high E/BA/GMA operating point (B). Therefore, these two mechanisms are coupled in the reaction model.

The details of the reaction model as well as the flow model are presented elsewhere (see Garge *et. al*, 2005). The salient features of the two are discussed below.

Flow Model*:* The flow model is split into two parts: (i) a RTD model and (ii) a 'Degree of Fill' (DOF) Model. Such a distinction, although not essential for developing a transient flow model, considerably

simplifies the coupling between the flow and reaction models. The DOF model demarcates the *partially filled* and the *fully filled* regions. The RTD model predicts the mean residence time and time delay that is used, along with the degree of fill profile obtained from the DOF model, to calculate the average axial velocity in different regions of the extruder.

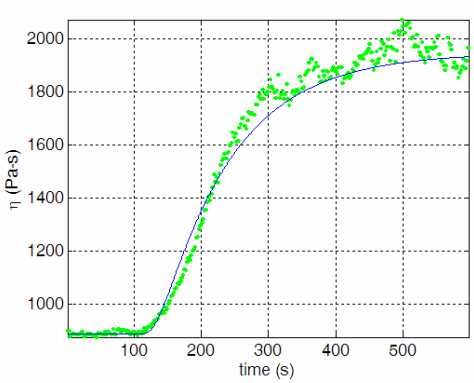


Fig. 6. Measured viscosity (dots) and the transient process model predictions (solid line) for a step change in E/BA/GMA feed-rate.

Reaction Model: The main objective of the reaction model is to quantify the effect of the reaction on the quality variables, **q**, specifically on the product composition, viscosity and the average molecular weight ($\overline{M}_w$). It can be divided into three components: (i) a reactor model to describe how the reaction occurs in the extruder. The one-dimensional axial dispersion model was found suitable for the example system. Non-isothermal terms were included to incorporate heat transfer effects (Eq. 2). (ii) a $\overline{M}_w$ model to represent the effect of the reaction on the average molecular weight, and (iii) a rheological model to represent the effect of the reaction on the viscosity.

Concentration for the species i (C_i):

$$\frac{\partial C_i}{\partial t} + \frac{\partial (v_x C_i)}{\partial x} = \frac{\partial}{\partial x}\left[D_L \frac{\partial C_i}{\partial x} \right] - R_{C_i} \qquad (2a)$$

Temperature (T):

$$\rho C_p \left[\frac{\partial T}{\partial t} + \frac{\partial (v_x T)}{\partial x} \right] = \frac{\partial}{\partial x}\left[\lambda \frac{\partial T}{\partial x} \right] + Q \qquad (2b)$$

In the above equations, v_x stands for average axial velocity, D_L is axial dispersion coefficient in the zone, R_{ci} is the reaction rate, ρ is the mean density of the polymer melt, C_p is its average specific heat, and Q represents the heat generation terms. The transient model, obtained by coupling the reaction and flow models, was validated with the viscosity measurements for step changes in the inputs, u (Fig. 6).

3. SUMMARY AND CONCLUSIONS

A modeling and control framework has been proposed for effective control of the product end-use properties of a reactive extrusion process. The modeling framework consists of a network of models that mathematically represent the relationships between the different classes of variables across the manufacturing chain. Due to practical reasons, the introduction of an additional class of unmeasured variables, called internal quality variables, is essential for this modeling scheme. The proposed multivariable cascade control scheme will then use this network of models, in addition to the all the available measurements, to guarantee acceptable end use performance.

The paper briefly describes two important components of the modeling scheme: (i) M_{uy} – a model relating the manipulated variables, u, to process output variables, y; (ii) M_{uq} – a process model relating the manipulated variables, u, to the internal product quality variables, $\hat{q}$. Based mainly upon the anticipated computational effort and applicability, it is suggested to use an empirical model obtained from carefully obtained input/output data for M_{uy}, and a first-principles model for M_{uq}. The development of these models was illustrated for an example process consisting of a reaction between two functionalized polymers in a co-rotating twin-screw extruder.

A systematic experimental approach for obtaining input/output data, consisting of a pre-test and a final test based on GBN signals, was employed for the identification of the model M_{uy}; a multi-model structure, consisting of a MISO BJ model for each output, was found suitable in this case. The first principles model, M_{uq}, was based on an experimental investigation carried out to probe the important physical mechanisms of the system. A divided model structure, consisting of a reaction model and a flow model, was developed and validated against the viscosity data for transient step changes in the manipulated inputs.

Our current effort is focussed on developing the models M_{qw} and M_{wz} in the modeling scheme of Fig. 1. The melting process strongly influences the product morphology, and thus plays an important role in determining some of the end-use product properties such as tensile strength. It is, therefore, important to incorporate this "melting effect" in the model M_{qw} that relates the quality variables to the end-use properties. For this purpose, we propose to use a two-stage procedure consisting of: (i) developing an empirical model that will provide system parameters related to the melting process, and (ii) quantitatively relating these parameters, along with appropriate quality variables, to the end-use properties (Garge *et. al*, 2006).

The primary purpose of the model M_{wz} is to relate customer feedback to the end-use properties, with customer feedback as a binary variable, z_w, that is '1' for acceptable product performance and '0' for

unacceptable performance. Based on this representation, we propose to use a *binary logistic regression model* for representing the relationships between z_w and the end-use properties.

Acknowledgements: The authors acknowledge assistance from Susan Latimer and Donald Denelsbeck of The Dupont Company in conducting the experiments.

REFERENCES

Broadhead T. O., W. I. Peterson, and J. M. Dealy (1996). Closed loop viscosity control of reactive extrusion with a in-line rheometer. *Polym. Eng. Sci.*, **36(23)**, 2840–2851.

Garge S. C., M. D. Wetzel., and B. A. Ogunnaike (2005). Control relevant modeling of reactive extrusion processes. *Manuscript in preparation*.

Garge S. C., M. D. Wetzel., and B. A. Ogunnaike (2006). An empirical model for melting in a co-rotating twin-screw extruders. *SPE ANTEC Tech Papers*.

Tullenken, H.J.A.F. (1990). Generalized binary noise test-signal concept for improved identification experiment design. *Automatica*, **26 (1)**, 37-49.

Zhu Y. C. (2001). *Multivariable system identification for process control*. Elsevier Science Ltd., Oxford, UK.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

FACTORS AFFECTING ON-LINE ESTIMATION OF DIASTEREOMER COMPOSITION USING RAMAN SPECTROSCOPY

Sze-Wing Wong [a], Christos Georgakis [a, *], Gregory Botsaris [a], Kostas Saranteas [b], and Roger Bakale [b]

[a] *System Research Institute for Chemical & Biological Processes and Department of Chemical & Biological Engineering, Tufts University, Medford MA, USA*
[b] *Chemical Process Research and Development, Sepracor Inc., Marlborough MA, USA*

Abstract: This paper addresses the estimation of fractional composition of two diastereomers during crystallization. The estimation is obtained through a Partial Least Square (PLS) model that utilizes on-line Raman spectroscopy and additional process information such as temperature and slurry density. Several PLS models are developed that incorporate conditions that either neglect or account for variability in the additional process variables. It is argued that the model that incorporates both temperature and slurry density is the most accurate. *Copyright © 2005 IFAC*

Keywords: Calibration; Prediction Method; Regression Analysis; and Spectra Correlation

1. INTRODUCTION

The manufacturing of pharmaceuticals often involves separation of enantiomers, which are chiral molecules that are mirror images of each other. Since the physical properties of both enantiomers (R = right handed and S = left handed) are the same, a traditional separation method such as crystallization by seeding is feasible but it becomes very sensitive to the experimental conditions (Qian and Botsaris, 1997). Thus, the pharmaceutical industry usually relies on achiral synthesis of the enantiopure product or reacts the enantiomers free base with another chiral acid to produce different diastereomers. A diastereomer is a molecule that has more than two chiral centres. The resulting diastereomers have different physical properties such as solubility and often crystallize into different crystal structures. With these differences, crystallization can effectively separate the desired product with very high purity.

With strict government regulation in place, the purity of the final product is of crucial importance. Thus, the capability of on-line monitoring of the optical purity of the crystals will help to develop a robust crystallization procedure. The on-line Raman spectroscopy is suitable for this application since Raman can detect the lattice vibrations corresponding to the translatory and rotatory motion of the entire molecule within the lattice structure of the crystal (Ferraro, 1971). As a result, Raman spectroscopy is capable of differentiating similar molecules with different crystal lattice structures. Several authors have demonstrated the ability of monitoring the changing compositions of two different crystals on-line during a solvent-mediated polymorphic transformation with Raman (Berglund, et al., 2000; Glennon, et al., 2003; Ono, et al., 2004; Myerson, et al., 2005). In addition, chemometric techniques can be applied to the Raman spectra to detect slight peak shifts and to remove noise from the signals (Falcon and Berglund, 2004; Rades, et al., 2002; Starbuck, et al., 2002).

The use of fiber optic to collect data through an immersion probe allows analysis of solid phase composition in real-time. However, the Raman intensity of the solids depends on the amount of inelastic scattering of the solids detected by the analyzer within the detection zone. As a result, the relative Raman intensity corresponding to the diastereomers in the slurry will be impacted by a number of solid-state factors. Several authors have suggested that Raman intensity with respect to different polymorphs may be a function of particle size and shape (Glennon, et al., 2003 and Wang, et. al., 2002). This is based on the assumption that Raman signals primarily come from the surface of the crystals. Additionally, slurry density may be another solid-state factor since the number of crystals inside the detection zone will influence the Raman intensity of the solids. In theory, the Raman spectrum will be affected by the amount of solvent and solids detected. Thus, slurry density

* Corresponding author.
Tel.: 617-627-2573; Fax: 617-627-3991
E-mail address: Christos.Georgakis@tufts.edu

should impact the Raman signal intensity of the solid phase.

In the present work, we examine whether the information provided by Raman spectroscopy is sufficient or whether it needs to be complemented by additional process measurements in order to provide an accurate estimation, through a Partial Least Square (PLS) model, of the solid composition of one of the two diastereomers involved in the production of an active pharmaceutical ingredient, denoted here as compound A. The selection of factors was based on the cooling crystallization procedure of compound A. Since the changing temperature, slurry density, and percent composition of the diastereomers in solid phase would affect the peak position and peak intensity, those were the variables selected in our modelling task. Partial Least Square regression (PLS) was used to quantify the composition of the diastereomers mixture.

2. EXPERIMENTAL SECTION

2.1. *Materials*

HPLC grade solvents were used as received from commercial suppliers without further purification. The starting materials (racemic free base of compound A and a chiral acid, denoted here as D) that met the specifications defined by Sepracor Inc. were used as received from qualified suppliers without further purification. Sepracor Inc. provided all of the materials.

2.2. *Preparation of Pure S-D Diastereomer*

Racemic free base of compound A was reacted with D (a chiral acid) in solvent and the solution was heated and held at 5 degrees above the saturated temperature to allow for complete dissolution. The solution was then slowly cooled and seeded with 2% by weight of the S-D diastereomer at the specified seeding temperature. The seeded slurry was cooled to a target isolation temperature. It was then followed by a filtration and drying step. The end product was analyzed by a chiral HPLC method resulting in optical purity of at least 97%.

2.3. Preparation of Pure R-D Diastereomer

The preparation of pure R-D diastereomer first involved purifying R enantiomer from the racemic free base of compound A. The pure R enantiomer would then react with D to form the R-D diastereomer. For the purification step, racemic free base of compound A was reacted with a chiral acid, denoted as L, in solvent and the solution was heated and held at 5 degrees above the saturated temperature to allow for complete dissolution. The solution was then slowly cooled and seeded with 2% by weight of the R-L diastereomer at a specified seeding temperature. The seeded slurry was cooled to a target isolation temperature. It was then followed by a filtration and drying step. The pure R-D diastereomer was produced by first letting the dry R-L crystal go through a free basing step to obtain pure R enantiomer. The pure R enantiomer then reacted with D and crystallized to form R-D diastereomer. The product was analyzed by a chiral HPLC method, resulting in optical purity greater than 98%.

2.4. Raman Spectroscopy

A RamanRxn1™ analyzer (Kaiser Optical System, Inc.) coupled with an immersion fiber optic probe was used for the in-situ measurements. Raman spectra were recorded using NIR excitation radiation at 785nm and the spectroscopy incorporates the TE-cooled CCD detector technology. All collected spectra were averaged over five accumulations collected over 8 seconds each.

Raman spectra were analyzed using either PLS_Toolbox 3.5 by Eigenvector Research, Inc. (Manson, WA) or the Unscrambler Chemometrics Software from Camo Inc. (Trondheim, Norway).

2.5. Calibration Experiments

In order to obtain spectra of a known amount of solid in suspension (slurry density) and percent composition of the S-D diastereomer in solid phase, the solution was first pre-saturated with respect to both diastereomers at specified temperatures (Table 1). The saturated solution was prepared by adding excess amounts of racemic free base of compound A and D in the solvent system and heated until dissolution. After nucleation occurred upon cooling, the slurry was under constant stirring for two hours to ensure it reached equilibrium at the specified temperature and finished by a filtration step. The saturated solution was kept in a jacketed round-bottom flask to maintain constant temperature.

Spectra of the standards were obtained for each pure diastereomer and of different binary mixture (from 0% to 100% of S-D) in the saturated solution. The spectra were collected at different temperatures (from 0^{o}C to 40^{o}C) and with different slurry densities (from 13.3 g/L to 80 g/L) that were within the range of the crystallization procedure. A total of 65 standards were used with varying conditions (Table 1) and were divided into two groups -- training and testing groups. 55 standards were selected and used to construct the model while the remaining 10 standards were used to test the accuracy of the model. The 10 standards from the testing group were randomly selected to cover the whole experimental space. The Raman probe

was inserted top-down into the 15mL vial and all the spectra were collected under constant stirring with magnetic stir bar to suspend the slurry.

Table 1 Experimental Condition for Standards

Fixed Variable	Fixed Variable	Changing Variable	# of samples
20 ^{0}C	13.3 g/L	0-100 % S-D	16
40 ^{0}C	13.3g/L	0-100 % S-D	6
30 ^{0}C	13.3g/L	0-100 % S-D	6
10 ^{0}C	13.3g/L	0-100 % S-D	6
0 ^{0}C	13.3g/L	0-100 % S-D	8
0 ^{0}C	33.3g/L	80-95 % S-D	4
0 ^{0}C	20g/L	70-100 % S-D	4
15 ^{0}C	26.7g/L	50-100 % S-D	3
15 ^{0}C	40g/L	65-85 % S-D	3
15 ^{0}C	53.3g/L	80-100 % S-D	3
5 ^{0}C	66.7g/L	75-95 % S-D	3
5 ^{0}C	80g/L	80-100 % S-D	3

2.6. Crystallization Experiment

5g of compound A (racemic free base) was reacted with 1.7g of D in 238mL of solvent. The solution was kept in a jacketed round-bottom flask under constant stirring and the jacket was connected to a temperature-controlled chiller. The Raman probe was inserted at a 45-degree angle into the reaction flask and spectra were collected at 1-minute intervals.

The solution was first heated to 40 ^{0}C and held for 20 minutes to allow for complete dissolution. It was then cooled at a rate of 5 ^{0}C/min to 0 ^{0}C and the temperature remained constant for four hours. After nucleation occurred, samples of the slurry were drawn for HPLC analysis until the end of experiment. A total of six samples were drawn throughout the experiment.

The collected samples were first filtered and the mother liquor was sent for HPLC analysis for solute concentration and percent composition of the diastereomers in solution phase. The wet cake did not go through solvent wash to avoid dissolution of the crystals during the wash. The cake was then weighed and vacuum dried at 40 ^{0}C overnight. The dry cake was again weighed and the solid contents sent for HPLC analysis for percent composition of the diastereomers in solid phase. Since the evaporated solvent contained residual diastereomers in the solution phase, the amount of solute in the evaporated solvent was calculated with the weight difference of the cakes and multiplied by the concentration of solute in solution from HPLC analysis. The percent composition of the S-D diastereomer was then corrected with the residual solute from the solution phase.

2.7. *Data Pre-treatment*

Data pre-treatment using first derivative was performed to correct any scattering effect from the crystals. The first derivative of all the spectra was computed with the Savitzky-Golay method using second-degree polynomial fit and 11 points window (Madden, 1978).

The data matrix of the three PLS models composed of different combinations of measurements such as the entire Raman spectrum (spectrum range: 75cm^{-1} to 3300cm^{-1}), temperature, and/or slurry density as shown in Equation 1-3. The percent composition of the diastereomers was the independent variable of the PLS model.

D_1 = [Spectra] (1)

D_2 = [Temperature Spectra] (2)

D_3 = [Temperature Slurry Density Spectra] (3)

In order to compare the contribution of each factor equally, temperature (T), slurry density (D), percent composition (%), and spectra (S) were first scaled to have zero mean and variance of one as shown in Equation 4-9.

$$\tilde{T}_i = \frac{T_i - \hat{T}}{\sigma_T} \quad (4) \qquad \tilde{D}_i = \frac{D_i - \hat{D}}{\sigma_D} \quad (5)$$

$$\tilde{\%}_i = \frac{\%_i - \hat{\%}}{\sigma_{\%}} \quad (6) \qquad F_i = \int f_i(w)dw \quad (7)$$

$$\hat{F} = \frac{\sum_{i=1}^{n} F_i}{n} \quad (8) \qquad \tilde{S}_i = \frac{f_i(w) - \hat{F}}{\sigma_F} \quad (9)$$

where the scaled values ($\tilde{T}$, $\tilde{D}$, $\tilde{\%}$, and $\tilde{S}$) were subtracted by the average of all the standards ($\hat{T}$, $\hat{D}$, $\hat{\%}$, and $\hat{F}$) and divided by the standard deviation (σ). However, the spectra data was scaled slightly differently in which $\hat{F}$ (Eq. 9) is the average of the under curve area of all spectra (Eq. 7) and σ_F is the standard deviation of the under curve area of all spectra, F. It should be noted that $f_i(w)$ is the spectrum function with respect to w denoted as the wave number.

3. RESULT & DISCUSSIONS

3.1. Raman Spectra of Pure Diastereomers

The Raman spectra of the pure diastereomer in Fig 1 shows that there was only a slight difference between the two diastereomers. Hence, chemometric techniques need be employed to account for the subtle differences in the whole spectrum. In addition, the spectra of the pure S-D diastereomer were compared at different temperatures (Fig.2) and with different slurry densities (Fig.3). While the relative intensity differs slightly with temperature and there was no peak shift observed due to temperature effect, the Raman spectra of different slurry density showed differences in peak positions and peak shape. The

denser sample showed in Fig. 3 had more distinct shape peaks that resembled Raman spectrum of pure solid. It was an indication that the Raman analyzer detected higher amount of solids in the slurry.

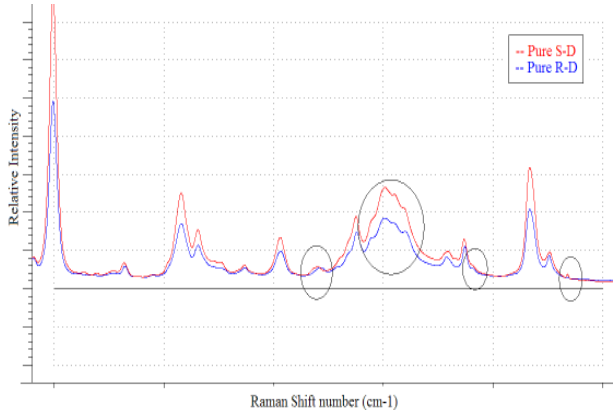


Fig. 1. Raman Spectra of pure R-D and S-D at the same temperature and slurry density. The circled regions of spectra are the slight differences between the diastereomers.

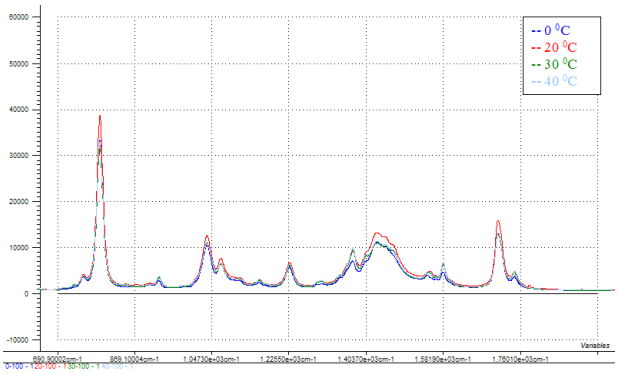


Fig. 2. Raman Spectra of pure S-D with the same slurry density and different temperature

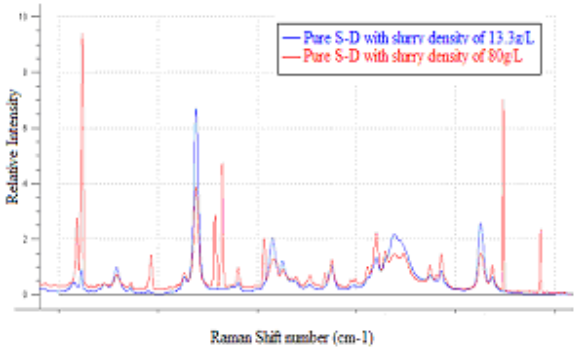


Fig. 3. Raman Spectra of pure S-D with different slurry density.

3.2. PLS Calibration Model

Three PLS models were developed to investigate whether additional process measurements (i.e. temperature and slurry density) would improve the accuracy of the estimation model. Each of the PLS models used the same training set that incorporates conditions that either neglect or account for variability in the additional process variables. The data matrix of each PLS model (Eq. 1-3) included 55 data points from Table 1.

The three PLS models were first validated with the testing set (10 standards points) and then compared the model estimation with the result from the HPLC analysis of the crystallization experiment. The testing set included standard points taken from different days of the experiment and with varying experimental conditions (Table 4). In addition, the testing set data points were not used in developing the models.

The estimation models were calculated by regressing the data of the training set with the percent composition using Partial Least Square Regression (PLS). Each training sets were divided into 4 subgroups for cross-validation. The cross-validation method used three of the four subgroups to build the model and tested with the last subgroup as validation step. The process was repeated with all combinations of subgroups as training and testing sets. The number of Latent Variable (LV) picked for each PLS models was based on the result from cross-validation and the percent variance captured by the LVs. The criteria were that the total number of LV would capture at least 80% of the y-block variance along with the lowest Root-Mean-Square-Error (RMSECV) from the Cross-Validation Error (Fig. 4). Finally, the Root-Mean-Square-Error (RMSE) and the relative percent error (% Error) would be used to compare the different PLS models (Eq. 10 and 11) with the testing set and the experimental result.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{i\,estimate} - y_{i\,experiment}\right)^2}{n}} \qquad (10)$$

$$\%\,Error = \left|\frac{\left(y_{i\,estimate} - y_{i\,experiment}\right)}{y_{i\,experiment}}\right| \times 100\% \qquad (11)$$

Where n is the number of the total data points.

From the results of the cross-validation, the number of Latent Variable was picked as described above. The detailed information of the models was summarized in Table 3. In addition, the predicted values of Y were plotted with the measured values of Y (Fig 5 – 7) to check the accuracy of the models. It was observed that the data points all lie close to the line indicating small prediction error form the models.

Table 3 Detailed Information about 3 PLS models

	Model 1	Model 2	Model 3
RMSECV	7.60	8.01	7.84
% Y Variance	96%	94%	95%
# of LV	7	7	7

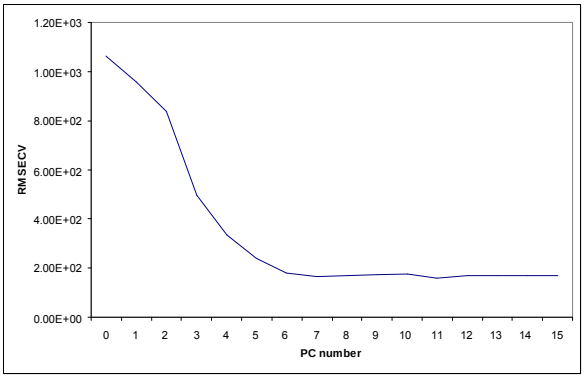


Fig. 4. RMSECV Vs. number of Latent Variable for the PLS model 2.

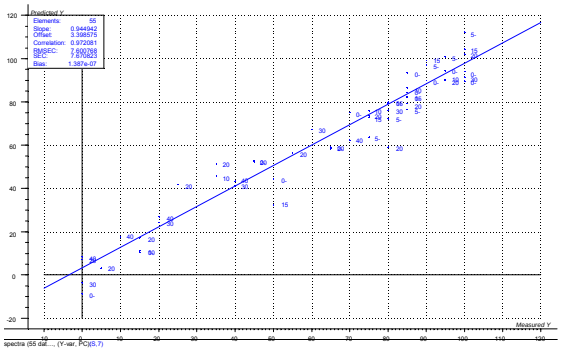


Fig. 5. Predicted Value of Y Vs. Measured Value of Y for Model 1

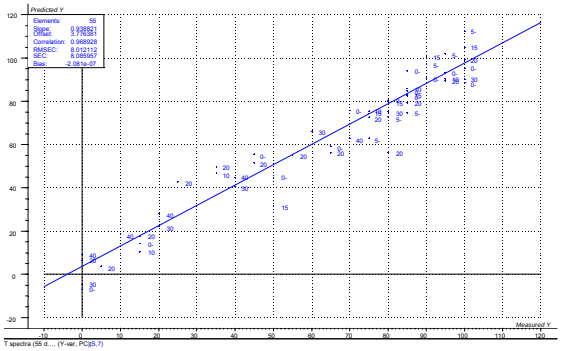


Fig. 6. Predicted Value of Y Vs. Measured Value of Y for Model 2

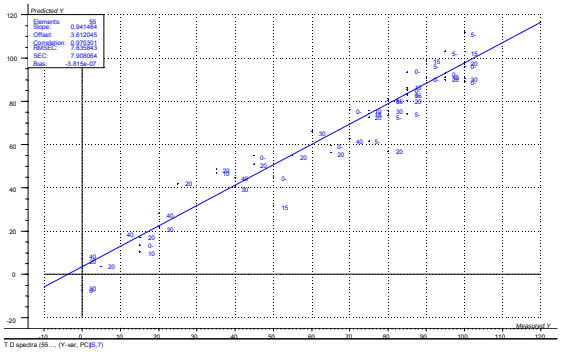


Fig. 7. Predicted Value of Y Vs. Measured Value of Y for Model 3

The final step in developing the PLS model was to validate the prediction result with the testing set. Each of the three PLS models was used to predict the percent composition of the S-D diastereomer and validated with the same testing group (Table 4). The result showed that the third PLS models that included all the process measurements was performing better than the other models (Table 5-6).

Table 4: % Composition of SD Diastereomer - Validation result from the Testing Set

Sample	Measured	Model 1	Model 2	Model 3
0C-75%-20g/L	75	74	75	76
0C-95%-13.3g/L	95	90	90	96
10C-55%-13.3g/L	55	62	62	60
15C-65%-40g/L	65	67	66	67
15C-80%-53.3g/L	80	81	82	83
15C-100%-26.7g/L	100	96	96	100
20C-30%-13.3g/L	30	31	30	29
20C-50%-13.3g/L	50	46	45	44
20C-90%-13.3g/L	90	91	88	87
40C-60%-13.3g/L	60	56	57	57

Table 5: % Error of the Validation result from the Testing Set

Sample	Model 1	Model 2	Model 3
0C-75%-20g/L	1.3	0.0	1.3
0C-95%-13.3g/L	5.3	5.3	1.1
10C-55%-13.3g/L	12.7	12.7	9.1
15C-65%-40g/L	3.1	1.5	3.1
15C-80%-53.3g/L	1.3	2.5	3.8
15C-100%-26.7g/L	4.0	4.0	0.0
20C-30%-13.3g/L	3.3	0.0	3.3
20C-50%-13.3g/L	8.0	10.0	12.0
20C-90%-13.3g/L	1.1	2.2	3.3
40C-60%-13.3g/L	6.7	5.0	5.0

Table 6: RMSE of the Validation result from the Testing Set

	Model 1	Model 2	Model 3
RMSE	3.6	3.65	3.08

3.3. Result from Crystallization Experiment

In order to verify their accuracy, the PLS models were used to predict the percent composition of the S-D diastereomer in a new crystallization experiment. While the standards from the calibration set were pre-mixed slurries in 15mL vials, the new crystallization experiment was run in a 250mL round-bottom jacketed flask. Six samples were draw sequentially after the onset of nucleation and submitted for HPLC analysis.

According to the existing knowledge of the experimental system, the R-D diastereomer has a wider metastable zone and a slower growth rate compared with the S-D diastereomer. The crystallization experiment aimed to investigate whether the R-D diastereomer would crystallize out simultaneously with the S-D diastereomer in an un-seeded environment. The HPLC analysis revealed that as nucleation occurred, both diastereomers

crystallized out. However, due to the slower crystallization kinetics of the R-D diastereomer, the percent composition of the S-D diastereomer slowly increased over time.

The results from the model prediction were compared with the HPLC analysis as shown in Table 7 and the RMSE was used as a means to compare the models (Table 8). Although all three models performed quite accurately with the testing set, the accuracy in predicting the data of the new crystallization experiment is understandably smaller. However, the superiority of Model 3 is demonstrated by including slurry density in the model. When nucleation occurred, there was only a thin layer of slurry in the solution. As the experiment progressed, more material crystallized out and the system tried to reach equilibrium between the two solids. It should also be noted that another PLS model (not included in Table 8) utilizing only slurry density and spectra data had a RMSE of 10.1 for the crystallization experiment. This clearly confirmed that the inclusion of slurry density into the PLS model improves the accuracy of estimation.

Table 7: % Error of the % Composition of SD Diastereomer Prediction in Crystallization Experiment

Sample	Model 1	Model 2	Model 3
1	15.6	11.3	1.9
2	18.6	15.6	4.9
3	26.0	23.2	13.4
4	15.2	13.9	3.8
5	29.0	25.4	16.8
6	24.9	21.1	11.7

Table 8: RMSE between Models and Experiment

	Model 1	Model 2	Model 3
RMSE	18.9	16.4	9.0

4. CONCLUSION

Three PLS models were constructed with the same 65 calibration standards using Raman spectra, temperature and/or slurry density data. The models were further tested and compared against data from a 250mL scaled crystallization experiment. This paper has shown that the in-situ Raman spectroscopy is capable of differentiating diastereomers in a crystallization slurry, provided the changing process parameters of temperature and slurry density are included in the calibration model. Because slurry density is not easily measured on-line without a sampling loop, it is essential to find an alternative on-line measurement from which to infer slurry density.

5. ACKNOWLEDGMENT

The financial support from Sepracor Inc in the form of an internship to Sze-Wing Wong and the availability of their experimental facilities is greatly appreciated.

REFERENCES

Berglund, K.A., Wang, F., Wachter, J.A., and Antosz, F.J. (2000). An Investigation of Solvent Mediated Polymorphic Transformation of Progesterone Using in Situ Raman Spectroscopy. *Organic Process Research & Development*, **Vol 4**, pp. 391-395

Falcon, J.A., and Berglund, K.A. (2004). In Situ Monitoring of Antisolvent Addition Crystallization with Principle Components Analysis of Raman Spectra, *Crystal Growth & Design*, **Vol. 4**, pp. 457-463

Ferraro, J.R. (1971). In: *Low-Frequency Vibrations of Inorganic and Coordination Compounds*, Plenum Press, New York

Hu, Y., Liang, J.K., Myerson, A.S., and Taylor, L.S., (2005). Crystallization Monitoring by Raman Spectroscopy: Simultaneous Measurement of Desupersaturation Profile and Polymorphic Form in Flufenamic Acid System. *Industrial and Engineering Chemistry Research*, **Vol 44**, pp. 1233-1240

Madden, H. H., (1978). Comments on the Savitzky-Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data. *Analytical Chemistry*, **Vol 50,** pp. 1383 - 1386

Ono, T., Horst, H.T., and Jansens, P.J. (2004). Quantitative Measurement of the Polymorphic Transformation of L-Glutamic Acid Using In-Situ Raman Spectroscopy. *Crystal Growth & Design*, **Vol. 4**, pp. 465-469

O'Sullivan, B., Barrett, P., Hsiao, G., Carr, A., and Glennon, B. (2003) In Situ Monitoring of Polymorphic Transitions, *Organic Process Research & Development*, **Vol. 7**, pp. 977-982

Qian, R. Y., and Botsaris, G. B., (1997) A New Mechanism for Nuclei Formation in Suspension Crystallizers: The Role of Interparticle Forces. *Chemical Engineering & Science*, **Vol 52**, pp. 3429-3440

Rades, T., Pratiwi, D., Fawcett, J.P., and Gordon, K.C. (2002) Quantitative Analysis of Polymorphic Mixtures of Ranitidine Hydrochloride by Raman Spectroscopy and Principal Components Analysis, *European Journal of Pharmaceutics and Biopharmaceutics*, **Vol. 54**, pp. 337-341

Starbuck, C., et al. (2002) Process Optimization of a Complex Pharmaceutical Polymorphic System via in Situ Raman Spectroscopy, *Crystal Growth & Design*, **Vol. 2**, pp. 515-522

Zhou, G., Wang, J., Ge, Z., Sun, Y., (2002) Ensuring Robust Polymorph Isolation Using In-Situ Raman Spectroscopy, *American Pharmaceutical Review*, Winter 2002

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

MODELING AND IDENTIFICATION OF NONLINEAR SYSTEMS USING SISO LEM-HAMMERSTEIN AND LEM-WIENER MODEL STRUCTURES

P. Bolognese Fernandes, D. Schlipf, J. O. Trierweiler

Group of Integration, Modeling, Simulation, Control and Optimization of Processes (GIMSCOP)
Department of Chemical Engineering, Federal University of Rio Grande do Sul (UFGRS)
Rua Luiz Englert, s/n CEP: 90040-040 – Porto Alegre – RS – Brazil
Fax: +55 51 3316 3277, Phone: + 55 51 3316 4072
Email: pedro@enq.ufrgs.br, david@enq.ufrgs.br, jorge@enq.ufrgs.br

Abstract: This paper applies the concept of linearization around the equilibrium manifold (LEM) already presented in the literature in order to construct model structures that can be viewed as extensions of the conventional Wiener and Hammerstein models. Instead of linear time-invariant subsystems in association with static nonlinearities, these extensions exhibit variable dynamic character and can therefore model a broader class of systems than the conventional cited approaches. Moreover, the identification strategy already used with LEM systems can be applied in order to construct such models from experiments, and the techniques destined for analysis and control of Wiener and Hammerstein systems can be applied promptly. To application of these concepts to the modeling and identification is demonstrated with a numerical example, considering a heat exchange system.

Keywords: System Identification; Nonlinear Models; Linearization; Interpolation; Wiener Systems; Hammerstein Systems.

1. INTRODUCTION

In order to control satisfactorily a nonlinear plant, two main approaches exist: either the use of "inherent" nonlinear control techniques or the use of robust linear methods to guarantee stability and adequate performance in despite of the nonlinear effects. In the first approach, it is necessary that nonlinear dynamic models are available, what is very often not the case. This is mainly due to the cost of nonlinear modeling and/or identification, but also to the fact that universal and fail-free methods allowing for the identification of accurate nonlinear models are still missing.

In order to describe the nonlinear characteristics that are encountered in the practice, it is often adequate to consider a given dynamic system as the composition of a linear dynamic block followed by a static nonlinearity, the so-called Wiener system. By reversing the order of the blocks, the result is the Hammerstein model. There is a plenty of literature on specific methods for identification of either Hammerstein or Wiener models, or both. A good survey on these model structures can be found in (Pearson, 1995).

Although interesting from the practical point of view, these approaches may be too simple if the description of a nonlinear dynamics is sought. Therefore, the concept of *linearization around the equilibrium manifold* (LEM) can be used to include such characteristic in the model representation. The advantage of the LEM systems is that they can be

constructed in a straightforward manner and result in simple, transparent model structures.

This paper is organized as follows: Section 2 reviews the concept of LEM systems already discussed in the literature, which is the basis for the two proposed model structures. Extended Hammerstein and Wiener models, in which the dynamics are dependent of the operating point, are shown in Sections 3 and Section 4, respectively. The models are then applied in Section 5 in the modeling and identification of a nonlinear system in a numerical example. Concluding remarks can be found in Section 6.

2. LEM SYSTEMS

Consider a continuous SISO nonlinear dynamic system of the form

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{r}(\mathbf{x},u) \\ y &= h(\mathbf{x}) \end{aligned} \tag{1}$$

where $\mathbf{r}: X \times U \to \Re^n$ is at least once continuously differentiable on $X \subseteq \Re^n$, $U \subseteq \Re$, and $h: X \to \Re$ is at least once continuously differentiable. The output equation will be frequently omitted in the sequel for shortness. The equilibrium manifold of (1) is defined as the family of constant equilibrium points

$$\Xi = \left\{ (\mathbf{x}_S, u_S, y_S) \in \Re^n \times \Re \times \Re : \; \mathbf{r}(\mathbf{x}_S, u_S) = \mathbf{0}, \;\; y_S = \mathbf{h}(\mathbf{x}_S, u_S) \right\} . \tag{2}$$

Similarly, the family of Taylor linearizations of (1) at the set of equilibrium points determined by (2) is given by

$$\dot{\mathbf{x}} = \left[\frac{\partial \mathbf{r}(\mathbf{x},u)}{\partial \mathbf{x}} \right]_{\mathbf{x}_S, u_S} (\mathbf{x} - \mathbf{x}_S) + \left[\frac{\partial \mathbf{r}(\mathbf{x},u)}{\partial u} \right]_{\mathbf{x}_S, u_S} (u - u_S) \tag{3}$$

and similarly for the output equation. Under the condition that the rank of $[\partial \mathbf{r}(\mathbf{x}_S,u_S)/\partial \mathbf{x}]$ is n for all points in the set Ξ (Wang and Rugh, 1987, Fernandes 2005), the equilibrium manifold and consequently the family of linearizations of (1) will be specified by one among the $n + 1$ variables $(\mathbf{x},u)$. Therefore, if this matrix is full rank, the input fully parameterizes both families of equilibrium points and linearizations. Calling the steady-state map $\mathbf{\Omega}: \Re \to \Re^n$, such that $\mathbf{r}(\mathbf{\Omega}(u),u) = \mathbf{0}$ (that is, the function $\mathbf{\Omega}$ gives the steady-state $\mathbf{x}_s$ corresponding to the constant input u_s), the input-parameterized linearization around the equilibrium manifold (LEM) of (1) is defined as the system (Fernandes 2005, Fernandes and Engell, 2005).

$$\dot{\mathbf{x}} = \mathbf{A}(u)(\mathbf{x} - \mathbf{\Omega}(u)) \tag{4}$$

where $\mathbf{A}(u)$ represents the evaluation of the Jacobian matrix $[\partial \mathbf{r}(\mathbf{x},u)/\partial \mathbf{x}]$ on $(\mathbf{\Omega}(u),u)$. Observe that the term arising from the second part of (3) is dropped by letting $u = u_s$. The output equation can be linearized in an analogous fashion, considering the stationary output mapping $\mathbf{\Psi}: \Re \to \Re$. The output function $\mathbf{\Omega}(u)$ can be obtained on the basis of the family of parameterized linearizations by integration of

$$\frac{d\mathbf{\Omega}(u)}{du} = -\mathbf{A}(u)^{-1}\mathbf{B}(u) . \tag{5}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the Jacobian matrices of $\mathbf{r}(\mathbf{x},u)$ with respect to $\mathbf{x}$ and u, respectively, evaluated on the equilibrium manifold. The model (4) has to be interpreted as *a* (state-affine) nonlinear system that possesses the same family of equilibrium points (2) and the same linearization family (3) as the nonlinear system (1). Following the discussion in (Fernandes, 2005), the LEM system can constitute also a good approximation of (1) in transient regimes away from the equilibrium manifold, depending on the "degree" of nonlinearity of the original system. Obviously, other representations that are equivalent on the equilibrium manifold can be constructed on the basis of a single parameter, other than u. Moreover, these representations can be easily interchanged, provided that the inverses of the corresponding elements in $\mathbf{\Omega}(u)$ and $\mathbf{\Psi}(u)$ exist.

The focus on input parameterization is due to the fact that identification experiments are carried out by exciting the plant with a designed input signal. In this sense, if one assumes that the local models can be identified by perturbing the plant around isolated equilibrium points, it is natural to use the input in order to parameterize the linearization family. Additionaly, since the exact LEM system (4) involves the infinite family of linearizations and of the equilibrium points of (1), described by the matrix functions $\mathbf{A}(u)$ and $\mathbf{\Omega}(u)$, in the identification context just a finite and probably small number of the members of these families are known, but one can still use approximation or interpolation methods (for example, polynomials, *splines* and so on) in order to "reconstruct" these functions from the available members. Therefore, an approximation to (1) can be constructed by means of a finite number of linear local models that are considered as members of its linearization family, obtained by means of a few "local" identification experiments. In order to solve the problem of constructing a state-space representation from local models obtained from input-output experiments, these can be transformed to a linear canonical normal form prior to the constructions of the approximate functions $\widetilde{\mathbf{A}}(u)$ and $\widetilde{\mathbf{\Omega}}(u)$ (Fernandes and Engell, 2005). In the absence of the values of all steady-states, the last function can be obtained by integration of $-\mathbf{A}(u)^{-1}\mathbf{B}(u)$ (Fernandes, 2005).

3. SISO LEM-HAMMERSTEIN MODELS

The LEM concept can be used to construct a Hammerstein-like model of (1) in which the dynamics depends on the operating point instead of

the LTI dynamics encountered in the usual Hammerstein structure (see Fig. 1).

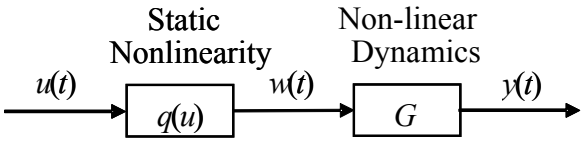


Fig. 1. LEM-Hammerstein model

A generic model with this structure can be defined in state-space form by

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}) + \mathbf{b}w = \mathbf{f}(\mathbf{x}) + \mathbf{b}q(u) \\ y &= \mathbf{c}\mathbf{x} \end{aligned} \tag{6}$$

where **b** and **c** are vectors of proper dimensions. A possibility of constructing a model of the form of Eq. (6) on the basis of the LEM models is to separate the static nonlinear gain function from the family of transfer functions (Pearson and Pottmann, 2000), that is,

$$G(s;\delta) = k(\delta)\cdot\frac{\beta_{n-1}(\delta)s^{n-1}+\ldots+\beta_1(\delta)s^1+1}{\alpha_n(\delta)s^n+\alpha_{n-1}(\delta)s^{n-1}+\ldots+\alpha_1(\delta)s^1+1} \tag{7}$$

where δ is a scalar parameterizing the set of equilibrium points/linearizations (u_s in this case). The resulting LEM-Hammerstein system is of the form

$$\begin{aligned} \dot{\mathbf{x}} &= \widetilde{\mathbf{A}}(w)(\mathbf{x}-\widetilde{\mathbf{\Omega}}(w)), w = q(u) \\ y &= \mathbf{c}x \end{aligned} \tag{8}$$

with

$$q(u) = \int_u k(\delta)\,\mathrm{d}\delta \tag{9}$$

such that the overall family of transfer functions correspond to that of (6). The LEM system (8) can be constructed with realizations of the parameterized transfer function of Eq. (7) in a suitable chosen coordinate basis, as for example a canonical or normal form. Obviously, Eq. (8) depends on the new input w, but an equivalent state- or output-parameterization can be easily constructed, as discussed above. These are nevertheless dynamically "worse" than the input-parameterized version (Fernandes, Engell and Trierweiler, 2004). This model can be obtained from experiments using the LEM approach as follows:

- identification of local linear models around some isolated operating points;
- transformation of the family of local models into a family of unit-gain linearizations;
- integration of $k(u_s) = -\mathbf{C}(u_s)\mathbf{A}(u_s)^{-1}\mathbf{B}(u_s)$ in order to obtain $q(u)$;
- interpolation of **A** and **B** in some suitable canonical form and integration of in $-\mathbf{A}(w)^{-1}\mathbf{B}(w)$ order to generate $\widetilde{\mathbf{\Omega}}(w)$.

Alternatively, since the "local" gain is the derivative of the stationary mapping with respect to u at a given operating point, q can be directly obtained by means of observations of the stationary output. This procedure can also be used iteratively, that is, values of y_s can be used to refine the interpolation of k and vice-versa.

4. SISO LEM-WIENER MODELS

In parallel to the Hammerstein-type structure considered above, it is also possible to define an "extended" Wiener model by replacing the linear block with an element exhibiting variable dynamics (Fig. 2).

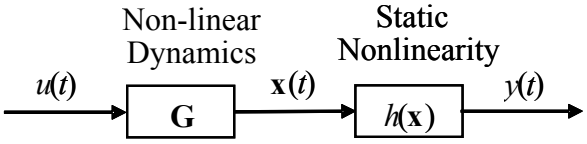


Fig. 2. LEM-Wiener model

Note that this model is not obtained by simply reversing the order of the blocks in Fig. 1, since the function h is a scalar valued function of n arguments whereas q is a n-valued function of *one* argument (that is, a collection of scalar functions).

This model can be defined in the state-space in the same fashion as in Eq. (6). Nevertheless, due to the nonlinear dependence of h on **x**, the input-parameterized LEM model would exhibit a direct feedthrough characteristic, what is not desirable for simulation (Fernandes, 2005). This problem can be avoided by constructing an output-parameterized LEM system, provided that the adequate conditions hold (Wang and Rugh, 1987); in the SISO case, for example, this implies that there is no change of the sign of the stationary gain. In any case, the LEM-Wiener model is given by

$$\begin{aligned} \dot{x}_1 &= x_2 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= f_n(\mathbf{x}) + g_n(x_1)(u-\phi(x_1)) \\ y &= h(\mathbf{x}) \end{aligned} \tag{10}$$

where

$$\begin{aligned} f_n(\mathbf{x}) &= -\sum_{j=2}^{n} a_{j-1}(x_1)x_j \\ h(\mathbf{x}) &= x_1 + \sum_{j=2}^{n} b_{j-1}(x_1)x_j \end{aligned} \tag{11}$$

where the functions $a_i(x_1)$, $b_j(x_1)$, $j, i = 0,\ldots, n-1$, $j \neq 0$, correspond to the coefficients of the parameterized transfer function

$$G(s;\delta) = \frac{b_{n-1}(\delta)s^{n-1}+\ldots+1}{s^n+a_{n-1}(\delta)s^{n-1}+\ldots+a_0(\delta)}\cdot g_n(\delta) \tag{12}$$

where δ is a scalar parameterizing the set of equilibrium points/linearizations ($x_{1,s}$ in this case), and $a_0(x_1) = -g_n(x_1)\cdot d\phi(x_1)/dx_1$. The "advantage" of this form for identification is that all involved functions are scalar and can be therefore identified by means of the variation of one single parameter. Moreover, since the steady-states of this representation are of the form $x_{1,s} = y_s$, $x_{j,s} = 0$, $j = 2,\ldots, n$, these functions can be obtained by means of local linear models parameterized by the output.

Another advantage of the LEM-Wiener model structure is that it can be further extended by including a second-order term in the output equation, in order to improve the accuracy of the model away from the equilibrium manifold, that is,

$$h(\mathbf{x}) = x_1 + \sum_{j=2}^{n} b_{j-1}(x_1)x_j + \mathbf{\Phi}(\mathbf{x}) \qquad (13)$$

where $\mathbf{\Phi}(\mathbf{x})$ is such that $\mathbf{\Phi}(\mathbf{x}_s) = \mathbf{0}$ and $[\partial\mathbf{\Phi}(\mathbf{x})/\partial\mathbf{x}]_{\mathbf{x}_s} = \mathbf{0}_{1\times n}$. In particular, one possibility for $\mathbf{\Phi}(\mathbf{x})$ is

$$\mathbf{\Phi}(\mathbf{x}) = \left(\mathbf{x} - \mathbf{\Omega}(x_1)\right)^T \mathbf{H} \left(\mathbf{x} - \mathbf{\Omega}(x_1)\right) \qquad (14)$$

where the $n \times n$ matrix $\mathbf{H}$ has to be adjusted from experiments, and $\mathbf{\Omega}(x_1) = [\, x_1 \;\; 0 \;\; \ldots \;\; 0\,]^T$. The advantage is that $\mathbf{H}$ does not affect the dynamics of (10) and consequently avoids several problems. Moreover, since the output depends linearly on $\mathbf{H}$, it can be adjusted by means of computationally simple methods (least-squares, for example).

5. NUMERICAL EXAMPLE: HEAT EXCHANGE SYSTEM

The model structures presented in the previous sections will be tested in the modeling and simulation of the heat exchange system (Duraiski, 2001) depicted in Fig. 3.

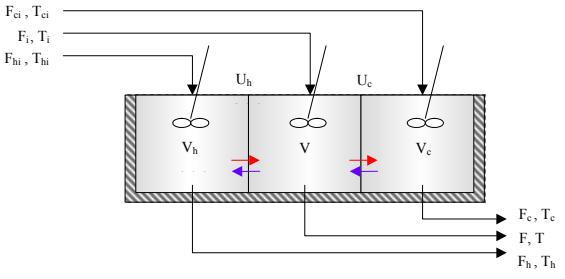


Fig. 3. Heat exchange system

This system is constituted by an insulated tank divided in three separate chambers that are allowed to transfer heat but not mass. The central chamber is in contact with both hot (h) and cold (c) chambers, but these are in contact just with the central one. The volumes of the chambers V_h, V and V_c, are constant, and all chambers are well-mixed. Water is fed to and removed from each chamber separately. Under these assumptions, the system can be described by means of the following differential equations:

$$\begin{aligned}
\frac{dT_h}{dt} &= \frac{F_{hi}\cdot(T_{hi}-T_h)}{V_h} - \frac{U_h\cdot A_h}{V_h\cdot Cp\cdot\rho}(T_h - T)\\
\frac{dT_c}{dt} &= \frac{F_{ci}\cdot(T_{ci}-T_c)}{V_c} - \frac{U_c\cdot A_c}{V_c\cdot Cp\cdot\rho}(T_c - T)\\
\frac{dT}{dt} &= \frac{F_i\cdot(T_i - T)}{V} + \frac{U_h\cdot A_h}{V\cdot Cp\cdot\rho}(T_h - T) + \frac{U_c\cdot A_c}{V\cdot Cp\cdot\rho}(T_c - T)
\end{aligned} \qquad (15)$$

where T_h, T_c and T are the temperatures of each chamber, C_p and ρ are the specific heat and specific mass of water (considered to be independent of the temperature), U_h (U_c) and A_h (A_c) are respectively the overall heat exchange coefficient and heat exchange area between the corresponding chambers. A more detailed description of this system can be found in (Duraiski, 2001). In this example, the input variable is considered to be the feed flowrate of hot water, $F_{h,i}$, which has constant temperature $T_{h,i}$. The output is the temperature of the central chamber, T. The values considered for the physical parameters and other inflows can be found in the Appendix. The variation of the dynamic character is obvious from the analysis of Fig. 4 and Fig. 5.

5.1 Constructing an approximated model in LEM, LEM-Hammerstein and LEM-Wiener forms

The original LEM, LEM-Hammerstein and LEM-Wiener models described in the previous sections can be constructed analytically on the basis of the model (15). In the first case, we have a system in the form of Eq. (4) with

$$\mathbf{A}(u) = \begin{bmatrix} -0.0033u - 0.239 & 0 & 0.239 \\ 0 & -0.0797 & 0.0797 \\ 0.239 & 0.0797 & -0.322 \end{bmatrix} \qquad (16)$$

$$\mathbf{\Omega}(u) = \frac{1}{1.272u + 1.254}\cdot\begin{bmatrix} 470.4u + 376.2 \\ 469.2u + 376.2 \\ 469.2u + 376.2 \end{bmatrix}$$

and $y = x_3$. For the LEM-Hammerstein model, it is first necessary to convert the matrices above to a normal form in order that the individual transfer functions from w to y in Fig. 1 have unit gain. The system is of the form:

$$\begin{aligned}
\tilde{\mathbf{A}}(u) &= \\
&10^{-2}\begin{bmatrix} 0 & 100 & 0 \\ -0.105u - 1.35 & -0.33u - 56.2 & -0.101u - 0.002 \\ 100 & 0 & -7.974 \end{bmatrix}\\
\tilde{\mathbf{\Omega}}(w) &= \begin{bmatrix} w \\ 0 \\ 12.54w \end{bmatrix}
\end{aligned} \qquad (17)$$

where u has to be substituted by $q^{-1}(w)$ for implementation, with

$$w = q(u) = \frac{469.2u + 376.2}{1.27u + 1.25}. \qquad (18)$$

The LEM-Wiener model (10)-(11) can be constructed similarly, giving

$$f_n(\mathbf{x}) = -0.507\frac{(-\ 0.00342\ \ x_1 + 1.27\ )}{(11.2\ -\ 0.0304\ x_1)}x_2 - 0.507\frac{(-0.0383x_1\ + 14.1)}{(11.2\ -\ 0.0304\ x_1)}x_3 \quad (19)$$

$$g_n(\ x_1) = 0.0238\ -\ 0.0000645\ x_1$$

$$\phi_1 = -71.8\frac{0.125\ -\ 0.000418\ x_1}{11.2\ -\ 0.0304\ x_1}$$

$$h(\mathbf{x}) = x_1 + 12.5x_2$$

The systems described above were simulated in Matlab with respect to the input function shown in Fig. 4; the responses are plotted in Fig. 5. The response of the linearized model at the operating point determined by $u_s = 1$ is also shown for comparison. Excepting this system, the other curves are practically indistinguishable.

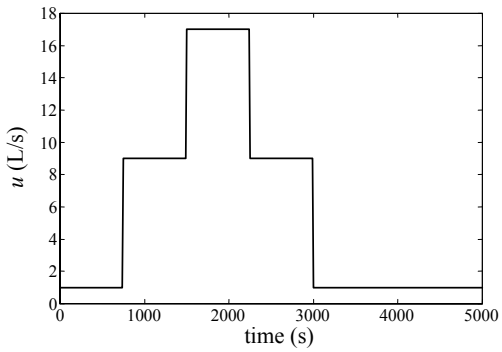


Fig. 4. Test input signal

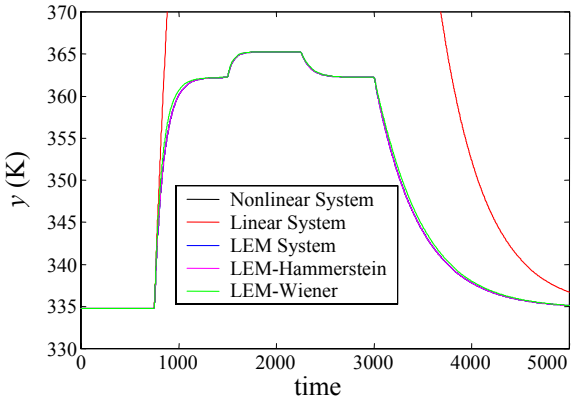


Fig. 5. Responses of the several systems to the signal in Fig. 3

5.2 Constructing the approximated models with identified local models

Approximated versions of the models derived in the previous section can be constructed with local models obtained either from linearizations or from identification experiments; only the last approach is exemplified here. The following procedure was adopted: three linear local models corresponding to the operating points defined by $u_{s,1} = 1$ L/s ($y_{s,1}$ = 334.76 K), $u_{s,2} = 8$ L/s ($y_{s,2}$ = 361.46 K), $u_{s,3} = 15$ L/s ($y_{s,3}$ = 364.78 K) were identified by means of "local experiments", that is, with identification signals of small amplitude around these operating points. No special methodology was employed to select the number or the location of these points; they were simply distributed over a desired range of the manipulated input. For each operating point, an identification signal u_{id} of the form depicted in Fig. 6 was designed. The switching period σ of the signal was determined as $t_{63}/20$, where t_{63} is the time needed from the step response to reach 63% of its steady-state value, what was obtained previously for each point by means of a step test with the nonlinear model (positive step of 0.2 L/s in u). The amplitude of the identification signal was fixed to 30% of $u_{s,i}$. An input sequence of the form $u_{s,i} - u_{id}$ was employed with validation purposes.

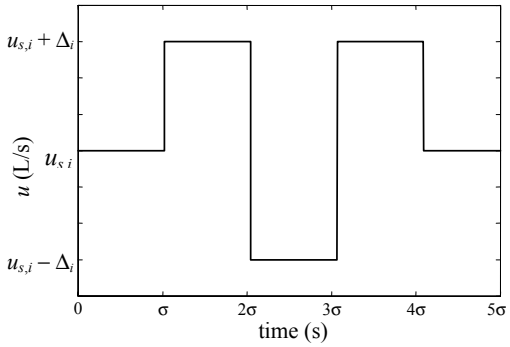


Fig. 6. Identification input signal

The response of the nonlinear model (15) was simulated in Matlab for the identification signal u_{id}. In order to simulate the effect of measurement error, a white-noise, Gaussian sequence with zero mean and standard deviation of 0.01 K was added to the output. A typical plot of the noisy output measurement is given in Fig. 7. The simulated signals were sampled with a convenient sample period in order to be used with the identification algorithms.

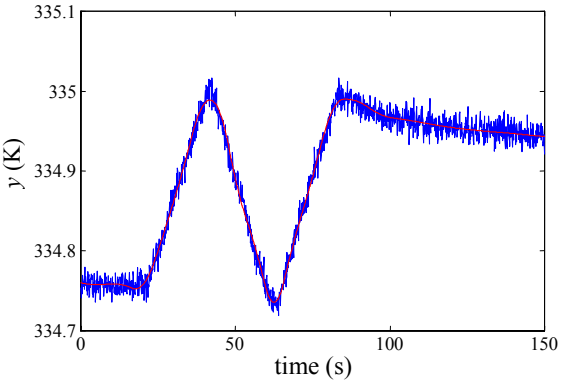


Fig. 7. Noisy and filtered output

Since an accurate representation of the local linearizations is necessary, the following procedure was adopted. First, a set of two runs was performed with u_{id} for each operating point and the average of the corresponding outputs y_{id} was taken as the

identification data; this has the objective of reducing the effect of noise. Second, the data was filtered by means of a least-squares smoothing cubic spline (Matlab function `spap2`). The best parameter set of the spline function was determined iteratively in function of the results of the identification procedure achieved in the subsequent step.

The linear local models in discrete form were identified through the combined use of subspace (Matlab functions `n4sid`/`subid`) and state-space prediction error methods (Matlab function `pem`). The subspace methods gave the initial estimates for the prediction error method and were also used for determining the order of the state-space models. As already suggested in the literature (Fernandes, 2005), the authors found that a good local identification is generally achieved when the order of the model is clearly evidenced by the singular value test provided by the subspace routines. Moreover, a frequent indication of excessive model order and poor identification by these methods is the generation of unstable poles, complex zeros, etc. The identification procedure was as follows: first, the filtered data was used in the subspace methods; the model order was selected and the estimates were passed to the `pem` routine. This result was then simulated and validated against the identification and validation data. If necessary, the parameters of the smoothing spline function were changed, the data was filtered again and the local models identified once more; this procedure was repeated until a good result was found.

The local models identified in this manner were used in the construction of the model structures presented in the sections 2, 3 and 4. The LEM and LEM-Hammerstein models were constructed with local models in observability form. The last one differs from the analytical case because the linear transformation to normal form depends on the relative degree which is not a "robust" quantity to be obtained from experiments. In all cases, proper spline or rational interpolation of the necessary functions was performed (the results are omitted due to the space limitations). The responses of the three structures with identified local models for the input signal in Fig. 4 are shown in Fig 8; the agreement with the nonlinear model is quite good. The most significant difference with respect to the analytical case refers to the Wiener model, due to the identification/interpolation of the b_i parameters that appear in the output function.

6. CONCLUSIONS

This paper presented new model structures based on the concept of linearization around the equilibrium manifold (LEM). These models extend the conventional Hammerstein and Wiener systems, in the sense that they allow for the inclusion of variable dynamics. These representations can be constructed on the basis of local models, which can be obtained for example by identification. A numerical example (bilinear system) showed that these structures are almost equivalent if the models are obtained analytically, but the effect of the errors in the estimated parameter can affect differently the distinct model classes.

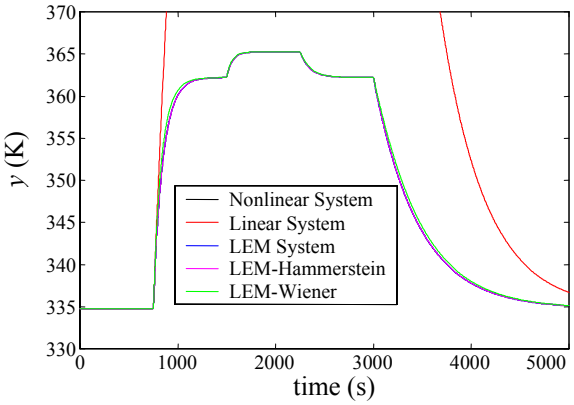


Fig. 8. Responses of the several systems constructed on the basis of identified local models

APPENDIX

Parameter values used in the example:
$\rho = 1000$ kg/m^3, $C_p = 4180$ J/kg/K, $V = V_c = V_h = 0.3$ m^3
$U_h A_h = 300.000$ J/K/s, $U_c A_c = 100.000$ J/K/s
$F_{ci} = 0$ m^3/s, $T_{ci} = 280$ K, $F_i = 0.001$ m^3/s, $T_i = 300$ K
$T_{hi} = 370$ K

AKNOWLEDGEMENTS

The authors thank FINEP and PETROBRAS for the financial support. The first and second authors thank respectively the German Academic Exchange Service (DAAD) and the Landesstiftung Baden-Württemberg.

REFERENCES

Duraiski, R. G (2001). *Controle Preditivo Não-Linear Utilizando Linearização ao Longo da Trajetória*. Masters' Thesis, Federal University of Rio Grande do Sul, Brazil (in Portuguese).

Bolognese Fernandes, P., S. Engell, J. O. Trierweiler (2004). A new approach to the local models networks technique. Proceedings of the Brazilian Congress of Engenharia Química, COBEQ04, Curitiba, Brazil.

Bolognese Fernandes, P. (2005). *The input-parameterized linearization around the equilibrium manifold approach to modeling and identification*. Phd Thesis, University of Dortmund (to be published).

Bolognese Fernandes, P., S. Engell (2005). Continuous Nonlinear SISO System Identification using Parameterized Linearization Families. *Proc. of the XVI IFAC World Congress*, *Prague*, *Tchech Republic*.

Pearson, R. K (1995). Gray-box identification of block-oriented nonlinear models. *Journal of Process Control,* **10**, 301-315.

Pearson, R. K. and M. Pottmann (2000). Gray-box identification of block-oriented nonlinear models. *Journal of Process Control,* **10**, 301-315.

Wang, J. and J. W. Rugh (1987). Parameterized Linear Systems and Linearization Families for Nonlinear Systems. *IEEE Transactions on Circuits and Systems,* **34**, 650-657.

MULTIVARIABLE FUZZY IDENTIFICATION APPROACH APPLIED TO COMPLEX LIQUID RESIDUES INCINERATION PROCESS

Felipe M. Almeida, Gilmar Barreto and Ginalber L. O. Serra

Control and Intelligent Systems Laboratory, State University of Campinas, 400, Cid. Univ. Zeferino Vaz. 13083-970, Campinas -SP, Brazil
Email: felipe@dmcsi.fee.unicamp.br, gbarreto@dmcsi.fee.unicamp.br and ginalber@dmcsi.fee.unicamp.br

Abstract: This paper proposes an identification scheme for a complex liquid effluent incinerator process. This scheme is developed to obtain a MIMO (*Multiple Input Single Output*) TS (*Takagi-Sugeno*) fuzzy model where the modified Gath Geva clustering algorithm is used to determine the antecedent part as the consequent parameters are estimated by RLS (*Recursive Least Square*) algorithm.

Keywords: Multivariable system identification, Liquid effluent incinerator, Fuzzy systems, Takagi-Sugeno fuzzy model, Recursive least square.

1. INTRODUCTION

Techniques of systems identification are widely used in control systems design and successful applications have appeared at last two decades. In a typical adaptive control design, a valid model of the dynamic system, in one of some operating conditions, is identified on-line, and the controller design is carried according to this model so that some performance specifications are satisfied (Serra, and Bottura, 2006a). In systems identification literature (Ljung, 1999; Soderstrom, and Stroica, 1989), the most approaches are concerned to linear modelling and control using continuous or discrete time equations as well as state space ones. Moreover, motivated by the fact of all dynamic system present a nonlinear behaviour, several approaches have been proposed for analysis, identification and control, where fuzzy systems are key elements in these application (Khalil, 2002; Isidori, 1995; Wang, 1996; Pedrycz, and Gomide, 1998; Serra, and Bottura, 2006b).

Fuzzy systems is an effective tool for uncertain nonlinear systems identification based on measured data (Hellendoorn and Driankov, 1997). Among different fuzzy modelling techniques, the Takagi-Sugeno fuzzy model has attracted the most attention (Takagi and Sugeno, 1985). This model consists of "if-then" rules with fuzzy antecedents and mathematical functions in the consequent part. The antecedents fuzzy sets partition the input space into a number of fuzzy regions, while the consequent functions describe the system's behavior in each region. The identification procedure of TS fuzzy models is usually done in two steps. Firstly, the antecedents parameters (membership functions parameters) are determined using knowledge of the process behavior or data-driven techniques. In the second step, the parameters of the consequent functions are estimated. As these consequent functions are linear in their parameters, the least-squares method can be applied.

A real world example of complex nonlinear dynamic system is the liquid residues incineration process (Cunha, 2003). It is part of the power unit at BASF industry, placed in Resende-Brazil. The necessity to study its dynamic behavior, which motivates a MIMO (*Multiple Input Single Output*) fuzzy identification scheme application, is due to the following reasons (Almeida and Barreto, 2004):

- Avoiding emission of gas from combustion out of ambient agency standards;
- Improving the reside burning efficiency to reduce the fuel consumption in the incinerator;
- Minimizing costs.

This paper proposes an identification scheme for a complex liquid effluent incinerator process. This scheme is developed to obtain a MIMO TS fuzzy model via modified Gath Geva clustering algorithm used to determine the antecedent parameters and RLS (*Recursive Lest Square*) method used to estimate the consequent parameters. Experimental results show the efficiency of the proposed scheme as well as the accuracy of the obtained models, so important characteristics in intelligent adaptive control design.

2. LIQUIDS EFFLUENTS INCINERATOR

The effluent liquids incinerator, whose study of its characteristics can be seen in (Cunha, 2003; Almeida and Barreto, 2005), receives residues from industrial plants. Basically, this incineration system consists in an unit which was developed by T-Thermal, Sub-X Down Fired type system, to incinerate liquid residues through oxidation in high temperature, as shown in Fig. 1. This unit is composed by: combustion chamber (1), oxidation chamber (2), cooling tank (3), initial separation tower (4), particle breaker (5), final separation tower (6) and gas washer (7).

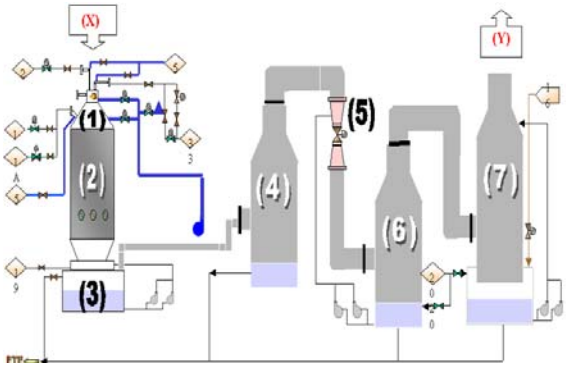


Fig. 1. Incinerator process

The capacity of the combustion chamber is of 6 million Kcal/h. The air/combustive relation is adjusted according to the stoichiometric computing. In these conditions, it is desired to obtain an efficient effluent toxics destruction at least of 99,99%. The combustion products are unloaded in the cooling tank. The gases leave the cooling tank for the duct of gases exit, passing to the initial separator, whose function is to minimize water transport, in the liquid state, presents in the gas. In the initial separator the gas follows to particle breaker. The recycled water through this washer is collected in the final separator; there are a constant draining of this water to prevent the extreme concentration of dissolved impurities. The gas leaves the final separator and follows to the gas washer. The gas washer is a tower with plastic filling where the gas flows to top, being washed and neutralized for a water solution with sodium hydroxide that is launched under sprayed form in the top of the tower, the gases leave for the chimney, located above of the gas washer gases. In this paper, we are concerned to identify the nonlinear relation between combustion chamber inputs (watery effluent, organic effluent, combustible oil, combustion air) and gas washer outputs (O_2, SO_2, CO) using a MIMO TS fuzzy model.

3. TAKAGI –SUGENO FUZZY MODEL

In the TS fuzzy model, proposed by Takagi-Sugeno in 1985, the antecedent is defined by linguistic terms of the input variables (linguistic variables), the consequent is a functional expression of these variables and the *i*-th IF-THEN rule has the following form:

$$R_i: \; IF \;\; x_1 \;\; \text{is} \;\; A_1^i \;\; AND \ldots AND \ldots \text{E} \;\; x_q \;\; \text{is} \;\; A_q^i$$
$$THEN \;\; y_i = f_i(\mathrm{x}), \;\; i = 1,2,..., \mathrm{c}. \qquad (1)$$

where c is the number of rules.

The vector $\mathbf{x} \in \Re^q$ contains the premise variables, which has its own universe of discurse that is partitioned into fuzzy regions by the fuzzy sets describing the linguistic variable $x_j \,|^{\,j=1,\ldots,q}$. The premise variable x_j belongs to a fuzzy set A_q with a truth value given by a membership function $\mu_{jk} : R \rightarrow [0,1]$ for $k = 1, 2, \ldots, s_j$ where s_j is the number of partitions of the linguistic variable x_j for premise variable j. The truth value h_i for the complete rule i is computed using the aggregation operator , or t-norm, denoted by $\wedge : [0,1]^2 \rightarrow [0,1]$:

$$h_i(\mathbf{x}) = \mu_1^i(x_1) \wedge \mu_2^i(x_2) \wedge \ldots \mu_q^i(x_q) \qquad (2)$$

Among the different t-norms available, in this work the algebrbraic product will be used:

$$h_i(\mathbf{x}) = \prod_{j=1}^{q} \mu_j^i(x_j) \qquad (3)$$

The activation degree for the rule *i* is normalized as:

$$\gamma_i(\mathrm{x}) = \frac{h_i(\mathrm{x})}{\sum_{r=1}^{c} h_r(\mathrm{x})} \qquad (4)$$

where *c* is the number of rules. This normalization implies that:

$$\sum_{i=1}^{c} \gamma_i(\mathrm{x}) = 1 \qquad (5)$$

The response of the TS model is a weighted sum of the consequent functions, i.e, a convex combination of the local functions (models) f_i:

$$y = \sum_{i=1}^{c} \gamma_i(\mathrm{x}) \; f_i(\mathrm{x}) \qquad (6)$$

1.1 Fuzzy Structure Model

In this paper is, the NARX (*Nonlinear Autoregressive with Exogenous Input*) structure, widely applied in fuzzy modeling, where the model output is a function of the past input-output data, is used:

$$\hat{y}(k+1) = f\,(y(k) \ldots y(k-n_y+1), u(k) \ldots \ldots u(\mathrm{k}-n_u+1)) \qquad (7)$$

where *k* denotes the time sampling, n_y and n_u are integers related to the system order, *u* e *y* are the input and output, respectively. The TS fuzzy model, in terms of IF-THEN rules, is given by:

$$R_i: IF \;\; y(k) \text{ is } A_1^i \;\; AND \ldots AND \;\; y(k-n_y+1) \text{ is } A_{n_y}^i$$
$$AND \; u(k) \text{ is } B_1^i \;\; AND \ldots AND \;\; u(k-n_u+1) \text{ is } B_{n_u}^i \;\; THEN$$

$$\hat{y}^i(k+1) = \sum_{j=1}^{n_y} a_{i,j}\, y(k-j+1) + \sum_{j=1}^{n_u} b_{i,j}\, u(k-j+1) + c_i \qquad (8)$$

where $a_{i,j}$, $b_{i,j}$ e c_i are consequent parameters to be estimated by the RLS (Recursive Least Square) method (Almeida and Barreto, 2005). The inference formula of the TS model is:

$$\hat{y}(k+1) = \sum_{i=1}^{l} \gamma_i(\mathrm{x}_k)\; y^i(k+1) \qquad (9)$$

$$\mathbf{x}_k = (y(k) \ldots y(k-n_y+1), u(k) \ldots u(\mathrm{k}-n_u+1)) \qquad (10)$$

4. RLS – RECURSIVE LEAST SQUARE

The basic idea of recursive least squares algorithm is to compute the new parameter estimate $\hat{\theta}(k+1)$ at the time k+1 by adding a correction vector to the previous parameter estimate $\hat{\theta}(k)$ at the time *k*. The estimation of the recursive weighted least squares algorithm for MISO (*Multiple Input Single Output*) systems, based on the global approach (all linear consequent parameters are estimated simultaneously), is given by:

$$\hat{\theta}(k) = \mathbf{P}(k)\, \boldsymbol{X}^{\mathrm{T}}(k)\, \boldsymbol{W}(k)\, \mathbf{y}(k) \qquad (11)$$

where $\boldsymbol{X}(k)$ is the regression matrix at the time k:

$$\boldsymbol{X}(k) = \begin{bmatrix} \mathrm{x}^{\mathrm{T}}(1) \\ \mathrm{x}^{\mathrm{T}}(2) \\ \vdots \\ \mathrm{x}^{\mathrm{T}}(k) \end{bmatrix}_{N \times 1+r} \tag{12}$$

and

$$\mathbf{P}(k) = (\boldsymbol{X}^{\mathrm{T}}(k)\, \boldsymbol{W}(k)\, \boldsymbol{X}^{\mathrm{T}}(k))^{-1} \tag{13}$$

The matrix $\boldsymbol{W}(k)$ is the weighting matrix:

$$\boldsymbol{W}(k) = \begin{bmatrix} w(1) & 0 & \dots & 0 \\ 0 & w(2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w(k) \end{bmatrix} \tag{14}$$

Similarly, the estimation of recursive weighted least squares algorithm for MISO systems, with local approach (the consequent parameters are for each rule *i*), is given by:

$$\hat{\theta}_i(k) = \mathbf{P}_i(k)\, \boldsymbol{X}^{\mathrm{T}}(k)\, \boldsymbol{W}_i(k)\, \mathbf{y}(k) \tag{15}$$

where the matrix $\boldsymbol{W}_i(k)$ is the weighting matrix:

$$\boldsymbol{W}_i(k) = \begin{bmatrix} w_i(\mathrm{x}(1)) & 0 & \dots & 0 \\ 0 & w_i(\mathrm{x}(2)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_i(\mathrm{x}(k)) \end{bmatrix} \tag{16}$$

and

$$\mathbf{P}_i(k) = (\boldsymbol{X}^{\mathrm{T}}(k)\, \boldsymbol{W}_i(k)\, \boldsymbol{X}^{\mathrm{T}}(k))^{-1} \tag{17}$$

The estimator equation for the time *k+1* is:

$$\hat{\theta}_i(k+1) = \mathbf{P}_i(k+1)\, \boldsymbol{X}^{\mathrm{T}}(k+1)\, \boldsymbol{W}_i(k+1)\, \mathbf{y}(k+1) \tag{18}$$

which can be rewritten as:

$$\hat{\theta}_i(k+1) = \mathbf{P}_i(k+1) \begin{pmatrix} X(k) \\ \mathrm{x}^T(k+1) \end{pmatrix}^T \begin{bmatrix} W_i(k) & 0 \\ 0 & w_i(\mathrm{x}(k+1)) \end{bmatrix} \begin{pmatrix} \mathrm{y}(k) \\ \mathrm{y}(k+1) \end{pmatrix} \tag{19}$$

$$= \mathbf{P}_i(k+1)[\boldsymbol{X}^{\mathrm{T}}(k)\, \boldsymbol{W}_i(k)\, \mathbf{y}(k) + \mathrm{x}(k+1)\, w_i(\mathrm{x}(k+1))\, \mathbf{y}(k+1)]$$

Substituting $\boldsymbol{X}^{\mathrm{T}}(k)\, \boldsymbol{W}_i(k)\, \mathbf{y}(k) = \mathbf{P}_i^{-1}(k)\, \hat{\theta}_i(k)$ in (19), adding and subtracting $\hat{\theta}_i(k)$ on the right side, results:

$$\hat{\theta}_i(k+1) = \hat{\theta}_i(k) + [\mathbf{P}_i(k+1)\, \mathbf{P}_i^{-1}(k) - I]\, \hat{\theta}_i(k) + \mathbf{P}_i(k+1)\, \mathrm{x}(k+1)\, w_i(\mathrm{x}(k+1))\, \mathbf{y}(k+1) \tag{20}$$

where according to (17):

$$\mathbf{P}_i(k+1) = (\mathbf{P}_i(k)^{-1} + \mathrm{x}(k+1)\, w_i(\mathrm{x}(k+1))\, \mathrm{x}(k+1)^{\mathrm{T}})^{-1} \tag{21}$$

Taking the inverse on both sides in (21), we obtain:

$$\mathbf{P}_i(k)^{-1} = \mathbf{P}_i(k+1)^{-1} - \mathrm{x}(k+1)\, w_i(\mathrm{x}(k+1))\, \mathrm{x}(k+1)^{\mathrm{T}} \tag{22}$$

Substituting (22) in (20), the recursive estimator equation is obtained by:

$$\hat{\theta}_i(k+1) = \hat{\theta}_i(k) + \mathbf{P}_i(k+1)\, \mathrm{x}(k+1)\, w_i(\mathrm{x}(k+1)) (\mathbf{y}(k+1) - \mathrm{x}(k+1)^{\mathrm{T}}\, \hat{\theta}_i(k)) \tag{23}$$

The RLS algorithm requires the inversion of the matrix **P**. Utilizing the matrix-inversion theorem, this procedure provides a lower computational cost and the equation (22) can be rewritten as:

$$\mathbf{P}_i(k+1) = \mathbf{P}_i(k) - \mathbf{P}_i(k)\, \mathrm{x}(k+1) (1/w_i(\mathrm{x}(k+1)) + \mathrm{x}(k+1)^{\mathrm{T}}\, \mathbf{P}_i(k)\, \mathrm{x}(k+1))^{-1} \mathrm{x}(k+1)^{\mathrm{T}}\, \mathbf{P}_i(k) \tag{24}$$

After some simplifications, results:

$$\mathbf{P}_i(k+1) = \mathbf{P}_i(k) - \frac{w_i(\mathrm{x}(k+1))\, \mathrm{P}_i(k)\, \mathrm{x}(k+1)\, \mathrm{x}(k+1)^T\, \mathrm{P}_i(k)}{1 + w_i(\mathrm{x}(k+1))\, \mathrm{x}(k+1)^T\, \mathrm{P}_i(k)\, \mathrm{x}(k+1)} \tag{25}$$

The recursive weighted least squares algorithm used for consequent parameters estimation is given by (20) and (25), where $w_i(\mathrm{x}(k+1))$ it is the activation degree for each rule.

In order to get the activation degree for each rule, to determine the antecedent parameters of the fuzzy model, is required.

5. MODIFIED GATH-GEVA ALGORITHM

The previous section has shown how the consequent parameters of the TS fuzzy model can be estimated by the recursive least squares algorithm when the antecedent parameters are given. In this section, in order to form an easily interpretable fuzzy model, the modified Gath-Geva clustering algorithm, which is based on the Expectation Maximization (EM) identification of Gaussian mixture models (Abonyi, *et al.*,2002; Abonyi and Szeifert, 2001) is presented. In this paper, this technique is extended for MIMO fuzzy models identification, that will be applied in the incineration system, where each cluster contains an input distribution, a local model and an output distribution:

$$p(\phi, y) = \sum_{i=1}^{c} p(\phi, y, n_i) = \sum_{i=1}^{c} p(y \mid \phi, n_i)\, p(x \mid n_i)\, p(n_i) \tag{26}$$

with $p(n_i)$ the *a priori* probability of the cluster, $p(x \mid n_i)$ the input distribution and $p(y \mid \phi, n_i)$ the output distribution. The clustering is based on the minimization of the sum of weighted squared distances between the data points $\mathbf{x}_k$ and the cluster centers $\mathbf{v}_i$

$$J_m(\mathbf{X}, \mathbf{U}, \mathbf{V}) = \sum_{k=1}^{n} \sum_{i=1}^{c} {w_{ki}}^m {d_{ki}}^2, 1 < m < \infty \tag{27}$$

where $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_c]$ contains the cluster centers and *m* is a weighting expoent that determines the fuzziness of the resulting clusters and it is often chosen as $m = 2$. The fuzzy partition matrix has to satisfy the following conditions:

$$U \in [0,1]^{nc} \mid w_{k,i} \in [0,1], \forall k, i;$$

$$\sum_{i=1}^{c} w_{ki} = 1, \forall k; 0 < \sum_{k=1}^{n} w_{ki} < n, \forall i \tag{28}$$

The minimization of (27) represents a non-linear optimization problem subject to constrains defined by (28) and can be solved by using a variety of available methods. The modified Gath-Geva algorithm is formulated as follows:

Initialization: Given a set of data matrix **X**, specify c, choose the weighting exponent m>1 and the tolerance ε > 0. Initialize the partition matrix such that (28) holds.
Repeat for l = 1,2,...

Step 1: Comput the parameters of the clusters:

- Center of membership functions

$$v_i^l = \frac{\sum_{k=1}^{n} [w_{ki}^{l-1}]^m x_k}{\sum_{k=1}^{n} [w_{ki}^{l-1}]^m}, i = 1,... c. \quad (29)$$

- Standard deviations of the Gaussian membership functions

$$\sigma_{j,i}^2 = \frac{\sum_{k=1}^{n} [w_{ki}^{l-1}]^m (x_{k,j} - v_{k,j})^2}{\sum_{k=1}^{n} [w_{ki}^{l-1}]^m}, i = 1,...c. \quad (30)$$

- Parameters of local models given by (20)

- A priori probabilities of the clusters

$$p(n_i) = \frac{1}{n} \sum_{k=1}^{n} [w_{ki}^{l-1}]^m \quad (31)$$

- Covariance matrix of the modeling error

$$F_i^y = \frac{\sum_{k=1}^{n} [w_{ki}^{l-1}]^m (y_k - \hat{y}_k)(y_k - \hat{y}_k)^T}{\sum_{k=1}^{n} [w_{ki}^{l-1}]^m} \quad (32)$$

Step 2: Compute the distance measure $d_{k,i}^2$:

The distance measure consists in two terms. The first one is the distance between the cluster centers and **x**, while the second one quantifies the performance of the local linear models.

$$\frac{1}{d_{ki}^2} = p(n_i) \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{j,i}^2}} \exp\left(\frac{1}{2} \frac{(x_{j,k} - v_{i,j})^2}{\sigma_{i,j}^2} \right) \cdot$$

$$\frac{\exp(-(y_k - \hat{y}_k)^T (F_i^y)^{-1} (y_k - \hat{y}_k))}{(2\pi)^{\frac{no}{2}} \sqrt{|F_i^y|}} \quad (33)$$

Step 3: Update the partition matrix

$$w_{ki}^l = \frac{1}{\sum_{j=1}^{c} (d_{ki} / d_{kj})^{2/(m-1)}}, \quad (34)$$

until $\| U^l - U^{l-1} \| < \varepsilon$.

6. IDENTIFICATION OF THE INCINETARION PROCESS

The identification of the liquid effluent incineration process will be made by the Takagi-Sugeno fuzzy model using the RLS algorithm for consequent parameters estimation and the modified Gath-Geva algorithm to obtain the antecedent parameters, partitioning the multivariable input space in valid fuzzy regions for the consequent sub-models, both presented in sections 4 and 5, respectively.

6.1 Process Characteristics

In order to get a better structure of TS fuzzy model for the process, we verify some pertinent characteristics of the incineration system (Cunha, 2003), such as:

- ***MIMO System***: 4 inputs and 3 outputs:

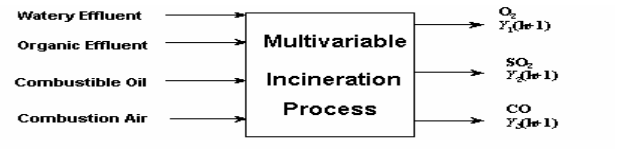


Fig. 2. Incineration system

- ***Correlation in the input variables***: The four input variables are all correlated, the burning reason according to stoichiometric parameters is of 4kg of watery effluent for 1 kg of fuel (combustible oil + organic effluent) and 11,32 m^3/h of combustion air. All the inputs variables influence in the outputs variables.

- ***Correlation in the output variables***: There are certain particularitities, such as:

- The first output variable, O_2 concentration, haven't correlation with the others two output variables, its value is given directly for the gas analyzer;

- The second output variable, SO_2 concentration, is obtained by computing in Feema-RJ (State Foundation of Environment Engineering) resolution for analysis SO_2 for dry base in 11% O_2:

$$SO_2 \text{ corrected} = \frac{SO_2 \text{ analysed.}(O_2 \text{ atmosphere} - 11\ \%)}{(O_2 \text{ atmosphere} - O_2 \text{ analyzed})} \quad (35)$$

where we can observe that second output is correlated with first output; therefore, the computing of SO_2 concentration depends to the O_2 value;

- The thirth output variable, CO concentration, is obtained by the calculation in Feema-RJ resolution for analysis CO for dry base in 11% O_2:

$$CO \text{ corrected} = \frac{CO \text{ analysed.}(O_2 \text{ atmosphere} - 11\ \%)}{(O_2 \text{ atmosphere} - O_2 \text{ analyzed})} \quad (36)$$

where we can observe that thirth output is correlated with first output; therefore the computing of the CO concentration depends to the O_2 value;

6.2 Structure of TS Fuzzy Model

Due to the characteristics in item 6.1, we could define the TS fuzzy model in the form of MIMO structure as 3 connected MISO fuzzy models, where we verify the correlation among the data of the system. We search this form to optimize the identification process. The multivariable fuzzy model is shown in Fig.3.

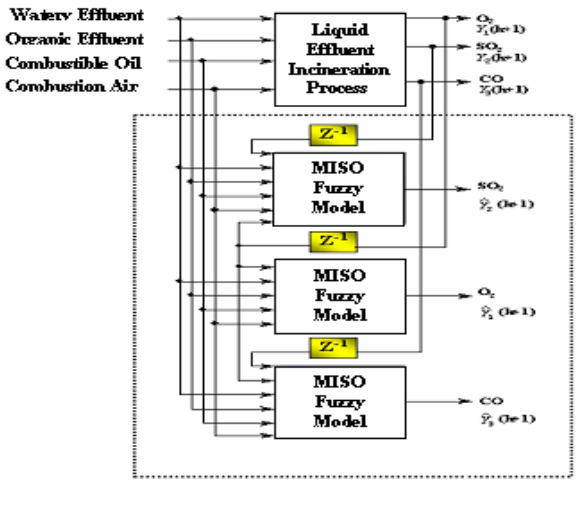


Fig. 3. MIMO fuzzy model

Once structure is known the fuzzy sets (the regions of operation of the local model) are defined in the domain of the outputs of the incinerator. For each output the following configuration is:

Output (O_2): R_i: IF $y_1(k)$ IS A_1^i THEN $\hat{y}_1^{\,i}(k+1)=$

$$\alpha_1^i\, y_1(k) + \beta_1^i\, u_1(k) + \beta_2^i\, u_2(k) + \beta_3^i\, u_3(k) + \beta_4^i\, u_4(k) + \gamma^i \quad (37)$$

Output (SO_2): R_i: IF $y_2(k)$ IS A_2^i THEN $\hat{y}_2^{\,i}(k+1)=$

$$\alpha_1^i\, y_2(k) + \beta_1^i\, u_1(k) + \beta_2^i\, u_2(k) + \beta_3^i\, u_3(k) + \beta_4^i\, u_4(k) + \beta_5^i\, u_5(k) + \gamma^i \quad (38)$$

Output (CO): R_i: IF $y_3(k)$ IS A_3^i THEN $\hat{y}_3^{\,i}(k+1)=$

$$\alpha_1^i\, y_3(k) + \beta_1^i\, u_1(k) + \beta_2^i\, u_2(k) + \beta_3^i\, u_3(k) + \beta_4^i\, u_4(k) + \beta_5^i\, u_5(k) + \gamma^i \quad (39)$$

where i = 1,…,c, is the number of rules and A_1^i, A_2^i, e A_3^i are sets fuzzy of antecedent variables for each TS model. The consequent parameters for each rule α_j^i, β_j^i e γ^i are estimated by RLS algorithm.

7. EXPERIMENTATION AND RESULTS

The system identification using TS MIMO fuzzy model was realized. For the modeling stage of the parameters, 600 samples (100 hours of incineration process operation) are collected from experiment. For the validation stage, others 600 samples were used. Two criteria had been used for the validation of the fuzzy models:

-VAF : (*Variance Accounted For*)

$$\textbf{VAF}\ (\%) = 100 \times \left[1 - \frac{\mathrm{var}(Y - \hat{Y})}{\mathrm{var}(Y)}\right] \quad (40)$$

where Y is the nominal output of the incineration process, $\hat{Y}$ is the estimate output of the model and *var* is the variance of the signal.

-MSE(*Mean Square Error*)

$$\textbf{MSE} = \frac{1}{N}\sum_{K=1}^{N}(Y_k - \hat{Y}_k)^2 \quad (41)$$

where Y_k is the nominal output of the incineration process, $\hat{Y}_k$ is the estimate output of the model and N is the number of points.

Five different models were identified: (1) MIMO ARX model, (2) MIMO TS fuzzy model with FCM (*Fuzzy C-Means*) clustering algorithm, (3) MIMO TS fuzzy model with GK (Gustafson and Kessel, 1979) clustering algorithm, (4) MIMO TS fuzzy model with GG (Gath and Geva, 1989) clustering algorithm, (5) MIMO TS fuzzy model with modified GG clustering algorithm. A comparative analysis is established between these models. The Table 1,2, and 3, presents the efficiency of the models that had been used in the liquid incineration process identification system for each output variable:

Table 1 Efficiency of the models – Output (O_2)

Model	VAF(%)	MSE	Rules N.
MIMO ARX	90,52	0,751	-
MIMO TS (FCM)	94,95	0,411	4
MIMO TS (GK)	96,21	0,312	4
MIMO TS (GG)	95,24	0,375	4
MIMO TS (Mod. GG)	98,89	0,236	4

Table 2 Efficiency of the models – Output (SO_2)

Model	VAF(%)	MSE	Rules N.
MIMO ARX	91,34	3,286	-
MIMO TS (FCM)	95,45	2,617	4
MIMO TS (GK)	97,81	2,409	4
MIMO TS (GG)	95,33	2,620	4
MIMO TS (Mod. GG)	98,12	2,377	4

Table 3 Efficiency of the models – Output (CO)

Model	VAF(%)	MSE	Rules N.
MIMO ARX	90,87	1,394	-
MIMO TS (FCM)	93,59	0,977	4
MIMO TS (GK)	95,18	0,764	4
MIMO TS (GG)	94,11	0,883	4
MIMO TS (Mod. GG)	98,88	0,530	4

In Table 1,2 and 3, we can observe that MIMO TS fuzzy model with modified GG clustering algorithm, had a better performance than others ones. A comparative analysis between the real outputs and the estimate output for this model, is showed in Figures 4, 5 and 6:

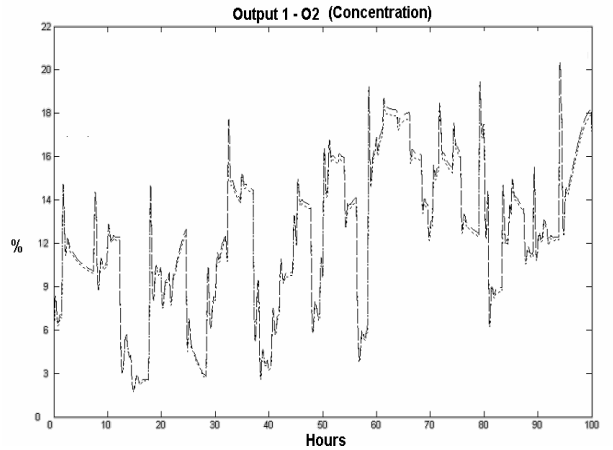


Fig. 6. Measured(-) and predicted(- -)process outputs

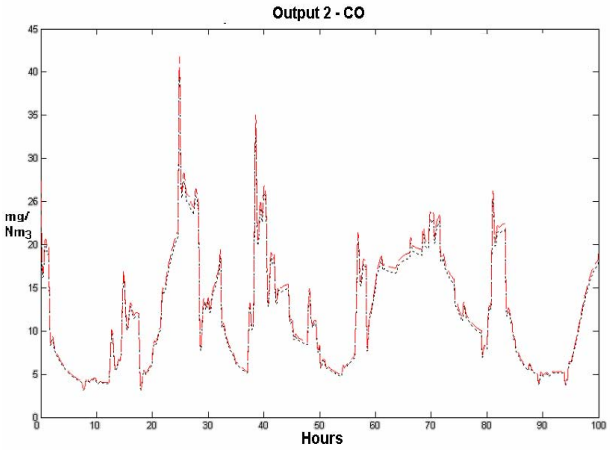


Fig. 7. Measured(-) and predicted(- -)process outputs

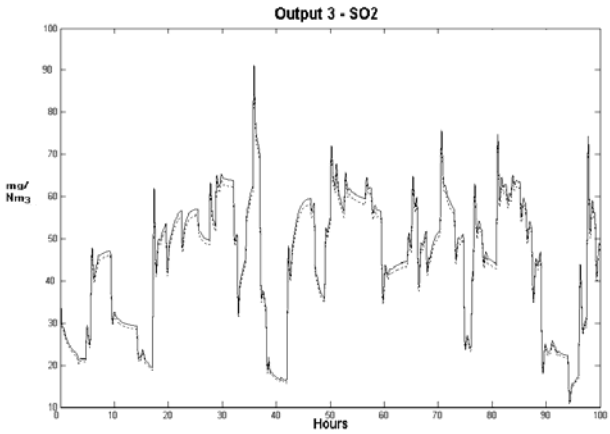


Fig. 8. Measured(-) and predicted(- -)process outputs

8. CONCLUSIONS

In this paper, the identification of complex nonlinear multivariable sistem is discussed. A fuzzy model structure has been proposed, where the liquids effluents incineration process at the BASF industry, is represented by a MIMO TS fuzzy model. The modified Gath-Geva clustering algorithm was used to determine the antecedent part of the MIMO fuzzy model and the consequent parameters were estimated by RLS algorithm. The obtained MIMO fuzzy model was able to represent the dynamic behaviour of the MIMO nonlinear dynamic system due to, mainly, the chosen structure based on the correlation analysis of input-output data. For future works, the development of a adaptive controller for the combustion system using obtained model of the incineration process makes necessary.

REFERENCES

Abonyi, J., R. Babuska and F.Szeifert (2002). "Modified Gath-Geva Fuzzy Clustering for Identification of Takagi-Sugeno Fuzzy Models". *IEEE Transactions on Systems, Man and Cybernetics*, Part B: 32:612-621.

Abonyi, J. and F. Szeifert (2001). "*Identification of MIMO Process by Fuzzy Clustering*". IEEE International Conference on Intelligent Systems, Finland.

Almeida, F.M. and G. Barreto (2004). "*Liquid Effluent Incineration: Contribution for its Modeling and Identification*". IEEE Sixth International Conference on Industrial Applications, October 12-15, Joinville-Brazil.

Cunha, J.R. (2003). "*Operation Handbook of the Effluents Incinerator*". Rev.06, UTL-001, BASF.

Gath, G. and A.B. Geva (1989). "Unsupervised Optimal Fuzzy Clustering". *IEEE Transactions on Pattern analysis and Machine Intelligence*,7:773-781.

Gomide, F. and W. Pedrycz (1998). *An introduction to fuzzy sets, analysis and design*. MIT Press.

Gustafson, D.E. and W.C. Kessel (1979). "*Fuzzy Clustering with Fuzzy Covariance Matrix*". In Proceedings of the IEEE CDC, San Diego, pages: 761-766.

Hellendoorn, H. and D. Driankov (1997). *Fuzzy Model Identification: Selected Approaches*, Springer.

Isidori, A. (1995). *Nonlinear Control Systems*. Springer Verlag.

Khalil, H. (2002). *Nonlinear Systems*. 3 [a] ed., Prentice Hall.

Ljung, L. (1999). *System Identification: Theory for the User*. 2 [a] ed., Prentice Hall.

Soderstrom, T. and P. Stoica (1989). *System Identification*. Prentice Hall.

Serra, G.L.O. and C.P. Bottura (2006a). "Multiobjetive Evolution Based Fuzzy PI Controller Design For Nonlinear Systems". *International Jounal Engineering Applications of Artificial Intelligence, 19: 157-167*.

Serra, G.L.O. and C.P. Bottura (2006b). "An IV-QR Algoritm for Neuro-Fuzzy Multivariable Identification". *IEEE Transactions on Fuzzy Systems*, In Press.

Takagi, T. and M. Sugeno (1985). "Fuzzy Identification of Systems and its Application to Modeling and Control". *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):116-132.

Wang, L. (1996). *A Course in Fuzzy Systems and Control*. Prentice Hall.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

IDENTIFICATION OF POLINOMIAL NARMAX MODELS FOR AN OIL WELL OPERATING BY CONTINUOUS GAS-LIFT

Pagano, D. J. * **Dallagnol Filho, V.** * **and Plucenio, A.** *

* *Departamento de Automação e Sistemas, Universidade Federal de Santa Catarina, 88040-900 Florianópolis-SC, Brazil*
e-mail: {valdemar, daniel, plucenio}@das.ufsc.br

Abstract: Two nonlinear models (polynomial NARMAX) are identified for a simulated oil well operating by continuous gas-lift. The chosen input/output pair (injected gas mass flow rate/pressure drop in the production tubing) used in the identification can be applied in a control strategy decoupling injection from production choke control. The model derived with data obtained by exciting the plant around three different operating points compares well with another using a more aggressive excitation.

Keywords: Nonlinear Identification, NARMAX models, oil well production, continuous gas-lift

1. INTRODUCTION

In order to control a well operating by continuous gas-lift, a mathematical model of the well is usually needed. However, physical modelling of the input and output relations is complex, encompassing partial differential equations, which are hard to manipulate. An alternative is to use *identification* techniques, which try to find mathematical relations between the input and the output series of a system, without prior knowledge of its internal behavior.

The ultimate goal is to control the wellhead flow-rate. In an effort to avoid using expensive multiphase flowmeters, this is obtained indirectly by controlling other variables like the pressure in front of the perforations. The idea is to control the pressure in the wellhead and the pressure drop in the production tubing in such a way as to have a desired pressure in front of the perforated zone. The control of the pressure in the wellhead is done with a local controller and is part of the setup used in the identification.

The system under analysis has a clearly nonlinear behavior, making any linear model valid only inside a narrow operating region. The specific type of nonlinear model chosen is the polynomial NARMAX (*Nonlinear AutoRegressive Moving Average model with eXogenous inputs*). An arsenal of simple and robust algorithms is available to estimate the parameters of this kind of models.

This paper is organized as follows: first of all, the polynomial NARMAX model is presented; then the system under analysis is described. Following the identification procedure is described and finally conclusions are drawn.

2. NARMAX MODELS

A NARMAX model is represented like follows (Leontaritis and Billings, 1985):

$$\begin{aligned} y(k) = F[&y(k-1),\ldots,y(k-n_y),\\ &u(k-1),\ldots,u(k-n_u),\\ &\nu(k),\nu(k-1),\ldots,\nu(k-n_\nu)], \end{aligned} \quad (1)$$

where F is a nonlinear function, $u(k)$ is the input signal, $y(k)$ is the output signal, $\nu(k)$ is the noise in the system, n_y, n_u and n_ν are the largest delays in y, u e ν, respectively. However the determination of the function F is a hard task.

A polynomial NARMAX model is an expansion of the function F in a polynomial function with degree of

nonlinearity ℓ. It is considered that the system does not have pure time delay and that none of the parameters to be estimated depends on $\nu(k)$. The polynomial approximation with degree of nonlinearity ℓ is given by (Chen and Billings, 1989):

$$\begin{aligned} y(k) = \theta_0 &+ \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(k) \\ &+ \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \theta_{i_1 i_2} x_{i_1}(k) \cdot x_{i_2}(k) + \dots \\ &+ \sum_{i_1=1}^{n} \cdots \sum_{i_l=i_{l-1}}^{n} \theta_{i_1 \dots i_l} x_{i_1}(k) \dots x_{i_l}(k) + \nu(k) \end{aligned} \tag{2}$$

where:

$$\begin{aligned} x_1 &= y(k-1) & x_{n_y+1} &= u(k-1) \\ &\vdots & &\vdots \\ x_{n_y} &= y(k-n_y) & x_n &= u(k-n_u) \end{aligned} \tag{3}$$

being $n = n_y + n_u$ and θ constant parameters.

The use of a polynomial NARMAX representation may be justified by the following reasons: it is a global representation, allowing the global dynamics of the system to be represented, and not only the dynamics around a certain equilibrium point; it is easy to quantify the complexity of the model, based on the degree of non-linearity, number of terms and maximum delay used; it may deal with moderated levels of noise; analytical information about the model is easy to acquire; it is possible to have NARMAX models with a good fit to the data, as long there are not abrupt variations in the signals (Leontaritis and Billings, 1985); simple and robust algorithms may be used to estimated the parameters (since the model is linear in the parameters).

3. SYSTEM DESCRIPTION

The continuous gas-lift works by reducing the gravity term of the production tubing pressure drop. This is accomplished by injecting gas inside the production tubing through a gas-lift valve. Gas, being much lighter than the liquid in the production tubing, moves up, gasifying the flowing fluid, reducing its average density and, consequently, the pressure in front of the perforated zone.

In most wells, several gas-lift valves are distributed along the production tubing in such a way as to permit gas to enter progressively from top to bottom valve when injecting gas in the annular tubing-casing. The deepest valve is the only one which remains in operation while the other valves are only used for the start-up of the well. This work proposes a different set-up in an effort to avoid the utilization of mechanical gas-lift valves. In this approach an orifice valve is installed downhole, substituting the classical gas-lift valves and

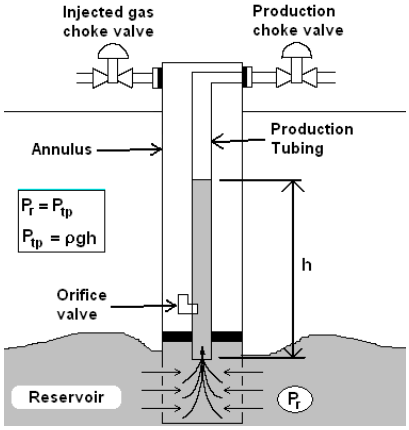


Fig. 1. Oil well

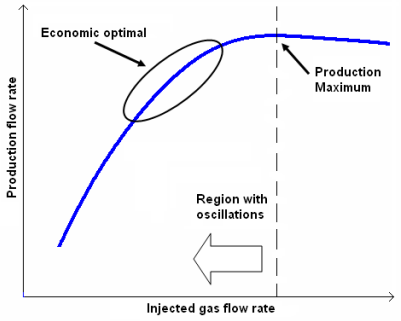


Fig. 2. Production flow x Injected gas flow curve, with the area of largest economic interest signaled

the control is done in the surface acting on the gas-lift and production chokes. Figure 1 shows the main components of the gas-lift oil well set-up considered in this work. The start-up procedure for this set-up is not studied but it could possibly be done with a high pressure compressor.

There is an optimal operating region for the well, economically speaking, shown in Figure 2, which is related to the fluid fraction flow-rates produced by the well, its current market prices, and the costs of gas-compression and so on. This region, however, has the inconvenient of presenting oscillations when the system operates in open-loop, reducing the well productivity and affects the oil, water, gas separation efficiency.

Several works have appeared in the literature (Eikrem *et al.*, 2004), proposing different strategies to stabilize the oscillations in wells operating via gas-lift using similar set-up acting in the production choke.

In (Plucenio, 2002) a control strategy is proposed using the mass flow-rate measured on the surface, and acting in the gas injection mass flow-rate. Linear ARX models are identified in three different points of operation in order to develop a robust control.

In this paper, the well is treated as a SISO system, with the mass flow rate of injected gas (Q_i) as the input and the pressure in the production tubing (P_{tp}) as the output (see Fig. (3)). The input of the system

Q_i is actually the setpoint of a controller actuating in the injection valve opening. This controller has the standard PI structure, with $K_p = 5 \times 10^{-3}$ and $K_i = 0.1$. The pressure in the production tubing may be decomposed as $P_{tp} = P_{wf} - P_h$, where P_{wf} is the pressure measured in the bottom of the well and P_h is the pressure measured in the head of the well. The main advantage of considering P_{tp} as the output of the system is that P_{tp} is relatively isolated of disturbances in the pressure on the boundaries of the system (in the separator). P_{wf} and P_h will react similarly to these disturbances and compensate for these disturbances when P_{tp} is calculated. The pressure in the head of

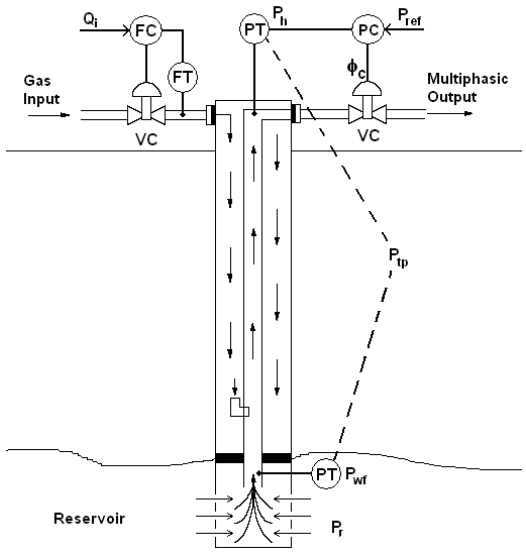


Fig. 3. Measuring and actuation points in the oil well

the well is also controlled by a local controller that, by acting on the opening of the production choke, guarantees that P_h remains constant (which is desirable). The setpoint for the P_h controller is 2.24 MPa, being the structure a standard PI with $K_p = -1 \times 10^{-5}$ and $K_i = 0.01$.

This definition of input and output variables have the advantage of allowing easy implementation, since instrumentation for measuring the pressure in the head and bottom of the well is common in modern wells (Veneruso *et al.*, 2000). Besides that, measuring pressure is trivial, on the contrary to the instrumentation needed for measuring the flow rate of a multiphasic fluid, which is very expensive.

The system possesses an obvious nonlinear behavior, which can be observed in Figure 4, showing the output corresponding to the application of a sequence of steps in the input of the system. It may be observed that not only the transitory response changes depending of the region of operation, but also the steady state response, and the signal of the static gain, which changes from negative to positive when the injected gas flow rate increases beyond a certain point. The desired operating region lies in a region with negative gain.

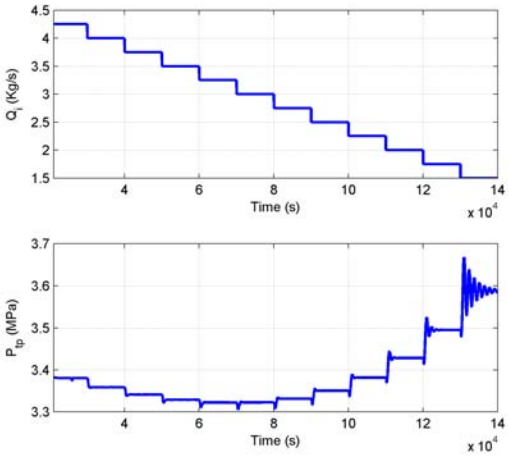


Fig. 4. Top: sequence of steps applied in the input of the system (Q_i). Bottom: Corresponding output (P_{tp}) showing the nonlinearity of the system

Besides the nonlinear characteristic, the system possesses a non-minimum phase response (see Fig.(4)), which makes harder the synthesis of a controller. [1]

It must be noted that the model identified to quantify the relation $P_{tp} \times Q_i$ is influenced by the choice of the parameters of the local controllers (for gas injection and for the pressure in the head of the well). Any change in the structure or the parameters of this controllers demand a new identification of the system.

The data used for identification was generated with the software OLGA® 2000, by Scandpower Co., version 4.10.1. The system used in the simulator is a modification of a model supplied by Scandpower, representing a real well operating in deep waters in the Mexican Gulf. The well has the following characteristics:

- Reservoir static pressure = 33.094 MPa
- Reservoir temperature = 82.2^oC
- Reservoir productivity index = 2×10^{-6} kg/s/Pa
- Pressure in the separator = 2.585 MPa
- Temperature in the separator = 26.7^oC
- Gas pressure at compressor output= 9.652 MPa
- Gas temperature at compressor output = 20^oC

4. NONLINEAR IDENTIFICATION

First of all, the original model in (2) was changed, including in the candidate terms those containing $u(k)$. The presence of a term containing $u(k)$ indicates that there may exist a direct transfer of information from the input to the output of the system, in other words, a part of the dynamic may be fast enough to reflect

[1] The term "non-minimum phase" is generally used in the context of linear systems meaning the presence of a zero outside the unit circle (in the discrete case). This notion was extrapolated here, where the term "non-minimum phase response" was used to state that the response of the system presents a behavior similar to the one that could be found in a linear system with non-minimum phase

"immediately" in the output. The model is therefore given by:

$$y(k) = \theta_0 + \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(k) + \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \theta_{i_1 i_2} x_{i_1}(k) \cdot x_{i_2}(k) + \ldots + \sum_{i_1=1}^{n} \cdots \sum_{i_l=i_{l-1}}^{n} \theta_{i_1 \ldots i_l} x_{i_1}(k) \ldots x_{i_l}(k) + e(k) \quad (4)$$

where:

$$\begin{aligned} x_1 &= y(k-1) & x_{n_y+1} &= u(k) \\ x_2 &= y(k-2) & x_{n_y+2} &= u(k-1) \\ &\vdots & &\vdots \\ x_{n_y} &= y(k-n_y) & x_n &= u(k-n_u) \end{aligned} \quad (5)$$

being $n = n_y + n_u + 1$, n_y the maximum delay in y and n_u the maximum delay in u.

As input signals for the system, two strategies where used: the first one used an "aggressive" signal, with more abrupt variations, which tries to excite a large range of frequencies and reach different operating regions of the system. The second signal is more "well behaved", using small variations around three operating points, reducing the risk of damage to the plant.

4.1 *Aggressive signal*

The "aggressive" signal was obtained by keeping the input signal constant at $Q_i = 2.15$ kg/s, until the initialization transitory of the system was over. After it, it was added to the constant signal a random signal with zero mean and unitary variance, being each step kept for 200 seconds. The use of this random signal tries to excite a broad range of frequencies. Before adding the random signal to the constant, it is multiplied by a crescent value, such that the system starts operating around the operating point and move away from it as time passes. Figure 5 shows the input signal applied and Figure 6 shows the corresponding output. The test duration was 15000 seconds, with a sampling rate of 40 seconds.

Another signal with the same characteristics but with another realization of random numbers was used as input of the system to produce data to validate the identified models. The desired model has a degree of nonlinearity $\ell = 2$, $n_y = n_u = 5$, resulting in 78 candidate terms. Besides these terms, 10 linear noise moving average terms were added to avoid biasing of the estimates.

Among the candidate terms, there are 6 term clusters ($\Omega_0, \Omega_y, \Omega_{y^2}, \Omega_u, \Omega_{u^2}$ and Ω_{yu}). The term cluster Σ_{y^2} was eliminated from the candidate terms set, because

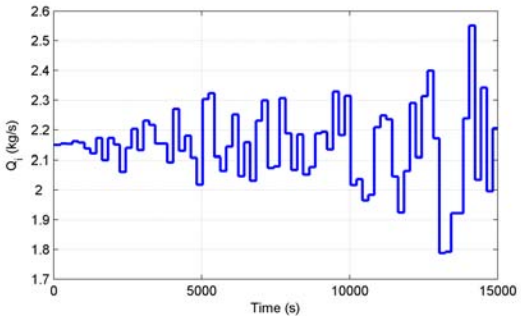


Fig. 5. Aggressive input signal of the system used to estimate the parameters of the nonlinear model

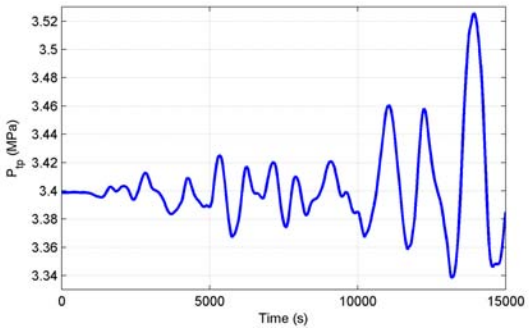


Fig. 6. Output signal resulting from the aggressive input

the desired polynomial NARMAX model should have only one fixed point. The location of the fixed points of the model (which has degree $\ell = 2$) is the solution of the equation:

$$(\Sigma_{y^2})\bar{y}^2 + (\Sigma_y + \Sigma_{yu}\bar{u} - 1)\bar{y} + (\Sigma_0 + \Sigma_u \bar{u} + \Sigma_{u^2}\bar{u}^2) = 0. \quad (6)$$

Therefore, by eliminating the term cluster Ω_{y^2}, there will exist only one fixed point located in:

$$\bar{y} = -\frac{(\Sigma_0 + \Sigma_u \bar{u} + \Sigma_{u^2}\bar{u}^2)}{(\Sigma_y + \Sigma_{yu}\bar{u} - 1)} \quad (7)$$

The Error Reduction Ratio (ERR) criterium (Chen and Billings, 1989) was used to sort the sequence that the terms should be included in the model, but the actual number of terms in the final model was determined by using the Akaike Information Criterium (AIC) (Akaike, 1974). The parameters of the estimated model were then checked for statistical significance, by comparison with the standard deviation of the estimate. A 99% level of significance was used, meaning that each parameter should satisfy $-3\sigma_i \leq \hat{\theta}_i \leq 3\sigma_i$, where σ_i is the standard deviation of the estimate of the parameter i and $\hat{\theta}_i$ is the estimate of the parameter i. An iterative process was then performed, were the terms that were not significant (but were still included in the model by the ERR criterium) were excluded from the set of candidate terms and a new model was identified and checked for significance.

The final model identified has 6 deterministic terms, shown in table 1, plus 10 linear moving average terms

Table 1. NARMAX model terms (aggressive input) ordered by the ERR value.

Order	Term	$\hat{\theta}_i$	σ
1	$y(k-1)$	$+2.01356$	$+2.23668\times10^{-2}$
2	$y(k-2)$	-1.01002	$+2.24378\times10^{-2}$
3	$u(k-4)y(k-2)$	-3.49843×10^{-2}	$+2.93181\times10^{-3}$
4	$u(k-4)y(k-5)$	$+2.37093\times10^{-2}$	$+3.52717\times10^{-3}$
5	$u(k-4)$	$+3.09134\times10^{-2}$	$+2.84186\times10^{-3}$
6	$u^2(k)$	$+8.50101\times10^{-4}$	$+7.68678\times10^{-5}$

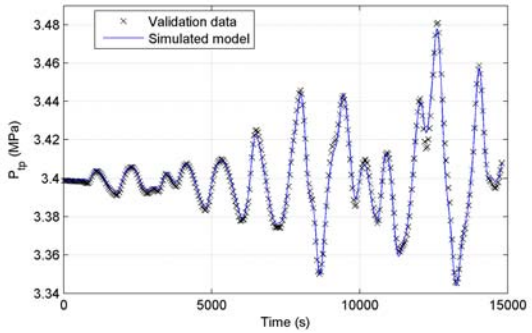


Fig. 7. Free simulation of the NARMAX model identified with the data from an aggressive input, compared with validation data

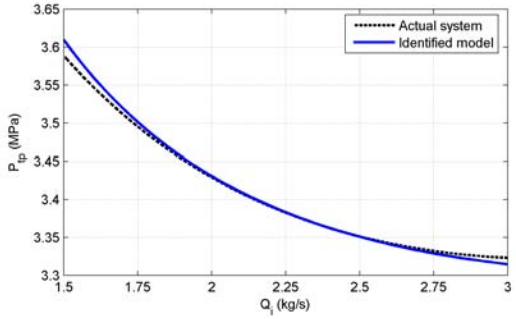


Fig. 8. Fixed points of the NARMAX model identified for the aggressive input

of the noise signal, which had the sole purpose of avoiding biasing of the estimates, being discarded afterwards. To quantify the quality of a model, it was used the fit index defined by (Ljung, 2004):

$$\text{fit} = 100 * \left(1 - \frac{\|\hat{\boldsymbol{y}} - \boldsymbol{y}\|}{\|\boldsymbol{y} - \text{mean}(\boldsymbol{y})\|}\right), \qquad (8)$$

where $\hat{\boldsymbol{y}}$ is the vector with the output of the model and $\boldsymbol{y}$ is the vector with the real output of the system. The equation (8) compares the quality of prediction of a model with the mean of the data as a trivial predictor.

By using the validation data to evaluate this model, the output of the model had a fit $= 87.78\%$, as seen in Figure 7. The steady-state characteristic of the model, compared to the steady-state characteristic of the system may be seen in Figure 8. It may be seen that the model represents well the system under analysis in the defined operating region (from $Q_i = 1.5$ kg/s to $Q_i = 3$ kg/s).

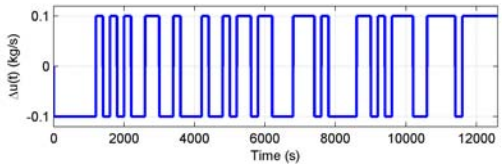


Fig. 9. PRBS Signal $\Delta u(t)$ applied in the system

4.2 *Well behaved signal*

The model identified in the previous section was able to reproduce adequately the dynamics of the system under analysis. However, the input used to generate the identification data is too "aggressive", presenting big changes and may be risky to use in the real plant. In order to avoid this risks, a new NARMAX model was identified, using data acquired from the use of a "well-behaved" input (with smaller changes in the signal).

The system was carefully brought to three operating points and when in steady-state, a PRBS signal (*Pseudo Random Binary Signal*) was applied in the input. Figure 9 depicts the PRBS signal used ($\Delta u(t)$). The actual signal applied in the input is $u(t) = \Delta u(t) + u_0$, where u_0 is the operating point. The three chosen operating points where $u_0 = 1.5$, $u_0 = 2.15$ and $u_0 = 2.8$ kg/s.

Obviously the data used for estimating the parameters of the model is not ideal in a theoretical point of view, because the input is restricted to three small operating regions, not passing through all the desired operating region (from 1.5 kg/s to 3 kg/s). However, the use of this data set has two advantages over the "aggressive" signal used in the previous section:

- it is less risky to the plant, for having less abrupt variations;
- production can still be carried on during the execution of the tests, because there is only a slight disturbance over the steady-state inputs.

The candidate models searched have the same characteristics of the ones searched in the previous section, being the degree of nonlinearity $\ell = 2$, $n_y = n_u = 5$, and 10 linear moving average terms used to avoid biasing of the estimates. From the set of candidate terms the term cluster Ω_{y^2} was eliminated too.

After repeating an iterative procedure which includes: sorting the remaining candidate terms with the use of the ERR criterium, defining the number of terms to be included in the final model with the Akaike Information Criterium, verifying the statistical significance of the estimated parameters and validating statistically the model by residual analysis, the model shown in table 2 was found. The model has a fit $= 91.2\%$ to the validation data, which is an excellent performance. Figure 11 shows the steady-state characteristic of the model identified compared with the actual steady-state

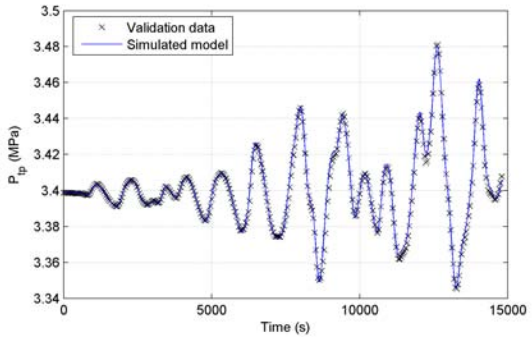


Fig. 10. Free simulation of the NARMAX model identified with the data from an well-behaved input, compared with validation data

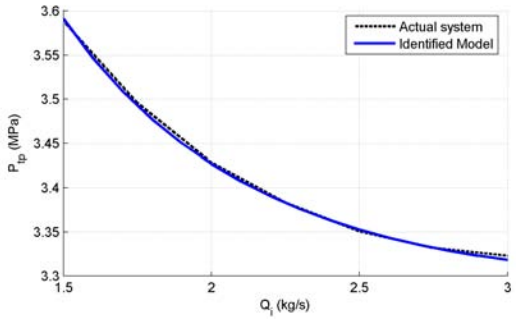


Fig. 11. Fixed points of the NARMAX model identified for the well-behaved input

characteristic of the system. It is seen in the figure that the model is a good representation of the oil well.

Table 2. NARMAX model Terms (well-behaved input) ordered by the ERR value.

Order	Term	$\hat{\theta}_i$	σ
1	$y(k-1)$	$+2.84619$	$+3.94354\times10^{-2}$
2	$y(k-2)$	-2.23221	$+5.70014\times10^{-2}$
3	$y(k-4)$	$+3.88183\times10^{-1}$	$+1.83896\times10^{-2}$
4	$u(k-2)y(k-1)$	-7.26036×10^{-3}	$+5.33763\times10^{-4}$
5	$u(k)$	$+5.29649\times10^{-3}$	$+1.78812\times10^{-4}$
6	$u(k-5)$	$+2.78799\times10^{-3}$	$+6.50890\times10^{-4}$
7	$u(k-3)y(k-1)$	-2.67362×10^{-1}	$+2.23946\times10^{-2}$
8	$u(k-2)$	$+1.25873\times10^{-2}$	$+2.18152\times10^{-3}$
9	$u(k-5)u(k-3)$	-1.22951×10^{-3}	$+2.48621\times10^{-4}$
10	$u(k-3)y(k-2)$	$+3.64800\times10^{-1}$	$+3.17758\times10^{-2}$
11	$u(k-3)y(k-4)$	-9.75874×10^{-2}	$+9.86630\times10^{-3}$
12	$u(k-2)u(k-2)$	$+1.73758\times10^{-3}$	$+2.33335\times10^{-4}$

5. CONCLUSIONS

In this paper, two models of an oil well operating by continuous gas-lift were identified, relating the pressure in the production tubing (output) with the mass flow rate of injected gas (input). The presented strategy has the advantage of allowing an easy implementation on existing oil wells, where the needed instrumentation is widely available (Veneruso *et al.*, 2000).

The two polynomial NARMAX models identified showed to represent adequately the system, which would be impossible to do with linear models. The absence of a stronger nonlinearity, in the considered range of gas-lift injection flow rate, made it possible to use a well behaved input signal, which is not ideal in a nonlinear identification viewpoint, but is preferred for presenting less risk to the plant during the test procedure.

The model identified with the well-behaved signal showed better performance when near the boundaries of the operating region, because two of the three operating points chosen to apply the PRBS signal are at the boundaries. The aggressive signal, in the other hand, concentrates the input in the middle of the operating region and so the model identified with has a slightly worse performance near the boundaries of the operating region.

As a next step in research, the models identified will be used to design a controller to the simulated plant in the OLGA simulator, as a previous step to the implementation of this control strategy in a real oil well.

6. ACKNOWLEDGMENTS

V. A. Dallagnol Filho was funded by CNPq. D. J. Pagano and A. Plucenio were funded by Agencia Nacional do Petroleo (ANP), Brazil, under project aciPG-PRH No 34 ANP/MCT. The authors would also like to acknowledge Scandpower for providing an academical OLGA software license.

7. REFERENCES

Akaike, Hirotugu (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Chen, S. and S. A. Billings (1989). Representation of non-linear systems: the narmax model. *International Journal of Control* **49**(3), 1013–1032.

Eikrem, G. O., B. Foss and L. Imsland (2004). Stabilization of gas lifted wells based on state estimation. *IFAC ADCHEM2004, Hong Kong*.

Leontaritis, I. J. and S. A. Billings (1985). Input-output parametric models for non-linear systems. *International Journal of Control* **41**(2), 303–344.

Ljung, Lennart (2004). *System Identification Toolbox For Use With Matlab 6.0*. The MathWorks, Inc.. Natick, MA.

Plucenio, A. (2002). Stabilization and optimization of an oil well network operating with continuous gas-lift. *SPE ATCE, San Antonio, Texas*.

Veneruso, A. F., S. Hiron, R. Bhavsar and L. Bernard (2000). Realiability qualification testing for permanently installed wellbore equipment. *SPE 62955, ATCE, Dallas, Texas*.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

COMPARSION BETWEEN PHENOMENOLOGICAL AND EMPIRICAL MODELS FOR POLYMERIZATION PROCESSES CONTROL

Tiago F. Finkler [1], Gustavo A. Neumann[2], Nilo S. M. Cardozo[3], Argimiro R. Secchi[4]

1,3,4 – Grupo de Modelagem, Simulação, Controle e Otimização de Processos (GIMSCOP)
Departamento de Engenharia Química – Universidade Federal do Rio Grande do Sul
Rua Sarmento Leite, 288/24 – CEP: 90050-170 – Porto Alegre –RS – Brazil
Phone: +55-51-3316-3528 – Fax: +55-51-3316-3277
2- Braskem S/A, III Pólo Petroquímico, Via Oeste, Lote 5 – CEP: 95853-000 – Triunfo – RS – Brazil
Phone: +55-51-457-5495 – Fax: +55-51-457-5450
E-mail: {[1]tiago, [3]nilo, [4]arge}@enq.ufrgs.br, [2]gustavo.neumann@braskem.com.br

Abstract: In this work, linear, quadratic, and nonlinear empirical models were built and compared with a dynamic nonlinear phenomenological model with respect to the capability of predicting the melt index and polymer yield rate of a low density polyethylene production process. Based on steady-state gains and on known first and second order time constants of the process, the empirical models were generated using PLS, QPLS, and BTPLS methods in order to predict the system dynamics. As the quadratic model provided more reliable predictions, it was used as melt index virtual analyzer of an advanced control strategy for an industrial plant, improving the controller action and the polymer quality by reducing significantly the process variability.

Keywords: partial least squares, empirical modelling, parameter estimation, phenomenological modelling, LLDPE.

1. INTRODUCTION

Correctly validated multivariate models are useful tools for the development of reliable predictive controllers in polymerization processes.

Depending on their nature, empirical or phenomenological, these models may provide different levels of information about the process. When those kinds of models are compared, phenomenological models are supposed to show higher extrapolation capability. However, empirical models require much less investments in modelling, especially when little is known about the physical and chemical phenomena underlying the process.

Concerning specifically to the modelling of polyolefin polymerization processes, many studies have been developed in the last decades. Sato et al (2000) studied the modelling and simulation of an industrial gas-phase ethylene polymerization process, based on the phenomenological model of McAuley (1991), for using in nonlinear controller design for melt index and density. Many published papers deal with modelling and parameter estimation for nonlinear model predictive controller design in industrial applications, like Zhao (2001) and Soroush (1998). Bindlish et al. (2003) studied the parameter estimation problem for industrial polymerization processes. In their work, two kinetic parameters were estimated for Exxon's homo and copolymerization to use in monitoring and feedback control systems of these processes.

In this work phenomenological and empirical models for the prediction of yield and melt index of an industrial process for the production of linear low-

density polyethylene (LLDPE) are compared. The studied process is composed by two gas-phase reactors connected in series. For both reactors, the considered operational variables (model inputs) are the ethylene (C2), butene (C4) and hydrogen (H2) concentrations, catalyst flow rate (Cat), the bed temperature (T), total pressure (P), and fluidized bed level (L). The response variables (model outputs) are the polymer melt index (MI) and polymer yield rate (YR) at the outlet of each reactor. The studied process is schematized in Figure 1:

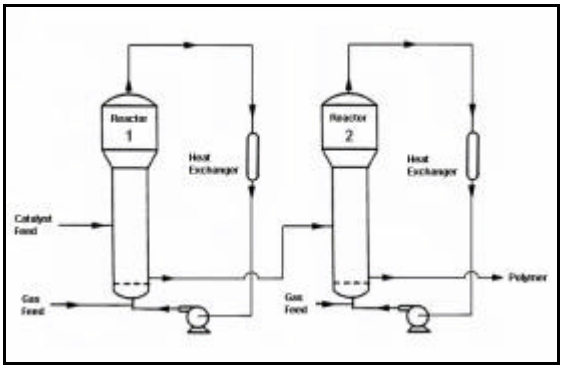


Figure 1: Scheme of the polymerization process.

Two ten-days dataset containing measurements of all considered variables were collected. The sampling rate varies from one variable to another. For the variables measured on line (T, P, C2, C4, H2 and Cat) it is in the order of minutes while for MI it is in the order of hours. In this text, the first dataset will be treated as dataset A and the second dataset will be treated as dataset B. These data sets are presented in Figure 2 and in Figure 3. The vertical axis of the plots correspond to the coded variables measurements and the horizontal axis correspond to the time window where these variables measurements ware made.

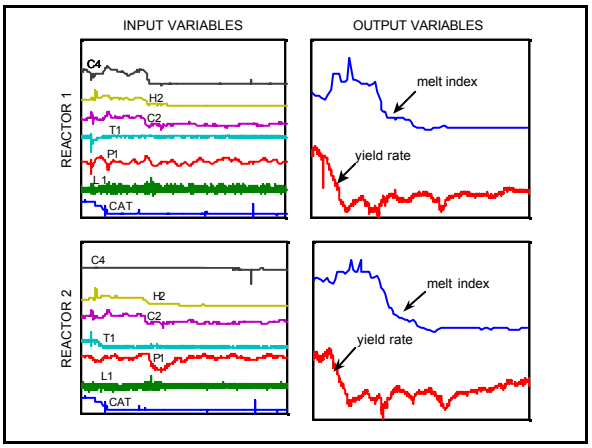


Figure 2: Process first dynamics dataset (A).

2. PHENOMENOLOGICAL MODEL

Industrial fluidized-bed reactors have been modelled by several authors, see for example Kunii and Levenspiel (1991) and Choi and Ray (1985). According to the model proposed by Gambetta et al. (2001), the fluidized bed can be divided in two regions: an emulsion phase and a bubble phase, connected by heat and mass transfer between them. The emulsion phase has a solid phase (polymer and catalyst), a gas phase at the minimal fluidization condition, and a gas phase adsorbed by the solid phase. The bubble phase is composed by the excess of gas required to keep the emulsion phase at the minimal fluidization condition. In the disengagement section, it is only considered the gas phase.

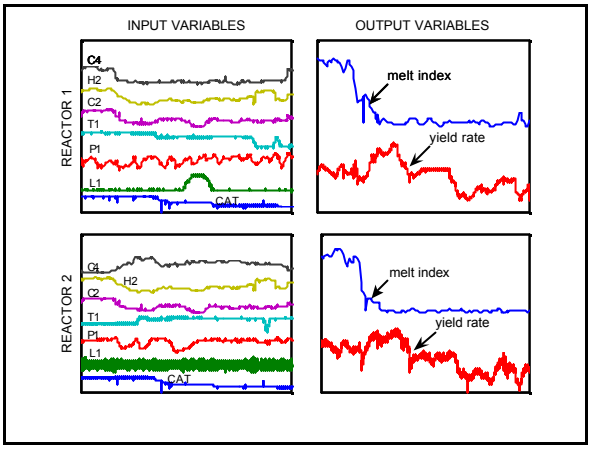


Figure 3: Process second dynamics dataset (B).

The kinetic model was developed for Ziegler-Natta catalysts, considering the following reactions: spontaneous activation of sites, chain initiation by monomer, chain propagation, chain transfer to hydrogen, and spontaneous and by hydrogen deactivations. The equations of these reactions are presented in Table 1, where C_p denotes a potential site, $P_d{}^k$ a site of type k, $P_{d,i}{}^k$ a initiated chain with monomer type i and site type k, $P_{n,i}{}^k$ a live polymer chain with n monomers with end group i and active site k, M_i a monomer molecule of type i, C_d a dead site and $D_n{}^k$ a dead polymer chain with n monomers of site k.

Table 1: Reactions considered in the kinetic model.

Spontaneous site activation	$C_P \xrightarrow{k^k_{aSp}} P^k_{\bar{0}}$
Chain initiation by monomer i	$P^k_{\bar{0}} + M_i \xrightarrow{k^k_{P0i}} P^k_{\bar{d}_i,i}$
Chain propagation by monomer j	$P^k_{\bar{n},i} + M_j \xrightarrow{k^k_{Pij}} P^k_{\bar{n}+\bar{d}_j,j}$
Chain transfer to hydrogen	$P^k_{\bar{n},i} + H_2 \xrightarrow{k^k_{cHi}} P^k_{\bar{0}} + D^k_{\bar{n}}$
Deactivation by hydrogen	$P^k_{\bar{n},i} + H_2 \xrightarrow{k^k_{dH}} C_d + D^k_{\bar{n}}$ $P^k_{\bar{0}} + H_2 \xrightarrow{k^k_{dH}} C_d$
Spontaneous chain deactivation	$P^k_{\bar{n},i} \xrightarrow{k^k_{dSp}} C_d + D^k_{\bar{n}}$ $P^k_{\bar{0}} \xrightarrow{k^k_{dSp}} C_d$

Mass balances for the main gases (ethylene, co-monomers, solvent, and hydrogen) and polymeric species were used to obtain the gas phase and the polymer compositions. The momentum technique for the bulk polymer (the sum of live and dead polymer) was used to determine the molecular weight distribution. Empirical correlations previously adjusted with experimental data were used to obtain the melt index as a function of the weight average molecular weight predicted by the mo del.

A differential-algebraic equations (DAE) system arises when the kinetic model equations and the melt index empirical correlations are inserted in the mass balances. The resulting DAE system was solved using the integrator DASSLC[1] and the Matlab/Simu link[2] environment for input and output data manipulation. Each reactor model has 22 states and the simulation time was about 25 seconds for 11 days of plant data using a Pentium III with 800 MHz and 128 MB RAM.

The dataset A was then used to adjust some of model parameter to the studied process. A sensitivity analysis was carried out to select the key parameters to be adjusted. According to this analysis, the selected parameters to adjust yield and melt index were the pre-exponential coefficients and the activation energies of the following reactions: cross propagation, chain transfer to hydrogen, and hydrogen and spontaneous deactivation of active sites. The seven inputs of the model were: monomer, hydrogen and solvent concentrations, catalyst flow rate, height of the fluidized bed, reactor temperature, and total pressure. The dataset B was reserved to validate the model.

Figure 4 shows the comparison between the data set A for polymer yield rate values and the values predicted by the model. When comparing the model predictions with plant data, it becomes clear that the model dynamics must be improved for the second reactor. This could be achieved by including a term of tendency in the objective function.

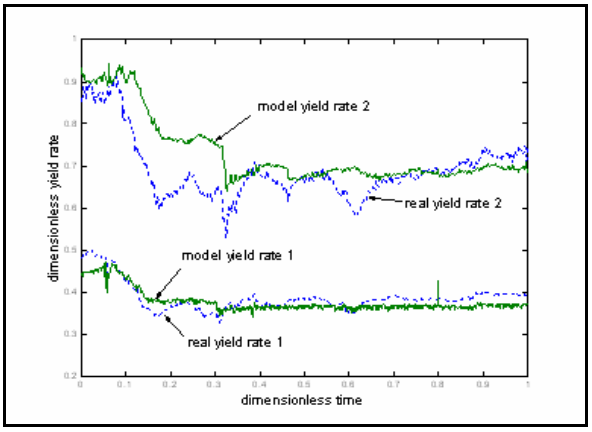


Figure 4: Plant data versus phenomenological model yield rate predictions for dataset A.

[1] http://www.enq.ufrgs.br/enqlib/numeric/numeric.html
[2] Copyright The Mathworks, Inc.

In Figure 5 the same comparison between plant data and model predicted values is presented for melt index. It can be noted that the model dynamics seems to be good, but a considerable offset can be observed.

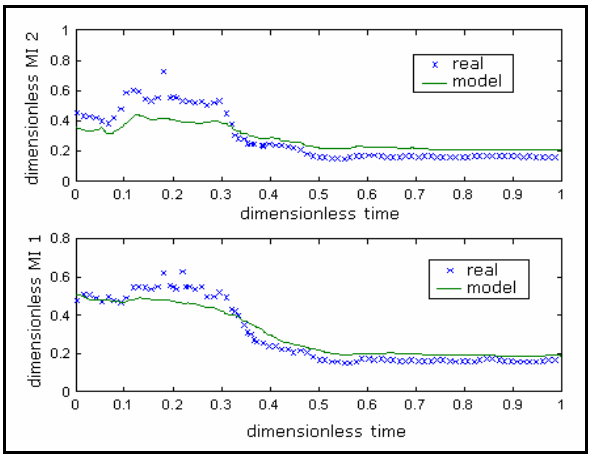


Figure 5: Plant data versus phenomenological model melt index predictions for dataset A.

The model validation will be presented in Section 4, where the phenomenological and the empirical models will be compared.

3. EMPIRICAL MODEL

For a given sample data, let the process input variables be collected as columns of an $\boldsymbol{X}$ matrix of rank r, whose rows represent different process observations. Let also the corresponding output variable values be collected as elements of a y vector. The dimension reduction methods perform the regression procedure in a subspace $\boldsymbol{T}$ extracted from the original $\boldsymbol{X}$ matrix. This subspace is constituted by at most r independent directions (latent structures or components), which are linear combinations of the original explanatory variables. The ability of building a model with the correct number of directions eliminates the collinearity problem and allows some noise filtration. The different dimension reduction methods are basically distinguished from one to another by the criteria considered to extract the latent structures from the original matrices. This work is focused in the linear and nonlinear PLS methods, which decompose the $\boldsymbol{X}$ matrix searching for the directions that better describe the response variable.

The linear PLS method proposed by Wold (1984) and its nonlinear extensions, Wold et al. (1989), Baffi et al. (1999), and Li et al. (2001), are based on the NIPALS (nonlinear iterative partial least squares) algorithm, which determinates the subspace $\boldsymbol{T}$ where the regression is performed. Actually, the NIPALS algorithm extracts the latent structures $\boldsymbol{t}$'s ($\boldsymbol{T}$ columns) from $\boldsymbol{X}$ one by one. Starting from the original $\boldsymbol{X}$ matrix, the algorithm determinates the first weight vector $\boldsymbol{w}$, extracts the direction $\boldsymbol{t} = \boldsymbol{Xw}$ and maps the $\boldsymbol{y}$-$\boldsymbol{t}$ relationship using a general mapping function $\hat{y} = f(\boldsymbol{t})$. The direction $\boldsymbol{t}$ must provide the best fit according to the considered mapping function.

The $\boldsymbol{X}$ matrix and the $\boldsymbol{y}$ vector are then orthogonalized with respect to $\boldsymbol{t}$ and $\hat{\boldsymbol{y}}$. This procedure is then repeated using the orthogonalized $\boldsymbol{X}$ and $\boldsymbol{y}$ until the optimal number of extracted dimensions is achieved. Cross-validation tests or statistical criteria can be used to determine the optimal number of dimensions (Höskuldsson, 1996).

In this work two different versions of NIPALS algorithm were considered. The main difference between them is the applied mapping function. The algorithm proposed by Wold et al. (1984), the linear PLS method, is based on a linear mapping function $\hat{\boldsymbol{y}} = b\boldsymbol{t}$. Aiming to consider the existence of curvature in the $\boldsymbol{X}$-$\boldsymbol{y}$ relationship, Wold et al. (1989) developed a nonlinear NIPALS algorithm that is based on a general mapping function f. In particular, the authors proposed the QPLS method, which employs a quadratic polynomial as mapping function: $\hat{\boldsymbol{y}} = b_0 + b_1\boldsymbol{t} + b_2\boldsymbol{t}^2$. Afterwards, Baffi et al. (1999) suggested some modifications in the nonlinear NIPALS algorithm. Recently, Li et al. (2001) proposed the BTPLS method which resorts a highly flexible mapping function: $\hat{\boldsymbol{y}} = b_0 + b_1.[\mathrm{sgn}(\boldsymbol{t})]^{\delta}|\boldsymbol{t}|^{a}$.

In order to model the studied system input-output relationship, several stationary points were identified in the dynamics datasets A and B. These stationary points were used to estimate the steady-state gains for the melt index and polymer yield rate. Based on these steady-state gains and on known first and second order time constants of the process, empirical models were generated using PLS, QPLS and BTPLS methods in order to predict the system dynamics. The computations were carried out using the software Matlab[3].

For PLS, QPLS and BTPLS methods, the melt index and yield rate variability explained by each component ($j = 1, 2, \ldots, 7$) is presented in Table 2 for both reactors. The significance of the increase that each component causes in the cumulative explained variance was tested by a standard t-test. In Table 2, the significant and the insignificant components are respectively presented in bold and grey numbers. The insignificant components were neglected to avoid overfitting.

As it can be noted, QPLS and BTPLS exhibit higher capacity in explaining the output variables variability. For both reactors, the single significant component of both nonlinear methods is able to explain more than 95% of melt index and yield rate.

Once it could not be observed considerable difference in fitting performance between the nonlinear methods, QPLS was chosen to be compared with the phenomenological model because it is expected to provide more reliable predictions when the original model space is extrapolated.

Table 2: PLS modelling summary.

		PLS	QPLS	DTPLS	PLS	QPLS	DTPLS
	j	Cumulative Melt Index Explained Variability (%)			Cumulative Yield Rate Explained Variability (%)		
REACTOR 1	1	**93,53**	**97,15**	**96,97**	**81,01**	**95,90**	**95,82**
	2	96,97	97,33	98,95	84,08	97,33	97,68
	3	96,29	98,00	97,63	00,00	98,40	90,02
	4	96,47	99,76	97,76	92,36	98,61	98,38
	5	96,57	98,53	97,79	93,43	98,73	98,57
	6	96,70	98,35	97,93	94,76	98,84	98,68
	7	96,00	99,74	97,82	96,61	98,47	90,76
REACTOR 2	1	**04,16**	**97,45**	**90,09**	**73,30**	**97,10**	**97,14**
	2	89,12	97,39	99,01	**92,58**	98,13	97,36
	3	**97,10**	98,33	99,05	95,94	98,32	97,65
	4	97,79	98,45	99,12	96,72	98,70	90,09
	5	98,86	99,53	99,14	97,09	98,81	98,11
	6	98,80	98,35	99,16	97,10	98,90	98,19
	7	98,82	98,70	99,16	97,11	98,00	98,26

4. COMPARSION BETWEEN MODELS

In order to compare the phenomenological and empirical methodologies, the previously generated phenomenological and QPLS models were used to predict the transient behaviour of data set B. The results are reported in Figure 6, Figure 7, Figure 8, and Figure 9.

When the phenomenological and QPLS predictions for polymer yield rate are compared, it becomes clear that the empirical model have superior capability in describing the process dynamics. The empirical model also exhibits a considerably smaller bias. The analysis of the melt index predictions is presented in Figure 8 and Figure 9. Again, the QPLS method exhibited outstanding predictive performance.

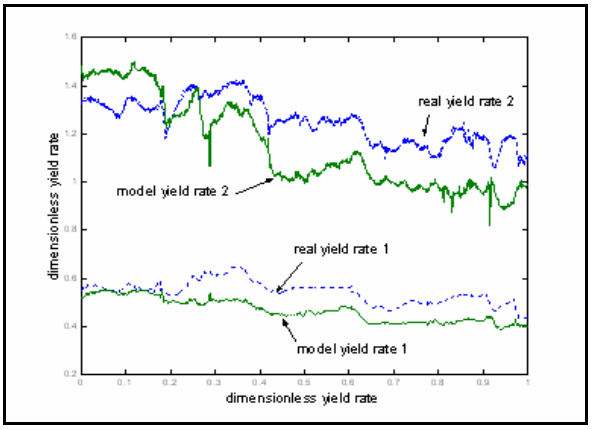


Figure 6: Plant data versus phenomenological model yield rate predictions for dataset B (validation).

[3] Copyright The Mathworks, Inc.

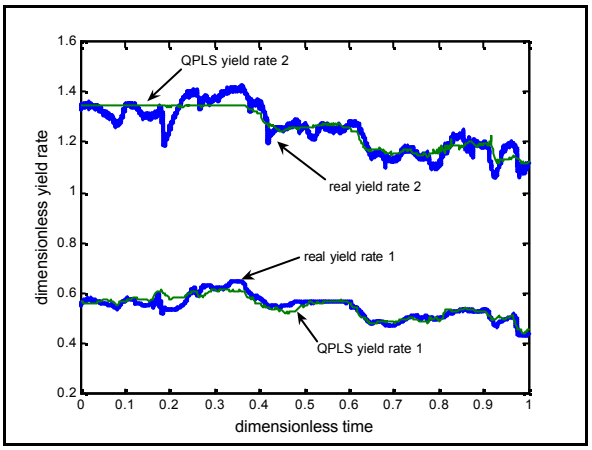


Figure 7: Plant data versus one-component QPLS model yield rate predictions for dataset B (validation).

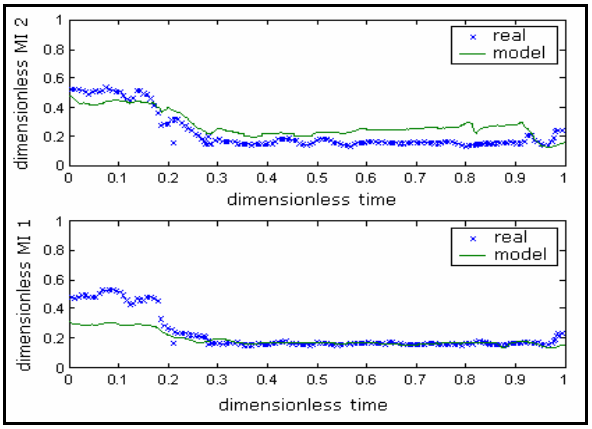


Figure 8: Plant data versus phenomenological model melt index predictions for dataset B (validation).

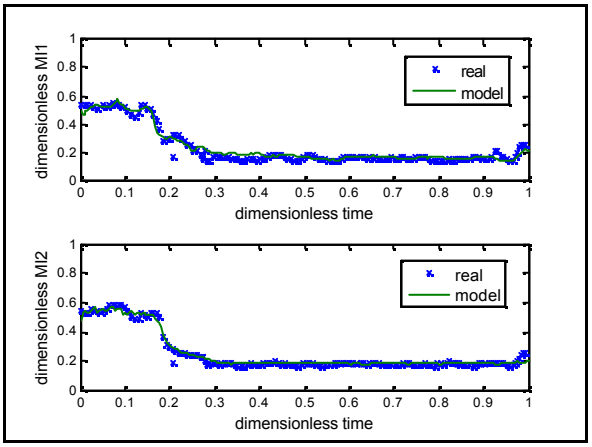


Figure 9: Plant data versus one-component QPLS model melt index predictions for dataset B (validation).

5. INDUSTRIAL APPLICATION

To illustrate the applicability of the developed models, the QPLS model for the melt index was used as virtual analyzer of a predictive controller (MPC) for the MI. Figure 10 and Figure 11 show historical data of MI in opened and closed loop. The MI data in these figures correspond to measurements performed in laboratory from samples taken at each two hours.

The virtual analyzer used in the closed loop provides predicted values of MI to the controller at time intervals of one minute, improving the controller action and the polymer quality, as observed in Figure **11** where the dashed line at normalized MI = 1 is the setpoint.

As can be observed in Figure 12, which shows the normal distribution curves built with the means and standard deviations of opened and closed loop data, the melt index variability was significantly reduced by the controller.

It is important to observe that the dashed lines at normalized MI equal to 0.8 and 1.2 in Figure 11 and Figure 12 correspond to the lower and upper MI specification limits. Consequently, these figures indicate that the closed loop strategy reduced the out of specification products. These results are confirmed by evaluating the process capability index (CPK), defined as the ratio between permissible deviation, measured from the mean value to the nearest specific limit of acceptability and the actual one-sided 3σ spread of the process, Montgomery (1991), and taking into account that larger values of CPK mean higher product quality. The CPK for the opened loop was 0.40 and for the closed loop was 1.00.

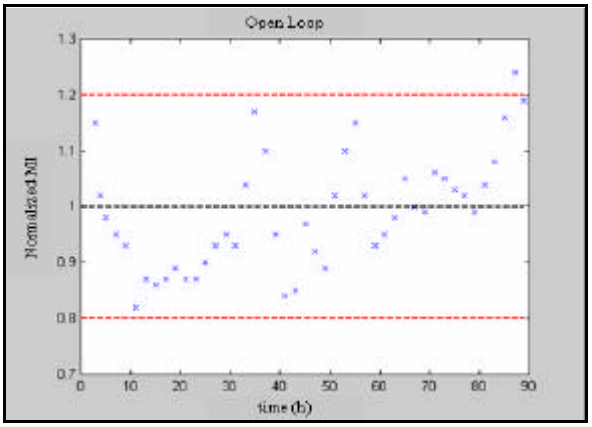


Figure 10: Historical melt index opened loop data.

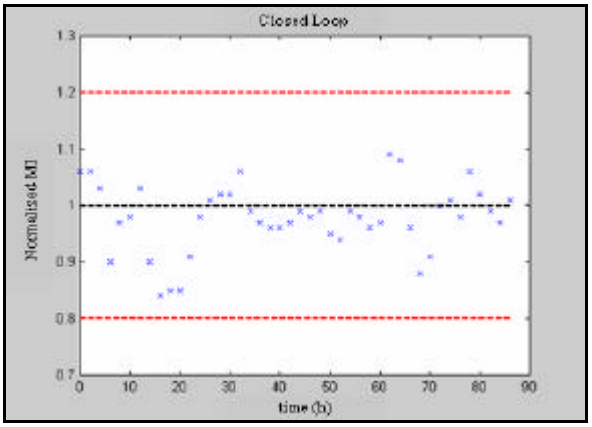


Figure 11: Historical melt index closed loop data.

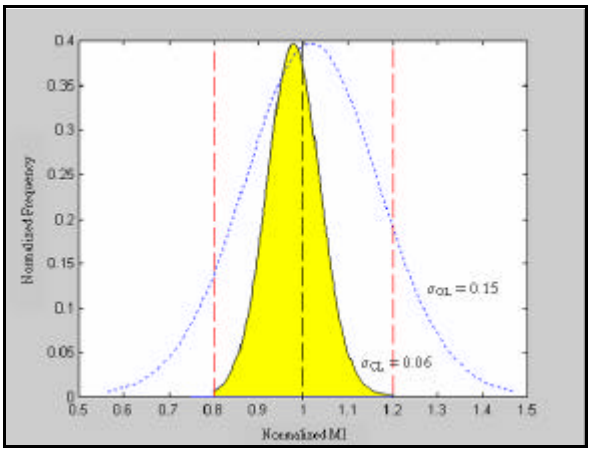


Figure 12: Distribution curves for the open loop and the closed loop.

6. CONCLUSION

Models of different types for ethylene polymerization reactors were adjusted with process industrial data. The comparison between these models showed better results for the empirical models with nonlinear steady-state gains and linear dynamics.

The empirical model for the melt index was successfully used as virtual analyzer of an advanced control strategy for an industrial plant, improving the controller action and the polymer quality by reducing significantly the process variability.

REFERENCES

Baffi, G., Martin, E.B and Morris, A.J. (1999), Non-Linear projection to latent structures revisted: the quadratic PLS algorithm, *Computers and Chemical Engineering*, **23**, 395.

Bindlish, R., Rawlings, J.B. and Young, R.E. (2003), Parameter Estimation for Industrial Polymerization Processes, *AIChE J.*, **49(8)**, 2071.

Choi, K.Y. and Ray, W.H. (1985), The Dynamic Behaviour of Fluidized Bed Reactors for Solid Catalized Gas-Phase Olefin Polymerization, *Chem. Eng. Sci*., **40**, 2261.

Finkler, T.F. (2003), *Desenvolvimento de uma Ferramenta para Obtenção de Modelos Empíricos*, Master Thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.

Gambetta R., Zacca J.J. and Secchi A.R. (2001) Model for Estimation of Kinetics Parameters in Gas-Phase Polymerization Reactors. *In proceedings of the 3rd Mercosur Congress on Process Systems Engineering* Santa Fé, Argentina, vol. II, 901-906.

Höskuldsson, A. (1996), *Prediction Methods in Science and Technology*, Thor Publishing, v.1, 1996.

Kunii, D. and Levenspiel, O. (1991), *Fluidization Engineering*, 2nd Edition, Butterworth-Heinemman, New York.

Li, B., Martin, E.B and Morris, A.J. (2001), Box-Tidwell based partial least squares regression, *Computers & Chemical Engineering*, **25**, 1219.

McAuley, K.B. and MacGregor, J.F. (1991), On-line inference of polymer properties in an industrial polyethylene reactor, *AIChE J.*, **37(6)**, 825.

Montgomery, D. (1991), *Introduction to Statistical Quality Control*, John Wiley & Sons, New York

Sato, C., Ohtani, T. and Nishitani (2000), H., Modeling, simulation and nonlinear control of a gas-phase polymerization process, *Comp. Chem. Eng*., **24**, 945.

Soroush, M. (1998), State and parameter estimations and their applications in process control, *Comp. Chem. Eng*., **23**, 229.

Wold, S. Ruhe A., Kettaneh N. and Skagerberg B. (1989), Nonlinear PLS modeling, *Chemometrics and intelligent laboratory systems*, **7**, 53.

Wold, S. Ruhe A. and Wold H. (1984), Dunn, W.J., The Collinearity Problem in Linear Regression: The Partial Least Squares Approach to Generalized Inverses, *Siam J. Sci. Stat*. Comput., **5**, 735.

Zhao, H., Guiver, J., Neelakantan, R. and Biegler, L.T. (2001), A nonlinear industrial model predictive controller using integrated PLS and neural net state-space model, *Control Eng. Practice*, **9**, 125.

Session 8.3

Performance Assessment of Closed-Loop Systems

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

PERFORMANCE ASSESSMENT OF RUN-TO-RUN EWMA CONTROLLERS

Amogh V. Prabhu*, Thomas F. Edgar* and Robert Chong†

**Dept of Chemical Engineering, The University of Texas at Austin, TX 78712*
†Advanced Micro Devices, Austin, TX 78751

Abstract: An iterative method is developed to determine a performance criterion for best achievable performance for discrete integral controllers. Using the performance criterion, optimal performance of the controller in place is also indicated. An analytical expression is derived so that a realistic assessment of the given integral controller is obtained. Using the theoretical equivalence of discrete integral and exponentially weighted moving average (EWMA) controllers, the method is then extended to performance assessment of EWMA controllers. A semiconductor manufacturing example is used to illustrate the utility of the method. *Copyright © 2005 IFAC*

Keywords: performance assessment, feedback control, delay, integral control, single loop

1. INTRODUCTION

For any feedback control system in a manufacturing process, variation from the desired output can occur due to two reasons: Either the process state has changed or the controller performance has degraded. A change in process state occurs whenever any of the major process parameters change by an amount which cannot be corrected without a change in the controller tuning. But if the controller performance is degraded without any change in the state, then the controller itself must be analyzed to verify that it is behaving optimally under the given conditions.

1.1 Minimum variance control (MVC)

The first effort towards developing a performance index for feedback control systems was made by Harris (1989). This work proposed that minimum variance control represents the best achievable performance by a feedback system. All other kinds of control behave sub-optimally as compared to it. The method is applicable only to SISO systems and involves fitting a univariate time series to process data collected under routine control. This is compared to the performance of a minimum variance controller. However it has certain drawbacks:

– If controller performance is close to that of minimum variance, it indicates that it is behaving optimally. But if the deviation from minimum variance performance is large, it does not imply that the controller is sub-optimal. Under the given setup, it may the best that the controller can do. Therefore, a different benchmark may be required in such a case.
– The minimum variance index does a good job of indicating loops that have oscillation problems. Unfortunately it considers loops that are sluggish to be fine. This particularly happens when the controller has been detuned to a large extent, making controlling the loop slow.
– Minimum variance index is only a theoretical lower bound on the best possible performance. If applied in a real system, it can lead to large variations in input signals, and the closed loop often has poor robustness properties. Therefore it is not recommended to be applied to a system, but just serve as a benchmark.

1.2 Alternative methods

While the minimum variance control concept proposed by Harris (1989) was initially developed for feedback and feedforward-feedback controlled univariate systems (Desborough and Harris (1992, 1993)), the idea was further extended to multivariate systems. Stanfelj *et al.* (1993) have diagnosed the performance of single loop feedforward-feedback systems based on the MVC criteria. Eriksson and Isaksson (1994) have analyzed the MVC index and pointed out several drawbacks in the index similarly to those listed earlier. Huang *et al.* (1995) have introduced a useful method for monitoring of MIMO processes with feedback control, known as Filtering and Correlation (FCOR) analysis. This concept is further developed by Huang *et al.* (1997) to estimate

a suitable explicit expression for the feedback controller invariant term of the closed-loop MIMO process from routine operating data. Harris *et al.* (1996a) have extended the MVC index to multivariable feedback processes in a manner similar to Huang *et al.* (1995) but without the filtering approach. Ko and Edgar (1998) have proposed a method to determine achievable PI control performance when the process is being perturbed by stochastic load disturbances. This is further extended to multivariable feedback control by Ko and Edgar (2000) using a finite horizon MV benchmark with specified horizon length. Salsbury (2005) has formulated statistical change detection procedures which can be used for processes subject to random load changes. The method is applicable to SISO feedback systems and uses a normalized index, which is similar to the damping ratio in a second order process.

Apart from these articles, Qin (1998) and Harris *et al.* (1999) have reviewed most methods up to 1998.

1.3 Performance Monitoring in Semiconductor Manufacturing

Most of the major processes involved in semiconductor manufacturing are done in a batch manner (Edgar *et al.*, 2000), so that any process change involves changes in the batch recipe. Run-to-run control is the most popular form of control wherein the controller parameters can be tuned after each lot, based on the data from the previous lot. Statistical process control is widely used, with most processes adopting an Exponential Weighted Moving Average (EWMA) algorithm. None of the above listed methods were developed for control systems used in the semiconductor industry. But a best achievable PID control performance bound was proposed by Ko and Edgar (2004). This was an iterative algorithm which optimized the controller parameters. Using the theoretical equivalence of EWMA controllers with discrete integral controllers, this iterative algorithm was adapted to run-to-run EWMA controllers, commonly used in semiconductor manufacturing.

In this article, we derive an iterative solution method for the calculation of achievable performance bound of a run-to-run EWMA controller, where the iterative solution uses the process input-output data and the process model. This iterative solution is based on an analytic solution for closed-loop output. A normalized performance index is then defined based on the best achievable performance. An example of a process controlled by such a controller is employed to illustrate the effectiveness of the proposed method.

2. THEORY DEVELOPMENT

The following theory explains in a step-wise manner how the performance monitoring method for a discrete integral controller (based on Ko and Edgar (2004)) can be used to monitor EWMA controllers.

2.1 Discrete Integral Controller

The process output is represented by the following discrete-time model

$$y_{k+1} = \overline{b_{k+1}} u_{k+1} + c_{k+1} \tag{1}$$

Where y is the output, u is the input, b is the gain and c is the disturbance driven by white noise.

The feedback integral controller is given by

$$K = \frac{k_I}{1 - q^{-1}} \tag{2}$$

The output u_k is obtained as

$$u_{k+1} = K\left(y_{sp} - y_k\right) = -\frac{k_I}{1 - q^{-1}} y_k \tag{3}$$

The above equation results from setting $y_{sp} = 0$. If there is no set-point change, the output of the process can now be simplified to

$$y_k = \frac{c_k}{1 + \overline{b_k} K} \tag{4}$$

From the given data, we develop an ARMAX (Auto-Regressive Moving Average with eXogenous input) model. Using a prediction horizon p, we calculate the step response coefficients of the model (which is equivalent to the gain of the process in this case). Thus,

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_p \end{bmatrix} = -\begin{bmatrix} 0 & \cdots & \cdots & 0 \\ s_1 & 0 & \ddots & \vdots \\ \vdots & s_1 & 0 & \vdots \\ s_p & \cdots & s_1 & 0 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_p \end{bmatrix} k_I + \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{bmatrix} \tag{5}$$

or more simply put

$$Y = \left(I + Sk_I\right)^{-1} C \tag{6}$$

This forms the model of the given data, which can be used to calculate the optimal response. The output data impulse response is then determined, so that

$$y_k = \sum_{i=0}^{p} \Psi_i c_{k-i} \tag{7a}$$

$$\begin{bmatrix} \Psi_0 \\ \Psi_1 \\ \vdots \\ \Psi_p \end{bmatrix} = \left(I + Sk_I\right)^{-1} C \tag{7b}$$

Thus, knowing the impulse response coefficients, the disturbance vector C can be calculated.

2.2 Optimal Controller Gain

The variance of the output is given by

$$V = C^T \left(I + S^T k_I\right)^{-1} \left(I + Sk_I\right)^{-1} C \tag{8}$$

Then the optimal k_I can be obtained using Newton's method so that

$$k_{I_{new}} = k_{I_{old}} - \frac{\left(\partial V / \partial k_I\right)_{old}}{\left(\partial^2 V / \partial k_I^2\right)_{old}} \tag{9}$$

The first and second derivatives are given by

$$\frac{\partial V}{\partial k_I} = -2C^T \left(L^{-1}\right)^T S L^{-2} C = 0 \tag{10}$$

$$\frac{\partial^2 V}{\partial k_I^2} = 2C^T \left(L^{-2}\right)^T S^T S L^{-2} C + 4C^T \left(L^{-1}\right)^T S^2 L^{-3} C \tag{11}$$

The first derivative becomes zero for the optimal gain and $L = I + Sk_I$

The performance index is now given by the ratio of the variance of optimal and actual response

$$\zeta = \frac{Y_{opt}^T Y_{opt}}{Y^T Y} \tag{12}$$

and the optimal response is calculated by

$$y_{k_{opt}} = \left(\frac{1 + \overline{b_k} k_I}{1 + \overline{b_k} k_{I_{opt}}}\right) y_k \tag{13}$$

The normalized performance index has the range of $0 < \zeta \le 1$, and $\zeta = 1$ indicates the best performance under Integral Control. With this definition, $1 - \zeta$ indicates the maximum fractional reduction in the output variance.

2.3 EWMA Controller

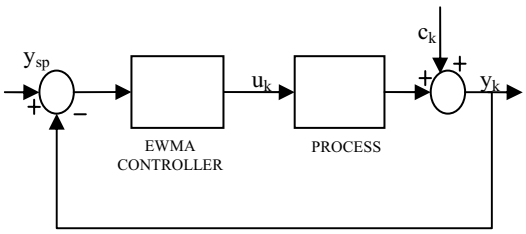


Fig. 1: EWMA controlled run-to-run process

The system shown above in Figure 1 is one controlled by a standard EWMA controller (Campbell *et al.*, 2002). The equations are as follows (with similar notations):

$$\overline{y_{k+1}} = \overline{b_{k+1}} u_{k+1} + c_{k+1} \tag{14}$$

The observer updates the disturbance c_{k+1} using an EWMA formula which is

$$c_{k+1} = \lambda \times \left(y_k - \overline{b_k} u_{k+1}\right) + \left(1 - \lambda\right) \times c_k \tag{15}$$

The input is now given by (with y_{sp} as the target)

$$u_{k+1} = \frac{y_{sp} - c_{k+1}}{\overline{b_{k+1}}} \tag{16}$$

The actual gains are determined before the lot is processed using historical data.

$$b_k = \frac{y_k}{u_k} \tag{17}$$

For a pure gain system, the EWMA controller is equivalent to a discrete integral controller with gain k_I (Box, 1993), with

$$k_I = \frac{\lambda}{b_{mean}} \qquad (19)$$

Thus by representing the process data as one controlled by a discrete integral process, the performance index of an EWMA controlled process may be obtained.

3. EXAMPLE

An etch process at AMD[1] was considered for analysis. The governing equations and input – output variables are also defined. The process model used for this process is as follows

EtchDepth = EtchRate*EtchTime + Bias (20)

The rate is updated by EWMA as given in the previous section. Accordingly, the manipulated variable is time, while the controlled variable is (EtchDepth – Bias). The algorithm for calculating the performance index essentially utilizes the moving window approach, i.e., considering only the last 'n' data points in time so that the performance index calculated represents the current state of the process. The data considered was for etch processes run with different equipment each time. Thus each type of etch process was analyzed separately to evaluate which equipment performed better than others. About 29 different etch processes at AMD were compared.

4. RESULTS

The following three types of results could be obtained by the above developed method. Not only can the method be used to compare different processes, the effect of delay is also demonstrated. Also the performance of a process can be tracked over time.

4.1 Distribution of performance indices

The etch processes showed a distribution of performance indices. The performance index usually lies between 0.8 and 1. Figure 2 shows the distribution of the processes considered in each range of performance index.

Although most processes were found to lie in the 0.9 to 1 range, the remaining processes were found to be uniformly distributed in the 0.1 to 0.8 region. Thus, majority of the processes were found to be operating sub-optimally.

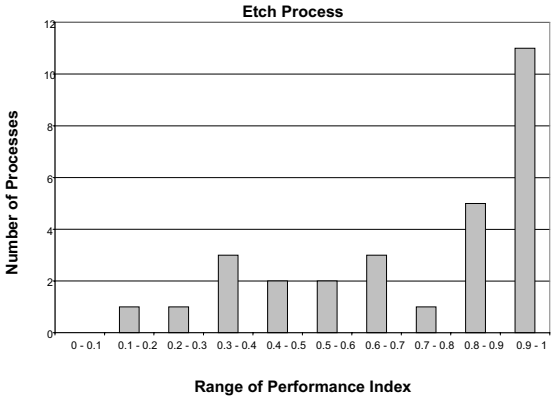


Fig. 2: Distribution of performance

4.2 Effect of delay

When no delay is considered in the calculation of the performance index, the algorithm assumes that the only reason for suboptimal performance is the controller itself. But if we do consider a delay of one or more, the algorithm takes into account that this delay is responsible for some degradation in performance. This is because the delay is considered during the selection of the ARMAX model for the data. Thus, with increasing delay, the performance index goes asymptotically to 1.0. This is because, as the metrology delay increases, it becomes the primary reason for suboptimal performance. In other words, the controller cannot work efficiently beyond a certain threshold. Consider the example shown in Figure 3.

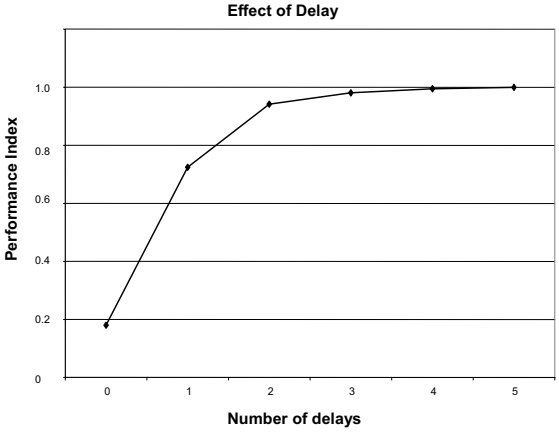


Fig. 3: Change in performance index with delay

[1] Advanced Micro Devices, Inc.

A performance index of one in this case does not indicate optimality but instead points to the delay in the process. Thus if the process has a significant amount of delay, expectations of optimal performance from the process must be greatly reduced.

4.3 Change in performance over time

Moving windows were used to study the change in performance of the process. Following is a sample chart which tracks the performance index over time for a moving window of 50 points. In Figure 4, the dots are the actual values of the index while the continuous line is the graphical trend for the thread with a 5-point moving average.

Figure 4 shows the decline in performance of the thread with time. A sudden degradation in performance is seen to have occurred mid-way in the process. Thereafter the performance is on the decline.

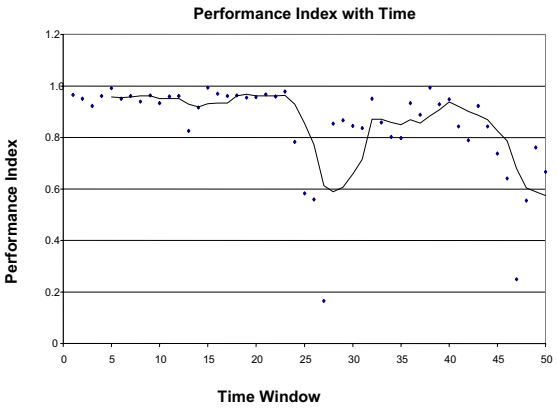


Fig. 4: Change in performance with time

5. CONCLUSIONS AND FUTURE WORK

The achievable performance bound was proposed for use in assessing and monitoring single-loop EWMA control loop performance. For this purpose, an iterative solution was derived that gives the best achievable performance in terms of the closed-loop input-output data and the process model. An explicit solution was derived as a function of EWMA settings. A performance index was defined based on the best achievable performance for use as a realistic performance measure in the single-loop EWMA control systems. An example showed the utility of the proposed method for the effective performance assessment of the existing controller as also for comparing the performance of different processes.

This work is one of the first applications of performance assessment techniques to run-to-run control systems. In the future, methods for non-EWMA processes can be developed. Also, most run-to-run processes in semiconductor manufacturing tend to have variable time delays. This aspect could be further explored and new techniques formulated to incorporate this variable delay. Also the next step in performance assessment needs to be suggested, viz. having determined which control loops perform sub-optimally, remedial steps must be outlined.

NOMENCLATURE

b_k = Actual gain
$\overline{b_k}$ = Predicted gain
b_{mean} = Average gain used
c_k = Disturbance
C = Vector of disturbance estimates
d_k = Actual measurement
$\overline{d_k}$ = Predicted measurement
h_k = Bias
I = Identity Matrix
k_I = Integral controller gain
$K\ (q^{-1})$ = Integral controller
L = $I + Sk_I$
q^{-1} = Backward shift operator
s_i = Step response coefficient
S = Matrix of step response coefficients
T = Target
u_k = Input used
V = Variance
y_k = Normalized output
y_{sp} = Set-point for normalized output
Y = Vector of normalized output values
Y_{opt} = Vector of optimal output values

Subscripts

I = Integral controller
k = Time
mean = Average from a set of given values
new = Value for current iteration
old = Value from previous iteration
opt = Optimal value
p = Prediction horizon
sp = Set-point

Greek Symbols

λ = EWMA weighting
ζ = Performance Index
Ψ_i = Impulse response coefficient

REFERENCES

Box G. E. P. (1993). Process Adjustment and Quality Control. *Total Quality Management*, **4**, 2.

Campbell, J. W., S. K. Firth, A. J. Toprac and T.F. Edgar (2002). A Comparison of Run-to-Run Control Algorithms. *Proceedings of the American Control Conference*, 2150.

Desborough, L., and T. J. Harris (1992). Performance Assessment Measures for Univariate Feedback Control. *Canadian Journal of Chemical Engineering*, **70**, 1186.

Desborough, L., and T. J. Harris (1993). Performance Assessment Measures for Univariate Feedforward/Feedback Control. *Canadian Journal of Chemical Engineering*, **71**, 605.

Edgar, T. F., S. W. Butler, W. J. Campbell, C. Pfeiffer, C. Bode, S. B. Hwang, K. S. Balakrishnan, and J. Hahn (2000). Automatic Control in Microelectronics Manufacturing: Practices, Challenges and Possibilities. *Automatica*, **36**, 1567.

Eriksson, P.-G., and A. J. Isaksson (1994). Some Aspects of Control Loop Performance Monitoring. *Proceedings of the IEEE Conference on Control Applications*, 1029.

Harris, T. J. (1989). Assessment of Control Loop Performance. *Canadian Journal of Chemical Engineering*, **67**, 856.

Harris, T. J., F. Boudreau, and J. F. MacGregor (1996a). Performance Assessment of Multivariable Feedback Controllers. *Automatica*, **32**, 1505.

Harris T. J., C. T. Seppala, and L. D. Desborough (1999). A Review of Performance Monitoring and Assessment Techniques for Univariate and Multivariate Control Systems. *Journal of Process Control*, **9**, 1.

Huang, B., S. L. Shah, and E. K. Kwok (1995). Online Control Performance Monitoring of MIMO Processes. *Proceedings of the American Control Conference*, 1250.

Huang, B., S. L. Shah, and E. K. Kwok (1997). Good, Bad or Optimal? Performance Assessment of Multivariable Processes. *Automatica*, **33**, 1175.

Ko, B.-S., and T. F. Edgar (1998). Assessment of Achievable PI Control Performance for Linear Processes with Dead Time. *Proceedings of the American Control Conference*, 1548.

Ko, B.-S., and T. F. Edgar (2000). Performance Assessment of Multivariable Feedback Control Systems. *Proceedings of the American Control Conference*, 4373.

Ko, B.-S., and T. F. Edgar (2004). PID Control Performance Assessment: The Single-Loop Case. *AIChE Journal*, **50**, 1211.

Qin, S. J. (1998). Controller Performance Monitoring – A Review and Assessment. *Computers and Chemical Engineering*, **23**, 173.

Salsbury, T. I. (2005). A Practical Method for assessing the Performance of Control Loops subject to random load changes. *Journal of Process Control*, **15**, 393.

Stanfelj, N., T. E. Marlin, and J. F. MacGregor (1993). Monitoring and Diagnosing Process Control Performance: The Single-Loop Case. *Industrial Engineering Chemistry Research*, **32**, 301.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

Modified Independent Component Analysis for Multivariate Statistical Process Monitoring

Jong-Min Lee[a], S. Joe Qin[a]* and In-Beum Lee[b]

[a]*Department of Chemical Engineering, The University of Texas at Austin, Austin, TX78712, USA*
[b]*Department of Chemical Engineering, Pohang University of Science and Technology, San 31 Hyoja-Dong, Pohang, 790-784, Korea*

Abstract: In this paper, a modified independent component analysis (ICA) and its application to process monitoring are proposed. The basic idea of this approach is to use the modified ICA to extract some dominant independent components from normal operating process data and to combine them with statistical process monitoring techniques. The proposed monitoring method is applied to fault detection and identification in the Tennessee Eastman process and is compared with the conventional PCA based monitoring method. The monitoring results demonstrate that the proposed method outperforms PCA in terms of the fault detection rate while attaining comparable false alarm rate.

Keywords: Fault Detection; Fault Identification; Statistical Process Control; Independent Component Analysis; Principal Component Analysis

1. INTRODUCTION

In order to extract useful information from a large amount of process data and to detect and diagnose various faults in an abnormal operating situation, a number of multivariate statistical process monitoring (MSPM) approaches based on principal component analysis (PCA) have been developed. PCA is a second-order method, considering only mean and variance of the data. It gives only uncorrelated components, not independent components. PCA performs well in many cases, but gives limited meaningful representations for non-Gaussian data, which can be typical in industrial measurement data (Kermit and Tomic, 2003).

More recently, several MSPM methods based on independent component analysis (ICA) have been proposed (Kano *et al*., 2003, 2004; Lee *et al*., 2003, 2004, Yoo *et al*., 2004; Albazzaz and Wang, 2004). The goal of ICA is to decompose observed data into linear combinations of statistically independent components. In comparison to PCA, ICA involves higher-order statistics, i.e., not only does it decorrelate the data (second order statistics) but also reduces higher order statistical dependencies (Lee, 1998). However, conventional ICA-based monitoring method has some drawbacks for MSPM. First, it is not easy to determine how many independent components (ICs) should be extracted in order to establish a stable ICA model (Kermit and Tomic, 2003). Generally, ICs are extracted up to the dimension of given data, which causes high computational load. Second, the extracted ICs are not ranked in any order as is the case for PCA. In addition, random initialization of the demixing matrix in the whitened space can give different solutions when performing the ICA algorithm.

In this paper, a modified ICA algorithm is proposed to extract dominant ICs from multivariate data. The basic idea is to estimate the variance and the axes of dominant ICs using PCA and then perform ICA to update the dominant ICs while maintaining the variance. This article is organized as follows. The original ICA algorithm is introduced, followed by a modified ICA algorithm and its application to process monitoring. Then, the performance of process monitoring using the modified ICA is illustrated through the Tennessee Eastman process. Finally, a conclusion is given.

* To whom correspondence should be addressed.
E-mail: qin@che.utexas.edu

2. ORIGINAL ICA

The model of ICA is given by

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_m]^T$ is an m-dimensional observation vector, $\mathbf{A} \in R^{m \times p}$ is an unknown mixing matrix and $\mathbf{s} = [s_1, s_2, \ldots, s_p]^T$ is a p-dimensional independent component vector. The objective of ICA is to estimate both $\mathbf{A}$ and $\mathbf{s}$ from only $\mathbf{x}$. This solution is equivalent to finding a demixing matrix $\mathbf{W}$ whose form is such that the elements of the reconstructed vector $\hat{\mathbf{s}}$, given as

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} \tag{2}$$

become as independent of each other as possible.

In the original ICA algorithm, it is assumed that m equals p and all ICs have unit variance for convenience. The initial step in ICA is to remove all the cross-correlation of $\mathbf{x}$, given as

$$\mathbf{z} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T\mathbf{x} = \mathbf{Q}\mathbf{x} \tag{3}$$

where $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$, $\mathbf{Q} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T$, and $\mathbf{U}$ and $\mathbf{\Lambda}$ are eigenvector and eigenvalue matrix, respectively, generated from the eigen-decomposition of $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$. Then, Eq. (3) can be expressed as

$$\mathbf{z} = \mathbf{Q}\mathbf{x} = \mathbf{Q}\mathbf{A}\mathbf{s} = \mathbf{B}\mathbf{s} \tag{4}$$

where $\mathbf{B} = \mathbf{Q}\mathbf{A}$ is an orthogonal matrix since

$$\mathbf{I} = E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{B}E\{\mathbf{s}\mathbf{s}^T\}\mathbf{B}^T = \mathbf{B}\mathbf{B}^T. \tag{5}$$

Thus, $\mathbf{s}$ can be estimated from Eq. (4)

$$\hat{\mathbf{s}} = \mathbf{B}^T\mathbf{z} = \mathbf{B}^T\mathbf{Q}\mathbf{x}. \tag{6}$$

From Eq. (2) and Eq. (6),

$$\mathbf{W} = \mathbf{B}^T\mathbf{Q}. \tag{7}$$

To calculate $\mathbf{B}$, each column vector $\mathbf{b}_i$ is randomly initialized and then updated so that the i-th independent component $\hat{s}_i = (\mathbf{b}_i)^T\mathbf{z}$ may have maximized non-Gaussianity. As a measure of non-Gaussianity, negentropy, the difference of the differential entropy between the given data and Gaussian distribution data, has been used. Hyvärinen and Oja (2000) introduced a reliable approximation of negentropy:

$$J(y) \approx \left[E\{G(y)\} - E\{G(v)\}\right]^2 \tag{8}$$

where y is assumed to be of zero mean and unit variance, v is a Gaussian variable of zero mean and unit variance, and G is any non-quadratic function. Hyvärinen and Oja (2000) suggested three functions for G:

$$G_1(u) = \frac{1}{a_1}\log\cosh(a_1 u) \tag{9}$$

$$G_2(u) = \exp(-a_2 u^2/2) \tag{10}$$

$$G_3(u) = u^4 \tag{11}$$

where $1 \le a_1 \le 2$ and $a_2 \approx 1$. G_2 and G_3 are more suitable for super-Gaussian and sub-Gaussian components, respectively. G_1 is a good general-purpose contrast function and is therefore selected for use in this paper.

Hyvärinen (1999) introduced a highly efficient fixed-point algorithm for ICA based on the approximate form for the negentropy. The algorithm, called FastICA, calculates each column of the matrix $\mathbf{B}$ one by one and allows the identification of each independent component. More details on the FastICA algorithm are well described in Hyvärinen and Oja (2000), Hyvärinen (1999), Hyvärinen *et al*. (2001).

3. MODIFIED ICA

ICA not only decorrelates the data (second order statistics) but also reduces higher order statistical dependencies; hence it can extract underlying hidden factors efficiently and capture the essential structure of the data. Based on this merit, some researchers have illustrated that applying ICA to process monitoring is useful to detect and identify various faults generated from abnormal situations (Kano *et al*., 2003; Lee *et al*., 2004).

However, the conventional ICA-based monitoring method has some drawbacks. A fundamental assumption behind original ICA is that the number of ICs equals that of variables of given data. In case that the number of measured variables is very large, it has high computational load and may extract additional ICs which are unimportant for detecting faults. Of course, one can reduce data dimension in advance using PCA before performing ICA (Hyvärinen *et al*., 2001). However, much information needed to extract essential ICs is ignored by data reduction with PCA. The second problem in the original ICA algorithm is that ICs are not ordered in the same fashion as with PCA since the variance of extracted ICs is assumed to be all one (Kermit and Tomic, 2003). There is no standard criterion to order ICs. Furthermore, random initialization of demixing matrix $\mathbf{B}$ in the whitened space can lead to different solutions when performing ICA algorithm (Kermit and Tomic, 2003). In order to solve these problems, a modified ICA algorithm is suggested in this paper. The modified ICA algorithm can extract a few dominant ICs, determine the order of ICs, and give a consistent solution. The basic idea is to first use PCA to estimate initial ICs where the variance of each IC is the same as that of each PC and then to update a few dominant ICs using FastICA algorithm. Here, it is reasonable to expect the space spanned by the major ICs to be essentially similar to the ones associated to the largest principal components (PCs) because ICA can be viewed as a modified PCA (centering and whitening) and an additional iterative process (Kocsor *et al*., 2004).

The objective of the modified ICA can be defined as follows: to find a demixing matrix $\mathbf{W} \in R^{p \times m}$ whose form is such that the elements of the extracted vector $\mathbf{y}$, given as

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{12}$$

become as independent of each other as possible and have been ordered by their variances that are the same as the variances of the corresponding PCs.

To solve above problem, first of all, all score components are extracted from PCA

$$\mathbf{t} = \mathbf{U}^T\mathbf{x} \tag{13}$$

where $\mathbf{t}$ is the score vector with $E\{\mathbf{t}\mathbf{t}^T\} = \mathbf{\Lambda} = diag\{\lambda_1,\ldots,\lambda_m\} \in R^{m\times m}$ and $\mathbf{U} \in R^{m\times m}$ is the loading matrix obtained from $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, respectively. In some cases, the last a few eigenvalues in $\mathbf{\Lambda}$ are so small that they are close to zero. In that case, the eigenvalues and the corresponding eigenvectors can be excluded. However, it is important to retain as many eigenvalues as possible because the extracted score components give additional information to find essential ICs even though their variances are small.

Eq. (13) can be changed as follows through the whitening transform:

$$\mathbf{z} = \mathbf{Q}\mathbf{x} \tag{14}$$

where $\mathbf{z}$ is the normalized score vector, $\mathbf{z} = \mathbf{\Lambda}^{-1/2}\mathbf{t}$, and $\mathbf{Q} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T$.

From $\mathbf{z} \in R^m$, a few dominant ICs, $\mathbf{y} \in R^p$ satisfying $E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{D} = diag\{\lambda_1,\ldots,\lambda_p\}$, should be found such that the elements of $\mathbf{y}$ are as independent of each other as possible, using

$$\mathbf{y} = \mathbf{C}^T\mathbf{z} \tag{15}$$

where $\mathbf{C} \in R^{m\times p}$, $\mathbf{C}^T\mathbf{C} = \mathbf{D}$. $E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{D}$ reflects that the variance of each element of $\mathbf{y}$ is the same as that of scores in PCA, hence ICs can be ordered according to their variances.

Eq. (15) can be arranged as a simpler model by multiplying $\mathbf{D}^{-1/2}$ to each side:

$$\mathbf{y}_n = \mathbf{C}_n{}^T\mathbf{z} \tag{16}$$

where $\mathbf{y}_n = \mathbf{D}^{-1/2}\mathbf{y}$, $\mathbf{D}^{-1/2}\mathbf{C}^T = \mathbf{C}_n{}^T$, $\mathbf{C}_n{}^T\mathbf{C}_n = \mathbf{I}$, and $E\{\mathbf{y}_n\mathbf{y}_n{}^T\} = \mathbf{I}$. Consequently, the problem of finding an arbitrary demixing matrix $\mathbf{W}$ is reduced to the simpler problem of finding a matrix $\mathbf{C}_n$ which has fewer parameters to estimate as a result of the orthogonality. Note that $\mathbf{z}$ is the normalized score vector generated from PCA, that is uncorrelated and had been ordered by its original variance. The first p components of $\mathbf{z}$ can be a good initial value of $\mathbf{y}_n$ since statistical dependencies of data have been removed up to the second order (mean and variance) by PCA. To do this, the initial matrix of $\mathbf{C}_n{}^T$ should be set to be

$$\mathbf{C}_n{}^T = \left[\mathbf{I}_p \vdots \mathbf{0}\right] \tag{17}$$

where $\mathbf{I}_p$ is the *p*-dimensional identity matrix and $\mathbf{0}$ is *p* by *m*-*p* zero matrix. This initialization is based on the assumption that extracted PCs are good initial estimates of ICs, and thereby can give a consistent solution unlike random initialization.

The detail procedures to find a few dominant ICs are:

1) Determine *p*, the number of ICs to estimate. Set counter $i \leftarrow 1$.
2) Denote $\mathbf{c}_{n,i}$ as the *i*-th column vector of $\mathbf{C}_n$ and take the initial vector $\mathbf{c}_{n,i}$ to be *i*-th column vector of $\mathbf{I}_p$ in Eq. (17)
3) Let $\mathbf{c}_{n,i} \leftarrow E\{\mathbf{z}g(\mathbf{c}_{n,i}{}^T\mathbf{z})\} - E\{g'(\mathbf{c}_{n,i}{}^T\mathbf{z})\}\mathbf{c}_{n,i}$, where g is the first derivative and g' is the second derivative of *G*, where *G* takes the form of Eq. (9), (10) or (11). This step is an approximate Newton iteration procedure for the maximization of the negentropy given in Eq. (8).
4) Do the orthogonalization:
$$\mathbf{c}_{n,i} \leftarrow \mathbf{c}_{n,i} - \sum_{j=1}^{i-1}(\mathbf{c}_{n,i}{}^T\mathbf{c}_{n,j})\mathbf{c}_{n,j}.$$
This step removes the information contained in the solutions already found.
5) Normalize $\mathbf{c}_{n,i} \leftarrow \mathbf{c}_{n,i} / \left\|\mathbf{c}_{n,i}\right\|$
6) If $\mathbf{c}_{n,i}$ has not converged, go back to Step 3).
7) If $\mathbf{c}_{n,i}$ has converged, output the vector $\mathbf{c}_{n,i}$. Then, if $i \le p$ set $i \leftarrow i+1$ and go back to Step 2).

Once $\mathbf{C}_n$ is found, then final demxing matrix $\mathbf{W}$ and mixing matrix $\mathbf{A}$ can be obtained from

$$\mathbf{W} = \mathbf{D}^{1/2}\mathbf{C}_n{}^T\mathbf{Q} \tag{18}$$

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{C}_n\mathbf{D}^{-1/2} \tag{19}$$

At last, we can obtain some dominant ICs from Eq. (12). The extracted ICs reveal the majority of information and represent a meaningful representation about the observed data $\mathbf{x}$.

4. MODIFIED ICA FOR MONITORING

In the proposed monitoring method, two types of statistics are considered: the *D*-statistic to monitor the systematic part change of the process variation and the *Q*-statistic to monitor the residual part of the process variation. The *D*-statistic, also known as the Hotelling's T^2 statistic, is the Mahalanobis distance defined as follows:

$$T^2 = \mathbf{y}^T\mathbf{D}^{-1}\mathbf{y} \tag{20}$$

where $\mathbf{y}$ is obtained from Eq. (12) and $\mathbf{D}$ is the diagonal matrix of the eigenvalues associated with the retained dominant ICs. In this paper, kernel density estimation is used to define the control limit for T^2 because $\mathbf{y}$ is not Gaussian (Silverman, 1986; Martin and Morris, 1996; Lee *et al*., 2004).

The *Q*-statistic, also known as the *SPE* statistic is defined as follows:

$$SPE = \mathbf{e}^T\mathbf{e} = \left(\mathbf{x}-\hat{\mathbf{x}}\right)^T\left(\mathbf{x}-\hat{\mathbf{x}}\right) \tag{21}$$

where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ and $\hat{\mathbf{x}}$ can be calculated as follows:

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{W}\mathbf{x}. \tag{22}$$

If the number of ICs is chosen such that the majority of non-Gaussianity is included in the ICs, the residual subspace will contain mostly random noise which can be treated as normal distribution. The upper control limit of *SPE* can then be calculated from Jackson and Mudholkar (1979).

The contribution based approach is simple to identify faults and can be generated without prior knowledge (Qin, 2003). In the proposed method, the T^2 statistic can be decomposed as:

$$T^2 = \mathbf{y}^T\mathbf{S}^{-1}\mathbf{y} = \mathbf{y}^T\mathbf{S}^{-1}\mathbf{W}\mathbf{x} = \mathbf{y}^T\mathbf{S}^{-1}\sum_{j=1}^{m}\mathbf{w}_j x_j$$
$$= \sum_{j=1}^{m}\mathbf{y}^T\mathbf{S}^{-1}\mathbf{w}_j x_j = \sum_{j=1}^{m} c_j \quad (23)$$

Therefore, the contribution to the T^2 statistic for a data $\mathbf{x}$, is given as follows (Westerhuis *et al*., 2000):

$$c_j\left(T^2\right) = \mathbf{y}^T\mathbf{S}^{-1}\mathbf{w}_j x_j \quad (24)$$

where $c_j\left(T^2\right)$ is the contribution of the *j*-th variable to the T^2 statistic, x_j is the *j*-th element of $\mathbf{x}$, and $\mathbf{w}_j$ is the *j*-th row of the demixing matrix $\mathbf{W}$.

Similarly, the contribution of process variable *j* at given time to the *SPE* statistic is defined as follows:

$$c_j\left(SPE\right) = e_j{}^2 \quad (25)$$

where e_j is the *j*-th variable of $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$.

In this paper, the upper control limits for $c_j\left(T^2\right)$ are calculated as the mean of the contributions plus three standard deviations of the contributions for each process variable (Westerhuis *et al*., 2000). Control limits for $c_j\left(SPE\right)$ are calculated the same way as the Q-statistic control limit (Westerhuis *et al*., 2000).

5. CASE STUDY

In this section, the proposed method is applied to the Tennessee Eastman process simulation data and is compared with PCA monitoring results. The details on the process description are well explained in Chiang *et al*. (2001). A total of 33 variables listed in Table 1 are used for monitoring in this study. A sampling interval of 3 minutes was used to collect the simulated data. Both the training and testing data sets for each fault are composed of 960 observations. A set of programmed faults (Fault 1-21) is listed in Table 2. All faults in the test data set were introduced from sample 160. The data can be downloaded from http://brahms.scs.uiuc.edu (Chiang *et al*., 2001).

All the data were auto-scaled prior to the application of PCA and the modified ICA. In the modified ICA, 30 whitened vectors are extracted from Eq. (13) to update and find ICs. 9 PCs are selected for the PCA by cross-validation and the same number of ICs is chosen for fair comparison.

The false alarm rates and the fault detection rates of the two multivariate methods, PCA and modified ICA, for all 21 fault data were computed and tabulated in Table 3. For the data obtained after the fault occurrence, the percentage of the samples outside the 99% control limits was calculated in each simulation and termed as detection rate. Maximum detection rate achieved for each fault is marked with a bold number. With 9 PCs and 9 ICs, false alarm rates of PCA and modified ICA are comparable though they are different for each fault data. As shown in Table 3, the modified ICA can detect most faults more effectively than PCA except Fault 4 and 11. For Faults 10 and 16, the detection rate of the proposed method is more than twice as high as that of PCA, which shows that the modified ICA with acceptable false alarm rate can detect small events that are difficult to detect by PCA. One thing that needs to be noted is T^2 ability of the proposed method for detecting faults. For all cases, the detectability of T^2 is considerably enhanced by the proposed method. It means the proposed method can extract essential features in a process much more sensitively than PCA. This result demonstrates that the proposed method is expected to be more effective than PCA to diagnose fault patterns in the feature space.

The monitoring charts of PCA and modified ICA in the case of Fault 10 are shown in Fig. 1. PCA can detect the fault from about sample 200, however, there are lots of samples below the 99% control limit despite the presence of the fault. On the other hand, the modified ICA detects the fault earlier than PCA by 11 samples and gives a consistent fault alarm up to the end of the processing time. Also, the random pattern changes caused by the fault are reflected well in the proposed method. The results of this example indicate the proposed method has a superior capability in detecting faults that are difficult to detect by the conventional method. Fig. 2 shows contribution plots to T^2 and *SPE* at sample 195, respectively, in the case of Fault 10. From this figure, variables 16 (Stripper pressure) and 18 (Stripper temperature) make the largest contribution to the T^2 statistic while variables 19 (Stripper steam flow) and 31 (Stripper steam valve) give dominant effects on *SPE* statistic. This contribution plot correctly indicates the major variable groups affected by the fault. Thus, the fault detection and identification ability of the proposed method is much worthy of consideration.

Table 1 Variables in the Tennessee Eastman process

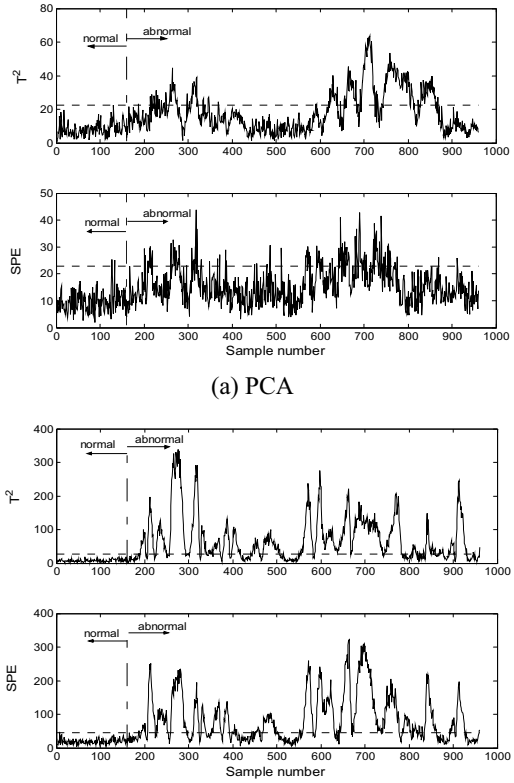
1	*A* feed	18	stripper temperature
2	*D* feed	19	stripper steam Flow
3	*E* feed	20	compressor work
4	total feed	21	reactor cooling water outlet temperature
5	recycle flow	22	separator cooling water outlet temperature
6	reactor feed rate	23	*D* feed flow valve
7	reactor pressure	24	*E* feed flow valve
8	reactor level	25	*A* feed flow valve
9	reactor temperature	26	total feed flow valve
10	purge rate	27	compressor recycle valve
11	product separator temperature	28	purge valve
12	product separator level	29	separator pot liquid flow valve
13	product separator pressure	30	stripper liquid product flow valve
14	product separator underflow	31	stripper steam valve
15	stripper level	32	reactor cooling water flow
16	stripper pressure	33	condenser cooling water flow
17	stripper underflow		

Table 2 List of Process faults for the Tennessee Eastman process

No.	Description	Type
1	*A*/*C* feed ratio, *B* composition constant	Step
2	*B* composition, *A*/*C* ratio constant)	Step
3	*D* feed temperature	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	*A* feed loss	Step
7	C header pressure loss - reduced availability	Step
8	*A*, *B*, *C* feed composition	Random
9	*D* feed temperature	Random
10	*C* feed temperature	Random
11	Reactor cooling water inlet temperature	Random
12	Condenser cooling water inlet temperature	Random
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16 ~ 20	Unknown	
21	The valve for Stream 4 was fixed at the steady state position	Constant Position

Table 3 Representative detection rates of PCA and modified ICA

Faults	False alarm rate				Detection rate			
	PCA		Modified ICA		PCA		Modified ICA	
	T^2	*SPE*	T^2	*SPE*	T^2	*SPE*	T^2	*SPE*
1	0	1.25	0	1.88	99	**100**	**100**	**100**
2	0.63	0	0	0.63	**98**	96	**98**	**98**
3	0.63	0.63	0	0	2	1	1	1
4	0.63	0	0	1.88	6	**100**	65	96
5	0.63	0	0	1.88	**24**	18	**24**	**24**
6	0.63	0	0	0	99	**100**	**100**	**100**
7	0	0.63	0	0.63	42	**100**	**100**	**100**
8	0	1.88	0	0	97	89	97	**98**
9	1.25	0.63	0.63	3.75	1	1	1	2
10	0.63	1.25	0	0.63	31	17	**70**	64
11	0.63	0	0	0	21	**72**	43	66
12	0.63	1.25	0	0	97	90	**98**	97
13	0	0.63	0	0	93	**95**	**95**	94
14	0	1.88	0	0.63	81	**100**	**100**	**100**
15	0.63	1.25	0.63	0	1	2	1	2
16	3.13	0.63	1.25	1.25	14	16	**76**	73
17	0	1.88	0	1.25	74	93	87	**94**
18	0	1.88	0.63	1.88	89	**90**	**90**	**90**
19	0	0	0	0	0	**29**	25	**29**
20	0	1.88	0	0	32	45	**70**	66
21	0	0	1.25	0.63	33	46	**54**	19



(a) PCA

(b) Modified ICA

Fig. 1. Monitoring charts of a) PCA and b) Modified ICA for Fault 10.

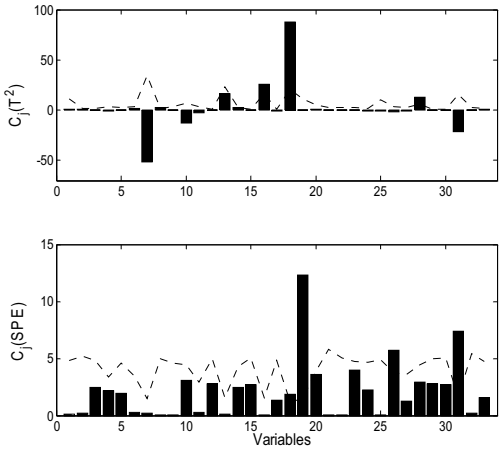


Fig. 2. Variables contribution plots to T^2 and *SPE* at sample 195 for Fault 10

6. CONCLUSION

This paper proposes a novel approach to process monitoring that uses modified ICA. Some problems of original ICA are analyzed and a modified ICA algorithm is developed and applied to MSPM. Compared to original ICA, the proposed algorithm has the following advantages: (1) It extracts a few dominant factors needed for process monitoring; (2) High computational load is attenuated by extracting a few dominant ICs, not all ICs; (3) The ordering of ICs is considered; (4) It gives a consistent solution.

The proposed method was applied to the fault detection and identification of Tennessee Eastman process. The fault detection performance was evaluated and compared with that of conventional PCA-based monitoring. This example demonstrates that the proposed method can detect various faults more efficiently than PCA. In particular, the extracted dominant ICs are expected to be more useful to diagnose fault patterns in the feature space. In addition, contribution plots of the proposed method can reveal the group of process variables responsible for the process to go out of control.

ACKNOWLEDGEMENT

This work was supported by the Korea Research Foundation Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund) (KRF-2005-214-D00027) and the Texas-Wisconsin Modeling and Control Consortium.

REFERENCES

Albazzaz, H., & Wang, X. Z. (2004). Statistical Process Control Charts for Batch Operations Based on Independent Component Analysis", *Industrial and Engineering Chemistry Research*, **43**(21), 6731-6741.

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2001). *Fault detection and diagnosis in industrial systems*, Springer, London

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, **10**, 626 -634.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, **13**(4-5), 411-430.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons, Inc., New York.

Jackson, J. E., & Mudholkar, G. S. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, **21**(3), 341-349.

Kocsor, A., & Tóth, L. (2004) Kernel-Based Feature Extraction with a Speech Technology Application. *IEEE Transactions on Signal Processing*, **52**(8), 2250-2263

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., & Ohno, H. (2003). Monitoring independent components for fault detection. *A.I.Ch.E. Journal*, **49**(4), 969-976.

Kano, M., Hasebe, S., Hashimoto, I., & Ohno, H. (2004). Evolution of multivariate statistical process control: independent component analysis and external analysis. *Computers & Chemical Engineering*, **28**(6-7), 1157-1166.

Kermit, M., & Tomic, O. (2003). Independent Component Analysis Applied on Gas Sensor Array Measurement Data. *IEEE Sensors Journal*, **3**(2), 218-228.

Lee, T. (1998). *Independent Component Analysis: Theory and Applications*. Kluwer Academic Publishers, Boston.

Lee, J.-M., Yoo, C. K., & Lee, I.-B. (2003). New monitoring technique with ICA algorithm in wastewater treatment process. *Water Science and Technology*, **47**(12), 49-56.

Lee, J.-M., Yoo, C. K., & Lee, I.-B. (2004). Statistical process monitoring with independent component analysis. *Journal of Process Control*, **14**, 467-485.

Martin, E. B., & Morris, A. J. (1996). Non-parametric confidence bounds for process performance monitoring charts. *Journal of Process Control*, **6**(6), 349-358.

Qin, S. J. (2003). Statistical Process Monitoring: Basics and Beyond. *Journal of Chemometrics*, **17**, 480-502 .

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, UK.

Westerhuis, J. A., Gurden, S. P., & Smilde, A. K. (2000). Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, **51**, 95-114.

Yoo, C.K., Lee, J.-M., Vanrolleghem, P. A., & Lee, I.-B. (2004). On-line Monitoring of Batch Processes Using Multiway Independent Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **71**, 151-163.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DETECTION AND DIAGNOSIS OF PLANT-WIDE OSCILLATIONS USING THE SPECTRAL ENVELOPE METHOD

Hailei Jiang [*] **M.A.A Shoukat Choudhury** [**]
Sirish L. Shah [*,1] **John W. Cox** [***]
Michael A. Paulonis [***]

[*] *Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada, T6G 2G6*
[**] *Matrikon Inc., Edmonton, AB, Canada, T5J 3N4*
[***] *Eastman Chemical Company, PO Box 431, Kingsport, TN 37662, USA*

Abstract: Plant-wide oscillations are common in many processes. Their effects propagate to many units and may impact the overall process performance. It is important to detect and diagnose the oscillations early in order to rectify the situation. This paper proposes a new procedure to detect and diagnose plant-wide oscillations. A technique called spectral envelope is used to detect the oscillations. Two kinds of plots - scaling and power plots - are proposed to identify the variables exhibiting common oscillation(s). These plots are also useful in isolating the key variables as the candidates of the root cause. An industrial case study is presented to demonstrate the applicability of the proposed procedure.

Keywords: Chemical industry, Oscillation, Fault Diagnosis, Power Spectrum, Control loop, Process Monitoring

1. INTRODUCTION

Detection and diagnosis of plant-wide disturbances is an important issue in many process industries (Qin, 1998). Oscillations are a common type of plant-wide disturbance whose effects propagate to many units and thus may impact the overall process performance. Increasing emphasis on plant safety and profitability strongly motivates the search for techniques to detect and diagnose plant-wide oscillations. Thornhill and Hägglund (1997) used the zero-crossings of the control error signal to calculate integral absolute error (IAE) in order to detect oscillation in a control loop. This method has poor performance for noisy error signals. Miao and Seborg (1999) suggested a method based on the auto-correlation function to detect excessively oscillatory feedback loop. The auto-covariance function (ACF) of a signal was utilized in Thornhill *et al.* (2003*a*) to detect oscillation(s) present in a signal. This method needs a minimum of five cycles in the auto-covariance function to detect oscillation, which is often hard to obtain, particularly in the case of a long oscillation (e.g., an oscillation with a period of 400 samples). Although the data set can be downsampled in such cases, downsampling may introduce aliasing in the data. Thornhill *et al.* (2002) proposed to perform spectral principal component analysis (SPCA) to detect oscillations and categorize the variables having similar oscillations. This method does not provide any diagnosis of the root cause of the oscillation which is generally the main objective of the exercise.

[1] Corresponding author. Tel.:+1-780-492-5162; fax:+1-780-492-2881.
E-mail address: sirish.shah@ualberta.ca

In this paper, a new procedure based on the spectral envelope method for detection and diagnosis of common oscillation(s) is proposed. The spectral envelope method needs neither a minimum number of cycles to be present in the signal nor filtering of the data to detect multiple oscillations. In terms of grouping the variables with common oscillation(s), the proposed procedure is more sensitive to oscillations and has a better resolution in identifying the variables oscillating at the same frequencies than the commonly used SPCA method. Furthermore, the proposed procedure can also deliver useful information about the root cause of common oscillation(s).

2. OSCILLATION DETECTION

In this section, the concept of spectral envelope is introduced. A simulation example is presented to demonstrate its ability to detect multiple oscillations. The performance comparison with SPCA method is also included.

2.1 Definition of the Spectral Envelope

The concept of spectral envelope was first proposed by Stoffer *et al.* (1993) to explore the periodic nature of categorical time series. In 1997, McDougall *et al.* (1997) extended the concept of spectral envelope to real-valued series. Here we provide an easy interpretation of the concept of spectral envelope.

Let $X(t) = [x_1(t), x_2(t), \cdots, x_m(t)]^T, t = 0, \pm 1, ...,$ be a vector-valued time series on $\Re^m$. $x_i(t), 1 \leq i \leq m$, is a univariate time series which can be a sequence of observations of a process variable. Denote the covariance matrix of $X(t)$ as V_X and the power spectral density matrix of $X(t)$ as $P_X(\omega)$. Here, ω represents frequency and is measured in cycles per unit time, for $-1/2 < \omega \leq 1/2$.

Let $g(t, \beta) = \beta^* X(t)$ be a scaled series from $\Re^m$ to $\Re$, where β is a column vector which may be real or complex. The * means complex conjugate. $g(t, \beta)$ is actually a linear combination of the rows of $X(t)$. Then the variance $V_g(\beta)$ of $g(t, \beta)$ can be expressed as $V_g(\beta) = \beta^* V_X \beta$, and the power spectral density $P_g(\omega, \beta)$ of $g(t, \beta)$ can be expressed as $P_g(\omega, \beta) = \beta^* P_X(\omega)\beta$.

The spectral envelope of $X(t)$ is defined to be:

$$\lambda(\omega) \triangleq \sup_{\beta \neq 0} \{\frac{P_g(\omega, \beta)}{V_g(\beta)}\} = \sup_{\beta \neq 0} \{\frac{\beta^* P_X(\omega)\beta}{\beta^* V_X \beta}\} \quad (1)$$

where $-\frac{1}{2} < \omega \leq \frac{1}{2}$ and the relationship between $P_g(\omega, \beta)$ and $V_g(\beta)$ is $V_g(\beta) = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_g(\omega, \beta) d\omega$.

The quantity $\lambda(\omega)$ represents the largest proportion of the power (or variance) that can be obtained at the frequency ω for any scaled series. The scaling vector that results in the value $\lambda(\omega)$ is called the optimal scaling vector at frequency ω, which is denoted as $\beta(\omega)$. Accordingly, the elements of the optimal scaling vector are called the optimal scalings. The optimal scaling vector $\beta(\omega)$ is not the same for all ω.

We prefer to limit β to the constraint, $\beta^* V_X \beta = 1$. Therefore the scaled series $g(t, \beta)$ is unit variance. This will make the calculated spectral envelope more interpretable and make the magnitude of the elements of $\beta(\omega)$ more comparable. Accordingly, the quantity $\lambda(\omega)$ represents the largest power(variance) that can be obtained at the frequency ω for any scaled series with unit variance.

With the optimal scaling vector $\beta(\omega)$, equation (1) can be rewritten as:

$$\lambda(\omega) V_X \beta(\omega) = P_X(\omega)\beta(\omega) \quad (2)$$

It follows that $\lambda(\omega)$ is the largest eigenvalue associated with the determinant equation:

$$|P_X(\omega) - \lambda(\omega) V_X| = 0 \quad (3)$$

$\beta(\omega)$ is the corresponding eigenvector satisfying equation (2).

2.2 Another Definition of the Spectral Envelope

Denote $V = diag(V_X)$, which only has the diagonal elements of V_X. We can use V instead of V_X in equation (1) and have a new expression for $\lambda(\omega)$:

$$\lambda(\omega) = \sup_{\beta \neq 0} \{\frac{\beta^* P_X(\omega)\beta}{\beta^* V \beta}\} \quad (4)$$

The resulting $\lambda(\omega)$ and $\beta(\omega)$ is different from those in the equation (1). Only under the condition that $\{x_1(t), x_2(t), ..., x_m(t)\}$ are mutually independent, V is equal to V_X and equation (4) is the same as equation (1).

We also prefer to limit β to the constraint such that $\beta^* V \beta = 1$, but this will not guarantee that the scaled series $g(t, \beta)$ is unit variance, except under the condition mentioned above.

2.3 Simulation Example

The following simulation example demonstrates the superiority of the performance of the spectral envelope method over the power spectrum and the SPCA method in detecting oscillation(s) in signals highly corrupted with noise.

2.3.1. Time Series Generation The example consists of 12 time series generated with various sinusoid oscillations. In these time series, $\varepsilon(t)$ is a white noise sequence with unit variance and $t = 1, ..., 512$.

The first four time series are corrupted by colored noise and have base oscillation at frequency $\omega_1 = 0.1Hz$:

$$x_1(t) = 0.8cos(2\pi\omega_1 t) + 2\varepsilon(t) + \varepsilon(t-1)$$
$$x_2(t) = 0.6cos[2\pi\omega_1(t-5)] + 2\varepsilon(t) + \varepsilon(t-1)$$
$$x_3(t) = 0.4cos[2\pi\omega_1(t-15)] + 2\varepsilon(t) + \varepsilon(t-1)$$
$$x_4(t) = 0.2cos[2\pi\omega_1(t-2)] + 2\varepsilon(t) + \varepsilon(t-1)$$

The next four time series are corrupted by colored noise and have base oscillation at frequency $\omega_2 = 0.3Hz$:

$$x_5(t) = 0.9cos(2\pi\omega_2 t) + 2\varepsilon(t) - \varepsilon(t-1)$$
$$x_6(t) = 0.7cos[2\pi\omega_2(t-7)] + 2\varepsilon(t) - \varepsilon(t-1)$$
$$x_7(t) = 0.5cos[2\pi\omega_2(t-10)] + 2\varepsilon(t) - \varepsilon(t-1)$$
$$x_8(t) = 0.3cos[2\pi\omega_2(t-20)] + 2\varepsilon(t) - \varepsilon(t-1)$$

The next two time series have oscillations at both frequencies $\omega_1 = 0.1Hz$ and $\omega_2 = 0.3Hz$:

$$x_9(t) = 0.4cos[2\pi\omega_1(t-6)] + 0.5cos[2\pi\omega_2(t-8)] + 2\varepsilon(t) + \varepsilon(t-1)$$
$$x_{10}(t) = 0.8cos[2\pi\omega_1(t-16)] + 0.6cos[2\pi\omega_2(t-4)] + 2\varepsilon(t) - \varepsilon(t-1)$$

The last two time series are simple colored noise sequences in a form of moving average:

$$x_{11}(t) = \varepsilon(t) + 0.5\varepsilon(t-1)$$
$$x_{12}(t) = \varepsilon(t) - 0.5\varepsilon(t-1)$$

Before doing further analysis, all the time series are normalized to be zero-mean and unit variance. Figure 1 shows the time trends and power spectra of the 12 time series. As shown in Figure 1, the signals are highly corrupted with noise. The power spectra of the time series do not highlight any oscillations at $0.1Hz$ or $0.3Hz$.

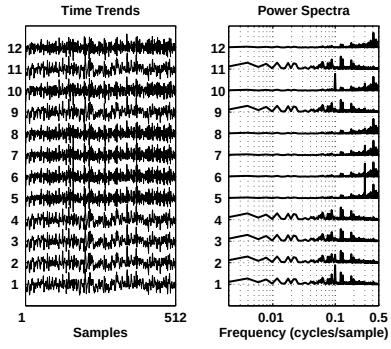


Fig. 1. Time trends and power spectral of the 12 time series

2.3.2. SPCA Analysis Figure 2 shows the first two principle components (PCs) plot. These two PCs explain over 95% of the variability of the spectra. However, these two PCs do not clearly indicate any oscillation at $0.1Hz$ or $0.3Hz$. Further clustering based on these two PCs could not give any useful information about the two oscillations as well. Therefore, SPCA fails to detect the oscillations at frequencies $0.1Hz$ and $0.3Hz$.

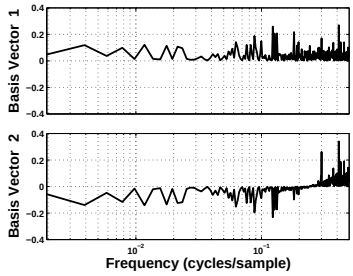


Fig. 2. SPCA PCs plot of the 12 time series

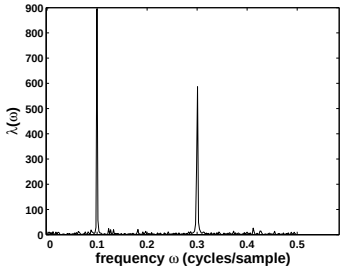


Fig. 3. Spectral Envelope of the 12 time series

2.3.3. Oscillation Detection Using the Spectral Envelope Method Figure 3 shows the spectral envelope calculated by equation (1) with the constraint $\beta^* V_X \beta = 1$. There are two significant peaks at $0.1Hz$ and $0.3Hz$, which means that the scaled series could have much more energy at these two frequencies than any other frequencies. It further implies that some of (or all of) the 12 time series may have significant energy at $0.1Hz$ and $0.3Hz$. Therefore, it can be concluded that the spectral envelope can clearly detect the multiple oscillations present in the time series.

3. VARIABLE CATEGORIZATION

After detecting the oscillation(s), the next step is to group the variables oscillating together at a common frequency. Here we propose two plots, a scaling plot and a power plot, to perform this task.

3.1 Scaling Plot

The first proposed plot is called the scaling plot, which is the bar plot of the magnitude of the optimal scalings calculated by equation (4) at the oscillation frequency in a descending sequence. The variables that have large scaling magnitudes at a oscillation frequency are the ones contributing most to the spectral envelope peak at that frequency, and thus are the ones participating in the oscillation. For example, Figure 4 is the scaling plot of the 12 time series (of the example in section

2.3) at the frequency $0.1Hz$. This plot clearly identifies that the time series 1, 10, 2, 9, 3 and 4 have oscillation at $0.1Hz$. The scaling plot of the 12 time series at the frequency $0.3Hz$ can also identify the variables oscillating at $0.3Hz$. Due to lack of space, we do not present it here.

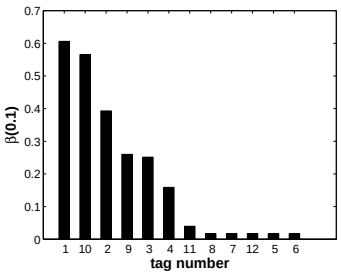


Fig. 4. Scaling plot of the 12 time series at $0.1Hz$

3.2 Power Plot

Another proposed plot is called the power plot, which is the bar plot of the power of each variable at the oscillation frequency in a descending sequence. The variables that have significant energy at the oscillation frequency are definitely the ones participating in the oscillation. For instance, Figure 5 is the power plot of the 12 time series (of the example in section 2.3) at the oscillation frequency $0.1Hz$. This plot clearly identifies that the time series 1, 10, 2, 9, 3 and 4 have oscillation at $0.1Hz$ since they have much more energy than the other time series at this frequency. The power plot of the 12 time series at the frequency $0.3Hz$ can also identify the variables oscillating at $0.3Hz$. Due to lack of space, we do not present it here.

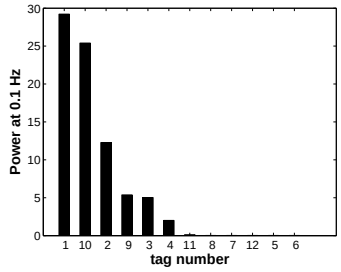


Fig. 5. Power plot of the 12 time series at $0.1Hz$

3.3 Comparison of Scaling Plot and Power Plot

Comparison between the scaling plot and the power plot reveals that the tag numbers of the variables appear in the same order in these two plots (see Figures 4 and 5). In other words, the series that has more energy at the oscillation frequency will have larger scaling magnitude in the optimal scaling vector. Therefore, these two plots are interchangeable in identifying and categorizing the variables at the oscillation frequency.

4. ROOT CAUSE DIAGNOSIS

Root cause diagnosis is a challenging problem in the area of detection and diagnosis of plant-wide oscillations. The main contribution of current data based root cause diagnosis techniques (Thornhill *et al.*, 2003*b*) is to isolate the few key variables as the candidates of the root cause, or at least identify those variables close to the root cause. This will reduce the workload and the cost of plant test to determine the real root cause.

Power plot (or scaling plot) may also be used to serve the same purpose. The main idea is to pick up the first few variables that contribute the most energy at the oscillation frequency as the key variables. The root cause probably lies within these few variables. The reason is that chemical processes are usually low pass filters. The process gain typically decreases as the frequency increases, as observed in most Bode plots. Therefore, as the oscillation propagates through different control loops, the energy at the oscillation frequency will decrease because of the low pass filtering effect of the chemical processes. The variables close to the root cause should exhibit more energy at the oscillation frequency than the other variables. Thus, we take the few variables that contribute the most energy at the oscillation frequency as the candidates of the root cause. The industrial case study in a later section will demonstrate the efficiency of using this idea to isolate the key variables.

5. NEW PROCEDURE TO DETECT AND DIAGNOSE PLANT-WIDE OSCILLATIONS

The following steps in a new procedure to detect and diagnose plant-wide oscillation are proposed:

- I. Normalize the data matrix that each variable is zero-mean and unit variance;
- II. Calculate the spectral envelope using equation (1) or (4) to find out the major oscillation frequencies;
- III. Use power plot (or scaling plot) at those oscillation frequencies identified in step II to categorize the variables having similar oscillations.
- IV. Use power plot (or scaling plot) to isolate the key variables having significant oscillations. The root cause probably lies within these variables.

6. AN INDUSTRIAL CASE STUDY

An industrial data set was provided courtesy of the Advanced Controls Technology group of Eastman Chemical Company. Figure 6 shows the process schematic of the plant. The Advanced Controls Technology group had identified a need

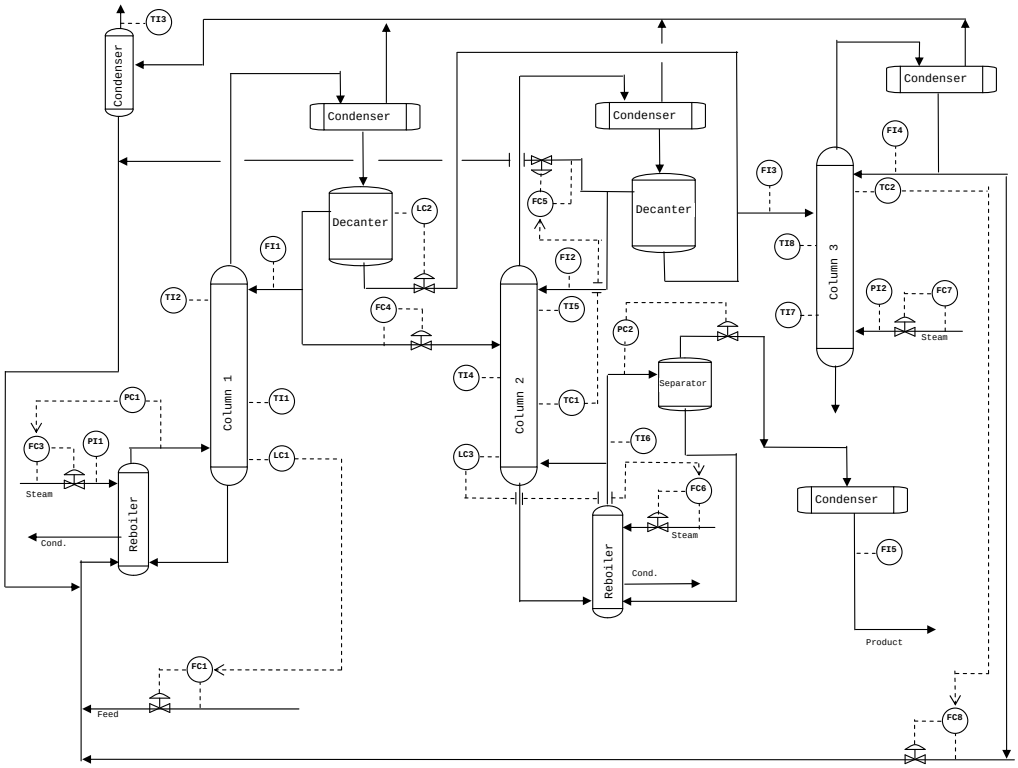


Fig. 6. Process schematic

for diagnosis of a common disturbance with an oscillation period of about 2 hours. In this section, the newly proposed procedure is applied to this data set to demonstrate its efficiency in detection and diagnosis of the plant-wide disturbance.

6.1 Data Description

The provided data set contains 48 variables: 14 process variables (*pv*'s), 14 controller outputs (*op*'s), 15 indicator variables and 5 cascade loop setpoints (*sp*'s). Each variable has 8640 observations with a sample interval 20s, which corresponds to data over 2 days of operation.

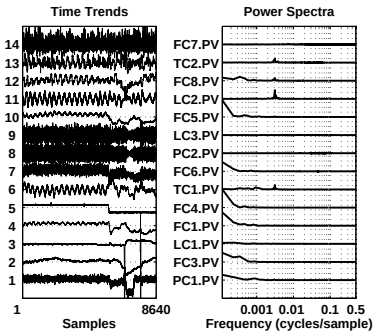


Fig. 7. Time trend and power spectra of 14 *pv*'s

Figure 7 shows the time trends and power spectra of the first 14 *pv* variables. The power spectra indicate the presence of oscillation at the frequency 0.003 cycles/sample (or about 333 samples/cycle, nearly a period of 2 hours). This oscillation affects many variables in the process and is considered as a plant-wide oscillation.

6.2 SPCA Analysis

The first two principle components (PCs) explained 87.47% variability of the spectra. The second PC has a small peak around the frequency 0.003 cycles/sample which indicates the interesting oscillation. However, the two-dimensional scores plot has no meaningful clustering. It is hard to analyze the frequency features of each variable. To save space, we do not present the PC and score plots here.

6.3 New Analysis Procedure

6.3.1. Oscillation Detection Figure 8 shows the spectral envelope (from equation (4)) of the 48 variables. In the spectral envelope, there is clear low frequency features. This is probably because the data is from a long term operation and there exists extremely long period influences like diurnal weather effects that impact the process. Beside the low frequency feature, there is a clear peak at the frequency of 0.0031 cycles/sample, indicating a oscillation with a period of 320 samples/cycle, or approximately 1.78 hours/cycle. This is exactly the oscillation that the Advanced Controls Technology group of Eastman Chemical Company wanted to detect and diagnose.

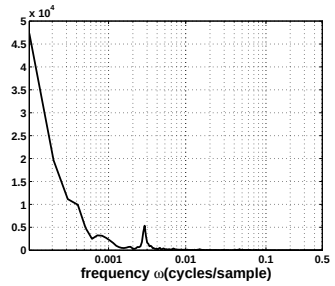


Fig. 8. Spectral Envelope of the 48 variables

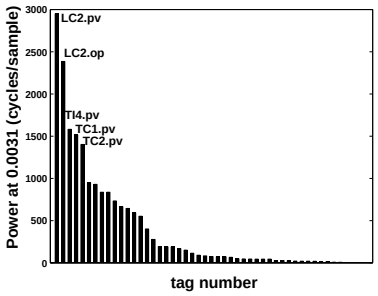


Fig. 9. Power plot of the 48 variables at the oscillation frequency 0.0031 cycles/sample

6.3.2. Variable Categorization

Figure 9 is the power plot of the 48 variables at the frequency 0.0031 cycles/sample. To make the figure clear, we only show the tag numbers of the five variables that contribute the highest energy at this frequency. They are the key variables that can be taken as the candidates of the root cause. Besides these five variables, the plot also clearly shows that many other variables are affected by this oscillation.

6.3.3. Oscillation Diagnosis Among all the variables, the *pv* and *op* of the control loop LC2 have the biggest energy at the oscillation frequency. This result indicates that the oscillation in loop LC2 is most severe and we should take this loop as the first candidate of the root cause.

Actually, the control loop LC2 was exactly the root cause found out by Thornhill *et al.* (2003*b*). It was reported that the control valve of the loop LC2 had a deadband of 4% and it was the root cause. For more information, refer to Thornhill *et al.* (2003*b*).

7. CONCLUSIONS

In this paper, the concept of spectral envelope is modified such that it is easy to apply for detecting plant-wide oscillations. This method is good at detecting oscillations whether single or multiple. Also the calculation of the spectral envelope is straightforward and no calculation parameter needs to be specified. In comparison to the ACF based method, the spectral envelope method does not suffer any limitation on minimum number of oscillation cycles and it does not require designing any filter. It can detect all oscillations in one step.

Scaling and power plots have been proposed for the purpose of grouping the variables participating in the common oscillation(s). The proposed plots can also deliver useful information about the root cause of a plant-wide oscillation.

Finally, an industrial case study was presented to demonstrate the efficacy of the method.

8. ACKNOWLEDGEMENT

The authors are grateful for the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC), Matrikon Inc. and the Alberta Science and Research Authority (ASRA), through the NSERC-Matrikon-ASRA Senior Industrial Research Chair program at the University of Alberta. The authors also would like to thank Dr. Bhushan Gopaluni for pointing out the papers by Stoffer and co-workers for this study.

REFERENCES

McDougall, A.J., D.S. Stoffer and D.E. Tyler (1997). Optimal transformations and the spectral envelop for real-valued time series. *Journal of Statistical Planning and Inference* **57**, 195–214.

Miao, T. and D.E. Seborg (1999). Automatic detection of excessive oscillatory feedback control loops. In: *Proc. of IEEE international Conference on Control Applications*. Kohala Coast, Hawai'i.

Qin, S.J. (1998). Control performance monitoring - a review and assessment. *Computer and Chemical Engineering* **23**, 173–186.

Stoffer, David S., David E. Tyler and Andrew J. McDougall (1993). Spectral analysis for categorical time series: Scaling and spectral envelope. *Biometrika* **80**, 611–622.

Thornhill, N.F. and T. Hägglund (1997). Detection and diagnosis of oscillation in control loops. *Control Engineering Practice*.

Thornhill, N.F., B. Huang and H. Zhang (2003*a*). Detection of multiple oscillations in control loops. *Journal of Process Control* **13**, 91–100.

Thornhill, N.F., John W. Cox and Michael A. Paulonis (2003*b*). Diagnosis of plant-wide oscillation through data-driven analysis and process understanding. *Control Engineering Practice* **11**, 1481–1490.

Thornhill, N.F., S.L. Shah, B. Huang and A. Vishnubhotla (2002). Spectral principal component analysis of dynamic process data. *Control Engineering Practice* **10**, 833–846.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

DETECTION OF PLANT-WIDE DISTURBANCES USING A SPECTRAL CLASSIFICATION TREE

Nina F. Thornhill* and Hallgeir Melbø+

**Department of Electronic and Electrical Engineering, UCL, UK*
+ABB Corporate Research Centre, Billingstad, Norway

Abstract: This article demonstrates the use of agglomerative hierarchical clustering to detect the structure within a data set. When combined with spectral principal component analysis to capture the main spectral features of a data set it allows visualization of the structure of a model with an optimum number of principal components. The paper presents the theory and methods for construction of the tree and gives an example using industrial data. *Copyright © 2006 IFAC.*

Keywords: Clustering; fault detection; hierarchical classification; performance analysis; plant-wide; principal component analysis; process operation; process monitoring.

1. INTRODUCTION

Large data bases are being accumulated by companies operating oil, gas and chemical processes. When these data are used in a plant audit, the aim is to find groups of measurements having similar characteristics so that the propagation paths of disturbances can be tracked through a process. With large data sets, however, it becomes challenging to present the results of a multivariate analysis. While principal component analysis (PCA) might reduce several hundred measurements to, say, ten principal components there then remains an issue of presentation of the ten-dimensional model to the analyst. This paper demonstrates a method for visualization of the clusters in a high-dimensional spectral PCA model by means of a hierarchical classification tree. The key elements in the procedure are an agglomerative hierarchical clustering algorithm and a recursive algorithm to create the tree.

Multivariate analysis using spectral methods has recently been reported for the case of dynamic disturbances (Thornhill *et. al.*, 2002; Xia and Howell, 2005). The benefits for plant audit application are that the power spectra are insensitive to time delays or phase lags between different measurement points and therefore bypass the need for time shifting and other methods needed for correlation-based analysis in the time domain.

The next section of the paper gives a review of related work while Section 3 gives the formulation of spectral multivariate data analysis. A distance measure for clustering analysis is also discussed in Section 3 together with the automated algorithm for creation of the hierarchical classification tree. An industrial data set is then analyzed to illustrate the concepts showing the clustering patterns present before and after a plant shutdown in which maintenance was carried out. The paper ends with a conclusion section.

2. BACKGROUND AND CONTEXT

2.1 PCA for cluster analysis

Descriptions of principal component analysis may be found from many sources, for example Chatfield and Collins (1980) and Wold *et. al.* (1987). In analytical chemistry, near infrared (NIR) and nuclear magnetic resonance (NMR) spectroscopy data are routinely analysed by PCA (Alam and Alam, 2005; Ozaki et. al., 2001) and Seasholtz (1999) described the industrial application of multivariate calibration in NIR and NMR spectroscopy at Dow Chemical Company. Principal component analysis has proved useful in other diverse areas such as paint colour analysis (Tzeng and Berns, 2005), in the analysis of the relationship between the crispness of apples and recorded chewing sounds (De Belie *et al*, 2000) and

in water quality analysis (e.g. Brodnjak-Voncina *et al*, 2002). Industrial uses include principal component analysis for monitoring of machinery and process equipment (Wu *et. al*. 1999; Malhi and Gao, 2004; Flaten *et al*., 2005). All these applications have the common aim to discover structure within the data set, to ascertain the items within the data set that belong together and to relate the results to underlying mechanisms. They are normally run off-line.

A prevalent area in process monitoring [Wise *et al*., 1990; Kresta *et. al*., 1991; Wise and Gallagher, 1996; Qin, 2003] is on-line multivariate statistical process control in which new measurements are projected into a PCA calibration model that was developed during normal operation. Multivariate warning and alarm limits are set which test whether a new set of measurements is within the normal bounds captured by the calibration model [Jackson and Mudholkar, 1979; Martin and Morris, 1996].

The work in this paper is aimed towards the first type of application and concerns the visualization of structures in a data set and ascertains the items that belong together, where the *items* are the power spectra of time trends at a given measurement point.

2.2 Visualization of high dimension PCA models

The visualization of a high dimension multivariate PCA model has previously been examined by Wang *et. al.*, (2004) for the purposes of a multivariate statistical process monitoring. They used parallel coordinates to display multiple dimensions of the score space. Each day of running in a data set was represented by one piecewise linear trend in the parallel coordinate plot, and these trends were overlaid on top of one another. Although it was not possible to see any structure within the plot it was possible to identify abnormal days of running by inspection of outliers in the parallel coordinates plot.

2.3 Hierarchical classification

Gordon (1987) gave a comprehensive review of hierarchical classification, distinguishing between agglomerative and divisive methods. Agglomerative hierarchical clustering is an unsupervised algorithm for building up groups of similar items from a population of individual items. The basic algorithm (Duda et. al., 2000) starts with *N* clusters each containing one item and proceeds as follows:

repeat

find the pair of nearest clusters

merge them into one cluster

until there is one cluster containing *N* items

The results of agglomerative hierarchical clustering may be visualized in a classification tree in which the items of interest are the leaves on the tree and are joined into the main tree and eventually to the root of the tree by branches. Industrial applications of clustering and/or classification trees have included methods for office buildings to detect days of the week with similar profiles of energy use (Seem, 2005), the presentation of results from an end-point detection method in a crystallization process (Norris *et. al*., 1997) and from analysis of illegal adulteration of gasoline with organic solvents (Wiedemann *et. al*., 2005). The method presented in this paper uses agglomerative classification in the score space of all significant principal components.

Classification trees are also used in divisive classification in which a large group of items is recursively split into subcategories. In the area of process analysis, divisive classification has been combined with PCA for detection of key factors that affect process performance in a blast furnace and a hot stove system generating hot air for the blast furnace (Lee *et. al*, 2004). Clusters of items appearing in the score space of the first two principal components were identified and then further divided into sub-clusters using PCA recursively.

3. METHODS

3.1 Spectral PCA

In spectral principal component analysis (PCA) (Thornhill *et. al*., 2002) the rows of the data matrix $\mathbf{X}$ are normalized power spectra $P(f)$:

$$N \text{ frequency channels } \rightarrow$$

$$\mathbf{X} = \begin{pmatrix} x_1(f_1) & \dots & x_1(f_N) \\ .. & .. & .. \\ x_m(f_1) & \dots & x_m(f_N) \end{pmatrix} \begin{matrix} m \\ \text{measurements} \\ \downarrow \end{matrix}$$

A PCA decomposition reconstructs the $\mathbf{X}$ matrix as a sum over *p* orthonormal basis functions $\mathbf{w}'_1$ to $\mathbf{w}'_p$ which are spectrum-like functions each having N frequency channels arranged as a row vector:

$$\mathbf{X} = \begin{pmatrix} t_{1,1} \\ \dots \\ t_{m,1} \end{pmatrix} \mathbf{w}'_1 + \begin{pmatrix} t_{1,2} \\ \dots \\ t_{m,2} \end{pmatrix} \mathbf{w}'_2 + \dots + \begin{pmatrix} t_{1,p} \\ \dots \\ t_{m,p} \end{pmatrix} \mathbf{w}'_p + \mathbf{E}$$

The *i*'th spectrum in $\mathbf{X}$ maps to a spot having the co-ordinates $t_{i,1}$ to $t_{i,p}$ in a *p*-dimensional space. The $t_{i,1}$ to $t_{i,p}$ are called scores and represent the weightings of the basis functions needed to approximately reconstruct the spectrum in the *i*'th row of the data matrix. Similar spectra have similar *t*-coordinates and form clusters in the score space.

The key to finding meaningful clusters is the choice of distance measure. In process performance analysis the angular measure discussed in Duda *et. al*., (2000) is often more suitable than Euclidian distances. The reason for this observation is that the PCA clusters frequently take the form of plumes radiating from the origin. Raich and Cinar (1997) also observed plumes in their analysis of simulated faults in the Tennessee Eastman benchmark model.

Let he vector $\mathbf{t}'_i = \left(t_{i,1},\ t_{i,2}, \dots\ t_{i,p} \right)$ be the *i*'th row of matrix $\mathbf{T}_p$ in $\mathbf{X} = \mathbf{T}_p \mathbf{W}_p^{\mathrm{T}} + \mathbf{E}$ in a *p* principal

component model. A measure for membership of a plume is that the direction of vector $\mathbf{t}'_i$ in the multidimensional score plot lies within the same solid angle as those of other $\mathbf{t}'$ – vectors belonging the plume. The angle between $\mathbf{t}'_i$ and $\mathbf{t}'_j$ may be determined through calculation of the scalar product:

$$\cos\left(\theta_{i,j}\right) = \frac{\mathbf{t}'_i \bullet \mathbf{t}'_j}{\left|\mathbf{t}'_i\right|\left|\mathbf{t}'_j\right|}$$

where:

$$\mathbf{t}'_i \bullet \mathbf{t}'_j = \sum_{k=1}^{p} t_{i,k}\, t_{j,k} \quad \text{and} \quad \left|\mathbf{t}'_i\right| = \sqrt{\sum_{k=1}^{p} t_{i,k}{}^2}$$

3.2 Clustering

A matrix $\mathbf{A}$, whose elements are $\theta_{i,j}$, is to be analyzed to find high-dimensional plumes in the PCA score plot. Two items in the score plot whose $\mathbf{t}'$ – vectors point in similar directions give a small value of $\theta_{i,j}$. The agglomerative hierarchical clustering algorithm is based on Chatfield and Collins (1980):

Algorithm: Agglomerative classification

Step 1: The starting point is the matrix of angular distances with elements $\theta_{i,j}$. A text vector of row and column headings is also defined which initially is $(1\ \ 2\ \ 3\ \ 4\ \ 5\ \ ....)$ to keep track of the items in the data set. For an process performance analysis application the items are the *N* plant profiles in the data set, for a process audit the items are the *m* tags.

Step 2: At the *k*'th iteration, the smallest non-zero value $\theta_{i,j}$ in the matrix is identified. Its row and column indexes *i* and *j* indicate the smallest angular separation and these are clustered together.

Step 3: A smaller matrix $\mathbf{A_k}$ is then generated from the original. It does not have rows and columns for the two similar items identified at step 2. Instead, it has one row and column that give the distances of all the other items from the cluster. The distances are $\min\left\{\theta_{i,n}, \theta_{j,n}\right\}$, i.e. the angular distance between the *n*'th item and whichever member of the cluster was closer. For instance, if $\theta_{9,15}$ is the smallest angular separation in the matrix then rows 9 and 15 would be deleted and replaced by a new single row, and likewise for columns 9 and 15.

Step 4: The row and column headings are redefined. The heading for the new row created at step 3 indicates the items that have been combined. For instance, if the smallest angular separation at Step 3 had been $\theta_{9,15}$ then the new heading would be (9 15).

Step 5: The results of the *k*'th step are written to a report showing the cluster size defined as the maximum distance between items in the cluster, the row heading for the cluster formed at iteration *k*, and the two sub-clusters within it.

Step 6: Steps 2 to 5 are repeated until all the items have been clustered. At any stage, the outcome of the next step is either another item added to a cluster already identified or the combining of two items to start a new cluster.

A feature of the agglomerative hierarchical classification procedure presented here is that it provides a text-based report which enables the detection of significant clusters as well as automated generation of the hierarchical tree plot.

3.3 Dealing with noise

There is an assumption underlying a process audit which is that any tag whose power spectrum has spectral features is being upset by unwanted dynamics which could be reduced by control action. The assumption is justified in a control systems study where the idea is that nothing but random noise should be present. The spectrum of random noise is broad band and in theory it is flat. Such a spectrum maps to the origin in spectral PCA (i.e. the elements in $\mathbf{t}'_i = \left(t_{i,1},\ t_{i,2}, \ldots\ t_{i,p}\right)$ are close to zero) because the p $\mathbf{w}'$ – vectors of the model reflect the spectral features in the data set. The broadband noise is captured by the remaining $m - p$ components and appears in the $\mathbf{E}$ matrix. Tags with small values of $\left\|\mathbf{t}'\right\|$ are therefore excluded from the hierarchical tree. In the case study presented below, the tags with the 90% longest $\mathbf{t}'$ – vectors were plotted. Therefore, out of 60 tags, 54 appear in the tree and the excluded ones are classified as broad-band noise.

3.4 Plotting of the hierarchical tree

The graphical representation of the hierarchical tree can be extracted from the report generated by the algorithm of Section 3.2. It utilises an algorithm which starts at the top and systematically searches down the left and then the right branches and sub-branches to parse the structure of the tree. The algorithm is recursive meaning it calls itself over and over again in a nested way until it reaches a leaf of the tree. The end result is a set of $x-$ and $y-$ coordinates tracing the path that joins each individual item on the horizontal axis to the master node at the top of the tree.

Algorithm: Path Search

At the current node,

Step 1: Search left if the next node to the left is not done
- find description of the next node to the left
- if the next node to the left is a leaf of the tree
  - set *label* equal to the item number
  - mark the path to that leaf as *done*
  - return (back to the next highest level of recursion)
- else if the next node to the left is not done yet
  - call Path Search (recursive call)
  - build the path by adding the y-coordinate of the node to the path (the path starts empty)
- else
  - mark the left node to the left as done.

Step 2: Search right if the next node to the right is not done:
- find the next node to the right
- If the next node to the right is a leaf of the tree
  - set *label* equal to the item number
  - mark the path to that leaf as *done*
  - return (back to the next highest level of recursion)

else if the next node to the right is not done yet
call Path Search (recursive call)
build the path by adding the y-coordinate of the node to the path (the path starts empty)
else
mark the node to the right as *done*
mark the current node as *done*.

Step 3: Plot paths for each leaf as a stairs plot to construct the tree from the leaf the tree

The result is a set of paths, one for each leaf of the tree. These paths may be plotted as stair plots to construct the tree.

Some nodes in the classification tree have more than two sub-branches, for example Tags 2, 9 and a subcluster including 21, 4, and others are joined by a single horizontal line at 7.6 degrees in the middle of Figure 3. The reason is that the overall maximum distance between items in the cluster does not always grow when a new item is added. Items join growing clusters in turn according to their distance from the nearest item that is already in the cluster. However their inclusion does not necessarily make the overall size of the cluster larger because items already in the cluster may be further apart from each other than they are from the new item.

4. CASE STUDY

4.1 Data sets

The aim of the case study is to use the hierarchical tree derived from spectral PCA to aid and evaluate the maintenance activity. The mean centred time trends of the data set for the case study are shown in Figure 1 and the spectra are in Figure 2. There are two panels for each because the data shown are from before and after a maintenance shutdown. In fact, part of the value of the study is in the presentation of these high density plots. Each represents one day of running and shows all tags. This is not a standard display on an operator's panel. The hierarchical tree aids the detection of tags with similar power spectra and the high density plots allow the engineer to visually confirm the findings.

The scaling used in the time trends is referenced to the *before* case. For instance, time trend 5 in the *before* panel has been scaled to unit standard deviation and the time trend 5 in the after panel is scaled with the same factor. The large deviation in Tag 4 in the *after* panel arises because that time trend moved more than the time trend of Tag 4 before maintenance.

The power spectra in the plots are scaled to the same maximum peak height for visualization purposes. In the spectral PCA computations, however, all spectra are scaled to unit power.

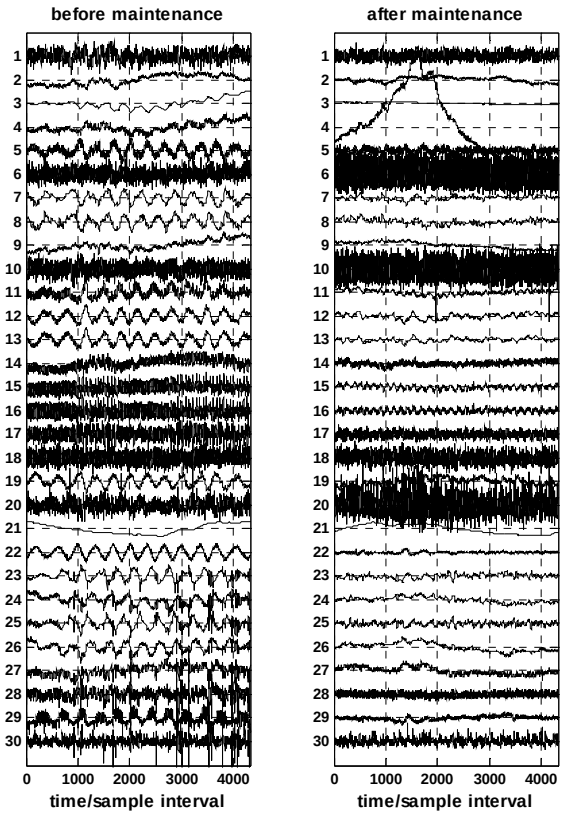


Figure 1. Time trends before and after maintenance

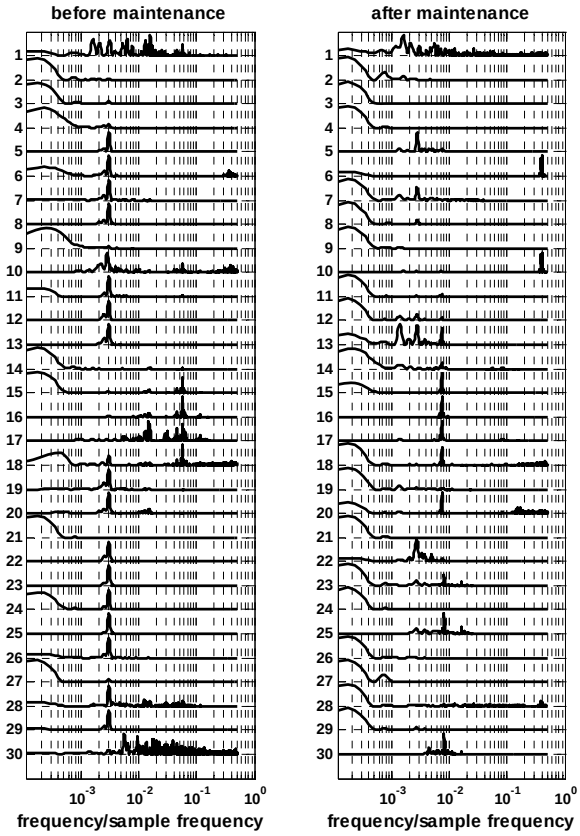


Figure 2. Power spectra before and after maintenance

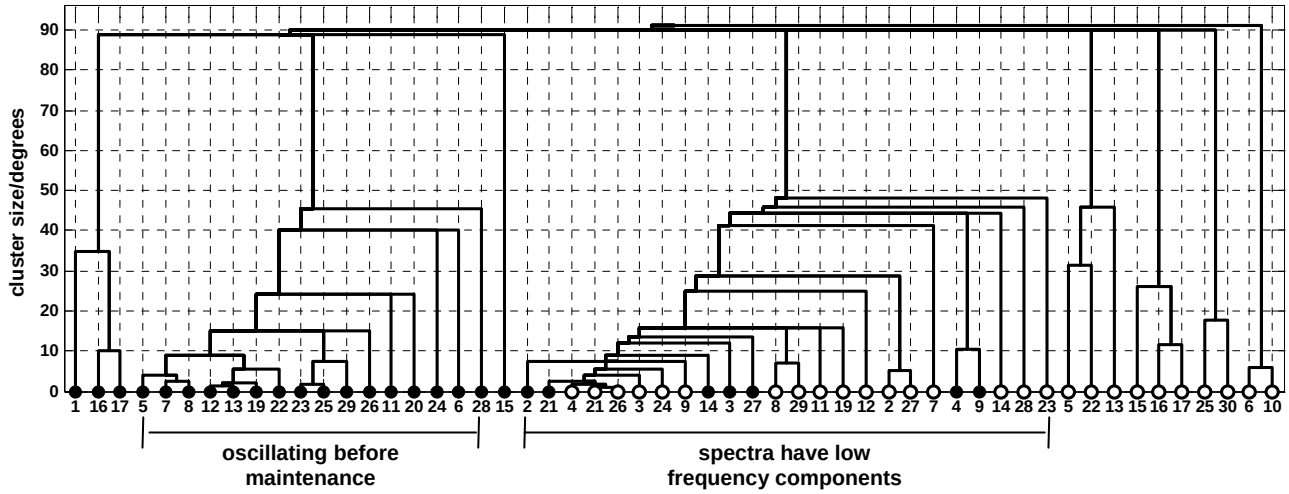


Figure 3. Hierarchical classification tree. Tags with black spots are from before maintenance data set, white spots are after maintenance.

4.2 Results

All 60 spectra (30 from before and 30 from after maintenance) were analysed together using spectral PCA and presented in Figure 3 as a hierarchical classification tree. The analysis needed 11 principal components to capture 99.5% of the variability in the data set. As discussed earlier, 54 of the 60 tags are represented in the tree, the remaining tags have broad band spectra similar to random noise and mapped close to the origin in the Spectral PCA score plot.

In Figure 3, each spot on the horizontal axis represents a complete spectrum from Figure 2. Black spots are the spectra before maintenance and the white are the spectra after maintenance. The numbers below indicate which tag generated the spectrum.

Clusters in the tree share a common branch into the main part of the tree and they represent tags which have similar power spectra and hence similar dynamic features in their time trends. Clusters are clearly visible in the tree; a cluster is a group of tags such as 25 and 30 on the extreme right, or the large group labelled as *spectra have low frequency components*. The *y*-axis shows the cluster size as the maximum angular separation between any two items within the cluster. Some sub-clusters also exist, such as tags 5, 7 and 8 which form a distinct sub-group in the cluster labelled *oscillating before maintenance*.

Here, the clusters have been identified by inspection, however they can also be detected automatically when the length of the branch joining a cluster to the main tree exceeds a pre-set fraction of the cluster size (both measured on the *y*-axis).

Oscillating before maintenance group: A group of tags that were oscillating before maintenance had a strong spectral peak at about 0.0028 on the normalized frequency axis (350 samples per cycle). They are Tags 5, 7, 8, 11-13, 19, and 22-29 which appear as a cluster. The tree shows tags 20 and 6 were also participating in the same oscillation before maintenance. It is not easy to tell that 20 was oscillating from its time trend, but spectral PCA detects the oscillation within the noise.

There are no tags from the after-maintenance data set in this group which demonstrates that maintenance successfully addressed the plant-wide disturbance.

Tags 15, 16 and 17: Tags 16 and 17 are clustered together in the before-maintenance data set, showing their spectra were more similar to each other than to any other spectra in the combined data set. The tree shows that Tag 1 is similar also. Although Tag 1 has spectral content across a broad range it also shares a prominent spectral peak with Tags 16 and 17. After maintenance, tags 16 and 17 lie in a different cluster and are joined by 15 showing that the dynamic behaviour of Tags 16 and 17 was changed by the maintenance activity The spectrum of tag 15 has some low frequency content which is stronger in the before maintenance data set. That is why Tag 15 did not join the {16, 17} cluster before maintenance. Tag 1 from the after maintenance data set is not in the tree because its spectrum was similar to random noise.

Low frequency components group: A cluster in the middle of the tree contains numerous tags both from before and after maintenance. Their spectra have low frequency components in common, and the time trends show that they all have slow drifting non-stationery behaviour. There are more tags in this group after maintenance than before. Some of them such as 7, 8, 11, 12, 19, 24, 27 and 29, migrated into this group once the main oscillating disturbance was removed. Many are indicators and are responding to long term drifts in ambient or operating conditions.

Small clusters: After maintenance there are several small clusters. These are:

- 5, 13 and 22: They have a broad peak at about 3×10^{-3} on the normalized frequency axis. No other tags share this spectral feature. The frequency of this spectral feature is very similar to that of the main oscillation in the before-maintenance data set. The classification tree shows, however, that it is a new frequency because the 5, 13 and 22 cluster form the after-maintenance data set is not connected to the cluster labelled *oscillating before maintenance*.

- 25 and 30: They have some spectral content at about 8×10^{-3} on the normalized frequency axis.
- 6 and 10: These tags have a high frequency spectral feature at 4×10^{-1} on the frequency axis. In fact, an inspection of the spectra shows that this feature was present before maintenance but was dominated by the main oscillation at 0.0028 on the frequency axis. It is more prominent after maintenance because the interference of the main oscillation has been removed.

Tags not in the tree: The tags excluded are 10, 18 and 30 in the before maintenance set and 1, 18 and 20 in the after maintenance set. Figure 1 shows that their time trends do not have any distinctive dynamic features, just noise. The benefit of the exclusion of tags with small $\mathbf{t}'$ – vectors from the tree is shown by considering Tags 6 and 10 in the after maintenance data set. They could be mistaken as random by visual inspection, however the spectral analysis shows that they have a distinctive high frequency peak and they therefore appear in the tree.

5. CONCLUSIONS

The paper has presented a hierarchical classification tree as a mean of visualization of the structure within a principal component model of arbitrary dimensions. Each item in the tree represents the power spectrum from one measurement point in the process and the vertical axis is an angle measure that indicates how similar the spectra are to one another. An industrial case study showed that the tree is useful in combination with high density plots of time trends and spectra for interpreting and understanding the impact of the maintenance activity.

6. ACKNOWLEDGMENTS

We are very grateful to John Cox and Michael Paulonis of the Eastman Chemical Company for making the data sets available for the case study. The first author gratefully acknowledges the support of the Royal Academy of Engineering (Global Research Award) and of ABB Corporate Research.

7. REFERENCES

Alam. T.M., and M.K. Alam (2005) Chemometric analysis of NMR spectroscopy data: A review, *Annual Reports on NMR Spectroscopy*, 54, 41-80.

Brodnjak-Voncina, D., D. Dobcnik, M. Novic, and J. Zupan (2002). Chemometrics characterisation of the quality of river water, *Analytica Chimica Acta*, 462,87-100.

Chatfield, C., and A.J. Collins (1980). *Introduction to multivariate analysis*, Chapman and Hall, London, UK.

De Belie, N., V. De Smeldt and J. De Baerdemaeker (2000). Principal component analysis of chewing sounds to detect differences in apple crispness, *Postharvest Biol. Technol.*, 18., 109-119.

Duda, R.O., P.E. Hart and D.G. Stork (2000). *Pattern Classification* (*2nd Edition*), Wiley-Interscience.

Flaten, G.R., R. Belchamber, M. Collins A.D. Walmsley (2005). Caterpillar - an adaptive algorithm for detecting process changes from acoustic emission signals, *Analytica Chimica Acta*, 544, 280-291.

Gordon, A.D., (1987). A review of hierarchical classification, *Journal of the Royal Statistical Society*, 150, 119-137.

Jackson, J.E., and G.S. Mudholkar (1979). Control procedures for residuals associated with principal components analysis, *Technometrics*, 21, 341-349.

Kresta, J.V., J.F. MacGregor and T.E. Marlin (1991). Multivariate statistical monitoring of process operating performance, *Canadian Journal of Chemical Engineering*, 69, 35-47.

Lee, Y-H., K.G. Min, C. Han, K.S. Chang T.H. Choi (2004). Process improvement methodology based on multivariate statistical analysis methods, *Control Engineering Practice*, 12, 945–961.

Malhi. A., and R.X. Gao (2004). PCA-based feature selection scheme for machine defect classification, *IEEE Transactions on Instrumentation and Measurement*, 53, 1517-1525.

Martin, E.B., and A.J. Morris (1996). Non-parametric confidence bounds for process performance monitoring charts, *Journal of Process Control*, 6, 349-358.

Norris, T., P.K. Aldridge and S.S. Sekulic (1997). Determination of end-points for polymorph conversions of crystalline organic compounds using on-line near-infrared spectroscopy, *Analyst*, 122, 549–552.

Ozaki, Y., S. Sasic and H.H. Jiang (2001). How can we unravel complicated near infrared spectra? Recent progress in spectral analysis methods for resolution enhancement and band assignments in the near infrared region, *Journal of Near Infrared Spectroscopy*, 9, 63-95.

Qin S.J. (2003). Statistical process monitoring: basics and beyond, *Journal of Chemometrics* 17, 480-502.

Raich, A., and A. Çinar (1997). Diagnosis of process disturbances by statistical distance and angle measures, *Computers & Chemical Engineering*, 21, 661-673.

Seasholtz, M.B., (1999). Making money with chemometrics, *Chemometrics and Intelligent Laboratory Systems*, 45, 55-64.

Seem, J.E., 2005, Pattern recognition algorithm for determining days of the week with similar energy consumption profiles, *Energy and Buildings*, 37, 127–139.

Thornhill, N.F., S.L. Shah, B. Huang, and A. Vishnubhotla, (2002). Spectral principal component analysis of dynamic process data, *Control Engineering Practice*, 10, 833-846.

Tzeng, D.Y., and R.S. Berns (2005). A review of principal component analysis and its applications to color technology, *Color Research and Application*, 30, 84-98.

Wang, X.Z., S. Medasani, F. Marhoon and H. Albazzaz (2004). Multidimensional visualization of principal component scores for process historical data analysis, *Industrial Engineering and Chemistry Research*, 43, 7036-7048.

Wiedemann, L.S.M., L.A. d'Avila and D.A. Azevedo (2005). Adulteration detection of Brazilian gasoline samples by statistical analysis, *Fuel*, 84, 467–473.

Wise, B.M., and N.B. Gallagher (1996). The process chemometrics approach to process monitoring and fault detection, *Journal of Process Control*, 6, 329-348

Wise, B.M., N.L. Ricker, D.F. Veltkamp and B.R. Kowalski (1990). A theoretical basis for the use of principal components models for monitoring multivariate processes, *Process Control and Quality*, 1, 41-51

Wold, S., K. Esbensen and P. Geladi (1987). Principal Component Analysis, *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.

Wu, H.D., M. Siegel and P. Khosla (1999). Vehicle sound signature recognition by frequency vector principal component analysis, *IEEE* Transactions on Instrumentation and Measurement, 48 1005-1009

Xia, C. and J. Howell, (2005). Isolating multiple sources of plant-wide oscillations via spectral independent component analysis. *Control Engineering Practice*. **13**, 1027-1035.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

ROOT CAUSE ANALYSIS OF OSCILLATING CONTROL LOOPS

R. Srinivasan * **M. R. Maurya** **
R. Rengaswamy *,1

* *Dept of Chemical Engineering, Clarkson University, Potsdam, NY 13699-5705*
** *San Diego Supercomputer Center, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0505*

Abstract:
Oscillation in a single control loop can propagate to many units and can cause several control loops to oscillate. In this work, an approach that uses detailed oscillation characterization in combination with signed digraphs is proposed for isolating the source loop that causes plant-wide oscillation. The success of this approach is built on a new oscillation characterization technique that identifies the zero-crossings of each oscillating measurement. A signed digraph that embeds the temporal information obtained from the zero-crossings of the data is analyzed to isolate the root cause for oscillation. A simulation case study illustrates the applicability of the proposed approach.

Keywords:
signed digraphs, control loop, oscillation diagnosis, performance monitoring

1. INTRODUCTION

A number of surveys on the performance of control loops (Desborough and Miller, 2001) indicate that a majority of control loops in process industries perform poorly. It was observed that performance degradation in control loops result in: (i) poor set point tracking, (ii) oscillations, (iii) poor disturbance rejection, and/or (iv) high excessive final control element variation. Reducing or removing such oscillations can yield substantial commercial benefits. Desborough and Miller (2001) claim that a 1% improvement in either energy efficiency or controller performance would save up to $300 million dollars per year. Sustained oscillations in control loops can be due to multiple reasons: (1) Valve non-linearity due to causes such as stiction, dead band and hysteresis, (2) Poorly tuned controller in a nonlinear processes, (3) Insufficient digital resolution (quantizing effects), (4) Controller saturation, (5) Interacting loops, (6) Oscillations that are external to the loop or (7) a combination thereof.

Diagnosing the cause for oscillation may involve separating the source loop from other secondary loops when plant-wide oscillations are present. Plant-wide oscillations occur when an oscillation in a single loop propagates to many units. Diagnosis of plant-wide oscillations has received considerable attention in the recent past. Thornhill *et al.* (2003b) use the detection of measurements oscillating at similar frequencies to perform root cause analysis. They assume that the source loop is oscillating due to the presence of a nonlinearity such as stiction in the control valve. The presence of stiction is confirmed through a nonlinearity index computed for each loop. An extension to

[1] Corresponding author: raghu@clarkson.edu

this method is discussed in Thornhill (2005). Xia and Howell (2005) propose the use of independent component analysis to distinguish between the source loop and the secondary oscillating loops.

In this work, a methodology to identify the root cause for oscillations when one or more loops oscillate simultaneously is proposed. A signed digraph that embeds the temporal information obtained from the zero-crossings of the data is analyzed to isolate the problem loop and identify the root cause for oscillation. The zero-crossings in the measurements (with respect to their steady-state information) is obtained using a novel oscillation characterization algorithm (Srinivasan *et al.*, 2005c).

2. CHARACTERIZING OSCILLATIONS IN CONTROL LOOPS

Oscillations in industrial data seldom have constant frequency or amplitude. Also, the measurements, Controller output (OP) and Process output (PV), have non-constant mean due to changes in Set point (SP) or due to the presence of measured or unmeasured disturbances. An oscillation characterization algorithm outlined in Srinivasan *et al.* (2005c) can be applied to obtain the zero-crossings of the measurements. A brief explanation of this procedure is presented here. Figure 1a shows the time-series data of an industrial loop that has both non-constant mean and intermittent oscillations. A modified Empirical Mode Decomposition (EMD) procedure (Huang *et al.*, 1998) is employed for characterizing such oscillations. There are three basic steps in the proposed oscillation characterization method. These are listed below:

Step 1. The first step removes the non-constant mean (i.e. low-oscillation modes) from the signal. For the given time series data, upper and lower envelopes are constructed by connecting the maxima and minima points respectively (See Figure 1b). A modified empirical mode decomposition procedure is employed in this step. An average of these envelopes is then subtracted from the signal to generate the time series shown at the extreme right of Figure 1(1c).

Step 2. Cumulative area of the dominant oscillating mode separated out from Step 1 is computed. The cumulative area is a weighted mean of the data and it averages the effect of noise, thereby reducing the number of spurious zero-crossings that may be reported.

Step 3. Extrema points of the cumulative area capture the zero-crossing points. These extrema points are identified and are reported as the zero-crossing points of the dominant oscillation mode.

The data pre-processing, which is not discussed here, involves removing outliers and replacing missing data. Several attributes can be calculated based on the zero-crossings, such as: (i) Time period of each sweep of oscillation, (ii) Amplitude and strength of each oscillation mode, (iii) Time instances when oscillations are present and (iv) Start and end time for each sweep of oscillation. A detailed discussion on the oscillation characterization technique can be found in Srinivasan *et al.* (2005c). It will be shown later that the information about the zero-crossings when used with digraphs can isolate the root cause of oscillation(s). In the next section, a succinct discussion on the application of signed digraphs for fault diagnosis is presented. An interested reader is referred to (Maurya *et al.*, 2003*a*; Maurya *et al.*, 2003*b*) for additional details.

3. FAULT DIAGNOSIS USING SIGNED DIGRAPHS

A directed graph (digraph (DG)) consists of nodes representing variables and directed arcs between nodes representing the interaction among variables. When signs are placed on the nodes and the arcs of a DG, it is called a signed digraph (SDG). Signed directed graph based methods are widely used for fault diagnosis because SDG models provide a powerful representation to capture the cause-effect information about the process. SDG models do not require complete quantitative description and can be developed from partial information such as the structure of the equations and information about the normal operating conditions. Signed digraphs have been used to model control loops as well.

3.1 Background

Iri *et al.* (1979) were the first to use SDG for modeling chemical processes. Oyeleye and Kramer (1988) discuss SDG-based steady state analysis and prediction of inverse response (IR) and compensatory response (CR). Bhushan and Rengaswamy (2002) have used SDG analysis for sensor location for efficient fault diagnosis. Chen and Howell (2001) presented fault diagnosis of controlled systems where SDG has been used to model control loops. Maurya *et al.* (2003*a*) have recently presented algorithms and methods for the development and analysis of SDG models for systems described by differential equations (DE), algebraic equations (AE) and differential algebraic equations (DAE). Briefly, a digraph for a DE system is developed by drawing arcs from the variables that occur in the time-derivative function to the corresponding state variable. For an

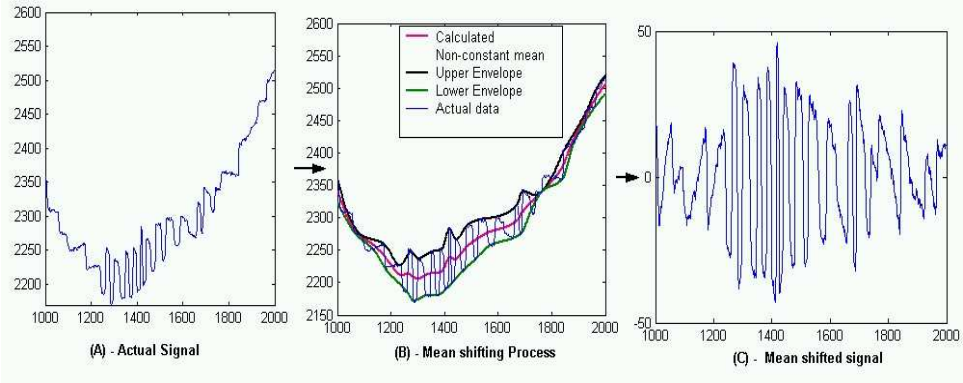


Fig. 1. Oscillation characterization algorithm - illustrative steps. Plot shows only a zoomed portion of data for clarity.

AE system, a digraph can be drawn after performing perfect matching between the algebraic equations and the variables. For a DAE system, the SDG corresponding to the DE and AE parts are combined. Maurya *et al.* (2003*b*) also proposed a unified SDG-model for control loops in which, both disturbances (e.g. sensor bias, bias in the manipulated valve) as well as structural faults (e.g. sensor failure and controller failure), can be easily modeled and analyzed. The SDG is developed using the topology of the control loops and a PI or PID approximation of the control algorithm. Since quantitative details are not needed, the required information is easily available for most controller configurations. The analysis of SDG depends on whether the SDG is for a steady-state (AE) system or for a dynamic system (DE or DAE) (Maurya *et al.*, 2006). The analysis of DE systems and DAE systems are relevant to the present work since steady-state is not reached in the time domain during oscillations.

For a chosen deviation (fault), the initial response of a system variable (dependent variables that are both measured and unmeasured) can be predicted by propagation through all the directed path(s) from the fault node to the system variable (see Maurya *et al.* (2003*a*) for certain exceptions for DAE systems). The inverse of this principle, i.e., back-propagation, is used for fault diagnosis. Ambiguity in qualitative simulation and diagnosis arise due to the presence of multiple paths with opposing signs. Hence, the use of quantitative information (e.g., through fuzzy-logic) has been suggested for dynamic diagnosis (Tarifa and Scenna, 2003; Chang and Chang, 2003).

3.2 Fault diagnosis using backward-reasoning

In any backward-reasoning based fault diagnosis methodology, the basic idea is to identify one or more paths from appropriate fault nodes to the measured nodes so that forward-propagation along these valid paths can explain the observed symptoms. Usually, depth-first search (DFS) is used to identify these paths (Tarjan, 1972). The given non-zero sign of a measured node and the signs of the incident arcs are used to infer the possible signs of the predecessor nodes. Any one of these incident arcs, propagation through which will explain the qualitative state of the observed node, is a valid branch. If a predecessor node is a measured node and its inferred sign contradicts the observed sign, then no further backward-reasoning is performed on this predecessor node. Thus, this branch of the search tree is terminated since it cannot be a part of a valid path. Otherwise, backward-reasoning is applied to the predecessor nodes successively. If a predecessor node is an exogenous variable then it is a candidate fault. Whenever a branch of the search tree is terminated, back-tracking is used and other predecessor nodes for the previous node are explored. This process is continued till all the predecessors to the measured node are exhaustively examined. Thus, every measured node ($y_j,\ j = 1,\ 2\ \ldots\ m$) yields a candidate fault set ($E_j = \{f_k\},\ k \in \{1,\ 2\ \ldots\ n\}$, where n = number of fault nodes). Intersection of these candidate fault sets is the actual candidate fault set. Whenever a candidate fault node (f_k) is reached, forward-propagation is used to verify that the measured pattern can be generated. This simple rule works well for those patterns which arise due to single faults alone. For patterns corresponding to multiple faults, the minimal combinations of faults ($\{f_{k_1}, f_{k_2} \ldots\}$), one from each set ($E_j$), are considered so that the union of the patterns generated by them ($\bigcup\{Y_{k_i}\}$, Y_{k_i} is the pattern generated by fault f_{k_i}) covers the measured patterns (ambiguity is allowed).

3.3 Incorporation of temporal order of start of oscillation in the SDG-based diagnosis

Onset of oscillations is similar to eliciting initial response after the occurrence of a fault. Hence, the temporal order in which oscillations start in

measured variables can be used to construct the paths through which faults propagate. This helps in pruning some of the propagation paths, resulting in an enhanced diagnostic resolution. This is the basic principle behind the utilization of the temporal order for fault diagnosis. The diagnostic procedure is:

(1) Start the search for root cause from the measured variable with the smallest oscillation start time.
(2) Use back-propagation till a fault node or a measured variable node is reached.
(3) If a measured variable with a larger start time of oscillation is reached, or a conflicting sign is inferred then this branch of the search tree is terminated. Back-track to the next unexplored node. Go to step 2.
(4) If a fault node is reached, use forward-propagation to verify that the measured patterns can be generated with the specified sign as well as the temporal order. Ambiguity is allowed in the predicted sign. The constraint on the temporal order is that if the start-time of oscillation of node 'B' is larger than that of node 'A' and there are no two separate paths between them, then node 'B' must be downstream of node 'A' on some path(s).
(5) Go to step 2 to explore any remaining unexplored nodes.

In the case study presented in the next section, it is shown that the use of temporal order results in a better diagnostic resolution.

4. RESULTS

4.1 Simulation set-up

A 2x2 interacting process with one cascade loop is considered for analysis. This is shown in Figure 2. The simulated system exhibits type-A interaction (Chen and Howell, 2001). There are totally 6 measurements, namely, Loop 1; Set-point (SP1), process (S1) and controller output (C1), Loop2: Set-point (C21), process (S22) and controller output (C22) and Loop 3: Set-point (SP2), process (S21) and controller output (C21). Following three scenarios are considered:

Case 1: External oscillations in Loop 1.
Case 2: External oscillations in Loop 3.
Case 3: Oscillations in Loop 1 due to stiction.

Figure 3 shows the data for case 1, with a sampling time of 0.1 seconds. Table 1 shows the start time of a sweep of oscillation from each case study. This information was obtained using the oscillation characterization algorithm outlined in section 2. Based on the start time, a temporal order is assigned. The direction of deviation of each measurement from its steady state is also provided; positive indicates an increase and the negative sign indicates decrease from the corresponding steady state values.

Table 1. Oscillation attributes for the three case studies.

Case No.	Measurements (variable name)					
	S1M	C1	S22	C22	S21M	C21
Case 1						
Start time of osc	2170	2175	2320	2270	2270	2270
Direction	+	-	-	-	-	-
Temporal Order	1	2	4	3	3	3
Case 2						
Start time of Osc	2300	2310	2240	2210	2200	2200
Direction	+	-	+	+	+	+
Temporal Order	4	5	3	2	1	1
Case 3						
Start time of Osc	1415	1415	1435	1418	1418	1418
Direction	+	-	+	+	+	+
Temporal Order	1	1	3	2	2	2

The start time is given in Sampling instants.

The signed digraph model of the controlled system is developed using the method presented by Maurya *et al.* (2003*b*) and is shown in Figure 4. S1M and S21M denote the measured values of the process variables, S1 and S21, respectively. B1 and B2 nodes represent the sensor biases in the respective loops. V1 and V2 are the valve positions and VB1 and VB2 are the corresponding valve-position biases.

Diagnosis of case 1 (Figure 4): Starting with S1M = '+', back-propagation identifies B1 = '+' as a candidate fault. VB1 = '+' is excluded since forward-propagation from VB1 = '+' violates signs of S21M, etc. Back-propagation to S21 and then to S22 leads to violation of the measured sign of S22, so this branch is also terminated. In this case, temporal order need not be utilized to get complete resolution.

Diagnosis of case 2 and 3: As listed in Table 1, the sign patterns in the two fault cases are the same. Hence, if one were to use only sign pattern then these two faults cannot be distinguished. However, by using the temporal-order information, for case 2, B2 = '+' is identified as a candidate fault. SP2 = '-' is ruled out since it violates the temporal order in the cascade control loop. VB1 = '+' can be considered a fault if one were not to differentiate between the temporal-order between S21M/C21 and S1M/C1 since they are on two different paths. However, since they are in different control loops and the oscillations show up first in the cascade control loop, in reality, VB1 = '+' is unlikely. For case 3, starting with S1M = '+', VB1 = '+' is identified as a candidate fault. B1 = '+' is ruled out since it violates sign of S21M, etc. B2 = '+' is ruled out since it violates the temporal-order between S21M and S1M. Thus, by using temporal-order information, better (com-

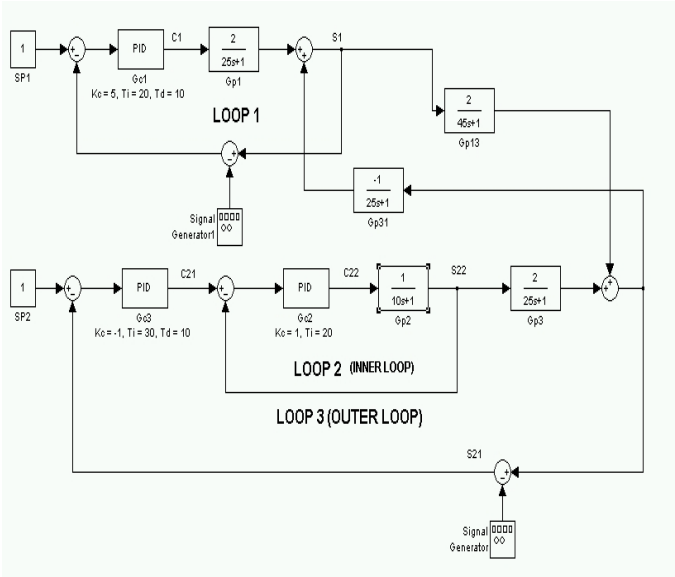


Fig. 2. Simulation case study.

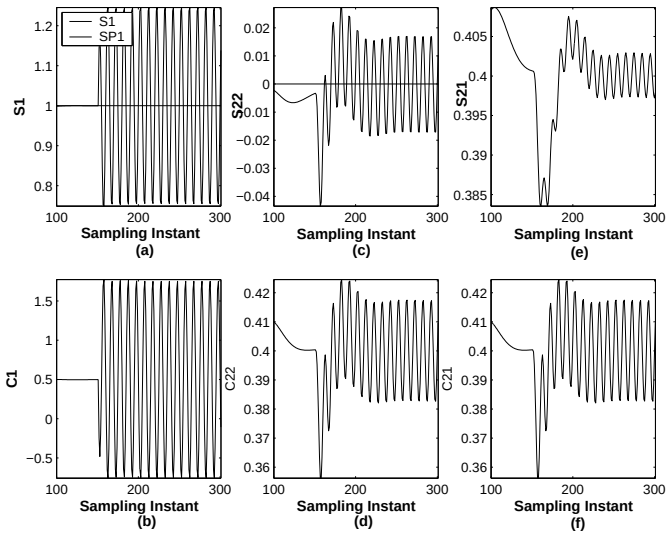


Fig. 3. Case 1: External disturbance in Loop 1 causing oscillations in other measurements.

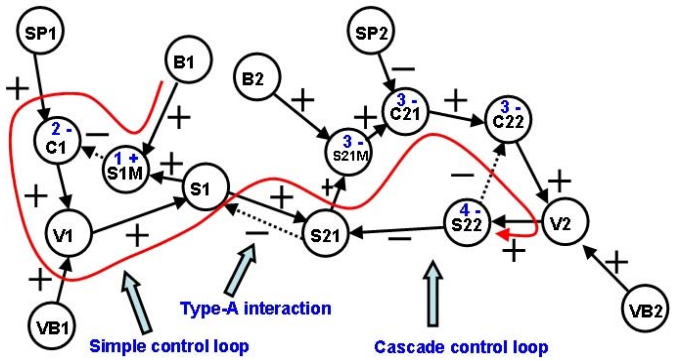


Fig. 4. Root cause analysis for case 1.

Table 2. Results of diagnosis using sign and temporal-order

No.	Fault induced	Fault diagnosed
1	B1 = '+'	B1 = '+' (sensor-bias in loop 1)
2	B2 = '+'	B2 = '+' (sensor-bias in loop 3) VB1 = '+' (valve-stiction in loop 1*)
3	VB1 = '+'	VB1 = '+' (valve-stiction in loop 1)

*: see text for explanation.

plete) diagnostic resolution is achieved. In fact, using the result of case 3 (i.e. if case 3 has occurred in the past and hence stored in a database), one can conclude complete resolution for case 2 as well. This is true provided that the process is not so nonlinear as to exhibit different temporal orders for different magnitudes of the same fault. In the present context, since the pattern and temporal-order for case 3 is known from the database, it is assumed that case 3 cannot result in another temporal-order (i.e. that of case 2). These results are summarized in Table 2.

5. CONCLUSIONS

Starting with a brief highlight on the benefits of advanced diagnosis, a brief discussion on oscillation characterization in control loops was presented. Then, a summary of the use of signed digraphs for fault diagnosis was presented. A procedure to incorporate the temporal-order of fault-propagation into the digraphs based diagnosis methodology was described. Finally a case study was presented to show how the proposed methodology, with the utilization of temporal-order, results in better diagnostic resolution to the extent that the source for malfunction in a control loop is uniquely located.

REFERENCES

Bhushan, M. and R. Rengaswamy (2002). Comprehensive design of a sensor network for chemical plants based on various diagnosability and reliability criteria: 1. Framework. *Ind. Engng Chem. Res.* **41**(7), 1826–1839.

Chang, S.Y. and C.T. Chang (2003). A fuzzy-logic based fault diagnosis strategy for process control loops. *Chem. Eng. Sci.* **58**(15), 3395–3411.

Chen, J. and J. Howell (2001). A self-validating control system based approach to plant fault detection and diagnosis. *Comput. & Chem. Engng.* **25**, 337–358.

Desborough, L. D. and R. M. Miller (2001). Increasing customer value of industrial control performance monitoring – Honeywell's experience. CPC-VI. Arizona, USA.

Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, E. H. Shih, Q. Zheng, C. C. Tung and H. H. Liu (1998). The empirical mode decomposition and the hilbert spectrum for non-linear and non-stationary time series. *Proceedings of Royal Soceity of London* **A**(454), 903–995.

Iri, M., K. Aoki, E. O'Shima and H. Matsuyama (1979). An algorithm for diagnosis of system failures in the chemical process. *Comput. & Chem. Engng.* **3**(1-4), 489–493.

Maurya, M. R., R. Rengaswamy and V. Venkatasubramanian (2003*a*). A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis. *Ind. Engng Chem. Res.* **42**(20), 4789–4810.

Maurya, M. R., R. Rengaswamy and V. Venkatasubramanian (2003*b*). A systematic framework for the development and analysis of signed digraphs for chemical processes. 2. Control loops and flowsheet analysis. *Ind. Engng Chem. Res.* **42**(20), 4811–4827.

Maurya, M. R., R. Rengaswamy and V. Venkatasubramanian (2006). A signed directed graph-based systematic framework for steady-state malfunction diagnosis inside control loops. *Chem. Eng. Sc.* **61**, 1790–1810.

Oyeleye, O.O. and M.A. Kramer (1988). Qualitative simulation of chemical process systems: Steady-state analysis. *AIChE J.* **34**(9), 1441–1454.

Srinivasan, R., R. Rengaswamy and R. M. Miller (2005c). A modified empirical model decomposition (emd) process for oscillation characterization in control loops. *Submitted to Control Engg. Practice.*

Tarifa, E. E. and N. J. Scenna (2003). Fault diagnosis for a MSF using a SDG and fuzzy logic. *Desalination* **152**(1-3), 207–214.

Tarjan, R.E. (1972). Depth-first search and linear graph algorithm. *SIAMJ. Comput.* **1**(2), 146–.

Thornhill, N. F. (2005). Finding the source of nonlinearity in a process with plant-wide oscillation. *IEEE Transactions on Control System Technology* **13**, 434–43.

Thornhill, N. F., W. J. Cox and M. A. Paulonis (2003b). Diagnosis of plant wide oscillation through data driven analysis and process understanding. *Control Engg.Practice* **11**, 1481–90.

Xia, C. and J. Howell (2005). Isolating multiple sources of plant-wide oscillations via independent component analysis. *Control Engg. Practice* **13**, 1027–35.

ADCHEM 2006

International Symposium on Advanced Control of Chemical Processes
Gramado, Brazil – April 2-5, 2006

QUANTIFICATION OF VALVE STICTION

Mridul Jain * **M.A.A. Shoukat Choudhury** *
Sirish L. Shah *,1

* *Department of Chemical and Materials Engineering,*
University of Alberta, Edmonton AB, Canada, T6G 2G6

Abstract: Oscillations in control loops lead to poor controller performance. Stiction in control valves is one of the major causes of such oscillations. Therefore, the correct diagnosis of stiction is important. There are several methods for detecting stiction, but quantification of stiction still remains a challenge. Two parameters are used to model the stiction phenomenon successfully, namely, deadband plus stickband, $'S'$, and slip-jump,$'J'$. It has been observed that the main cause of valve deterioration is the presence of slip-jump, $'J'$. The higher the value of $'J'$, the more severe is the level of deterioration of controller performance. Thus, in addition to the estimation of $'S'$, an estimate of $'J'$ is the main challenge in monitoring the condition of a control valve. In this work a method is proposed to estimate both $'S'$ and $'J'$ simultaneously unlike existing quantification methods where stiction is quantified as a single parameter.

Keywords: Control loops, Control valves, Process control, Nonlinearity, Stiction.

1. INTRODUCTION

Non-linear effects are often encountered in process plants. These non-linearities can be due to: (1) Valve non-linearity, for example due to stiction, deadband and hysteresis; (2) the presence of non-linear external oscillations, and/or (3) non-linearity in the process.

The presence of a non-linearity in a control loop often leads to oscillations in a control loop and hence poor performance. About 30% of the oscillations in control loops are due to valve problems (e.g. the presence of static friction or stiction). Therefore, detection and quantification of stiction in control valves is an important issue in the process industry. There are several stiction detection methods (Choudhury *et al.*, 2004*b*; Choudhury *et al.*, 2004*c*; Horch, 1999; Singhal and Salsbury, 2005; Stenman *et al.*, 2003; Srinivasan *et al.*, 2005*a*; Srinivasan *et al.*, 2005*b*). But quantifying stiction still remains a challenge.

Earlier work by (Choudhury *et al.*, 2004*c*) quantifies stiction by fitting an ellipse to the pv-op plot and the maximum width of the ellipse is reported as 'apparent stiction'. Recently, Srinivasan *et al.* (2005*a*) introduced another approach where they exploited the fact that the presence of stiction has distinct qualitative shapes or pattern in the controller output, op and the controller variable, pv signals. They have applied a Pattern Recognition technique using Dynamic Time Warping (DTW) on the pv and op data. First, the test patterns (for both op and pv) are generated using the zero crossing data from the raw signals. Then these test patterns are compared to the actual signal. If stiction is confirmed then the maximum peak-to-peak amplitude is reported as stiction. However, the maximum peak-to-peak amplitude is just the magnitude of limit cycle and cannot be attributed to real stiction. Another disadvantage with this approach is the apriori knowledge of the patterns in the op and pv due to stiction. The patterns described therein may not be always due to stiction. Some of those patterns in the pv and op signals may arise simply due to the presence of a tightly tuned controller or an oscillatory disturbance. In addition to these, asymmetric stiction, which is not uncommon, cannot be detected and quantified using this approach.

In another method proposed by (Srinivasan *et al.*, 2005*b*), a Hammerstein model identification approach

[1] Corresponding author, E-mail: sirish.shah@ualberta.ca, Tel: +1 (780) 492 5162, Fax: +1 (780) 492 2881.

is explored. A general structure of a Hammerstein model is shown in figure 1. The non-linear part of the Hammerstein model is described by a single parameter stiction model ((Stenman *et al.*, 2003)).

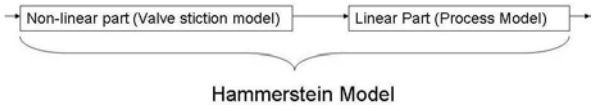


Fig. 1. *General Structure of a Hammerstein Model*

It has been observed that the single parameter stiction model does not depict the true stiction behavior (Choudhury *et al.*, 2004*a*), as discussed in section 2.

In this study, the proposed approach uses a two parameter stiction model proposed by (Choudhury *et al.*, 2004*a*) to model the non-linear component of the Hammerstein model.

The rest of the paper is organized as follows: In Section 2 a brief discussion of the two parameter stiction model is provided. This is followed by an example demonstrating the importance of slip-jump, J, in loop dynamics. Section 4 describes the proposed method. Sections 5 and 6 summarize simulation and experimental results respectively, followed by concluding remarks in Section 7.

2. WHY USE A TWO PARAMETER MODEL OF STICTION?

This section briefly discusses the adequacy of a two parameter stiction model for closed loop simulation of stiction. Also, the limitations of the one parameter stiction model proposed by (Stenman *et al.*, 2003) and used in (Srinivasan *et al.*, 2005*b*) are briefly discussed. Before discussing the data-driven stiction models, a case of an industrial example where a valve was sticky is presented in order to find the right pattern of stiction present in a valve operating under closed loop control configuration.

2.1 An industrial control loop with a sticky valve

Consider a level control loop which controls the level of condensate in the outlet of a turbine by manipulating the flow rate of the liquid condensate. The control valve of this loop is confirmed to have stiction. In total 8640 samples for each tag were collected at a sampling rate of 5 s. Figure 2 shows a portion of the time domain data. The left panel shows time trends for level (pv), the controller output (op) which is also the valve demand, and valve position (mv) which can be taken to be the same as the condensate flow rate. The plots in the right panel show the characteristics pv-op and mv-op plots. The bottom figure clearly indicates both the stickband plus deadband and the slip jump effects. The slip jump is large and visible from the bottom figure especially when the valve is moving in a downward direction. It is marked as 'A' in the figure. The pv-op plot does not show the jump behavior clearly because the process dynamics (i.e., the transfer function between mv and pv) destroys the pattern. The pattern shown in the actual valve position (mv) vs. controller output (op) can be taken as a typical signature of valve stiction because it clearly shows the deadband plus stickband and the slip-jump. Similar patterns can be found in many industrial control valves suffering from stiction.

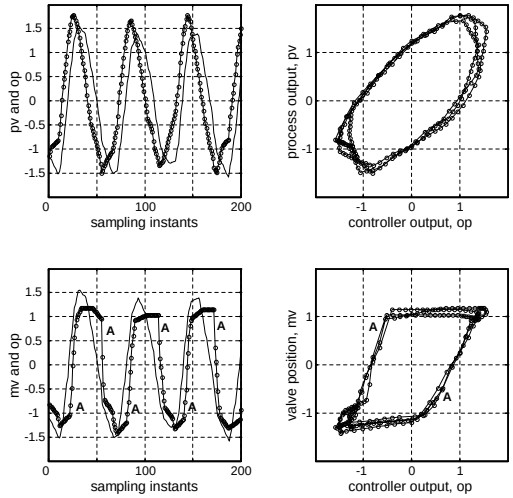


Fig. 2. *Flow control cascaded to level control in an industrial setting, the line with circles is pv and mv, the thin line is op*

2.2 One-parameter stiction model

A simple one parameter stiction model was proposed by (Stenman *et al.*, 2003). The model can be mathematically expressed by the following equation

$$x(t) = \begin{cases} x(t-1) & ,if \mid x(t) - d \mid < d \\ u(t) & \text{otherwise} \end{cases}$$

Where, *x(t)* and *x(t-1)* are the valve output (stem position) at time *'t'* and *'t-1'* respectively, *u(t)* is the controller output at time *'t'* and *'d'* is the valve stiction band. For details of this stiction model, interested readers are referred to (Stenman *et al.*, 2003).

2.3 Two-parameter stiction model

A two parameter model proposed by (Choudhury *et al.*, 2004*a*) captures the stiction phenomenon successfully. The two parameters are: S (Stickband + Deadband) and J (Slip-jump). For details on this stiction model interested readers are referred to (Choudhury *et al.*, 2004*a*).

2.4 Comparison between one-parameter and two parameter stiction model

Figure 3(a) shows a typical valve output (mv), vs. controller output (op) plot for the one parameter

stiction model described in (Stenman *et al.*, 2003; Srinivasan *et al.*, 2005*b*) while Figure 3(b) shows the same plot for the two parameter stiction model proposed in (Choudhury *et al.*, 2004*a*). Figure 3(a) is clearly different from the pattern of stiction shown in Figure 2. It suffices to say that the one parameter stiction model does not capture the true characteristic of stiction. Indeed it should not be called a stiction model, rather it should be defined as a quantization or a staircase function. On the other hand, the plot for two parameter stiction model (Figure 3(b)) clearly matches with the pattern in Figure 2. Thus the two parameter stiction model is able to adequately capture the characteristic of valve stiction (Choudhury *et al.*, 2004*a*).

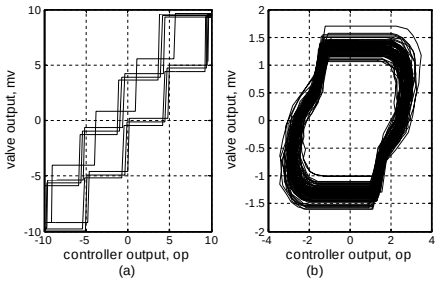


Fig. 3. *(a) mv-op for one parameter model ('d') (b)mv-op for two parameter model (S, J)*

3. ISSUES IN QUANTIFYING STICTION

3.1 Effect of controller dynamics and process dynamics on apparent stiction

Earlier work by (Choudhury *et al.*, 2004*c*; Choudhury *et al.*, 2005) quantifies stiction by fitting an ellipse to the *pv-op* plot and the maximum width of the ellipse is reported as 'apparent stiction'. Stiction is reported as 'apparent' because the estimate includes the effect of the process and controller dynamics. The following simulation example demonstrates the effect of the controller tuning on the estimation of apparent stiction.

Figure 4 shows the simulink block diagram used for generating stiction data. The process model is

$$G(z) = \frac{1.45z - 1}{z^4 - 0.8z^3} \tag{1}$$

The controller is implemented in the following form:

$$C(s) = K_c \left(1 + \frac{1}{\tau_i\ s}\right) \tag{2}$$

The reset time, τ_I, is fixed at 1 sec and the gain,K_c, is varied.The Stiction parameters 'stickband+deadband', S and 'slip jump', J are fixed at 3 and 1, respectively. Three cases, K_c = 0.05, 0.10 and 0.15, are considered and 1024 samples are generated for each case.

Figure 5 shows the *pv-op* plot and the fitted ellipse for the three cases. The apparent stiction reported are: for K_c = 0.05, 0.10 and 0.15, the estimated apparent stiction are 5.79, 3.06 and 1.62, respectively. Ideally,

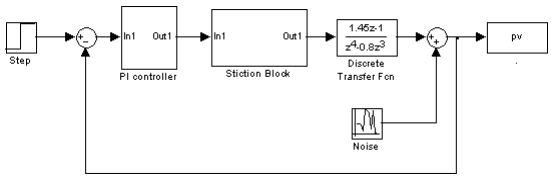


Fig. 4. *Simulink block diagram used for generating stiction data*

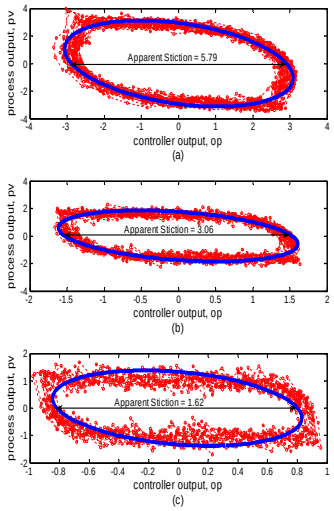


Fig. 5. *$mv - op$ plot and fitted ellipse (a)K_c = 0.05 (b)K_c = 0.10 (c)K_c = 0.15*

it should be same because the same amount of stiction was used for all cases (S=3 and J=1). A similar effect of the process dynamics can also be observed on the value of apparent stiction. Hence the width of the ellipse in the *pv-op* plot termed as 'apparent stiction' cannot be taken as an accurate estimate of stiction.

3.2 The importance of quantifying Slip-Jump (J)

Describing function analysis performed in (Choudhury *et al.*, 2004*a*) suggests that for processes without any integrator, limit cycles in a control loop may occur only in the presence of slip-jump (J) for the case of a sticky valve. Moreover, the amplitude and frequency of the limit cycles depend significantly on the slip-jump (J). The following simulation results show the effect of J on the amplitudes and frequencies of the limit cycles.

The system considered here is the same as in Section 3.1. In order to observe the impact of J clearly, the controller parameters are chosen as K_c = 0.15 and τ_I = 0.15 sec. Figure 6 shows the variation of the frequency and amplitude of limit cycles with slip jump (J) keeping S constant ($S = 6$). For each case, 1024 points were collected. No oscillations are observed for the case when there is no slip-jump, i.e. J=0. Periods of oscillation (T_p) are 250 s, 111 s and 72 s for values of $J = 1$, 3 and 6, respectively. From this simulation study, it is clear that both amplitude and frequency of limit cycles increase with the increase of

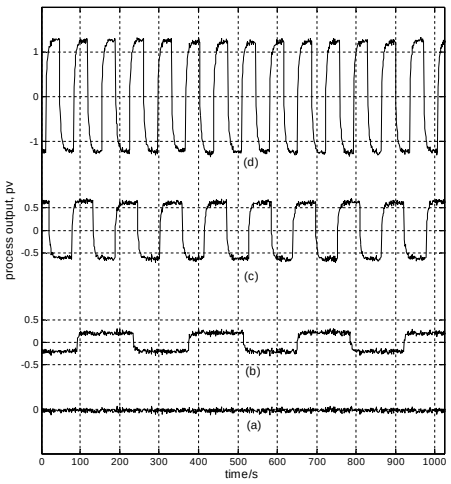


Fig. 6. *(a) J=0, no oscillations detected (b) J=1, T_p=250, amplitude=0.20 (c) J=3, T_p=111, amplitude=0.60 (d) J=6, T_p=72, amplitude=1.20.*

J. Therefore, the estimation of J is as important as the estimation of S.

4. METHODOLOGY FOR SIMULTANEOUS ESTIMATION OF S AND J

Figure 7 shows the detailed flow chart of the procedure for estimating $'S'$ and $'J'$. This is an iterative optimization procedure to identify both the stiction model parameters and the process model simultaneously. The controller output data (op) is supplied to the two parameter stiction model to obtain the actual valve output or valve-postion data, (vo), for a fixed value of S and J. Then, the predicted valve output, vo, and the process output data, (pv), are used to identify the process model using Akaike's Information Criteria (AIC). The procedure is repeated for various values of S and J obtained from a two dimensional grid search. The value of S and J that gives minimum mean square error for the controlled process variable (pv) is reported as stiction. The details of the algorithm are as follows:

- Import process ouput, pv and the controller output, op.
- Check for non-linearity in the system. In this work the bicoherence based method proposed by (Choudhury *et al.*, 2002) is used for non-linearity detection.
- Choose a value for (S_i,J_i) from a two dimensional grid of S and J.
- Use the controller output, op, data and the two-parameter stiction model with chosen (S_i, J_i) to compute the valve output, vo. This is the non-linear part of the Hammerstein model.
- Identify the process model (linear part of the Hammerstein model) using the valve output, vo, and the process output, pv.
- Then the process output is predicted $(pv^{'})$ using the identified process model and the computed valve output, vo.
- Compute the Mean Squared Error between the predicted and the actual process output

$$MSE(S_i, J_i) = \sum_{i=1}^{N} (pv_i - pv_i^{'})^2 \qquad (3)$$

- MSE is computed for all the points in the grid of S and J. The value (S_m, J_m), for which MSE is minimum, is reported as stiction.

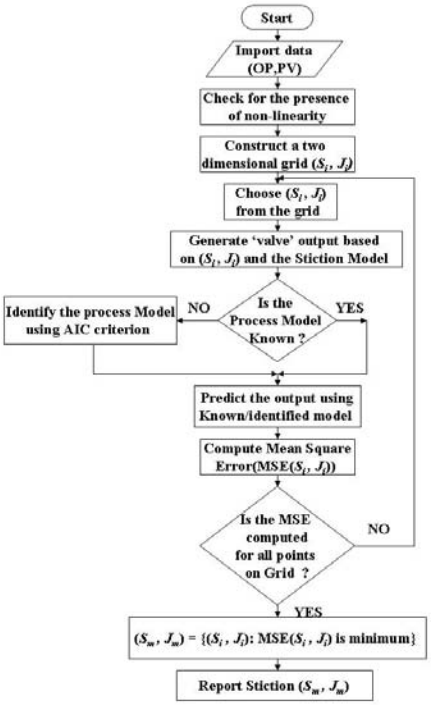


Fig. 7. *Logic flow diagram of the proposed method*

The following important points should be considered in the implementation of the method:

- The prediction, using the identified or known model and the valve output, is done using a one-step-ahead predictor. The purpose of using one-step-ahead predictor is that this makes the overall procedure less dependent on the process model, estimation of which is of less interest for this case.
- There is a possibility that for a particular value of (S, J) the computed valve output may be saturated. In this case the identification of the linear part of the Hammerstein model would be difficult because the input signal would not be persistently exciting. This may result in erroneous results. Therefore, before using the valve output (vo) for the identification of the process model, the signal should be should be examined for possible saturation.

5. RESULTS FROM SIMULATION STUDIES

All simulations were performed using the same system described in Section 3.1. The controller gain, K_c = 0.15 and τ_I = 1 are fixed.

Two scenarios are considered here. First, when the process model is known i.e. the linear component of

the Hammerstein model is known. Second, when the linear part of the Hammerstein model is unknown and estimated along with the nonlinear part.

Table 1 shows the estimation results using the proposed method. It is assumed that the process model is known. The estimated values are close to the actual values.

Table 1. *Comparison of actual and estimated S and J (known model case)*

S		J	
Actual	Estimated	Actual	Estimated
1	1	0	0
1	1	1	1
4	4	2	2
6	6	4	3.5
8	8	8	8
8	8	10	10
10	8	2	0

Table 2 shows the estimation results when an external disturbance is added to the system with sticky valve. A sinusoidal input with a frequency of 1 rad/sec and amplitude of 1 is used as external disturbance. The process model is assumed to be known. The estimation is exact in most cases. This indicates that the proposed method is able to quantify stiction even in presence of external oscillations.

Table 2. *Comparison of actual and estimated S and J in the presence of external oscillations (known model)*

S		J	
Actual	Estimated	Actual	Estimated
1	1	1	1
4	4	2	2
6	6	4	4.5
10	10	5	5
12	12	4	4
12	13	0	1

In Table 3 estimation results are shown when the data is corrupted by noise (random noise with zero mean). Signal to noise ratio (SNR) is computed as the ratio of the variance of the noise free signal to variance of the noise. For this the value of S and J are fixed to 6 and 4 respectively and the data is simulated with different noise levels in the system. The results show that the method is relatively insensitive to the presence of noise, and therefore it should work well when applied to real process data.

Table 3. *Prediction in presence of noise (S = 6 and J = 4) (known model)*

SNR	Estimated (S)	Estimated (J)
100	6	3.5
50	6	4
25	6	4
12.5	6	4
10	6	3.5

Table 4 shows the results for the case when it is assumed that the process model is not known. The algorithm was not supplied with the process model. For all cases, S has been estimated correctly except when $S < J$ $(S = 4, J = 8)$. But such cases, where $J > S$, are rarely encountered in real life. Slip-jump is also estimated correctly for most cases.

Table 4. *Comparison of actual and estimated S and J, unknown model case*

S		J	
Actual	Estimated	Actual	Estimated
1	1	0	0.5
4	4	2	2
6	6	4	4
10	10	10	7.5
10	10	8	8
10	10	2	2
4	2	8	8

6. RESULTS FROM PILOT PLANT EXPERIMENTS

For the verification of the proposed method, data was generated using a laboratory scale setup of a tank system in the Computer Process Control Laboratory in the Department of Chemical and Materials Engineering at the University of Alberta. Data is generated for two control loops: flow and level(cascade) control.

6.1 Flow Control Loop:

The schematic of the process is shown in Figure 8. First of all, the control valve was checked for possible presence of stiction using the so called bump test or the valve travel test and it was found to be stiction free. Then the two-parameter stiction model was used to introduce valve stiction within the software as shown in Figure 8. The signal from the flow controller (FC) is supplied to the stiction model (with already known S and J). The output of the stiction model is then provided to the flow control valve (FV).

Figure 9 show the process output (pv) and the controller output (op) for the system for $(S, J) = (2, 1)$. Clearly, stiction introduces limit cycle behaviour in the loop. The results of stiction estimation are provided in Table 5. Two cases are considered for this loop. For both cases, estimated S and J are in good agreement with the actual S and J.

6.2 Level Control Loop:

The schematic of the control loop is shown in figure 10. This is a cascaded loop. The level controller (LC) signal acts like a set point for the flow controller (FC).Process output (pv, the level) and the controller output (op) for $(S, J) = (1, 1)$ are shown in Figure 11.Results of stiction estimation are summarized in Table 5. The method successfully quantifies S and J.

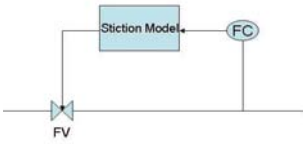


Fig. 8. *Schematic for the Flow loop*

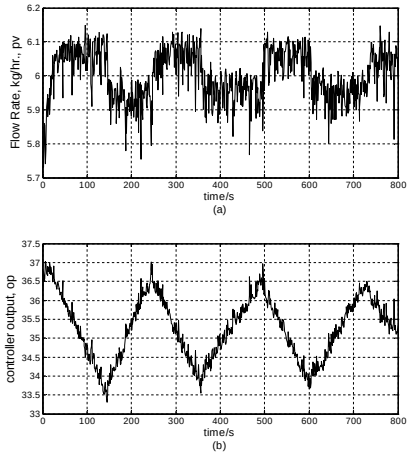


Fig. 9. *Process output (flow rate, PV) and controller output (OP) for the flow control loop*

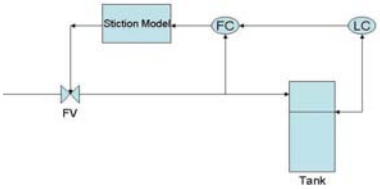


Fig. 10. *Schematic of the cascaded level loop control*

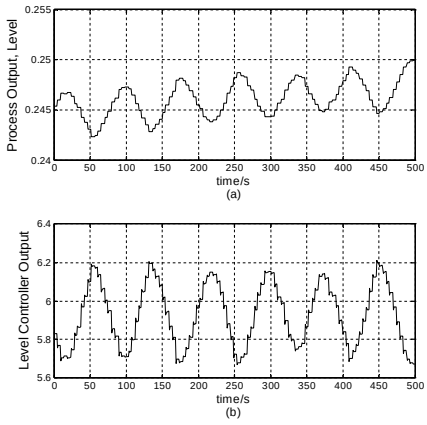


Fig. 11. *(a)Process output (Level, PV) (b)Level controller output (OP)*

7. CONCLUSIONS

In this work, the effect of controller dynamics on the apparent stiction and the impact of J on the frequency and amplitude of limit cycles due to stiction have been demonstrated using simulation examples. A method is proposed to simultaneously estimate both S (stickband+deadband) and J (slip-jump). The stiction model parameters and the process model are jointly identified using an optimization approach. The proposed method has been tested successfully on simulated and experimental data. The method needs only routine operating data from a control loop. The stiction model used in this method in its slightly modified form can also handle asymmetric stiction. Therefore it is possible to extend the method to estimate parameters of asymmetric stiction model.

Table 5. *Estimated S and J from experimental data*

	S		J	
Data	Actual	Estimated	Actual	Estimated
Flow Loop	1	1	1	1
	2	2	1	1.5
Level Loop	1	1	1	1
	2	1.5	1	0.5

The proposed method can also be extended to cases when the plant model is non-linear in itself. In such cases, to correctly estimate S and J, knowledge of the presence and structure of the non-linearity is required.

REFERENCES

Choudhury, M. A. A. S., N. F. Thornhill and Sirish L. Shah (2004*a*). Modelling valve stiction. *Control Engng. Prac., In press*.

Choudhury, M. A. A. S., S. L. Shah and N. F. Thornhill (2004*b*). Diagnosis of poor control loop performance using higher order statistics. *Automatica* **40**, 1719–1728.

Choudhury, M. A. A. S., Sirish L. Shah and Nina F. Thornhill (2002). Detection and diagnosis of system nonlinearities using higher order statistics. In: *15th IFAC World Congress*. Barcelona, Spain.

Choudhury, M. A. A. S., Sirish L. Shah and Nina F. Thornhill (2004*c*). Detection and quantification of control valve stiction. In: *The proceedings of DYCOPS 2004, July 5-7, 2004*. Cambridge, USA.

Choudhury, M. A. A. Shoukat, Sirish L. Shah, Nina F. Thornhill and David S. Shook (2005). An automatic method for detection and quantification of stiction in control valves. *Control Engng. Prac., to appear*.

Horch, A. (1999). A simple method for detection of stiction in control valves. *Control Engng. Prac.* **7**, 1221–1231.

Singhal, A. and T. I. Salsbury (2005). A simple method for detecting valve stiction in oscillating control loops. *Journal of Process Control* **15**, 371.

Srinivasan, R., R. Rengaswamy and R. Miller (2005*a*). Control loop performance assessment:i. a qualitative approach for stiction diagnosis. *Ind. Eng. Chem. Res.* **44**, 6708–66718.

Srinivasan, R., R. Rengaswamy, S. Narasimhan and R. Miller (2005*b*). Control loop performance assessment:ii. hammerstein model approach for stiction diagnosis. *Ind. Eng. Chem. Res.* **44**, 6719–66728.

Stenman, A., F. Gustafsson and K. Forsman (2003). A segmentation-based method for detection of stiction in control valves. *Int. J. Adapt. Control Signal Process.* **17**, 625–634.

AUTHOR INDEX